



Agentic AI and Autonomous LLM Systems: A Scientific Literature Review

S. Karthick

Assistant Professor

Department of IT & Cognitive Systems

Sri Krishna Arts and Science College

Coimbatore – 641008

karthickksenthil@gmail.com

A. Sherill

Assistant Professor

Department of SS and AIML

Sri Krishna Arts and Science College

Coimbatore – 641008

sherillantony@gmail.com

How to Cite this Article:

Karthick, S. & Sherill, A. (2026). Agentic AI and Autonomous LLM Systems: A Scientific Literature Review. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(10), 1-9. <https://doi.org/10.55041/ijcope.v2i10.194>

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.



<https://doi.org/10.55041/ijcope.v2i10.194>

Abstract—The shift of large language models (LLMs) from passive text generators to agentic systems marks a move toward autonomous intelligence. Agentic AI refers to systems that reason, plan, and execute actions in pursuit of complex goals with minimal human intervention. This paper reviews recent research on the architectural foundations, core functional modules, applications, and safety challenges of autonomous LLM systems. Current work emphasizes modular architectures in which an LLM serves as the central reasoning engine, augmented by components for perception, memory, planning, and tool interaction. Multi-agent systems have emerged as a dominant trend, enabling collaborative problem solving through structured orchestration and communication protocols. Although these systems show significant potential in software engineering, manufacturing, and cybersecurity, critical barriers remain in safety, reliability, and ethical governance. This review provides a structured overview of the state of the art and identifies key research directions for the coming years.

Index Terms—agentic AI, autonomous agents, large language models, multi-agent systems, retrieval-augmented generation, tool use, AI safety

I. INTRODUCTION

Large language models (LLMs) have significantly expanded the capabilities of intelligent software systems, enabling them to perform tasks that require language understanding, reasoning, and decision support. More recently, research has focused on integrating these models into autonomous frameworks that can pursue objectives through sequences of decisions and actions. These frameworks, often described as agentic systems, combine language-based reasoning with planning, memory, and interaction mechanisms to address complex tasks in dynamic environments [1], [2]. Unlike conventional language-model applications that generate responses to user prompts,

autonomous systems operate through continuous cycles of information gathering, goal decomposition, action selection, and evaluation [3]. By incorporating external knowledge sources, software tools, and memory structures, such systems can execute multi-step workflows and adapt their behavior as new information becomes available [4], [5].

Research in this area has accelerated with the development of architectures that support task planning, tool utilization, and feedback-driven refinement. Representative frameworks demonstrate how language models can coordinate multiple operations, maintain context over extended interactions, and address problems requiring long-term reasoning [6]. In parallel, multi-agent approaches have gained attention by distributing responsibilities across specialized agents that communicate and collaborate to accomplish shared objectives [7].

These developments have led to applications in software engineering, manufacturing, scientific discovery, business automation, and cybersecurity. However, several challenges remain unresolved, including reasoning errors, unreliable outputs, security risks, evaluation difficulties, and governance concerns [8]. Addressing these limitations is essential for the dependable deployment of autonomous systems in real-world environments.

This paper presents a literature review of agentic and autonomous LLM-based systems. It examines their architectural foundations, core functional components, application domains, and associated challenges. The review also highlights current research trends and identifies opportunities for future investigation in the development of reliable and effective autonomous intelligent systems.

II. FOUNDATIONAL FRAMEWORKS AND ARCHITECTURES

Architectural design is central to the performance and reliability of agentic systems. Researchers have proposed various frameworks to organize the complex interactions between an agent's reasoning core and its environment.

A. Theoretical Foundations and Taxonomies

Modern agentic architectures often draw inspiration from classical AI paradigms, such as the Belief-Desire-Intention (BDI) model, while evolving toward utility-based and LLM-centric designs [8], [9]. A unified taxonomy categorizes agentic systems into six primary dimensions: perception, brain (the core LLM), planning, action, tool use, and collaboration [10]. Another framework, the Agentic AI Systems Architecture (AISA), provides a layered conceptual model that integrates



cognitive design, tool execution, and governance into a single coherent structure [11].

