UNIVERSITY OF LEEDS

University of Sheffield

UNIVERSITY of York

Interpretable Attention-Based Multi-Agent PPO for Latency Spike Resolution in 6G RAN Slicing

Kavan Fatehi*, Mostafa Rahmani Ghourtani *, Amir Sonee †, Poonam Yadav *,
Alessandra M Russo †, Hamed Ahmadi *, Radu Calinescu*
*University of York, UK; †Imperial College London, UK.

Abstract—Sixth-generation (6G) radio access networks (RANs) must enforce strict service-level agreements (SLAs) for heterogeneous slices, yet sudden latency spikes remain difficult to diagnose and resolve with conventional deep reinforcement learning (DRL) or explainable RL (XRL). We propose *Attention-Enhanced Multi-Agent Proximal Policy Optimization (AE-MAPPO)*, which integrates six specialized attention mechanisms into multi-agent slice control and surfaces them as zero-cost, faithful explanations. The framework operates across O-RAN timescales with a three-phase strategy: predictive, reactive, and inter-slice optimization.

A URLLC case study shows AE-MAPPO resolves a latency spike in $18\,$ms, restores latency to $0.98\,$ms with 99.9999% reliability, and reduces troubleshooting time by 93% while maintaining eMBB and mMTC continuity. These results confirm AE-MAPPO's ability to combine SLA compliance with inherent interpretability, enabling trustworthy and real-time automation for 6G RAN slicing.

Index Terms—6G, RAN slicing, deep reinforcement learning, explainable AI, multi-agent systems, O-RAN, latency mitigation.

I. INTRODUCTION

Future 6G wireless networks are expected to support massive device connectivity while simultaneously accommodating services with heterogeneous and often conflicting performance requirements. Radio Access Network (RAN) slicing has emerged as a key architectural mechanism to address this challenge by enabling the logical partitioning of shared physical RAN infrastructure into multiple virtualised slices. Each slice can then be configured to satisfy specific service-level objectives, such as ultra-reliable low-latency communication (URLLC), enhanced mobile broadband (eMBB), or massive Internet of Things (mIoT) services [1]. By allocating resources on a per-slice basis, operators can flexibly meet heterogeneous demands. However, orchestrating multiple slices in dynamic environments remains challenging, as traffic patterns, user mobility, and radio conditions fluctuate rapidly, making it difficult to consistently meet service-level agreements (SLAs) [2].

To address these challenges, recent work has emphasized AI-driven automation, often termed zero-touch network and service management (ZSM). In particular, machine learning (ML) and deep reinforcement learning (DRL) have shown strong potential for adaptive and efficient slice management [3]. Several efforts have proposed AI-native slicing architectures that embed learning and optimization into the control loop. For example, Wu et al. [4] present an AI-native framework that integrates AI into both the design and operation of slicing, enabling intelligent orchestration and dynamic adaptation. Complementing this direction, the survey in [5] reviews the evolution of RAN toward B5G, comparing architectures across edge support, virtualization, control and management, energy efficiency, and AI integration. It also highlights both opportunities and challenges in applying deep learning to next-generation RANs.

Despite these advances, traditional AI models often behave as black boxes, providing little insight into their decision-making. In mission-critical scenarios with strict SLAs and safety requirements, this lack of transparency undermines trust and hinders deployment in operational networks. To address this limitation, the field of Explainable AI (XAI) develops methods to improve interpretability and trust in AI-driven solutions. A recent survey on XAI for AI-enabled Open RAN (O-RAN) [6] provides a comprehensive mapping of O-RAN research to XAI-supported solutions, discusses automated XAI pipelines for stable performance, and shows how explainability can strengthen security and trust. The study also identifies open challenges and research directions, underscoring the importance of transparent and trustworthy AI in 6G systems.

