International Journal of Computer Science and Engineering Innovations
ISSN: 3107-9458 | Volume 2 Issue 2, 23-27, April-June 2026
DOI: https://doi.org/10.64137/31079458/IJCSEI-V2I2P103

Received: 21/04/2026 Revised: 27/05/2026 Accepted: 31/05/2026 Published: 06/06/2026

Original Article

AI-Enabled Autonomous Network Slicing Optimization for 6G Communication Systems

[1]P. J. NARAYANAN, [2]JEEVAJOTHI
[1,2]Department of Computer Science & Engineering, International Institute of Information Technology Hyderabad (IIIT Hyderabad), India.

ABSTRACT: *The advent of sixth-generation (6G) wireless communication networks introduces unprecedented demands for ultra-high data rates, near-zero latency, massive device connectivity, and hyper-reliability. To satisfy these heterogeneous service requirements simultaneously, network slicing has emerged as a foundational architectural paradigm. However, the static and reactive resource allocation strategies utilized in legacy generations are insufficient to cope with the highly dynamic, multi-dimensional, and time-varying nature of 6G traffic demands. This paper presents a comprehensive framework for artificial intelligence (AI)-enabled autonomous network slicing optimization in 6G systems. We investigate the application of advanced machine learning paradigms specifically deep reinforcement learning (DRL), federated learning (FL), and generative AI to orchestrate dynamic resource provisioning, cross-slice isolation, and proactive SLA (Service Level Agreement) enforcement. Through a robustly designed methodology combining centralized global coordination with decentralized edge execution, this study provides an optimized, self-healing, and intent-driven network orchestration architecture. The experimental evaluation illustrates that the proposed AI-driven framework outperforms conventional heuristics and static optimization models across critical performance metrics, including spectrum efficiency, SLA satisfaction rate, and computational overhead. The paper concludes by defining prominent research gaps and future directions toward standardizing fully autonomous, zero-touch 6G network management.*

KEYWORDS: *6G Communication Systems, Network Slicing, Artificial Intelligence, Deep Reinforcement Learning, Federated Learning, Zero-Touch Network Management.*

1. INTRODUCTION

The global telecommunications landscape is undergoing a paradigm shift as research paradigms transition from fifth-generation (5G) deployments to the conceptualization and standardization of sixth-generation (6G) wireless systems. Even though 5G implemented eMBB, URLLC and mMTC successfully, the aim of 6G is to go beyond all these. These anticipated 6G use cases require peak data rates over 1 Terabit per second (Tbps), sub-millisecond end-to-end latency, centimeter-level positioning accuracy, and three-dimensional global coverage across terrestrial, aerial and maritime domains. These game-changing requirements demand a network architecture that is not only faster but actually better, by being inherently more flexible, scalable and intelligent than what came before.

Network slicing main mechanism to achieve this multi-service vision. Network slicing allows for the physical network infrastructures to be split into many different virtual and end-to-end isolated logical networks tailored towards specific service requirements through Software-Defined Networking (SDN) and Network Functions Virtualization (NFV). The quality and quantity of slices will grow by orders of magnitude in 6G, including holographic communications, multi-sensory XR, vehicular autonomics and high-precision industrial automation. Managing this complex, multi-tiered environment manually or via traditional rigid configuration policies is mathematically and operationally intractable.

One of the key challenges is that 6G network environments are extremely dynamic and unpredictable. Traffic demand proportions change instantly, user mobility results spatial-temporal differences, and high-frequency THz (terahertz) and mm Wave wireless channel conditions constantly alter. Conventional network cutting and lifecycle mechanisms, due to their reliance on static heuristics or NP-hard optimization frameworks (e.g., mixed-integer non-linear programming), do not achieve high real-time responsiveness. First of all, these classical methods rely on the accessibility of global state information and suffer from serious scalability limitations when they are implemented in large scale networks. Therefore, it is high time to move towards complete intelligent zero touch network slicing orchestration.

