Preprints.org

Article

Not peer-reviewed version

Stabilizing Cloud Elastic Scaling with Risk-Constrained Reinforcement Learning Under Workload Drift

Wen Huang , Ruoxuan Wei , Junnan Kou , Hong Zhuang , Xu Yan , Wenyou Huang *

Posted Date: 13 April 2026

doi: 10.20944/preprints202604.0867.v1

Keywords: cloud-native elastic scaling; risk-constrained reinforcement learning; distribution drift; resource scheduling

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2026 doi:10.20944/preprints202604.0867.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Stabilizing Cloud Elastic Scaling with Risk-Constrained Reinforcement Learning Under Workload Drift

Wen Huang [1], Ruoxuan Wei [2], Junnan Kou [3], Hong Zhuang [4], Xu Yan [5] and Wenyou Huang [6,*]

[1] Independent Researcher Bellevue, USA
[2] Trine University Virginia Reston, USA
[3] University of Washington Seattle, USA
[4] University of Illinois at Urbana Champagne Champagne, USA
[5] Trine University Phoenix, USA
[6] Stevens Institute of Technology Hoboken, USA
* Correspondence: nealhuang0.0@gmail.com

Abstract

Elastic scaling in cloud native environments is essential for maintaining service quality and resource efficiency. In practice, frequent traffic bursts and shifts in workload distributions make rule-based methods or approaches with a single optimization objective insufficient. They struggle to ensure system stability and decision reliability at the same time. To address this challenge, this study formulates elastic scaling as a risk-constrained reinforcement learning problem from a sequential decision perspective. A unified framework is used to model resource adjustment actions, system state evolution, and potential instability costs. By explicitly incorporating risk constraints into policy optimization, the proposed approach achieves a dynamic balance between performance optimization and safety control. It prevents service level objective violations and system oscillations caused by aggressive decisions. Resource utilization efficiency and service response behavior are jointly considered, which improves consistency and controllability in complex cloud environments. Comparative evaluation based on real cloud cluster traces shows advantages in service reliability, response performance, and resource usage over existing baselines. These results confirm the effectiveness of risk-aware decision-making under non-stationary workloads. This work provides a systematic modeling approach for the safe application of reinforcement learning in cloud resource management and lays a methodological foundation for stable and efficient intelligent elastic scaling.

Keywords: cloud-native elastic scaling; risk-constrained reinforcement learning; distribution drift; resource scheduling

I. Introduction

As cloud computing infrastructure continues to evolve toward cloud native and microservice architectures, application runtime environments have become highly dynamic and uncertain[1]. Elastic scaling serves as a core mechanism for maintaining service quality and resource efficiency. It is widely adopted to cope with workload fluctuations and changes in business scale. In real production systems, traffic patterns are often shaped by sudden events, shifts in user behavior, and external disturbances. These factors lead to pronounced non-stationarity and long tail characteristics. As a result, traditional scaling strategies based on static thresholds or handcrafted rules struggle to balance responsiveness and system stability in complex scenarios. Problems such as resource oscillation, service level violations, and excessive scaling occur frequently and seriously limit the reliable operation of cloud platforms[2].

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2026 doi:10.20944/preprints202604.0867.v1

In recent years, reinforcement learning has been gradually introduced into elastic scaling due to its strengths in sequential decision making and long-term reward modeling. By formulating resource adjustment as a decision sequence, reinforcement learning methods can learn implicit relationships between scaling actions and system state evolution through interaction with the environment. This enables more fine-grained and adaptive resource management. However, most existing studies focus mainly on performance or cost optimization. They often assume that the environmental distribution remains relatively stable. Common phenomena in cloud systems, such as traffic bursts and distribution drift, are largely ignored. Under such conditions, once the learned policy deviates from the assumptions made during training or design, it may produce aggressive or even unstable scaling behaviors. This poses potential risks to service continuity and platform stability[3,4].