Building on this line of work, recent efforts have proposed explainable reinforcement learning (XRL) frameworks for slicing. In [7], a multi-agent XRL scheme employs interpretable reward functions, guided by SHAP importance values, to encourage SLA-aware and explainable decision-making. This approach achieves lower latency and fewer packet drops compared with conventional DRL while enhancing trust in agent behavior. Similarly, [8] introduces EXPLORA, a lightweight framework that generates network-oriented explanations by linking DRL agent actions to wireless context via attributed graphs. Implemented on a near-real-time RIC, EXPLORA demonstrates not only improved transparency but also measurable bitrate gains through intent-based action steering. Beyond slicing, explainability has also been applied to radio resource management (RRM) problems in wireless networks. For instance, the study in [9] presents case studies on vehicular

networks and massive MIMO beam alignment, proposing a generic explainability pipeline and a credibility assessment tool to enhance both scalability and trustworthiness of AI-enabled RRM solutions [1].

These contributions illustrate the promise of combining XAI and DRL for reliable and interpretable RAN slicing. However, they primarily target average SLA compliance and policy transparency, leaving open the challenge of mitigating short-term latency spikes sudden, severe tail-latency excursions that critically affect slice reliability. Unlike gradual SLA violations, spikes can occur unpredictably due to cross-slice interference, congestion surges, or mobility events, and they demand rapid, interpretable responses. Failure to resolve such spikes within milliseconds can compromise safety-critical applications such as industrial automation or emergency communications.

To tackle this challenge, we propose an Attention-based Multi-Agent Proximal Policy Optimization (**AE-MAPPO**) framework that embeds interpretability directly into the policy optimization process. In contrast to post-hoc explanation methods (e.g., SHAP-based XRL), our approach employs multiple specialized attention heads that operate during inference to highlight the underlying causes of latency spikes. These include semantic attention (identifying critical states such as buffer overflows), temporal attention (detecting recurring patterns in time), cross-slice attention (capturing inter-slice interference), and counterfactual attention (evaluating alternative actions). By integrating these mechanisms into a multi-agent PPO setup, the framework produces inherently interpretable decisions while maintaining high responsiveness.

Our main contributions are as follows:

1. **Interpretable AE-MAPPO framework:** We design an attention-enhanced multi-agent PPO architecture for RAN slicing that generates real-time explanations at zero additional computational cost.
2. **Latency spike resolution:** We formulate the problem of SLA violations caused by sudden latency spikes and develop a decision pipeline for rapid mitigation.
3. **Explainability mechanisms:** We introduce six attention modules (semantic, temporal, cross-slice, confidence, counterfactual, and a meta-controller) that provide actionable, human-readable insights into agent decisions.
4. **Case study and evaluation:** Using a realistic 6G RAN slicing scenario with URLLC, eMBB, and mMTC slices, we demonstrate that AE-MAPPO resolves latency spikes in under 20 ms, reduces troubleshooting time by 93%, and provides explanations that operators can use to anticipate and prevent recurring anomalies.

II. PROBLEM FORMULATION AND SYSTEM MODEL

We consider a 6G O-RAN system where a near-RT RIC orchestrates distributed units (DUs) to serve three canonical
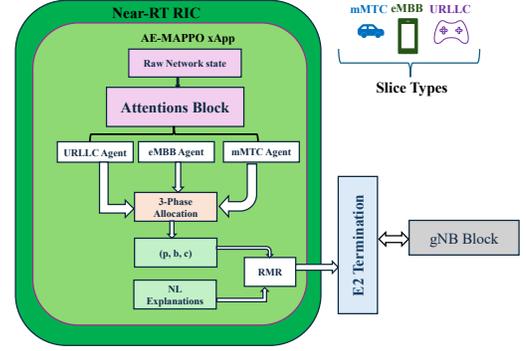

Fig. 1: AE-MAPPO architecture integrating six attention mechanisms with three-phase allocation strategy for explainable network slicing.

slices $\mathcal{N} = \{\text{URLLC}, \text{eMBB}, \text{mMTC}\}$. Each slice $n \in \mathcal{N}$ is associated with a QoS tuple $Q_n = \{\mathcal{L}_n, \mathcal{R}_n, \mathcal{T}_n, \mathcal{P}_n\}$ (latency, reliability, throughput, power). UE arrivals for slice n form non-stationary Poisson processes with rates $\lambda_n(t)$, and the active UE set is $\Omega_n(t)$. The near-RT RIC allocates per-interval power, bandwidth (PRBs), and edge computation to maximize end-to-end QoS under resource budgets.

