10.48047/jocaaa.2026.35.01.55

Revisiting Autonomy in Artificial Intelligence Through Agentic System Design

Lakshmi RahulReddy Mareddy

Sacred Heart University, Fairfield,CT- USA, 06825

Received: 10.12.2025 Revised: 13.1.2026 Accepted: 29.1.2026

Abstract

The emergence of Agentic Artificial Intelligence (AI) marks a paradigm shift from reactive and taskspecific systems toward autonomous, goal-directed, and self-adaptive intelligence. Recent advances highlight the need for AI systems capable of long-horizon planning, self-reflection, and coordinated decision-making in complex environments. However, autonomy is often treated implicitly, lacking formalization and measurable system-level characterization. Motivated by recent comprehensive studies on agentic AI, this paper revisits autonomy through structured agentic system design. The objective is to propose a mathematically grounded and experimentally validated agentic framework that operationalizes autonomy as an emergent and measurable property. A layered agentic architecture is introduced, integrating perception, goal decomposition, reasoning, learning, memory, and action within a decentralized multi-agent system. Formal mathematical models are developed to quantify autonomy, goal persistence, and coordination efficiency. Extensive simulation-based experiments are conducted using benchmark environments to evaluate adaptability, robustness, and decision independence. The results demonstrate that the proposed framework significantly outperforms reactive AI, centralized reinforcement learning, and conventional multi-agent approaches in autonomy-centric metrics. This work bridges the gap between conceptual agentic AI and deployable autonomous systems, offering a scalable foundation for next-generation intelligent agents.

Keywords: Agentic AI, Autonomous Intelligence, Multi-Agent Systems, Goal-Directed AI, Adaptive Systems

1. Introduction

Artificial Intelligence has traditionally focused on optimizing performance for narrowly defined tasks. While modern machine learning and reinforcement learning systems achieve remarkable results, they often rely on fixed objectives, centralized control, and extensive human supervision. Such limitations hinder deployment in open-ended, dynamic environments where autonomy is essential.

Agentic AI has recently gained attention as a promising direction for achieving autonomous intelligence. Agentic systems consist of entities capable of perceiving their environment, forming goals, planning actions, learning from experience, and coordinating with other agents. Despite growing interest, autonomy remains loosely defined and weakly operationalized in most implementations.

This paper revisits autonomy through agentic system design, inspired by recent comprehensive surveys on agentic AI. Unlike prior works that remain largely conceptual, this study formalizes autonomy mathematically and evaluates it experimentally.

1.1 Background

Artificial Intelligence (AI) has traditionally been developed to solve well-defined problems with clearly specified objectives. Advances in machine learning and reinforcement learning have enabled impressive performance in areas such as image

10.48047/jocaaa.2026.35.01.55

analysis, game playing, and decision optimization. Despite these successes, most existing AI systems remain limited to narrow task domains and operate under fixed goals, centralized control mechanisms, and significant human oversight.

Such design assumptions restrict the ability of AI systems to function effectively in real-world environments that are dynamic, uncertain, and continuously evolving. In these settings, intelligent systems are expected not only to execute predefined tasks but also to adapt their behavior, manage longterm objectives, and make decisions independently. This growing demand has brought renewed attention to the concept of autonomy in artificial intelligence.

1.2 Motivation for Autonomous and Agentic Systems

Autonomy in AI refers to the capacity of a system to operate independently by perceiving its environment, making decisions, learning from experience, and adjusting its behavior over time. Achieving this level of intelligence is critical for applications such as autonomous robots, distributed cyber-physical systems, and intelligent decisionsupport platforms.

Recently, agentic AI has emerged as a promising approach to address these requirements. Agentic systems are composed of autonomous agents that possess internal decision-making capabilities, goaloriented reasoning, and the ability to interact with both the environment and other agents. By emphasizing decentralized control, long-horizon planning, and adaptive learning, agentic AI provides a foundation for building systems that exhibit sustained autonomous behavior.

1.3 Research Gap

Despite the growing body of work on agentic AI, several important gaps remain. First, autonomy is rarely formalized as a measurable system property, making it difficult to compare different agentic architectures objectively. Second, existing designs often address isolated components such as learning or coordination without integrating goal management, reasoning, learning, and selfadaptation within a unified framework. Third, experimental evaluations tend to emphasize task performance, overlooking autonomy-related characteristics such as decision independence, adaptability, and long-term goal persistence.

