IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, October 2026
Impact Factor: 8.2

AI Agents: Architecture, Applications, Challenges And Future Directions

Prof. Santosh Pawar[1], Mr. Sarthak More[2], Prof. Snehal Phapale[3]
Vidya Niketan Institute of Management, Bota, Maharashtra, India

Abstract: *AI agents are intelligent software systems designed to pursue goals by perceiving information, reasoning about tasks, planning actions, using external tools, and interacting with users or environments. Recent advances in large language models (LLMs) have accelerated the development of LLM-based agents that can perform multi-step tasks with comparatively greater autonomy. This review paper examines the architecture, major components, applications, limitations, and future directions of modern AI agents. The review focuses on perception, planning, memory, reasoning, tool use, action, and multi-agent collaboration. Applications in software engineering, education, healthcare, finance, cybersecurity, and knowledge work are discussed. The paper also compares traditional rule-based agents, conventional machine-learning agents, and LLM-based agents in terms of autonomy, adaptability, interaction, and resource requirements. Key challenges include hallucination, reliability, evaluation, security, privacy, latency, coordination, and human oversight. Recent surveys indicate that trustworthy evaluation, secure tool integration, human-agent collaboration, and scalable multi-agent architectures are important directions for future research.*

Keywords: AI Agents, Agentic AI, Large Language Models, LLM Agents, Autonomous Systems, Planning, Tool Use, Multi-Agent Systems

I. INTRODUCTION

Artificial Intelligence has evolved from systems designed for narrowly defined tasks toward systems capable of interacting with dynamic environments. An AI agent can be viewed as a system that receives information from an environment, selects or generates actions, and attempts to achieve a specified objective. Traditional intelligent agents commonly relied on predefined rules, symbolic reasoning, or task-specific machine-learning models. The rapid progress of large language models has introduced a new class of agents in which an LLM can serve as a central reasoning component connected to memory, tools, planners, and external environments.

LLM-based agents extend the capabilities of ordinary conversational models by allowing them to plan multi-step tasks, retrieve information, call software tools, maintain context, and act on intermediate results. Recent literature organizes these systems around components such as perception, planning, memory, action, tool use, and evaluation. The growing research interest has also produced work on human-agent systems, multi-agent collaboration, security, privacy, and agent communication.

This review summarizes the current research landscape of AI agents, identifies their major architectural components and application areas, compares different agent paradigms, and highlights technical and ethical challenges. It also outlines future directions for building reliable, secure, explainable, and scalable agentic systems.

II. ARCHITECTURE AND KEY COMPONENTS OF AI AGENTS

2.1 Perception and Input Processing

The perception layer receives information from users, documents, APIs, databases, sensors, or other digital environments. In LLM-based systems, natural-language input may be combined with structured data or multimodal

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, October 2026
Impact Factor: 8.2

information. Effective perception is important because incorrect or incomplete input can propagate errors through later planning and action stages .

2.2 Reasoning and Planning

Reasoning converts a goal and available information into an actionable strategy. Planning may involve decomposing a complex objective into smaller tasks, selecting tools, determining an execution order, and revising the plan when an action fails. Research on autonomous agents emphasizes planning and reasoning as central mechanisms for handling multi-step problems.

2.3 Memory

Memory enables an agent to retain information beyond a single interaction. Short-term context can support the current task, while longer-term memory can store user preferences, previous interactions, retrieved knowledge, or task history. Memory design must balance usefulness with privacy, accuracy, storage cost, and the risk of retaining incorrect information.

2.4 Tool Use and Action

A major difference between a standard LLM response system and an agent is the ability to interact with external tools. Agents can retrieve information, execute code, access databases, call APIs, manipulate files, or operate software environments. Tool use expands capabilities but also introduces security and reliability risks because an incorrect decision may cause an external action with real consequences.

2.5 Multi-Agent Collaboration

Multiple specialized agents can collaborate by dividing a complex task into subtasks and exchanging intermediate results. Multi-agent systems can improve specialization and parallelism, but communication overhead, conflicting goals, coordination failures, and security risks become important concerns.

III. APPLICATIONS OF AI AGENTS

3.1 Software Engineering

AI agents are increasingly studied for software engineering tasks such as code generation, debugging, testing, repository navigation, and automated program repair. Agent-based workflows can decompose software tasks, use development tools, inspect outputs, and iteratively refine solutions. This area is promising but requires strong evaluation because generated code can contain functional and security defects.

3.2 Education and Knowledge Assistance

Educational agents can support personalized explanations, question answering, study planning, feedback, and information retrieval. Human-agent research suggests that incorporating human feedback and control can improve reliability and safety in tasks where fully autonomous behaviour is inappropriate.

3.3 Healthcare and Finance

In healthcare and finance, agents can assist with information retrieval, workflow coordination, documentation, decision support, and analysis. These domains require stronger safeguards than ordinary productivity applications because incorrect outputs can create significant consequences. Human oversight, traceability, privacy protection, and domain-specific evaluation are therefore essential.

3.4 Cyber security

AI agents can assist security teams with alert analysis, threat-information retrieval, investigation workflows, and defensive automation. At the same time, argentic systems can introduce new attack surfaces through malicious inputs, unsafe tool calls, prompt injection, data leakage, and compromised external services. Security must therefore be considered as part of the architecture rather than added only after deployment.