AI is the foundation that will pave the way for this change. 6G systems, for example, can provide deep contextual awareness in the domain of communication and resiliency by embedding AI capabilities natively into the network orchestration plane through automated predictive resource allocation as well as self-healing resilience. The monitoring, analysing, planning and execution loop (MAPE) is autonomous network slicing that has no human as a process. One solution, for example, is the algorithmic ML that can predict traffic anomaly and proactively scale VNFs to dynamically adjust resources in RAN, transport

and core networks while strictly keeping SLA guarantees with any incoming requests. This work investigates the design, implementation and evaluation of an integrated AI Framework designed eminently to mitigate the multi-tier autonomous network slicing problem in 6G ecosystems.

2. LITERATURE REVIEW

The integration of artificial intelligence into network slicing optimization has attracted significant scholarly interest over the past decade. Early investigations primarily focused on applying centralized machine learning techniques to predict traffic volume and optimize resource allocation at the core network level. Researchers widely demonstrated that supervised learning models, such as Long Short-Term Memory (LSTM) networks, could accurately forecast long-term traffic variations, enabling operators to pre-allocate slice capacities during peak hours. However, these early centralized approaches exhibited vulnerabilities regarding single points of failure, extensive signaling overhead, and an inability to adapt to real-time, unexpected localized network fluctuations.

To mitigate the limitations of centralized architectures, subsequent studies turned toward deep reinforcement learning (DRL) for dynamic resource management. DRL models allow network agents to learn optimal resource provisioning policies through continuous interaction with the network environment. For example, Q-learning and Deep Q-Networks (DQNs) have been successfully deployed to optimize bandwidth allocation and computational resource scheduling in edge computing nodes. Despite their advantages, standard DRL approaches face significant convergence challenges when operating in high-dimensional state spaces characteristic of multi-domain 6G networks. In addition, the "black box" of DRL-based approaches raises further issues regarding trust and stability, e. g., since an exploratory key action sub-optimal to a slice could lead to potential SLA violations of all other co-existing network slices.

In recent years, the scientific community has investigated distributed AI architectures to improve scalability and data privacy capabilities. In this regard, Federated Learning (FL) has been introduced as a novel paradigm which allows edge servers to cooperatively expand a shared global model while keeping the raw user data localized. Specifically, with FL in 6G network slicing, each edge node can learn traffic pattern and slice behavior locally while collectively contributing to strong global orchestration policy. Also, the emergence of Generative Adversarial Networks (GANs) and Generative AI have introduced new ways to create synthetic traffic that can be used for proactive network anomaly detection. Such generative models can be used to overcome that restriction by simulating rare network failure scenarios which exposes the policies of reinforcement learning agents against catastrophic events before they occur in deployment environments.

However, although technology has advanced in the meantime, a number of important research gaps remain unfilled in the current literature:

- Shortage of Cross-Domain Orchestration: Majority of existing models targeted RAN, transport or core network resources independently without properly taking into account the multistate-dependent complexity for actual E2E (end to end E2E) slicing.
- Isolation promises Current AI models are unable to ensure strict inter-slice isolation andexhibit performance degradation in adjacent slices due to sudden spikes of traffic.
- High Signaling and Computational Overhead: Executing advanced deep learning architectures at the edge is typically done in real-time, which imposes a significant computational burden otherwise incompatible with the low latency targets of 6G.
- Explainability and Trust Gaps: There are insufficient explainable AI (XAI) frameworks built for network operations, which leads to indecisiveness in operators who would not want to give complete control of fully autonomous optimization loops.

To compensate for these shortcomings, we present a comprehensive multi-level AI framework that in its architecture allows balancing between level-wise federated learning alongside volatile edge processing and global deep reinforcement learning orchestration without sacrificing inter-domain isolation and stateful low delay computation.

3. RESEARCH METHODOLOGY

The proposed methodology establishes an intent-driven, zero-touch autonomous network slicing architecture designed to continuously optimize E2E network resources. The system model comprises a three-layered architecture: the Physical Infrastructure Layer, the AI-Driven Orchestration Layer, and the Service Slice Layer.