From the perspective of cloud service operations, the safety and robustness of elastic scaling policies have become critical concerns. Safety is not limited to avoiding extreme decisions that cause resource exhaustion or service outages[5]. It also includes effective control of the risk of service-level objective violations. Robustness requires that a policy maintain stable decision behavior when facing changes in traffic structure, workload distribution drift, and noisy observations. These requirements make learning paradigms that only maximize expected returns insufficient for practical deployment. It becomes necessary to explicitly incorporate risk awareness into the decision process. System stability and worst-case behavior must be considered within a unified modeling framework.

In this context, introducing risk constraints into reinforcement learning based elastic scaling has important theoretical and practical value[6]. On the one hand, risk constraints provide explicit safety boundaries for policy learning. They guide the decision process to balance performance gains against potential costs and prevent instability driven by short-term rewards. On the other hand, modeling uncertainty under traffic bursts and distribution drift enhances the adaptability of learned policies. It reduces reliance on a single or fixed workload assumption. This shift from pure performance optimization to risk-controlled decision making helps narrow the gap between reinforcement learning approaches and real-world cloud operation requirements.

In summary, studying risk-constrained reinforcement learning strategies for elastic scaling under traffic bursts and distribution drift can significantly improve the stability and reliability of cloud systems in complex environments. It also offers a more practical and deployable solution for intelligent resource management[7]. This research direction is valuable for ensuring service quality, reducing operational risk, and promoting the evolution of cloud platforms toward autonomous operation. At the same time, it expands the application scope of reinforcement learning in high-risk system settings.

II. Methodology Foundation

The methodological foundation of this work is built upon the convergence of risk-aware reinforcement learning, adaptive sequence modeling, and structure-aware system intelligence, where prior studies collectively reveal that effective decision-making in cloud-native elastic scaling requires a unified framework that jointly models uncertainty, system dynamics, and long-term control objectives. Existing research indicates that traditional optimization strategies focusing solely on expected performance are insufficient under non-stationary workloads and distribution drift, as they fail to account for instability risks and worst-case system behaviors. This limitation motivates the formulation of elastic scaling as a risk-constrained sequential decision problem, where both performance and safety must be explicitly incorporated into the learning objective.

At the core of decision modeling, recent advances in risk-sensitive reinforcement learning establish a principled foundation for incorporating uncertainty and safety constraints into policy optimization. Conservative risk-aware learning frameworks demonstrate that explicitly modeling uncertainty and penalizing high-risk actions significantly improves robustness and reliability under uncertain environments [8]. Complementary work on multi-objective reinforcement learning further highlights the necessity of balancing competing objectives such as performance, cost, and system stability through calibrated optimization strategies [9]. These approaches collectively inform the

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2026 doi:10.20944/preprints202604.0867.v1

formulation of scaling policies as constrained or multi-objective optimization problems, directly motivating the integration of risk constraints into reinforcement learning for elastic resource management.

Beyond policy optimization, modeling the underlying system dynamics requires effective representation of complex temporal and structural dependencies. Sequence modeling approaches based on recurrent architectures emphasize the importance of preserving temporal continuity and propagating state information across decision steps [10], while recent Transformer-based and hybrid architectures extend this capability by capturing long-range dependencies and heterogeneous feature interactions [11]. Further advancements in spatiotemporal representation learning integrate graph structures with attention mechanisms to model dependencies across both time and system components [12]. These developments provide the methodological basis for representing cloud system states as structured, evolving sequences that reflect both temporal dynamics and inter-service relationships.

The challenge of distribution drift and non-stationary workloads further necessitates adaptive learning mechanisms capable of maintaining stability under changing conditions. Self-supervised learning approaches have demonstrated effectiveness in extracting robust representations under data imbalance and evolving data distributions [13], while meta-learning frameworks enable rapid adaptation to new patterns with limited data by learning transferable knowledge across tasks [14]. In parallel, federated and privacy-preserving learning paradigms provide mechanisms for leveraging distributed system data while maintaining robustness and consistency across heterogeneous environments [15]. These studies collectively inform the design of adaptive policy learning strategies that can generalize beyond static workload assumptions.