IV. COMPARISON OF AI AGENT PARADIGMS

Feature	Rule-Based Agents	ML-Based Agents	LLM-Based Agents	Multi-Agent Systems
Decision Making	Predefined rules	Learned	Reasoning + planning	Distributed decisions

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, October 2026
Impact Factor: 8.2

		policies/models		
Adaptability	Low	Moderate	High	High
Natural Language	Limited	Task dependent	Strong	Strong
Tool Integration	Manual/limited	Application specific	Extensive	Extensive
Coordination	Usually single agent	Usually single agent	Possible	Core capability
Main Limitation	Rigid behavior	Training/data dependence	Reliability and cost	Coordination and security

The comparison shows that LLM-based agents provide stronger natural-language interaction and flexible tool use than traditional rule-based systems, while multi-agent architectures extend these Capabilities through collaboration. However, greater autonomy also increases the need for robust evaluation, monitoring, access control, and human oversight.

V. CHALLENGES AND LIMITATIONS

- Reliability and Hallucination: Agents may produce incorrect reasoning, fabricated information, or invalid tool selections. Reliable evaluation remains an open research problem.
- Security and Privacy: Tool-enabled agents can expose sensitive data or perform unsafe actions when inputs, permissions, or external tools are compromised.
- High Computational Cost and Latency: Multi-step reasoning and repeated model/tool calls can increase response time and operational cost. Recent research specifically identifies latency as a major barrier to practical deployment.
- Evaluation Difficulty: Agent performance depends on both final answers and intermediate actions, making simple accuracy metrics insufficient for many workflows.
- Human Oversight and Alignment: Fully autonomous operation is not appropriate for every domain. Human feedback, intervention, and controllability are important for high-impact applications.
- Multi-Agent Coordination: Collaborative agents can experience communication overhead, conflicting objectives, and emergent failures. Secure coordination protocols remain an active research area.

VI. FUTURE DIRECTIONS

- Trustworthy Agent Evaluation: Future research is expected to develop standardized benchmarks that measure planning, tool use, factuality, safety, robustness, and long-horizon task completion.
- Human-in-the-Loop Agent Systems: Combining autonomous capabilities with human feedback and intervention can improve reliability, safety, and accountability.
- Secure Tool and Protocol Integration: Agents will require stronger permission models, sandboxing, auditing, authentication, and secure communication mechanisms for external tools and other agents.
- Efficient Agent Architectures: Reducing model calls, optimizing planning and memory, and improving inference infrastructure can lower latency and cost.
- Scalable Multi-Agent Ecosystems: Research is moving toward systems in which specialized agents can discover capabilities, communicate, coordinate tasks, and resolve conflicts reliably.
- Responsible and Explainable Agentic AI: Future systems should provide understandable decision traces, controllable autonomy, privacy protection, and mechanisms for detecting and correcting unsafe behaviour.

VII. CONCLUSION

AI agents represent an important development in artificial intelligence because they extend models from generating responses toward pursuing goals through planning, memory, tool use, and interaction with environments. Recent research shows rapid progress in LLM-based and multi-agent systems, together with growing applications in software

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, October 2026
Impact Factor: 8.2

engineering, education, healthcare, finance, cyber security, and knowledge work [1]–[6]. However, increased autonomy also introduces challenges related to reliability, hallucination, security, privacy, latency, evaluation, and human control. The future of AI agents therefore depends not only on improving reasoning and autonomy, but also on building systems that are measurable, secure, transparent, efficient, and aligned with human oversight.

REFERENCES

1. Luo, J., Zhang, W., Yuan, Y., et al. (2025). Large Language Model Agent: A Survey on Methodology, Applications and Challenges. arXiv preprint arXiv:2503.21460.
2. Zou, H. P., Huang, W.-C., Wu, Y., et al. (2025). A Survey on Large Language Model based Human-Agent Systems. arXiv preprint arXiv:2505.00753.
3. Ferrag, M. A., Tihanyi, N., & Debbah, M. (2025). From LLM Reasoning to Autonomous AI Agents: A Comprehensive Review. arXiv preprint arXiv:2504.19678.
4. IEEE Communications Surveys & Tutorials. (2025). Large Model-Based Agents: State-of-the-Art, Cooperation Paradigms, Security and Privacy, and Future Trends. 28, 1906–1949. DOI: 10.1109/COMST.2025.3576176.
5. IEEE. (2024). A Survey of LLM-based Agents: Theories, Technologies, Applications and Suggestions. 2024 3rd International Conference on Artificial Intelligence, Internet of Things and Cloud Computing Technology. DOI: 10.1109/AIoTC63215.2024.10748304.
6. IEEE Access. (2025). Agentic AI: A Comprehensive Survey of Technologies, Applications, and Societal Implications. 13, 151824–151837. DOI: 10.1109/ACCESS.2025.3585609.
7. IEEE Computer. (2025). Can AI Fix Buggy Code? Exploring the Use of Large Language Models in Automated Program Repair. Computer, 58(7), 122–128. DOI: 10.1109/MC.2025.3527407.
8. IEEE Access. (2026). Minimizing Response Latency in LLM-Based Agent Systems: A Comprehensive Survey. 14, 26140–26168. DOI: 10.1109/ACCESS.2026.3664226.