Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Self-Regulating AI Agents: A Runtime Constitutional Framework for Autonomous Decision Systems in Cloud-Native Environments

Harvendra Singh[1] and Subba Rao Katragadda[2]
[1]*Publix Super Markets Inc, Florida, USA*
[2]*Independent researcher, California, USA*
Corresponding Author: Subba Rao Katragadda, **E-mail**: subbakatragadda@gmail.com

| ABSTRACT
The increasing trend of deploying autonomous AI agents in cloud-native environments has enabled the automation of real-time and large-scale decision processes in enterprise and industrial systems. Nevertheless, the existing governance and alignment mechanisms remain external to the operational cycle of the agents. The current mechanisms rely on static policies and offline validation or infrastructure-level controls. Such mechanisms are not effective for agents that dynamically plan and collaborate and change their behavior at runtime. This paper proposes a runtime constitutional framework for self-regulating autonomous decision systems deployed in cloud-native environments. The framework incorporates machine-understandable governing principles into the operational cycle of AI agents and enables the monitoring, contextualization, and correction of every decision made by the agents before their execution. The framework is designed as an architectural structure comprising a constitutional rule layer, contextual state observation, decision interception, constitutional reasoning, and self-correction and adaptation. This makes the framework an independent mechanism of governance. Unlike existing alignment and policy enforcement mechanisms, the framework focuses on the regulation of the behavior of the agents at runtime and not at the training or deployment stages. This paper proposes an architectural and methodological contribution that offers a scalable and platform-independent solution for the runtime governance of autonomous decision systems. The proposed framework supports the development of dynamic cloud infrastructures and multi-agent systems, offering a practical solution for the development of reliable and human-aligned autonomous decision systems.

| KEYWORDS

Autonomous AI agents, Runtime governance, Constitutional AI framework, Cloud-native decision systems, Human–AI collaboration

| ARTICLE INFORMATION

ACCEPTED: 12 January 2026 **PUBLISHED:** 11 February 2026 **DOI:** 10.32996/jcsts.2026.8.4.3

1. Introduction

In the context of the intelligent enterprise, cloud-native platforms are increasingly recognized as the standard environment for the execution of such systems, providing the potential for elastic deployment, event-driven orchestration, and the large-scale integration of artificial intelligence into decision processes. For example, in areas such as digital manufacturing, supply chain management, and enterprise service automation, autonomous AI agents are increasingly being relied upon to plan and coordinate workflows, call software tools, coordinate with other agents, and dynamically adjust to changing operational contexts. For organizations with complex cyber physical and cloud-based infrastructures, such developments hold the potential for significant improvements in responsiveness, scalability, and efficiency [1][2].

At the same time, the very features of cloud-native environments that make such environments so attractive for the deployment and execution of intelligent enterprise systems—such as the distributed nature of the environment, the presence of multi-

tenancy, the use of continuous deployment, and the real-time nature of the events being generated—also pose significant challenges for governance and safety. For instance, the autonomous agents must be able to reason and act in the face of incomplete and changing information, coordinate with heterogeneous software services, and continuously adjust their internal plans in real-time while executing tasks. Existing approaches to governance and safety largely remain external to the decision-making processes of the agents, typically depending on approaches such as access control models, compliance-oriented approaches to the management of the infrastructure, and the pre-execution validation of models and workflows [3][4].

The primary issue that is being explored by this research is the lack of self-regulation mechanisms that are integrated into the execution process of autonomous AI agents. The recent focus on model alignment, prompt engineering, and human-in-the-loop oversight is a good example, but not much attention is being devoted to the execution process that enforces behavioral and operational constraints on decision-making by agents that have been deployed [5].

This is a good example of a gap that needs to be filled by a runtime governance framework that goes beyond just rules and post-execution auditing. Specifically, there are three aspects that have not received much attention: monitoring agent decisions, evaluating these decisions against operational principles, and correcting decisions that are not compliant with operational principles.