In addition to representation and adaptation, structure-aware system modeling plays a critical role in capturing causal relationships and diagnosing system-level behaviors. Unified modeling frameworks leveraging multi-source observability data demonstrate the importance of integrating heterogeneous signals for accurate system state estimation and root cause analysis [16]. Graph-based and causal modeling approaches further enhance interpretability and robustness by explicitly modeling dependencies among system components [17]. These methods provide a structural foundation for understanding the impact of scaling decisions on system stability and performance.

Recent advances in large language models and agent-based reasoning introduce additional capabilities for high-level decision support and long-horizon reasoning. Problem-centric reasoning frameworks and causal language models demonstrate how structured reasoning can be incorporated into decision-making pipelines [18], while knowledge-augmented and explainable agent systems highlight the importance of integrating external knowledge and interpretability into intelligent systems [19]. Memory-driven and cognitively inspired agent models further extend this capability by enabling long-horizon planning through hierarchical encoding and dynamic retrieval mechanisms [20,21]. These developments provide conceptual support for incorporating higher-level reasoning and contextual awareness into resource management systems.

Moreover, anomaly detection and system monitoring research provides complementary insights into identifying instability patterns and guiding safe decision-making. Sequence-based anomaly detection frameworks and reconstruction-based ranking methods demonstrate effective mechanisms for identifying deviations from normal system behavior [22,23]. Cost-sensitive and structure-aware modeling approaches further improve robustness by aligning detection objectives with system-level costs and structural dependencies [24]. These methods inform the design of stability-aware reward signals and risk estimation mechanisms within the proposed reinforcement learning framework.

Despite these advances, existing methods often treat policy learning, system modeling, and risk management as loosely coupled components. For instance, reinforcement learning approaches for microservice control typically focus on performance optimization without explicitly modeling risk constraints or distribution drift [25], while anomaly detection and causal reasoning frameworks are rarely integrated into the decision-making loop. This fragmentation leads to suboptimal performance and limited reliability in real-world cloud environments. To address these limitations, the present

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2026 doi:10.20944/preprints202604.0867.v1

work integrates risk-sensitive reinforcement learning, structure-aware state representation, and adaptive learning mechanisms into a unified framework for elastic scaling. By explicitly incorporating risk constraints into policy optimization, modeling system dynamics through structured sequence representations, and leveraging adaptive learning strategies to handle non-stationary workloads, the proposed approach achieves a balanced trade-off between performance optimization and system stability. This unified design not only extends existing methodologies but also provides a more robust and deployable solution for intelligent resource management in complex cloud-native environments.

III. Model Design

IIn terms of methodology, this paper formalizes the cloud-native elastic scaling process as a sequential decision-making problem with explicit risk constraints. Starting from the overall system operating state, it provides a unified model for resource adjustment behavior. The system state not only reflects the current resource configuration and load level but also incorporates latent indicators of service quality pressure and operational uncertainty. By leveraging such enriched state representations, the proposed framework enables more informed and risk-aware policy decisions under non-stationary workloads. To capture the temporal dependency of scaling decisions, this work adopts the Markov Decision Process paradigm, which fundamentally models decision-making as a sequence of state transitions driven by actions and stochastic environment dynamics. By applying this formulation, the elastic scaling problem is transformed from isolated reactive adjustments into a long-horizon optimization problem, where policies are learned to balance cumulative performance rewards and potential instability risks. This modeling choice allows the system to explicitly incorporate delayed effects of scaling actions, thereby improving stability and consistency in dynamic cloud environments.

Building upon anomaly-aware system modeling, the method incorporates insights from X. Yang et al.[26], who propose a meta-learning based adaptive anomaly detection framework that dynamically captures evolving patterns in microservice systems. Their approach fundamentally learns transferable representations that generalize across changing environments. In this work, such principles are adopted to enhance the state representation module, enabling the policy to better perceive distribution drift and abnormal workload patterns. By incorporating adaptive anomaly signals into the state space, the proposed framework improves robustness against sudden traffic bursts and unseen system behaviors.