A. System and QoS Targets

We adopt typical targets:

$$Q_{\text{URLLC}} = \{1\text{ms},\ 99.999\%,\ 50\text{Mbps},\ -\},$$
$$Q_{\text{eMBB}} = \{30\text{ms},\ 99\%,\ 1\text{Gbps},\ -\},$$
$$Q_{\text{mMTC}} = \{1\text{s},\ 99\%,\ 1\text{kbps},\ 1\text{mW}\}.$$

Let $\mathbf{p}_n = [p_i^n]_{i \in \Omega_n}$, $\mathbf{b}_n = [b_i^n]_{i \in \Omega_n}$, and $\mathbf{c}_n = [c_i^n]_{i \in \Omega_n}$, respectively, denote the transmit power, the number of PRB, and the computation per UE of a slice n.

B. Joint Resource Allocation

We propose a weighted single-objective optimization formulation that integrates slice utilities and an explainability utility. While existing network slicing formulations optimize performance metrics alone or employ multi-objective optimization [10], our formulation uniquely incorporates explainability directly into the objective function:

$$\max_{\{\mathbf{p},\mathbf{b},\mathbf{c}\}} \sum_{n \in \mathcal{N}} w_n U_n(\mathbf{p}_n, \mathbf{b}_n, \mathbf{c}_n) + w_{\text{xrl}} \mathcal{E}(\mathbf{p}, \mathbf{b}, \mathbf{c})$$

$$\text{s.t.} \quad \sum_n \sum_{i \in \Omega_n} p_i^n \leq P_{\text{total}} \quad \text{(C1)}$$

$$\sum_n \sum_{i \in \Omega_n} b_i^n \leq B_{\text{total}} \quad \text{(C2)}$$

$$\sum_n \sum_{i \in \Omega_n} c_i^n \leq C_{\text{total}} \quad \text{(C3)}$$

$$\Pr[\mathcal{L}_n^{\text{ach}} \leq \mathcal{L}_n^{\text{tar}}] \geq \mathcal{R}_n, \ \forall n \quad \text{(C4)}$$

$$\mathcal{T}_n^{\text{ach}} \geq \mathcal{T}_n^{\text{tar}}, \ \forall n \quad \text{(C5)},$$

(1)

where w_n and w_{xrl} represent weighting coefficients for slice utility U_n and explainability $\mathcal{E}$, respectively. This formulation differs from prior work by treating explainability not as a post-processing step but as an integral optimization objective, enabling the joint optimization of performance and interpretability. Slice utility U_n is considered as a weighted aggregation of QoS satisfaction, efficiency, and fairness:

$$U_n = \alpha_n U_n^{\text{QoS}} + \beta_n U_n^{\text{eff}} + \gamma_n U_n^{\text{fair}}, \quad \alpha_n + \beta_n + \gamma_n = 1, \quad (2)$$

with QoS satisfaction, efficiency, and fairness described below.

1) **QoS satisfaction:** Let $u_{\mathcal{X}}^n = \min\{1, \exp(-\lambda_{\mathcal{X}} v_{\mathcal{X}}^n)\}$ for $\mathcal{X} \in \{\mathcal{L}, \mathcal{R}, \mathcal{T}, \mathcal{P}\}$, with violations

$$v_{\mathcal{L}}^n = (\mathcal{L}_n^{\text{ach}} - \mathcal{L}_n^{\text{tar}})_+, \quad v_{\mathcal{R}}^n = (\mathcal{R}_n^{\text{tar}} - \mathcal{R}_n^{\text{ach}})_+,$$

$$v_{\mathcal{T}}^n = \frac{(\mathcal{T}_n^{\text{tar}} - \mathcal{T}_n^{\text{ach}})_+}{\mathcal{T}_n^{\text{tar}}}, \quad v_{\mathcal{P}}^n = \begin{cases} (\mathcal{P}_n^{\text{used}} - \mathcal{P}_n^{\text{tar}})_+, & \mathcal{P}_n^{\text{tar}} \neq \text{N/A} \\ 0, & \text{otherwise.} \end{cases}$$

The multiplicative form $U_n^{\text{QoS}} = \prod_{\mathcal{X}} u_{\mathcal{X}}^n$ drives the utility toward zero if *any* QoS dimension is violated, matching strict SLA semantics.