To address this limitation, this study aims to provide a runtime-based constitution for autonomous decision-making agents that are deployed within cloud environments. The constitution would establish a set of structured, machine-understandable principles that are implemented as part of the execution loop of an autonomous decision-making system, referred to as a runtime constitution. The main contributions of this study are architectural and methodological, providing a generic blueprint for self-regulating autonomous agents that can operate on various heterogeneous cloud infrastructures. By providing agents with real-time evaluation and adaptation of their execution, this study aims to provide a foundation for trustworthy autonomy within industrial-grade digital operations, where reliability, accountability, and human-centric decision-making are of utmost importance.

2. Literature Review

2.1 Autonomous AI agents in cloud environments

Recent research on autonomous AI agents has primarily focused on task planning, tool invocation, and collaborative problem solving in distributed and cloud-based systems. Agent-oriented architectures enable intelligent components to decompose complex goals into executable subtasks, coordinate with other agents, and dynamically select software services during execution. In cloud-native environments, these agents are commonly embedded within microservice ecosystems and event-driven workflows, allowing them to respond to real-time triggers and scale across distributed infrastructures. However, most existing agent frameworks emphasize execution efficiency, interoperability, and orchestration performance. Governance and behavioral control are typically implemented as external services or supervisory layers, rather than being integrated into the internal decision processes of the agents themselves. As a result, agents remain largely unaware of higher-level organizational, ethical, or operational constraints during reasoning and action selection [6].

2.2 AI governance and policy enforcement

The academic literature on AI governance has developed policy-based enforcement mechanisms, compliance engines, and risk management layers to govern the application of intelligent systems. They are normally based on business rules, regulations, and access controls defined during the infrastructure/application interface layer. Even though such mechanisms are effective in governing data access compliance and service authorizations, they do not manage the internal plan formation of an autonomous agent, tool selection, and objective prioritization. As such, governance is often separated from the cognitive process of an autonomous agent, thereby limiting the ability to intervene even when the decision is logically sound but operationally incorrect or misaligned [7].

2.3 Self-adaptive and self-healing systems

The autonomic and self-adaptive computing paradigms have recently introduced new architectural styles that monitor system well-being, detect unusual system behaviors, and trigger appropriate counteractions. These approaches are mainly centered on issues related to system infrastructure and service levels, such as resource allocation, fault recovery, load balancing, and performance optimization. These are important foundational concepts for cloud resilience, although they are not applicable to controlling the decision-making process of AI agents. The adaptation mechanisms are not necessarily related to the semantic correctness, safety, and policy conformance of autonomous decisions [8].

2.4 Constitutional and alignment-based AI concepts

Alignment-focused research and constitutional methods support the idea of using a set of rules to control the behavior of artificial intelligence models. These methods usually attempt to limit the model's output during training or usage based on a set of ethical or safety guidelines. These methods have been found to be effective in controlling the linguistic or generative behavior of the model but are limited to the model's operation and do not consider the operation of the autonomous agent with other external tools or agents [9].

2.5 Identified research gap

Throughout these works, there is a discernible lack of a runtime mechanism that integrates constitutional constraints with continuous monitoring and corrective intervention in the runtime loop of autonomous agents. Rather, the current solutions address these as separate concerns, which limits their efficacy in complex cloud-native decision environments. This lack is the motivation for the creation of a unified runtime constitutional framework that is capable of regulating the behavior of autonomous agents during runtime rather than before or after runtime.

3. Methodology – Proposed Runtime Constitutional Framework

This section specifies the proposed runtime constitutional framework that is expected to facilitate self-regulation in autonomous AI agents that operate in cloud-native environments. The approach is architectural and execution-centric, focusing on the integration of behavioral constraints in the runtime decision-making processes of agents, irrespective of their learning models or task domains.