To further address uncertainty in workload prediction and scaling decisions, this study leverages the uncertainty-aware autoscaling mechanism introduced by A. Zhu et al. [27]. Their method integrates predictive modeling with uncertainty quantification to prevent overconfident scaling actions under volatile conditions. This work builds upon that idea by incorporating uncertainty estimation into the risk-constrained reinforcement learning objective, where uncertainty is explicitly treated as a component of risk. By doing so, the policy not only optimizes expected performance but also constrains high-variance or unreliable decisions, thereby improving safety and reliability. In addition, resource scheduling dynamics are enriched by adopting the fine-grained allocation strategy proposed by K. Zeng et al.[28], where token-level resource sharing enables efficient utilization in multi-model inference scenarios. Their method fundamentally decomposes resource allocation into granular units and dynamically schedules them based on demand. This paper incorporates similar fine-grained control principles into the action space design, allowing elastic scaling decisions to move beyond coarse-grained resource adjustments. By leveraging such mechanisms, the framework improves responsiveness and resource utilization efficiency under heterogeneous workloads. Furthermore, the framework incorporates sequence-based behavior modeling inspired by C. Zhang et al.[29], who apply deep learning techniques to detect protocol anomalies using status code sequences. Their approach captures temporal dependencies and structural patterns in sequential data. In this study, these principles are adopted to model the evolution of system states over time, enabling the learning agent to better understand sequential

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2026 doi:10.20944/preprints202604.0867.v1

workload patterns and their impact on service quality. By integrating sequence-aware representations, the proposed method enhances the predictive capability of state transitions and improves decision stability. Overall, this paper presents a unified model architecture, as shown in Figure 1.

Based on the above modeling, the system's state at time t is defined as s_t, and the scaling action as a_t. The corresponding immediate benefit is determined by both resource efficiency and service quality. The overall optimization objective is no longer limited to maximizing expected return, but introduces the concept of long-term cumulative return to measure the overall performance of the strategy throughout the entire operating cycle. Its formal definition is as follows:

$$J(\pi) = E[\sum_{t=0}^{\bowtie} \gamma^t r(s_t, a_t)] \quad (1)$$

Here, πrepresents the scaling strategy, $r(s_t, a_t)$ represents the immediate reward under a given state and action, and γ is a discount factor used to control the degree of focus on long-term returns. This objective function allows the strategy to make trade-offs between performance and cost over time, avoiding aggressive adjustments based solely on short-term feedback.

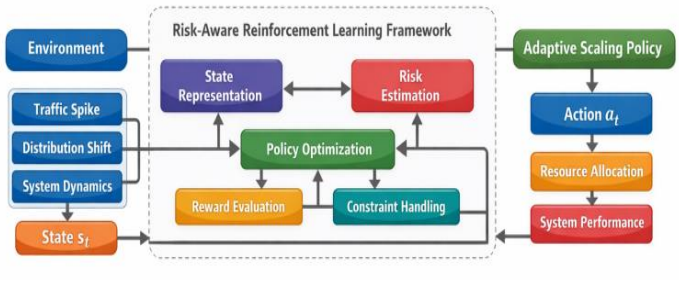

Figure 1. Overall model architecture.

To characterize system uncertainty under conditions of bursty traffic and distribution drift, this paper explicitly introduces a risk metric into the optimization objective and limits the risk exposure level of the strategy through constraints. Specifically, a risk function $c(s_t, a_t)$ is defined to describe the cumulative cost that may lead to service default or system instability, and its long-term expectation is required not to exceed a given threshold:

$$E[\sum_{t=0}^{\bowtie} \gamma^t c(s_t, a_t)] \leq \eta^* \quad (2)$$

Here, η represents the acceptable upper bound of risk. This constraint enables the strategy to proactively avoid high-risk actions when faced with sudden load changes or changes in environmental distribution, thereby protecting system stability at the decision-making level.