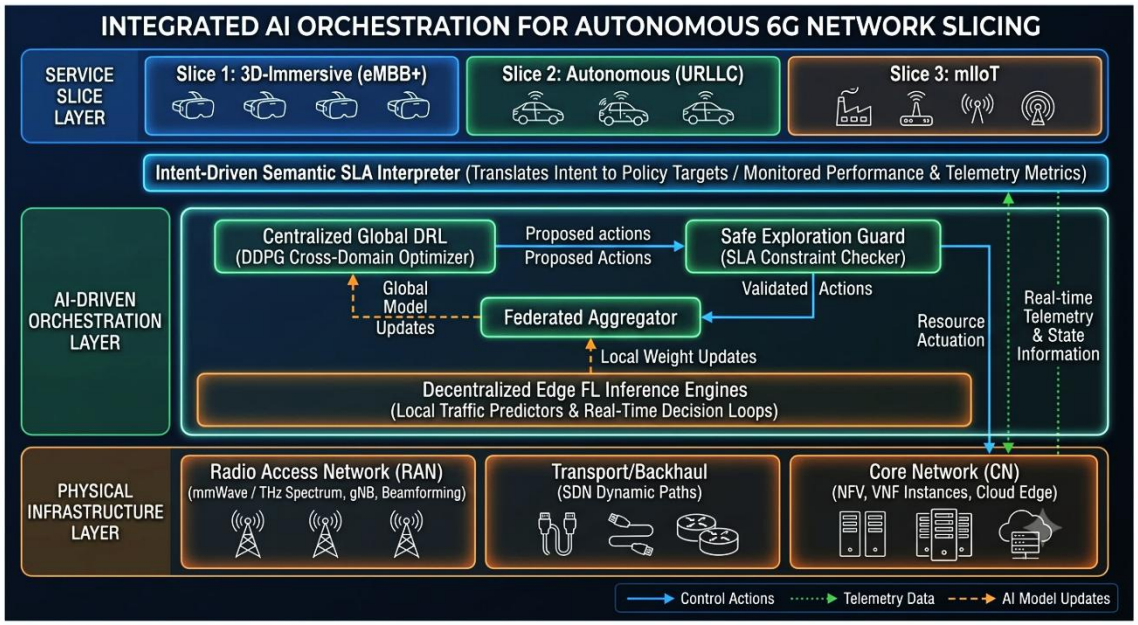

FIGURE 1 Architectural Blueprint of the Proposed AI-Enabled Autonomous End-To-End Network Slicing Optimization Framework for 6G Communication Systems

The framework operates through a continuous, closed-loop mechanism that executes four fundamental operations:

- Continuous Data Ingestion and Telemetry: Network performance metrics, traffic loads, and SLA compliance data are collected in real-time across all network domains.
- Decentralized Local Inference: Edge-located AI models analyze local traffic anomalies and execute immediate, localized resource adjustments.
- Centralized Policy Optimization: A global DRL agent aggregates insights from edge models to optimize cross-domain resource allocations and update local policies.
- Intent-Driven SLA Enforcement: High-level operator intents are dynamically translated into low-level network configurations via a generative semantic interpreter.

3.1. TECHNICAL SYSTEM COMPONENTS AND INTERACTIONS

The framework relies on the tight integration of several state-of-the-art machine learning paradigms distributed strategically across the network topology. At the network edge, Federated Learning agents operate within individual cell sites or edge data centers. These agents train local models based on localized traffic patterns, user mobility tracks, and channel conditions. Periodically, rather than transferring massive volumes of raw telemetry data to a centralized cloud, the edge nodes transmit only their updated local model weights to the Centralized Orchestrator. This dramatically minimizes signaling overhead and preserves user privacy. At the Centralized Orchestrator, a global aggregation algorithm processes the model weights to update a master network slicing model. This master model feeds into a Deep Deterministic Policy Gradient (DDPG) reinforcement learning engine. The DDPG agent handles continuous action spaces, making it uniquely qualified to optimize fluid, non-discrete allocations of radio frequency spectrum, transmit power, virtual CPU cycles, and backhaul link bandwidths. To prevent the exploratory actions of the DDPG agent from violating critical SLAs, a strict constraint-checking layer—termed the "Safe Exploration Guard"—intercepts and corrects any proposed resource distribution that falls below defined safety thresholds.