B. Modular and Unified Architectures

The prevailing design trend favors modularity, in which specific functions are decoupled to allow for iterative refinement and scalability [3], [12]. Common architectural layers include the execution layer, which handles physical or digital actions; the resource layer, which manages data and external API connections; the model layer, which comprises the underlying foundation model (an LLM or multimodal LLM); and the agent layer, which contains the reasoning logic and internal state management [13], as shown in Fig. 2. For enterprise-scale deployments, researchers suggest a composable architecture that integrates modular agents with a centralized data fabric and event-driven backbones [14]. Furthermore, specialized agent operating systems such as AIOS and KAOS have been developed to manage resource

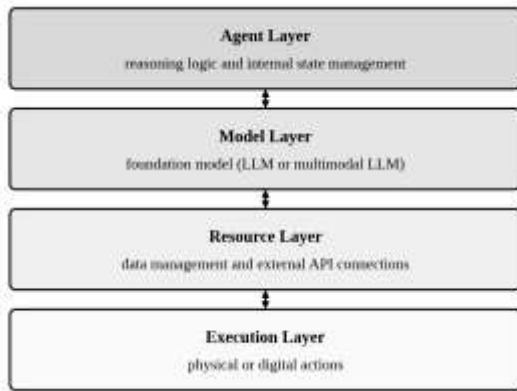


Fig. 2. Common layered architecture of agent services, after [13].

allocation and task scheduling for multiple agents, mirroring the function of traditional operating systems for software [15], [16]. The frameworks discussed in this section are compared in Table I.

TABLE I
REPRESENTATIVE ARCHITECTURAL FRAMEWORKS AND TAXONOMIES

FRAMEWORK	KEY CHARACTERISTICS	REF.
BDI-inspired and LLM-centric designs	Evolution from the classical BDI model toward utility-based and LLM-centric designs	[8], [9]
Unified taxonomy	Six dimensions: perception, brain, planning, action, tool use, and collaboration	[10]
AISA	Layered model integrating cognitive design, tool execution, and governance	[11]
Four-layer architecture	Execution, resource, model, and agent layers	[13]
Composable enterprise architecture	Modular agents with a centralized data fabric and an event-driven backbone	[14]
Core functional modules	Profiling, memory, planning, and action	[12]
AIOS and KAOS	Agent operating systems for resource allocation and task scheduling	[15], [16]

III. CORE FUNCTIONAL COMPONENTS

The degree of autonomy exhibited by an agentic system is largely determined by the integration of four key functional components: profiling, memory, planning, and action [12].

A. Reasoning and Planning Mechanisms

Planning is the process by which an agent breaks down a complex goal into executable steps. Techniques such as Chain-of-Thought (CoT) and Tree-of-Thought (ToT) enable agents to explore multiple reasoning paths before committing to an action [17], [18]. Advanced planning frameworks incorporate reflection and self-criticism, allowing agents to analyze their own errors and refine their strategies iteratively [4], [6]. Hierarchical reinforcement learning is also employed to handle tasks with long-term dependencies and sub-goal structures [6].

B. Memory Systems

Memory allows agents to maintain context and learn from past experiences. The literature distinguishes two primary forms. Short-term memory is often implemented through the model's prompt context window and is sometimes referred to as a "scratchpad" for intermediate reasoning [3], [19]. Long-term memory is facilitated by external storage solutions, such as vector databases (e.g., ChromaDB), which support retrieval-augmented generation and the storage of persistent agent knowledge [20], [21]. Effective memory management is critical; studies have shown that the way memory is structured directly impacts an agent's ability to follow complex experiences and instructions [22].

C. Tool Integration and Environment Interaction

A defining feature of autonomous agents is their ability to use external tools, such as web browsers, code interpreters, and specialized APIs [3], [6]. Tool integration layers allow agents to invoke functions using standard formats such as JSON schemas or the Model Context Protocol (MCP) [14], [21]. Recent advances even include "autonomous tool creation," in which agents develop their own tools to solve novel problems [23].