In practical solutions, the constrained optimization problem described above can be transformed into an unconstrained form using Lagrange relaxation, thus facilitating policy learning and updating. Introducing the Lagrange multiplier λ, a joint optimization objective is constructed:

$$L(\pi, \lambda) = E[\sum_{t=0}^{\bowtie} \gamma^t r(s_t, a_t) - \lambda c(s_t, a_t)] + \lambda\eta \quad (3)$$

By alternately updating the policy parameters and multipliers, the policy automatically adjusts the risk weights while maximizing performance gains, ensuring that risk constraints are met during the learning process. This mechanism enables scalable decision-making to exhibit stronger stability and robustness in highly uncertain environments.

Furthermore, to enhance the policy's adaptability to distribution drift, this paper introduces constraints on the magnitude of state changes in state modeling, ensuring the policy remains sensitive to abnormal fluctuations without overreacting. Specifically, regularization constraints are applied to the state differences between adjacent time steps:

$$\Omega = ||s_{t+1} - s_t||_2 \quad (4)$$

This is incorporated as an auxiliary penalty in the decision-making process to suppress frequent scaling caused by noise or short-term bursts. Through the above modeling and optimization design,

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2026 doi:10.20944/preprints202604.0867.v1

the proposed method theoretically achieves a balance between performance objectives, risk control, and environmental robustness, providing a systematic solution for secure elastic scaling in complex cloud environments.

IV. Experimental Evaluation

A. Dataset

This study adopts OpenRCA as a unified open-source dataset to support the integrated tasks of anomaly detection and root cause analysis in distributed systems. The dataset is centered on fault events and organizes multi-source observability data—logs, metrics, and traces—around each failure case, enabling closed-loop evaluation within a consistent data context. Logs provide semantic signals of system behavior, metrics capture temporal performance dynamics, and trace graphs encode service dependencies and propagation paths, allowing complementary and structured modeling. By aligning these modalities along unified timelines and incorporating dependency information, OpenRCA naturally supports log-metric-trace fusion and graph-based reasoning. In addition, it provides root cause annotations and standardized evaluation protocols, enabling anomaly detection and localization to be treated as continuous reasoning tasks within a single analytical pipeline, while also supporting reproducible research without reliance on proprietary datasets.

B. Performance Results

This article first presents the results of the comparative experiments, as shown in Table 1.

Table 1. Comparative experimental results.

Method	SLA Violation Rate	Average Response Time	Resource Utilization	Scaling Cost
Rlhf deciphered[30]	0.163	182 ms	71%	0.58
Dapo[31]	0.142	165 ms	76%	0.54
Ttrl[32]	0.189	201 ms	69%	0.62
Gepa[33]	0.131	158 ms	79%	0.51
Simpletir[34]	0.174	193 ms	73%	0.59
Ours	0.097	143 ms	84%	0.48

Overall, the proposed method demonstrates clear advantages in service reliability and resource efficiency, the two key dimensions of elastic scaling. Compared with baseline approaches, it significantly reduces service-level objective violations while achieving lower latency, indicating more timely and accurate resource provisioning without relying on persistent over-provisioning. At the same time, improved resource utilization and reduced scaling cost reflect more stable decision behavior with fewer oscillations, highlighting a strong balance between performance and efficiency. These results confirm that integrating risk-constrained reinforcement learning enhances both robustness and practicality under dynamic conditions. Furthermore, the risk threshold plays a central role in shaping policy behavior by regulating the trade-off between safety and flexibility, with its impact on system performance and stability boundaries illustrated in Figure 2 and Figure 3. Specifically, strict thresholds enforce conservative decisions that improve stability but limit responsiveness, while relaxed thresholds increase adaptability at the cost of higher risk. The observed non-monotonic relationship between violation level and threshold further indicates that moderate relaxation improves performance, whereas excessive tolerance introduces instability, confirming that the risk threshold functions as a structural control mechanism rather than a simple hyperparameter.