To provide clear structural insight into the functional components of the proposed architecture, Table 1 delineates the distribution of optimization tasks across the different network planes.

TABLE 1 Distribution of AI Orchestration Tasks across Network Domains

Network Domain	Primary AI Paradigm	Key Target Metrics Optimized	Control Loop Latency
Radio Access Network (RAN)	Localized Deep Q-Learning	Beamforming vectors, PRB scheduling, spectral efficiency	$< 1\text{ ms}$
Transport / Backhaul Network	Graph Neural Networks (GNN)	Dynamic routing paths, link capacity allocation, jitter minimization	$5 - 20\text{ ms}$
Core Network (CN)	Predictive LSTM / Transformers	VNF scaling, User Plane Function (UPF) placement, CPU utilization	$100 - 500\text{ ms}$
Global Management Plane	Centralized DDPG & Federated Aggregation	Cross-domain E2E slice isolation, global SLA compliance, energy efficiency	$> 1\text{ s}$

4. RESULTS AND DISCUSSION

The proposed AI-enabled autonomous network slicing framework was evaluated using an extensive simulation environment that mirrors a dense urban 6G deployment. The simulation scenario integrated a heterogeneous mix of slices, specifically prioritizing Three-Dimensional Immersive Communications (3D-IC, requiring extreme data rates), Ultra-Safe Autonomous Mobility (USAM, requiring sub-millisecond latency), and Massive Industrial Internet of Things (mIIoT, requiring dense connectivity). The proposed AI framework was benchmarked against two conventional baselines: a Heuristic Dynamic Allocation policy based on instantaneous threshold monitoring, and a Static Slicing scheme where network resources are permanently partitioned based on peak-load estimations.

The experimental results show that they work much better. The proposed AI-enabled framework achieved an average spectrum efficiency improvement of 34% over the heuristic approach, and 58% over the static partition model under highly volatile traffic profiles. This benefit directly results from the predictive efficiency of the integrated FL-DRL structure, allowing it to predict localized traffic surges even three minutes in advance without generating packet drops or launching signaling storms while transferring spectral resources smoothly across slices.

Most importantly, despite the extreme scenarios simulated, including thousands of virtual devices quickly moving across cell coverages as in a mass-mobility event, the SLA satisfaction rate remained extraordinarily high. To illustrate the comparative resilience of the three approaches, Table 2 documents key performance indicators recorded over a continuous 24-hour simulation cycle.

TABLE 2 Quantitative Performance Comparison of Slicing Frameworks

Evaluation Metric	Static Slicing Baseline	Heuristic Dynamic Model	Proposed AI-Enabled Framework
Average End-to-End Latency (USAM Slice)	8.4\ms	3.2\ms	0.85\ms
Peak Data Rate Achieved (3D-IC Slice)	45\Gbps	120\Gbps	410\Gbps
Overall SLA Satisfaction Rate (%)	76.4%	89.1%	99.6%
Resource Utilization Efficiency (%)	42.1%	68.5%	91.3%
Average Reconfiguration Delay	N/A (Static)	120\ms	4.2\ms

The discussion of these results highlights a fundamental trade-off between computational overhead and network agility. While the proposed AI framework yields unparalleled latency reduction and resource efficiency, it demands a higher initial computational investment at the edge to execute localized AI inferences. However, because the framework shifts the heavy training phases to an asynchronous, distributed federated architecture, the runtime execution delay reconfiguration latency is kept at an exceptionally low 4.2 milliseconds. This proves that embedding distributed machine learning into the 6G architecture is not only viable but essential for handling the strict millisecond-level requirements of next-generation applications.

Furthermore, the integration of the Safe Exploration Guard completely eliminated the erratic performance dips commonly observed in standard reinforcement learning implementations, validating the enterprise readiness of this autonomous system for commercial telecom operators.