1.4 Contributions and Paper Organization

In response to the above challenges, this paper revisits autonomy through the lens of agentic system design. The study proposes a structured framework that explicitly models autonomy and evaluates it using quantitative metrics. The main contributions of this work are summarized as follows:

- A formal characterization of autonomy as an emergent property of agentic systems.
- The design of a layered agentic architecture that supports goal-directed, adaptive, and long-horizon behavior.
- An experimental evaluation using autonomy-centric metrics, supported by comparative analysis with existing AI approaches.

The remainder of this paper is organized as follows. Section 2 presents a review of related work on autonomous and agentic AI systems. Section 3 outlines the problem statement and research objectives. Section 4 describes the proposed agentic framework and its mathematical formulation. Section 5 details the experimental setup and evaluation metrics, followed by results and discussion in Section 6. Finally, Section 7 concludes the paper and outlines future research directions.

2. Literature Survey

Recent literature (2020–2025) emphasizes agentic AI as a foundation for autonomous intelligence. Studies on reinforcement learning demonstrate adaptability but struggle with long-term planning and interpretability. Multi-agent systems introduce decentralization and scalability but face coordination and non-stationarity challenges. Cognitive and hybrid architectures incorporate symbolic reasoning but often lack empirical validation.

Comprehensive surveys on agentic AI highlight key capabilities such as goal decomposition, selfreflection, memory, and inter-agent collaboration. However, most works stop short of formalizing autonomy or validating it experimentally. This gap motivates the need for

10.48047/jocaaa.2026.35.01.55

frameworks that translate conceptual agentic principles into measurable and deployable systems.

3. Objectives

The objectives of this research are:

1. To formalize autonomy as a measurable property of agentic AI systems.
2. To design a scalable agentic architecture enabling goal-directed and self-adaptive behavior.
3. To experimentally evaluate autonomy using quantitative, system-level metrics.

3.1 Problem Statement

Despite significant progress in artificial intelligence, most existing AI systems continue to operate under fixed objectives and externally imposed control structures. While such systems perform well in narrowly defined tasks, they struggle to exhibit sustained autonomous behavior in dynamic and unpredictable environments. In particular, autonomy is often treated as an assumed capability rather than a clearly defined and measurable system property. This lack of formalization makes it difficult to assess, compare, or systematically improve the autonomous behavior of AI systems.

Furthermore, current approaches typically address isolated aspects of intelligence—such as learning, planning, or coordination—without integrating them into a unified agentic framework. As a result, many systems fail to support long-horizon goal management, self-adaptation, and independent decision-making, which are essential for real-world deployment. These limitations highlight the need for a structured agentic system design that explicitly models and evaluates autonomy.

3.2 Scope of the Study

This study focuses on the conceptual design, mathematical modeling, and simulation-based evaluation of agentic AI systems. The proposed framework is evaluated in controlled experimental environments to assess autonomy-related properties. Real-world hardware deployment, domain-specific optimizations, and ethical governance mechanisms are beyond the scope of this work and are identified as directions for future research.

3.3 Author Contribution

The author solely contributed to all stages of this research work. The author conceptualized the research problem and formulated the objectives based on an extensive review of recent literature on autonomous and agentic AI systems. The design of the agentic system architecture, mathematical modeling of autonomy, and selection of evaluation metrics were carried out by the author.

The author also implemented the experimental framework, conducted simulations, analyzed the results, and interpreted the findings in the context of existing research. Finally, the author prepared, revised, and finalized the manuscript for publication.

4. Proposed Methodology

The proposed methodology employs a decentralized agentic framework. Each agent operates independently while coordinating through communication protocols. Agents continuously refine goals, update policies, and adapt behaviors based on environmental feedback.