D. Agentic Retrieval-Augmented Generation (RAG)

Agentic RAG represents a sophisticated evolution of traditional RAG, in which the retrieval process is not static but is driven by an agent's reasoning core. These systems use single-agent routers or multi-agent hierarchies to decide when and where to search for information, improving the precision and relevance of retrieved data in complex workflows [24].

IV. MULTI-AGENT COORDINATION AND ORCHESTRATION

Multi-agent systems (MAS) represent a shift from single autonomous entities to collaborative societies of agents [25], [26]. In MAS, different agents may be assigned specific roles, such as "moderator," "executor," or "critic," to solve tasks through collective intelligence [27], [28], as depicted in Fig. 3.

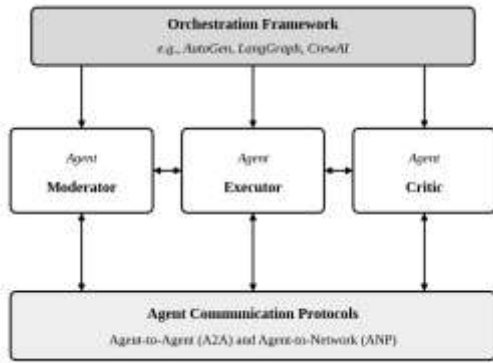


Fig. 3. Conceptual multi-agent orchestration with role-specialized agents and standardized communication protocols.

Key frameworks for MAS orchestration are summarized in Table II. AutoGen orchestrates multi-agent tasks through conversational patterns [1], [19]. LangGraph and CrewAI support stateful transitions and role-based collaboration, enabling structured workflows for complex task execution [5], [27], [29]. Agent communication protocols (ACPs) provide standardized methods for agent-to-agent (A2A) and agent-to-network (ANP) communication, ensuring interoperability between disparate agentic systems [14], [27].

TABLE II
MULTI-AGENT ORCHESTRATION FRAMEWORKS AND PROTOCOLS

FRAMEWORK	ORCHESTRATION APPROACH	REF.
AutoGen	Orchestrates multi-agent tasks through conversational patterns	[1], [19]
LangGraph / CrewAI	Stateful transitions and role-based collaboration for structured workflows	[5], [27], [29]
ACPs (A2A, ANP)	Standardized agent-to-agent and agent-to-network communication for interoperability	[14], [27]

Research into these systems also highlights the emergence of “sustainability behaviors” in agent societies, in which agents must learn to share resources and cooperate to avoid system collapse [28].

V. APPLICATIONS OF AGENTIC AI

Agentic AI is being applied across a broad spectrum of industries, transforming traditional workflows into autonomous processes. Table III summarizes the main application domains discussed in the literature.

A. Software Engineering and Data Science

In software engineering, agents are used for automated code generation, debugging, and project management [30]–[32]. Data science agents can automate the entire lifecycle of a project, from data cleaning to model evaluation, using modular execution structures and unit testing [33].

B. Enterprise and Industrial Systems

For enterprises, agentic AI supports scalable and resilient architectures by automating routine business processes and managing large-scale data flows [14], [34]. In manufacturing, “hybrid agentic AI” systems are being integrated into smart factories to enhance decision-making and production flexibility [35], [36].

C. Specialized Domains: Healthcare, Geospatial, and Robotics

In healthcare, agentic systems in radiology assist in diagnosis and design, although they require strict governance owing to the critical nature of the domain [37]. In remote sensing, agents are used for geospatial workflows, focusing on trajectory-aware reasoning and active tool orchestration for environmental monitoring [38]. In robotics, embodied agentic AI integrates large visual language models (VLMs) with robot autonomy, enabling physical agents to interact with the real world through perception–action loops [39], [40].