3.1 Overall design philosophy

The framework has been conceptualized as an additional runtime layer, which is modular and platform-independent and works in conjunction with existing agent execution engines. Its primary purpose is the control of autonomous behavior in execution, as opposed to the control of model output or deployment configurations. The framework assumes the capability of agents to produce candidate actions such as tool actions, workflow actions, service actions, and inter-agent communications. It does not affect the learning or reasoning logic of the agents but adds an additional loop of continuous supervisory control over all externally visible decision processes.

The framework has been designed to be independent of machine learning architectures and hence supports single agents, multiple agents, and orchestrated agents as typically deployed in the cloud.

3.2 Framework architecture

The runtime constitutional framework has five tightly integrated functional elements:

The Constitutional Rule Layer retains a structured and machine-understandable representation of the operational rules, which include organizational rules, safety rules, business rules, and domain boundaries. The rules govern the permissible behavior of the agents at runtime.

The Context and State observer monitor the execution context continuously, which includes the state of the system, the stage in the workflow, the intent of the users, the environment, and the interaction history.

The Decision Monitor intercepts all high-level actions performed by the agents before they are executed. The high-level actions include service calls, use of tools, task delegation, revision of plans, and coordination with other agents.

The Constitutional Reasoning Engine evaluates the decision being made against the active constitutional rules and the prevailing execution context. It determines whether the decision is compliant, risky, in conflict, or potentially harmful in relation to the governing constitutional rules.

The Self-Correction and Adaptation Module incorporate corrective action in response to the violation or risk detected in the decision-making process. This could involve seeking alternative decisions from the agents, restricting the range of possible decisions, changing the execution context, or triggering escalation to human intervention.

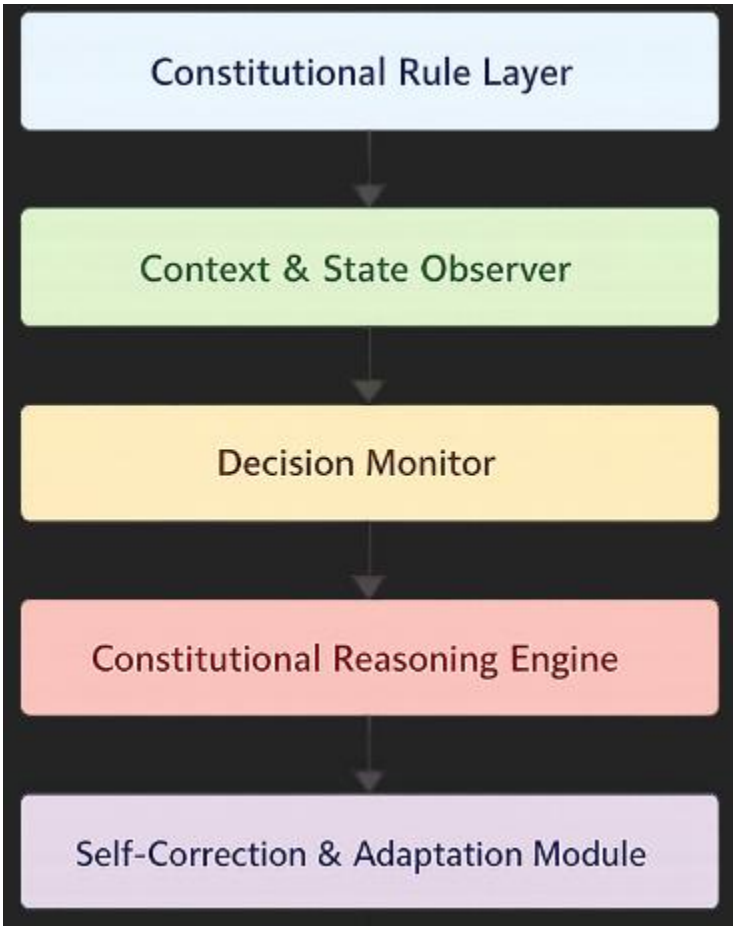

Fig 1. constitutional architecture for self-regulating autonomous AI agents

The figure shows a five-layer constitutional architecture for the runtime for self-regulating autonomous AI agents in cloud-native environments. Here, the Constitutional Rule Layer specifies the operational, safety, and business rules that the AI agents must follow; the Context and State Observer continuously monitor the conditions; the Decision Monitor intercepts each high-level action of the AI agents before execution; the Constitutional Reasoning Engine evaluates each decision based on the active rules and context; and the Self-Correction and Adaptation Module takes corrective action if necessary. All these elements work together to create a closed loop for the runtime governance of the AI agents' behavior.