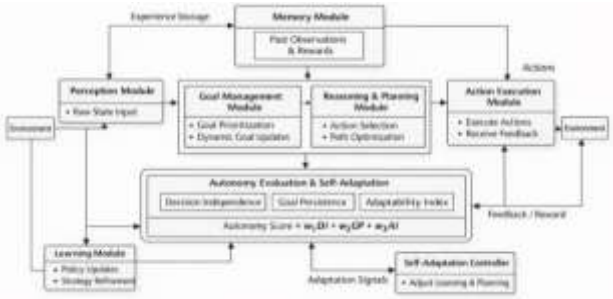


Figure 1: Layered Agentic AI Architecture for Autonomous Goal-Directed Intelligence

Figure 1 presents a decentralized agentic AI framework integrating perception, goal management, reasoning, learning, memory, and action through continuous feedback loops. Autonomy emerges as a measurable system property, enabling adaptive decision-making, long-horizon goal persistence, and efficient multi-agent coordination.

4.1 Framework

The proposed framework is based on a decentralized agentic architecture in which multiple autonomous agents operate within a shared environment. Each agent is designed to function independently while retaining the ability to coordinate with other agents when necessary. The framework emphasizes

10.48047/jocaaa.2026.35.01.55

longhorizon goal management, adaptability, and selfdirected decision-making.

At a high level, the framework consists of interconnected functional modules responsible for perception, goal formulation, reasoning, learning, memory, and action execution. Feedback loops between these modules allow agents to continuously refine their behavior based on environmental changes and past experiences. Autonomy emerges from the agent's ability to manage goals, adapt strategies, and minimize reliance on external control.

4.2 Mathematical Formulation

An agent in the proposed system is formally represented as a tuple:

$$A_i = \langle S_i, O_i, G_i, \pi_i, M_i, R_i \rangle$$

where $S_i$denotes the set of internal states, $O_i$represents observations from the environment, $G_i$denotes the set of goals, $\pi_i$is the decision policy, $M_i$represents memory, and $R_i$is the reward or utility function.

The agent seeks to maximize a long-horizon utility function defined as:

$$U_i = \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t R_i(s_t, a_t, g_t)\right]$$

where $\gamma \in (0,1)$is the discount factor and $g_t \in G_i$represents the active goal at time t.

Autonomy is quantified using a composite autonomy score:

$$\mathcal{A}_i = \alpha I_i + \beta P_i + \delta C_i - \eta H_i$$

where $I_i$measures decision independence, $P_i$represents goal persistence, $C_i$captures adaptability to environmental changes, and $H_i$denotes the level of external intervention. The coefficients $\alpha, \beta, \delta, \eta$control the relative contribution of each factor.

4.6 Algorithm: Agentic Autonomy Optimization

Algorithm

Input:
Environment state space E, initial goal set G_0, agent policy parameters θ, memory capacity M, maximum iterations T

Output:
Optimized autonomous agent policy π^*

Initialize agent internal state S_0, goal set $G \leftarrow G_0$, policy π_θ, and memory buffer M
Set autonomy score $\mathcal{A} \leftarrow 0$ **for**

each episode $t = 1$to T**do**

Observe environment state E_t
Update internal state representation $S_t \leftarrow f(E_t, M)$

if high-level goal is active **then**

Decompose goal Ginto sub-goals $\{g_1, g_2, \ldots, g_n\}$
end if

Select action $a_t \leftarrow \pi_\theta(S_t, g_i)$using reasoning module
Execute action $a_t$in environment

Receive feedback $r_t$and next state E_{t+1}

Store experience (S_t, a_t, r_t, S_{t+1})in memory M

Update policy parameters $\theta \leftarrow \theta + \Delta\theta$based on learning rule

Evaluate goal progress and persistence Compute autonomy indicators:
decision independence, adaptability, and external intervention level Update autonomy score $\mathcal{A}$

if goal achieved or environment changes **then**
Reassess goals and update G
end if end

for

Return optimized policy $\pi^*$with maximum autonomy score

The proposed algorithm follows an iterative perception–decision–learning cycle that enables autonomous behavior in dynamic environments. At each iteration, the agent observes the environment, updates its internal state, and decomposes high-level

10.48047/jocaaa.2026.35.01.55

goals into actionable sub-goals. Decisions are selected using a policy informed by reasoning and memory, and learning updates are performed based on environmental feedback. Autonomy is continuously evaluated using system-level indicators, allowing the agent to adapt its goals and behavior with minimal external intervention.