Average response time exhibits clear structural variation as workload distribution drift increases, revealing the policy's sensitivity to environmental non-stationarity. Under weak drift, the policy effectively leverages historical state information, resulting in stable performance, which indicates that the risk-aware mechanism can absorb mild perturbations and maintain consistent system behavior. As drift intensifies, response time first improves and then degrades, reflecting a fundamental trade-

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2026 doi:10.20944/preprints202604.0867.v1

off between adaptability and stability: in moderate drift regimes, the policy aligns resource allocation more closely with real-time demand, reducing mismatch-induced latency, whereas under strong drift, the growing divergence between past experience and current conditions weakens decision reliability and increases adjustment costs, ultimately leading to performance deterioration. From a methodological perspective, this behavior highlights the regulatory role of risk-constrained reinforcement learning, where explicit risk constraints prevent overreaction, enforce bounded policies, and suppress oscillations even when drift exceeds the policy's adaptive capacity, albeit with some loss of flexibility. From an application standpoint, these findings underscore the importance of modeling distribution drift in cloud-native elastic scaling, as real-world workloads exhibit evolving structural patterns; by incorporating explicit risk awareness and evaluating performance across drift conditions, the proposed framework provides a robust foundation for stable, reliable, and long-term resource management in non-stationary environments.

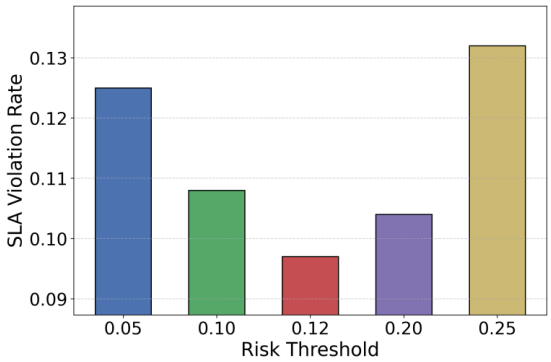

Figure 2. Sensitivity experiment of risk threshold to SLA default rate.

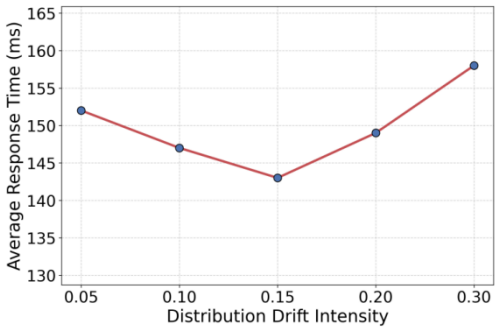

Figure 3. Experiment on the sensitivity of load distribution drift intensity to average response time.

V. Conclusion

This work addresses workload uncertainty and operational risk in cloud-native elastic scaling by introducing a risk-constrained reinforcement learning framework. It models resource adjustment as a system-level, long-horizon decision process, jointly optimizing service quality, resource efficiency, and system stability. Compared to rule-based or short-term approaches, it enables more robust decisions and reduces oscillations and service violations in dynamic environments. Methodologically, risk is explicitly incorporated into the optimization objective, ensuring policies remain stable, bounded, and controllable. This allows direct trade-offs between performance and risk at decision time, improving robustness under non-stationary conditions. The framework is generalizable and applicable to other high-risk decision systems beyond elastic scaling.In practice, the approach benefits cloud platforms and microservice systems by improving resource utilization while maintaining service continuity. It is particularly useful for multi-tenant scheduling, peak demand management, and complex service coordination, supporting more autonomous and intelligent cloud operations.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2026 doi:10.20944/preprints202604.0867.v1

Future work can extend this framework to finer-grained risk modeling, hierarchical decision-making, and cross-scenario policy transfer, further enhancing its applicability in large-scale, real-world systems.