TABLE III
APPLICATION DOMAINS OF AGENTIC AI

DOMAIN	REPRESENTATIVE USE	REF.
Software engineering	Automated code generation, debugging, and project management	[30]–[32]
Data science	Project lifecycle automation, from data cleaning to model evaluation	[33]
Enterprise systems	Routine business process automation and large-scale data flow management	[14], [34]
Manufacturing	Hybrid agentic AI in smart factories for decision-making and production flexibility	[35], [36]
Healthcare (radiology)	Diagnosis and design assistance under strict governance	[37]
Remote sensing	Geospatial workflows with trajectory-aware reasoning and active tool orchestration	[38]
Robotics	Embodied agents combining VLMs with robot autonomy through perception–action loops	[39], [40]

VI. EVALUATION, SAFETY, AND TRUSTWORTHINESS

As agents gain autonomy, the risks associated with their operation increase. Evaluating these systems requires moving beyond traditional language metrics to assess behavioral outcomes.

A. Performance Benchmarking

Standardized benchmarks have been developed to measure agent performance across diverse environments, as listed in Table IV. AgentBench evaluates agents across code, gaming, and web-based tasks [2], [13], [21]. GAIA is a benchmark for general AI assistants that focuses on complex, multi-step tasks [21], [41]. TheAgentCompany measures the efficiency and capability of agents in enterprise-like environments [2].

TABLE IV
BENCHMARKS FOR EVALUATING AGENTIC SYSTEMS

BENCHMARK	EVALUATION FOCUS	REF.
AgentBench	Agent performance across code, gaming, and web-based tasks	[2], [13], [21]
GAIA	General AI assistants on complex, multi-step tasks	[21], [41]
TheAgentCompany	Efficiency and capability in enterprise-like environments	[2]

B. Security Threats and Defenses

Agentic systems face unique security vulnerabilities, including prompt injection, malfunction amplification, and adversarial exploitation [10], [42]. Defensive strategies include the use of guardrails (e.g., GuardAgent), sandboxed emulation



environments (e.g., ToolEmu), and restricted pure functions to prevent malicious code execution [3], [13], [27]. AI security layers now employ anomaly detection and adversarial simulation to maintain system integrity [14].

C. Ethical Governance and Alignment

Ethical assurance is a major research focus, emphasizing the need for transparency and human alignment [4], [5]. Frameworks such as AGENTS SAFE aim to provide unified ethical governance for agentic AI [43]. Human-in-the-loop (HITL) and human-on-the-loop (HOTL) configurations remain critical for supervised development and for ensuring that autonomous decisions remain socially aligned [20], [44].

VII. RESEARCH GAPS AND FUTURE DIRECTIONS

Despite significant progress, several challenges remain. The first concerns grounding and reliability: agents still struggle with “hallucinations in action” and infinite loops when faced with novel environment states [10], [38]. The second concerns scalability and efficiency, as managing the resource demands of multi-agent systems, especially those using large-scale multimodal models, is a persistent technical gap [4], [25]. The third concerns standardized safety: there is a lack of universal metrics for agency and of dynamic alignment mechanisms that can adapt to changing ethical standards [2], [45].

Future research is expected to focus on decentralized, self-sovereign AI agents and on the development of “Agentic Webs,” in which autonomous entities can interact and trade services securely [46], [47].

VIII. CONCLUSION

Agentic AI and autonomous LLM systems represent a significant development in the evolution of artificial intelligence, extending the capabilities of language models beyond content generation toward autonomous reasoning, planning, and action execution. By integrating LLMs with complementary components such as memory systems, planning mechanisms, retrieval frameworks, and external tools, researchers have developed architectures capable of addressing complex and multi-step tasks across a wide range of domains. Recent studies demonstrate the growing effectiveness of these systems in software engineering, cybersecurity, healthcare, scientific research, manufacturing, and enterprise decision support.

The emergence of multi-agent frameworks has further expanded the scope of autonomous systems by enabling collaboration among specialized agents. Through coordinated communication, task decomposition, and role-based execution, these architectures have shown improvements in scalability, adaptability, and problem-solving performance. At the same time, advances in retrieval-augmented generation, long-term memory management, and tool-use capabilities continue to enhance the operational effectiveness of autonomous LLM systems in dynamic environments.