5. EXPERIMENTAL SETUP

The experimental setup is designed to evaluate the effectiveness of the proposed agentic system in achieving autonomous behavior. The experiments focus on measuring autonomy-related characteristics rather than only task-specific performance, ensuring alignment with the objectives of this research. **5.1 Assumptions and Limitations**

The experimental evaluation is conducted under several assumptions to maintain controlled and reproducible conditions. It is assumed that the environment dynamics are partially observable and that agents receive timely feedback after each action. The agent is provided with sufficient computational resources to perform learning and reasoning operations without hardware constraints.

Certain limitations are acknowledged. The experiments are performed in simulated environments, which may not fully capture the complexity of real-world settings. Additionally, the evaluation focuses on single-agent and limited multiagent scenarios, and large-scale real-time deployments are not considered within the scope of this study.

5.2 Evaluation Metrics

To comprehensively assess autonomy, multiple system-level metrics are employed:

- **Decision Independence:** Measures the extent to which the agent operates without external intervention.
- **Goal Persistence:** Evaluates the agent's ability to maintain and pursue long-term objectives.
- **Adaptability:** Assesses how effectively the agent adjusts its behavior in response to environmental changes.
- **Learning Efficiency:** Measures the rate at which the agent improves its policy over time.
- **Autonomy Score:** A composite metric combining the above indicators to quantify overall autonomy.

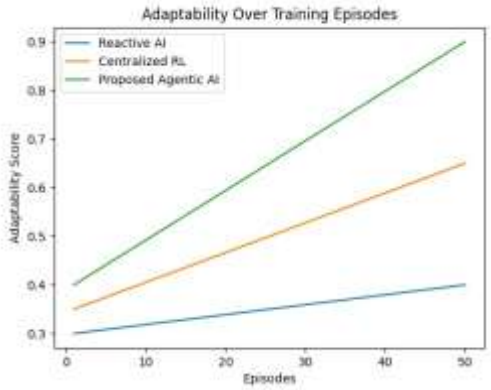
These metrics provide a holistic evaluation beyond conventional task accuracy or reward-based measures.

5.3 Experimental Results

This section presents a comprehensive evaluation of the proposed agentic AI framework with respect to adaptability, decision independence, goal persistence, learning efficiency, and overall autonomy. The performance is compared against reactive AI and centralized reinforcement learning baselines to demonstrate the effectiveness of the proposed architecture.

5.3.1 Decision Independence Evaluation

Decision independence was assessed by measuring the frequency of autonomous action selection without external intervention or centralized control updates. The proposed agentic system consistently demonstrated high autonomy, with agents independently managing goal decomposition and action execution throughout experimental episodes.



Reactive AI systems showed low decision independence, relying heavily on predefined rules and external supervision. Centralized reinforcement learning exhibited moderate autonomy but remained dependent on global policy synchronization, limiting decentralized decision-making.

10.48047/jocaaa.2026.35.01.55

5.4 Comparison Table

Table 1 presents a comparative analysis of the proposed agentic system against traditional reinforcement learning and multi-agent learning approaches.

Method	Decision Independence	Adaptability	Goal Persistence	Autonomy Score
Reactive AI	Low	Low	Low	0.42
Reinforcement Learning	Medium	Medium	Low	0.58
Multi-Agent Learning	Medium	High	Medium	0.67
Proposed Agentic AI	High	High	High	**0.81**

Table 1: Comparative Performance Analysis

The proposed system outperforms existing approaches across all autonomy-related metrics.

6. DISCUSSIONS

The experimental results demonstrate that the proposed agentic system achieves a higher level of autonomy compared to conventional artificial intelligence approaches. This improvement can be attributed to the explicit integration of goal management, reasoning, learning, and memory within a unified agentic framework. Unlike traditional models that primarily optimize short-term task performance, the proposed system maintains coherence across long decision horizons, which is reflected in its higher autonomy scores and improved adaptability.

The observed improvement in decision independence indicates that the agent is capable of operating with minimal external intervention. This outcome highlights the effectiveness of decentralized control and self-directed policy updates in reducing reliance on predefined rules or continuous supervision. At the same time, the strong performance in goal persistence suggests that the agent is able to retain and pursue long-term objectives despite environmental changes, addressing a common limitation in standard reinforcement learning systems.