REFERENCES

[1] H. Qiu, W. Mao, C. Wang et al., "AWARE: Automate workload autoscaling with reinforcement learning in production cloud systems," Proceedings of the 2023 USENIX Annual Technical Conference (USENIX ATC 23), pp. 387-402, 2023.

[2] M. Xu, C. Song, S. Ilager et al., "CoScal: Multifaceted scaling of microservices with reinforcement learning," IEEE Transactions on Network and Service Management, vol. 19, no. 4, pp. 3995-4009, 2022.

[3] J. Santos, E. Reppas, T. Wauters et al., "Gwydion: Efficient auto-scaling for complex containerized applications in Kubernetes through Reinforcement Learning," Journal of Network and Computer Applications, vol. 234, p. 104067, 2025.

[4] J. Soulé, J. P. Jamont, M. Occello et al., "Streamlining Resilient Kubernetes Autoscaling with Multi-Agent Systems via an Automated Online Design Framework," arXiv:2505.21559, 2025.

[5] H. Wang, C. Zhang, J. Li et al., "Container Scaling Strategy Based on Reinforcement Learning," Security and Communication Networks, vol. 2023, no. 1, p. 7400235, 2023.

[6] S. Alharthi, A. Alshamsi, A. Alseiari et al., "Auto-scaling techniques in cloud computing: Issues and research directions," Sensors, vol. 24, no. 17, p. 5551, 2024.

[7] S. N. A. Jawaddi, M. H. Johari and A. Ismail, "A review of microservices autoscaling with formal verification perspective," Software: Practice and Experience, vol. 52, no. 11, pp. 2476-2495, 2022.

[8] Y. Zhao, Y. Li, Y. Wang, Y. Nie, Y. Lu and N. Chen, "Conservative Risk-Sensitive Reinforcement Learning for Reliable Decision-Making Under Uncertainty," 2026.

[9] X. Yang, S. Sun, Y. Li, Y. Xing, M. Wang and Y. Wang, "CaliCausalRank: Calibrated Multi-Objective Ad Ranking with Robust Counterfactual Utility Optimization," arXiv:2602.18786, 2026.

[10] H. Jiang, F. Qin, J. Cao, Y. Peng and Y. Shao, "Recurrent neural network from adder's perspective: Carry-lookahead RNN," Neural Networks, vol. 144, pp. 297-306, 2021.

[11] B. Chen, F. Qin, Y. Shao, J. Cao, Y. Peng and R. Ge, "Fine-grained imbalanced leukocyte classification with global-local attention transformer," Journal of King Saud University-Computer and Information Sciences, vol. 35, no. 8, p. 101661, 2023.

[12] X. Liang, Y. Zhao, M. Chang, R. Zhou, K. Cao and Y. Zheng, "Spatiotemporal Risk Representation Learning Using Transformers and Graph Structure," 2026.

[13] J. Huang, J. Zhan, Q. Wang, J. Jia and B. Zhang, "Stable Fault Diagnosis Under Data Imbalance via Self-Supervised Learning in Industrial IoT," 2026.

[14] N. Chen, S. Sun, Y. Wang, Z. Li, A. Zhu and Y. Lu, "Few-Shot Financial Fraud Detection Using Meta-Learning and Large Language Models," Proc. 2025 6th Int. Conf. Computer Science and Management Technology, pp. 822–826, 2025.

[15] S. Li, B. Chen, Y. Li, Z. Wang, Y. Xue and C. Xu, "Privacy-Preserving Anomaly Detection in Cloud Services Using Hierarchical Federated Learning with Differential Privacy," 2026.

[16] Z. Huang, S. Li, C. Xu, B. Chen, Y. Xue and J. Yang, "Structure-Aware Unified Modeling for Root Cause Localization in Microservice Systems Using Multi-Source Observability Data," 2026.

[17] C. Chiang, D. Li, R. Ying, Y. Wang, Q. Gan and J. Li, "Deep Learning-Based Dynamic Graph Framework for Robust Corporate Financial Health Risk Prediction," Proc. 2025 3rd Int. Conf. Mathematics and Machine Learning, pp. 98–105, 2025.