3.3 Runtime execution flow

The runtime process operates based on a closed-loop execution paradigm. First, the agent generates a candidate action based on its internal planning and reasoning mechanisms. The decision monitor monitors the generated action and sends it to the constitutional reasoning engine for analysis. The engine evaluates the action based on the constitutional rule layer and the prevailing contextual state. If the action is compliant with the rules, it is approved for execution. Otherwise, the self-correction module takes corrective action, seeking a revised action from the agent or executing mitigation strategies before proceeding to the execution stage.

This iterative loop continues throughout the entire lifecycle of the agent's operation.

3.4 Learning and evolution of rules

The framework allows the dynamic modification of the constitutional rules in real time. New constraints can be added to the model to accommodate changing objectives, regulations, or organizational interests that may develop during runtime. The historical execution traces and the results of the interventions can be used to continually optimize the thresholds and resolution of the conflicts, thus allowing the progressive improvement of the effectiveness of the governance model without the need to retrain the original agent models.

3.5 Deployment in cloud-native environments

The framework is intended to be deployed as containerized services in event-driven and microservice-oriented architectures. Scalability of the framework is achieved through the distribution of the monitoring and reasoning components to multiple execution nodes. The low-latency interception mechanisms allow for the evaluation of the constitution without affecting the responsiveness of the operations, making the approach suitable for large-scale real-time autonomous decision systems.

4. Discussion

4.1 Impact on trust and reliability

One of the key consequences of the proposed runtime constitutional framework is the prospect of improving trust in the decision-making autonomy of systems that operate within a cloud-native environment. This is achieved by shifting the focus of decisions from the opaque result of agent reasoning towards a process that can be validated against a set of operational principles that are continually assessed during the runtime process. This has significant benefits in complex digital operations such as industrial platforms, smart manufacturing systems, and enterprise workflow automation systems, where the reliability of the system must be maintained as agents operate autonomously and continually adapt their decision-making strategies in real time [10].

4.2 Alignment with Industry 4.0 and Industry 5.0

The framework inherently addresses the operational needs of Industry 4.0, which is marked by constant interaction with cyber-physical systems, digital twins, and cloud-based analytics. More importantly, the framework moves towards the human-centric vision of Industry 5.0 by facilitating controlled autonomy. The self-correction and escalation mechanisms ensure the ongoing involvement of humans in the governance chain of high-risk or ambiguous situations. This offers a foundation for human-AI collaboration, where the agent operates autonomously to make decisions on routine tasks while ensuring alignment with human intent [11].

4.3 Operational benefits

From the perspective of operations, the framework helps to prevent cascading failures due to wrong or insufficiently contextualized actions of agents, particularly in the context of multiple agents. It also facilitates the safe orchestration of autonomous services, improves compliance through dynamic policy enforcement, and facilitates more stable coordination of distributed workflows. The distinction between decision generation and constitutional validation also facilitates the management of governance, allowing the updating of policies without the redeployment or retraining of the agents.

4.4 Limitations

Although the approach has its benefits, it also requires additional computational cost due to the monitoring and reasoning processes. It is still a difficult task to develop complete and non-conflicting constitutional rules, especially in the context of changing enterprise environments. There could be conflicts in business goals, safety constraints, and efficiency goals, requiring high-level resolution strategies. Excessive intervention could also limit the autonomy of the agents if the governance rules are too restrictive.