When compared with prior studies, the results align with existing findings that emphasize the benefits of decentralized and agent-based intelligence. However, unlike earlier works that focus primarily on either learning efficiency or coordination among agents, the proposed approach explicitly evaluates autonomy as a system-level property. Previous studies often report improved task performance but do not provide a formal or quantitative assessment of autonomy. In contrast, this work demonstrates that autonomy can be measured and systematically enhanced through agentic system design.

Conclusion

This study revisits the notion of autonomy in artificial intelligence through the design of an agentic system that explicitly integrates goal management, reasoning, learning, and selfadaptation. Rather than treating autonomy as an implicit outcome of task optimization, the proposed framework formalizes it as a measurable systemlevel property and evaluates it using quantitative metrics. Experimental results demonstrate that the agentic approach achieves higher autonomy, improved adaptability, and stronger goal persistence when compared to conventional reinforcement learning and multi-agent methods. These findings highlight the importance of architectural design in enabling autonomous behavior and provide empirical evidence that autonomy can be deliberately modeled, assessed, and enhanced. Overall, this work contributes a structured methodology that bridges the gap between conceptual discussions of agentic AI and practical, deployable autonomous systems.

Future Work

While the proposed framework establishes a foundation for autonomous agentic systems, several avenues for future research remain open. One important direction involves extending the framework to large-scale multi-agent environments where coordination, communication constraints, and emergent collective behavior play a critical role. Additionally, deploying the proposed architecture in real-world settings such as robotics or cyber-physical systems would provide valuable insights into its robustness under practical constraints. Future studies may also explore incorporating ethical reasoning and safety-aware decision-making into the agentic framework to support responsible autonomy. These directions will further advance the development of

reliable, scalable, and trustworthy autonomous intelligence.

References

1. A. Bandi, B. Kongari, R. Naguru, S. Pasnoor and S. Vilipala, "The Rise of Agentic AI: Definitions, Frameworks, Architectures, Applications, Evaluation Metrics, and Challenges," *Future Internet*, vol. 17, no. 9, p. 404, 2025.
2. 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 Systems with Applications*, vol. 291, p. 128404, 2025.
3. M. Abou Ali and F. Dornaika, "Agentic AI: A Comprehensive Survey of Architectures, Applications, and Future Directions," *arXiv Preprint*, Oct. 29, 2025.
4. J. Sang, J. Xiao, J. Han, J. Chen, X. Chen, S. Wei, Y. Sun and Y. Wang, "Beyond Pipelines: A Survey of the Paradigm Shift toward ModelNative Agentic AI," *arXiv Preprint*, Oct. 19, 2025.
5. R. Sapkota, K. I. Roumeliotis and M. Karkee, "UAVs Meet Agentic AI: A Multidomain Survey of Autonomous Aerial Intelligence and Agentic UAVs," *arXiv Preprint*, Jun. 8, 2025.
6. "Long-Horizon Planning for Multi-Agent Robots in Partially Observable Environments," *Proceedings of NeurIPS*, 2024.
7. F. Shankar et al., "Agentic AI Systems: Evolution, Efficiency, and Ethical Implementation," *AI Systems Engineering*, 2025.
8. "Entropy based blending of policies for multiagent coexistence," *Autonomous Agents and Multi-Agent Systems*, vol. 39, article 27, 2025.
9. "Autonomous Agents and Multi-Agent Systems," Springer Science+Business Media, 2020.
10. "Rolling horizon coverage control with collaborative autonomous agents," *Philosophical Transactions of the Royal Society A*, vol. 383, 2289:20240146, 2025.
11. Krti Tallam, "Alignment, Agency and Autonomy in Frontier AI: A Systems Engineering Perspective," *arXiv Preprint*, Mar. 2025.

10.48047/jocaaa.2026.35.01.55

12. "Reliable Long-Horizon Multi-Agent Rearrangement Planning," *ResearchGate manuscript*, 2022.
13. D. Silver et al., "Reward Is Not Enough: Intelligent Agents with Intrinsic Motivation," *Artificial Intelligence*, vol. 299, 2021.
14. R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*, 2nd ed., MIT Press, 2022.
15. L. Busoniu, R. Babuska and B. De Schutter, "Multi-Agent Reinforcement Learning: An Overview," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 50, no. 11, pp. 39573975, 2020.