Despite these achievements, several challenges remain unresolved. Issues related to hallucination, unreliable reasoning, memory inconsistencies, security vulnerabilities, and evaluation limitations continue to affect the reliability of autonomous agents. Furthermore, increasing levels of autonomy introduce important concerns regarding transparency, accountability, privacy protection, bias mitigation, and regulatory compliance.

Addressing these challenges requires the development of robust safety mechanisms, standardized benchmarking methodologies, and comprehensive governance frameworks capable of ensuring responsible deployment.

Future research is expected to focus on improving agent reliability, enhancing long-term planning capabilities, strengthening human-agent collaboration, and developing scalable multi-agent coordination strategies. Greater emphasis on interpretability, alignment, and risk assessment will also be essential as autonomous systems become more integrated into real-world applications. Ultimately, the successful advancement of agentic AI will depend not only on improvements in model capability but also on the establishment of trustworthy, secure, and human-centered design principles. As the field continues to mature, autonomous LLM systems are likely to become a foundational technology for next-generation intelligent systems, supporting complex decision-making processes across both industrial and societal contexts.

REFERENCES

- [1] A. Koubâa, “From pre-trained language models to agentic AI: Evolution and architectures for autonomous intelligence,” July 2025, doi: 10.20944/preprints202507.1809.v1.
- [2] S. T. Erukude, S. R. Veluru, and V. C. Marella, “Agentic AI: The rise of autonomous intelligent agents in the era of LLMs,” *Indian Journal of Computer Science and Engineering*, vol. 16, no. 2, pp. 9–16, Apr. 2025, doi: 10.21817/indjcs/2025/v16i1/251602024.
- [3] S. S. Chowh et al., “From language to action: A review of large language models as autonomous agents and tool users,” arXiv:2508.17281, Aug. 2025, doi: 10.48550/arxiv.2508.17281.
- [4] N. S. Sibai, Y. Ahmed, S. Sibae, S. AlHalawani, A. Ammar, and W. Boulila, “The path ahead for agentic AI: Challenges and opportunities,” arXiv:2601.02749, Jan. 2026, doi: 10.48550/arxiv.2601.02749.
- [5] M. Ricardo, “Inteligência artificial agêntica (agentic AI): Fundamentos, arquiteturas e aplicações,” Jan. 2026, doi: 10.5281/zenodo.18195261.
- [6] X. Qu et al., “A comprehensive review of AI agents: Transforming possibilities in technology and beyond,” arXiv:2508.11957, Aug. 2025, doi: 10.48550/arxiv.2508.11957.
- [7] Y. Jian, “Conceptual foundations of LLM-powered agents: From language processing to autonomous reasoning,” *Scientific Journal of Intelligent Systems Research*, vol. 7, no. 4, pp. 29–35, Apr. 2025, doi: 10.54691/atmxvb43.
- [8] M. Bharti, “AI agents: A systematic review of architectures, components, and evolutionary trajectories in autonomous digital systems,” *International Journal of Computer Engineering and Technology*, vol. 16, no. 1, pp. 809–820, Jan. 2025, doi: 10.34218/ijcet_16_01_065.
- [9] S. Deshmukh, M. Kuri, V. Vanarote, M. Nimbalkar, P. Chandre, and S. Yadav, “Agentic AI: A survey of autonomous agents, architectures, and emerging applications,” pp. 1–8, Dec. 2025, doi: 10.1109/ictbig68706.2025.11323264.
- [10] A. V. G. R. and R. Buyya, “Agentic artificial intelligence (AI): Architectures, taxonomies, and evaluation of large language model agents,” arXiv:2601.12560, Jan. 2026, doi: 10.48550/arxiv.2601.12560.
- [11] O. Nacar, A. Deema, and A. Mohammed, “AISA: A unified architecture for agentic AI systems,” Jan. 2026, doi: 10.5281/zenodo.18161879.
- [12] Y. Ren, Y. Liu, T. Ji, and X. Xu, “AI agents and agentic AI—Navigating a plethora of concepts for future manufacturing,” arXiv:2507.01376, July 2025, doi: 10.48550/arxiv.2507.01376.
- [13] W. Xu et al., “Deploying foundation model powered agent services: A survey,” arXiv:2412.13437, Dec. 2024, doi: 10.48550/arxiv.2412.13437.
- [14] P. Kumar, “Agentic AI-driven enterprise architecture: A foundational framework for scalable, secure, and resilient systems,” *International Journal of Computational and Experimental Science and Engineering*, vol. 11, no. 4, Mar. 2025, doi: 10.22399/ijcesen.4210.
- [15] Z. Zhuo et al., “KAOS: Large model multi-agent operating system,” arXiv:2406.11342, June 2024, doi: 10.48550/arxiv.2406.11342.
- [16] K. Mei, Z. Li, S. Xu, R. Ye, Y. Ge, and Y. Zhang, “AIOS: LLM agent operating system,” arXiv:2403.16971, Mar. 2024, doi: 10.48550/arxiv.2403.16971.