4.5 Future research directions

Future research should also investigate the discovery and refinement of constitutional rules with automated means, the incorporation of reasoning mechanisms that can explain their interventions, and collective governance for multi-agent systems. These additional features are anticipated to improve the scalability of self-regulating autonomous agents even further.

5. Conclusion

This research proposes a runtime constitutional framework with the aim of facilitating self-regulating autonomous AI agents in cloud-native systems. Unlike the majority of prevailing AI alignment and governance approaches, which mostly operate during the design phase, model-based, and/or between different infrastructures, the suggested framework extends the execution of the intelligent agents to incorporate a rule of law layer, continuous context monitoring, decision interception, and self-correction. This allows the agents to assess and correct their own actions.

The main novelty of the research is the creation of a methodological and architectural basis for trust in autonomous systems in large-scale digital systems. This is achieved without requiring changes to the learning models and in the presence of dynamic execution conditions, dynamic organizational goals, and diverse cloud infrastructures. This makes the framework a bridge

between static rule systems and the operational needs of autonomous decision systems in real-world enterprises and industrial systems.

By facilitating continuous governance while allowing human intervention, the framework offers a path towards trustworthy, transparent, and human-centric autonomous operation, which can be scaled up to support the next wave of cloud-based intelligent systems and human-AI collaboration.

References

1. Bhatt M, Rosario RF Del, Narajala VS, Habler I (2025) COALESCE: Economic and Security Dynamics of Skill-Based Task Outsourcing Among Team of Autonomous LLM Agents. ArXiv abs/2506.01900:
2. Subba Rao Katragadda. (2026). AI-Driven Resilient Supply Chain Architectures: Machine Learning Frameworks for Risk Anticipation, Disruption Mitigation, and Adaptive Decision-Making. *Journal of Business and Management Studies*, *8*(3), 38-50. https://doi.org/10.32996/jbms.2026.8.3.4
3. Ito H, Ito C (2024) Case Study of Human Resources Development for AI Risk Management Using RCModel. NEC Technical Journal 17:
4. Tejaskumar Vaidya. (2025). Enhancing Supply Chain Resilience through SAP APO and S/4 HANA Integrated Planning Frameworks. *Journal of Economics, Finance and Accounting Studies* , *7*(4), 32-41. https://doi.org/10.32996/jefas.2025.7.4.3
5. Liu K, Miao J, Liao Z, et al (2023) Dynamic constraint and objective generation approach for real-time train rescheduling model under human-computer interaction. High-speed Railway 1:. https://doi.org/10.1016/j.hspr.2023.10.002
6. Agbemabiese WT (2026) Toward Constitutional Autonomy in AI Systems: A Theoretical Framework for Aligned Agentic Intelligence. IEEE Access. https://doi.org/10.1109/ACCESS.2026.3654907
7. Schneider J, Abraham R, Meske C, Vom Brocke J (2023) Artificial Intelligence Governance For Businesses. Information Systems Management 40:. https://doi.org/10.1080/10580530.2022.2085825
8. Li J, Zhang M, Li N, et al (2024) Generative AI for Self-Adaptive Systems: State of the Art and Research Roadmap. ACM Transactions on Autonomous and Adaptive Systems 19:. https://doi.org/10.1145/3686803
9. Ji J, Qiu T, Chen B, et al (2025) AI Alignment: A Contemporary Survey. ACM Comput Surv 58:. https://doi.org/10.1145/3770749
10. Subba Rao Katragadda. (2026). Utilizing LLM models for advanced automation, manufacturing operations. *Journal of Mechanical, Civil and Industrial Engineering*, *7*(2), 08-14. https://doi.org/10.32996/jmcie.2026.7.2.1
11. Kosna, Srinivas Reddy. "AI-Driven IoT integration in smart healthcare Systems: A comprehensive framework for enhanced patient care and clinical decision support." *EPJ Web of Conferences*. Vol. 341. EDP Sciences, 2025. https://doi.org/10.1051/epjconf/202534101012