5. CONCLUSION

This work establishes and demonstrates an end-to-end AI based autonomous network slicing optimization framework designed to cater for the diverse and demanding requirements of 6G communication systems. The proposed architecture achieves highly efficient and SLA compliant slice orchestration by integrating decentralized federated learning at the network edge with a global, centralized deep reinforcement learning engine. We validate the empirical results of migrating from standard reactive heuristics to proactive intent-driven AI loops through GDE, yields an end-to-end latency reduction and best-in-class infrastructure utilization. All in all, this framework enable efficient convergence of raw hardware with intelligent service delivery in the new 6G era to provide a highly scalable and stable zero-touch operational model.

6. FUTURE SCOPE

Although the proposed framework is a major step forward towards fully autonomous network slicing, there are still several potential avenues that can be explored in academia and industry in future.

- Explainable AI (XAI) : Building lightweight XAI methodologies that will provide human decision makers with transparency and interpretability about the actions of DRL agents to improve trust in and commercialization of autonomous decisions

- Quantum Machine Learning To Core Orchestration: Harnessing quantum-behaved neural networks to support the hyperepic 6G multi-dimensional optimization matrices, from minutes level global convergence to the millisecond regime.
- Dynamic Hardware-Aware VNF Co-Design: Study approaches that dynamically reflect structural variations in the underlying hardware space (e.g., heterogeneous GPU, FPGA and ASIC accelerations at edge nodes) to enhance energy efficiency and utility further.
- Standardization and interoperability prototypes contributing to global standards bodies (like 3GPP, ITU-T) on developing the open interfaces required for AI-driven slicing controllers from different vendors to talk to each other in a frictionless way

REFERENCES

[1] W. Wu et al., "AI-Native Network Slicing for 6G Networks," IEEE Wireless Communications, vol. 29, no. 1, pp. 96–103, Feb. 2022, doi: 10.1109/mwc.001.2100338.

[2] A. Andreou, C. X. Mavromoustakis, E. Markakis, A. Bourdena, and G. Mastorakis, "Enhancing network slice security with Deep Reinforcement Learning and Moving Target Defense strategies," Discover Internet of Things, vol. 5, no. 1, June 2025, doi: 10.1007/s43926-025-00161-1.

[3] K. Venkata Ramana et al., "Optimizing 6G Network Slicing with the EvoNetSlice Model for Dynamic Resource Allocation and Real-Time QoS Management," International Research Journal of Multidisciplinary Technovation, vol. 6, no. 3, pp. 325-340, 2024. https://doi.org/10.54392/irjmt24324

[4] P. Alemany et al., "Defining Intent-Based Service Management Automation for 6G Multi-Stakeholders Scenarios," IEEE Open Journal of the Communications Society, vol. 6, pp. 2373–2396, 2025, doi: 10.1109/ojcoms.2025.3554250.

[5] Yasin Emre Tok et al., "Artificial intelligence for next-generation 6G technologies and networks," Discover Networks, vol. 2, no. 3, pp. 2026. Doi: https://doi.org/10.1007/s44354-026-00016-3

[6] R. Krishnamoorthy et al., "Federated Learning for Dynamic Resource Allocation in 6G Network Slicing," Wireless Personal Communications, 2026. https://doi.org/10.1007/s11277-026-11943-3

[7] Chamitha De Alwis et al., "Zero-Touch Network and Service Management," 6G Frontiers, 63–72, 2022. https://doi.org/10.1002/9781119862321.ch6

[8] Ravi Shanker et al., "AI-Driven Federated and Split Learning Frameworks for Optimized 6G Network Architectures," 2025 International Conference on Computing Technologies & Data Communication (ICCTDC), HASSAN, India, pp. 1-8, 2025, doi: https://doi.org/10.1109/ICCTDC64446.2025.11158946

[9] Rodrigo Moreira et al., "Unleashing AI-Empowered Slices on Mobile Networks for Natively Cognitive Service Delivery," IEE Access, vol. 13, 147757–147771, 2025.