- [17] G. H. Kahsay, L. Alexander, and K. Alois, "Fundamentals of building autonomous LLM agents," arXiv:2510.09244, Oct. 2025, doi: 10.48550/arxiv.2510.09244.
- [18] Z. Zhang et al., "Igniting language intelligence: The hitchhiker's guide from chain-of-thought reasoning to language agents," arXiv:2311.11797, Nov. 2023, doi: 10.48550/arxiv.2311.11797.
- [19] W. Zhou et al., "Agents: An open-source framework for autonomous language agents," arXiv:2309.07870, Sept. 2023, doi: 10.48550/arxiv.2309.07870.
- [20] A. P. Бідочко and Y. Vyklyuk, "LLMAgentNet: A collaborative network of autonomous AI agents for complex task execution," *Scientific Bulletin of UNFU*, vol. 35, no. 4, pp. 101–114, June 2025, doi: 10.36930/40350412.
- [21] C. Sypherd and V. Belle, "Practical considerations for agentic LLM systems," arXiv:2412.04093, Dec. 2024, doi: 10.48550/arxiv.2412.04093.
- [22] Z. Xiong et al., "How memory management impacts LLM agents: An empirical study of experience-following behavior," arXiv:2505.16067, May 2025, doi: 10.48550/arxiv.2505.16067.
- [23] X. Du, "Empowering LLM-based agents: Methods and challenges in tool use," *Applied and Computational Engineering*, vol. 203, no. 1, pp. 9–16, Nov. 2025, doi: 10.54254/2755-2721/2026.tj28954.
- [24] A. Singh, A. Ehtesham, S. Kumar, and T. T. Khoei, "Agentic retrieval-augmented generation: A survey on agentic RAG," arXiv:2501.09136, Jan. 2025, doi: 10.48550/arxiv.2501.09136.
- [25] H. Joshi and N. Singh, "Agentic AI, autonomous agents, and multi-agent systems: Concepts, challenges, and research pathways," Dec. 2025, doi: 10.55041/ijrsrem55390.
- [26] M. Špeletić and S. Cvetković, "Autonomous intelligence with LLMs: From simple conversations to multi-agent collaboration," doi: 10.46793/ieeestec18.017s.
- [27] H. Derouiche, Z. Brahmi, and H. Mazeni, "Agentic AI frameworks: Architectures, protocols, and design challenges," arXiv:2508.10146, Aug. 2025, doi: 10.48550/arxiv.2508.10146.
- [28] G. Piatti, Z. Jin, M. Kleiman-Weiner, B. Schölkopf, M. Sachan, and R. Mihalcea, "Cooperate or collapse: Emergence of sustainability behaviors in a society of LLM agents," arXiv:2404.16698, Apr. 2024, doi: 10.48550/arxiv.2404.16698.
- [29] Z. Duan and J. Wang, "Exploration of LLM multi-agent application implementation based on LangGraph+CrewAI," arXiv:2411.18241, Nov. 2024, doi: 10.48550/arxiv.2411.18241.
- [30] J. Liu et al., "Large language model-based agents for software engineering: A survey," arXiv:2409.02977, Sept. 2024, doi: 10.48550/arxiv.2409.02977.
- [31] Y. Dong et al., "A survey on code generation with LLM-based agents," arXiv:2508.00083, July 2025, doi: 10.48550/arxiv.2508.00083.
- [32] K. Cinkusz, J. A. Chudziak, and E. Niewiadomska-Szynkiewicz, "Cognitive agents powered by large language models for agile software project management," *Electronics*, vol. 14, no. 1, p. 87, Dec. 2024, doi: 10.3390/electronics14010087.