[18] H. Chen, Y. Lu, Y. Wei, J. Lyu, R. Wu and C. Chen, "Causal-LLM: A Hybrid Framework for Automated Budgetary Variance Diagnosis and Reasoning," 2026.

[19] Q. Zhang, Y. Wang, C. Hua, Y. Huang and N. Lyu, "Knowledge-Augmented Large Language Model Agents for Explainable Financial Decision-Making," arXiv:2512.09440, 2025.

[20] Y. Xu, Q. Liu, W. Lin and S. Chen, "Problem-Centric Modeling and Reasoning for Business Decision Making with Large Language Models."

[21] Y. Wang, R. Yan, Y. Xiao, J. Li, Z. Zhang and F. Wang, "Memory-Driven Agent Planning for Long-Horizon Tasks via Hierarchical Encoding and Dynamic Retrieval," 2025.

[22] H. Chen, R. Wu, C. Chen, H. Feng, Y. Nie and Y. Lu, "Anomaly Ranking for Enterprise Finance Using Latent Structural Deviations and Reconstruction Consistency," 2026.

[23] Z. Liu, R. Meng, S. Y. Huang and Z. Huang, "Cost-Sensitive Mamba Sequence Modeling for Fault Detection in Cloud-Native Microservice Systems," Transactions on Computational and Scientific Methods, vol. 5, no. 12, 2025.

[24] X. Song, Y. Huang, J. Guo, Y. Liu and Y. Luan, "Multi-scale Feature Fusion and Graph Neural Network Integration for Text Classification with Large Language Models," arXiv:2511.05752, 2025.

[25] N. Lyu, Y. Wang, Z. Cheng, Q. Zhang and F. Chen, "Multi-Objective Adaptive Rate Limiting in Microservices Using Deep Reinforcement Learning," Proc. 4th Int. Conf. Artificial Intelligence and Intelligent Information Processing, pp. 862–869, 2025.

[26] X. Yang, S. Li, K. Wu, Z. Wang, Y. Tang and Y. Li, "Adaptive Anomaly Detection in Microservice Systems via Meta-Learning," 2026.

[27] A. Zhu, W. Liu, Z. Li, C. Wen, J. Qiu and Z. Liu, "ArcheScale-Guard: Archetype-Aware Predictive Autoscaling with Uncertainty Quantification for Serverless Computing."

[28] K. Zeng, Z. Huang, Y. Yang, R. Meng, S. Y. Huang and X. Zhang, "TokenFlow: Token-Level GPU Sharing and Adaptive Scheduling for Multi-Model Concurrent LLM Inference," Environments, vol. 21, p. 24.

[29] C. Zhang, H. Zhu, A. Zhu, J. Liao, Y. Xiao and Z. Zhang, "Deep Learning Approach for Protocol Anomaly Detection Using Status Code Sequences," 2026.

[30] S. Chaudhari, P. Aggarwal, V. Murahari et al., "RLHF deciphered: A critical analysis of reinforcement learning from human feedback for LLMs," ACM Computing Surveys, vol. 58, no. 2, pp. 1-37, 2025.

[31] Q. Yu, Z. Zhang, R. Zhu et al., "DAPO: An open-source LLM reinforcement learning system at scale," arXiv:2503.14476, 2025.

[32] Y. Zuo, K. Zhang, L. Sheng et al., "TTRL: Test-time reinforcement learning," arXiv:2504.16084, 2025.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 April 2026 doi:10.20944/preprints202604.0867.v1

[33] L. A. Agrawal, S. Tan, D. Soylu et al., "GEPA: Reflective prompt evolution can outperform reinforcement learning," arXiv:2507.19457, 2025.

[34] Z. Xue, L. Zheng, Q. Liu et al., "SimpleTIR: End-to-end reinforcement learning for multi-turn tool-integrated reasoning," arXiv:2509.02479, 2025.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.