- [33] P. Wang, Y. Yu, K. Chen, X. Zhan, and H. Wang, "Large language model-based data science agent: A survey," arXiv:2508.02744, Aug. 2025, doi: 10.48550/arxiv.2508.02744.
- [34] A. Xu et al., "Agentic AI for enterprise: Emerging applications and real-world challenges," pp. 6300–6301, Aug. 2025, doi: 10.1145/3711896.3737871.
- [35] M. A. Farahani, M. I. Khan, and T. Wuest, "Hybrid agentic AI and multi-agent systems in smart manufacturing," arXiv:2511.18258, Nov. 2025, doi: 10.48550/arxiv.2511.18258.
- [36] Y. Xia, M. Shenoy, N. Jazdi, and M. Weyrich, "Towards autonomous system: Flexible modular production system enhanced with large language model agents," Sept. 2023, doi: 10.1109/etfa54631.2023.10275362.
- [37] B. Christian et al., "Agentic systems in radiology: Design, applications, evaluation, and challenges," arXiv:2510.09404, Oct. 2025, doi: 10.48550/arxiv.2510.09404.
- [38] N. A. Talemi, J. Boone, and F. Afghah, "Agentic AI in remote sensing: Foundations, taxonomy, and emerging systems," arXiv:2601.01891, Jan. 2026, doi: 10.48550/arxiv.2601.01891.
- [39] S. Salimpour et al., "Towards embodied agentic AI: Review and classification of LLM- and VLM-driven robot autonomy and interaction," arXiv:2508.05294, Aug. 2025, doi: 10.48550/arxiv.2508.05294.
- [40] E. K. Raptis, A. Ch. Kapoutsis, and E. B. Kosmatopoulos, "Agentic LLM-based robotic systems for real-world applications: A review on their agenticness and ethics," *Frontiers in Robotics and AI*, vol. 12, Aug. 2025, doi: 10.3389/frobt.2025.1605405.
- [41] L. Jiarun et al., "JoyAgent-JDGenie: Technical report on the GAIA," arXiv:2510.00510, Oct. 2025, doi: 10.48550/arxiv.2510.00510.
- [42] B. Zhang et al., "Breaking agents: Compromising autonomous LLM agents through malfunction amplification," arXiv:2407.20859, July 2024, doi: 10.48550/arxiv.2407.20859.
- [43] R. Khan, D. Joyce, and M. Habiba, "AGENTS SAFE: A unified framework for ethical assurance and governance in agentic AI," arXiv:2512.03180, Dec. 2025, doi: 10.48550/arxiv.2512.03180.
- [44] C. Anshuman and M. Prasant, "Agentic AI security: Threats, defenses, evaluation, and open challenges," arXiv:2510.23883, Oct. 2025, doi: 10.48550/arxiv.2510.23883.
- [45] S. N. Brohi, Q. Mastoi, N. Z. Jhanjhi, and T. R. Pillai, "A research landscape of agentic AI and large language models: Applications, challenges and future directions," *Algorithms*, vol. 18, no. 8, p. 499, Aug. 2025, doi: 10.3390/a18080499.
- [46] B. Hu, Y. Liu, and H. Rong, "Trustless autonomy: Understanding motivations, benefits and governance dilemma in self-sovereign decentralized AI agents," arXiv:2505.09757, May 2025, doi: 10.48550/arxiv.2505.09757.
- [47] Z. Guo, Y. Zhou, C. Wang, L. You, M. Bian, and W. Zhang, "BetaWeb: Towards a blockchain-enabled trustworthy agentic web," arXiv:2508.13787, Aug. 2025, doi: 10.48550/arxiv.2508.13787.