QoS resource relations. With SNR $\text{SNR}_n = \frac{\mathbf{p}_n g_n}{N_0 \mathbf{b}_n}$ and packet size L,

$$\mathcal{T}_n^{\text{ach}} = \mathbf{b}_n \log_2(1 + \text{SNR}_n), \quad (3)$$

$$\mathcal{L}_n^{\text{ach}} = \frac{L}{\mathcal{T}_n^{\text{ach}}} + \frac{1}{\mu \, \mathbf{c}_n}, \quad (4)$$

$$\mathcal{R}_n^{\text{ach}} = (1 - \text{BER}_n)^L \approx \left(1 - Q\left(\sqrt{2\,\text{SNR}_n}\right)\right)^L, \quad (5)$$

and $\mathcal{P}_n^{\text{used}} = \sum_{i \in \Omega_n} p_i^n + P_{\text{circuit}}$. These relations provide a PRB-level link abstraction that maps allocated PRBs, transmit power, and channel gains into achievable throughput, latency, reliability, and power consumption, enabling practical evaluation of resource allocation and slicing decisions under QoS constraints [11].

2) **Efficiency:** Given per-slice budgets $(P_n^{\text{alloc}}, B_n^{\text{alloc}}, C_n^{\text{alloc}})$, resource efficiency is:

$$U_n^{\text{eff}} = 1 - \frac{1}{3}\left(\frac{\sum_i p_i^n}{P_n^{\text{alloc}}} + \frac{\sum_i b_i^n}{B_n^{\text{alloc}}} + \frac{\sum_i c_i^n}{C_n^{\text{alloc}}}\right), \quad (6)$$

where each ratio represents utilization $\in [0, 1]$ for power, bandwidth, and computation. The formulation extends single-resource efficiency metrics to multi-dimensional resources, with $(1 - \text{utilization})$ ensuring $U_n^{\text{eff}} = 1$ for zero usage (maximum efficiency) and $U_n^{\text{eff}} = 0$ for full utilization, thus encouraging QoS satisfaction with minimal resource consumption.

3) **Fairness:** We measure intra-slice fairness using the Gini coefficient [12], [13], where $\text{Gini}(\mathbf{x}_n) \in [0, 1]$ quantifies inequality (0=perfect equality, 1=maximum inequality):

$$U_n^{\text{fair}} = \frac{1}{3}\Big[(1 - \text{Gini}(\mathbf{p}_n)) + (1 - \text{Gini}(\mathbf{b}_n)) + (1 - \text{Gini}(\mathbf{c}_n))\Big]. \quad (7)$$

The transformation $(1 - \text{Gini})$ converts inequality to fairness, where 1 represents perfect fairness. Equal weighting $(\frac{1}{3})$ across resource dimensions prevents fairness in one resource from masking inequality in others.

C. Explainability Utility

We integrate interpretability as a first-class objective, aligned with attention-driven decisions in AE-MAPPO. Let $\mathcal{A}_\pi(s) \in \mathbb{R}^d$ be the attention vector (state dimension d) produced by the policy π at state s. We define

$$\mathcal{E} = \eta_1 \mathcal{E}_{\text{sparse}} + \eta_2 \mathcal{E}_{\text{cons}} + \eta_3 \mathcal{E}_{\text{faith}}, \quad \eta_1 + \eta_2 + \eta_3 = 1, \quad (8)$$

with:

$$\mathcal{E}_{\text{sparse}} = 1 - \frac{H(\mathcal{A}_\pi)}{\log d} = 1 + \frac{\sum_i \alpha_i \log \alpha_i}{\log d}, \quad (9)$$

$$\mathcal{E}_{\text{cons}} = \mathbb{E}_{(s,s') \in \mathcal{S}_\epsilon}\left[\frac{\mathcal{A}_\pi(s) \cdot \mathcal{A}_\pi(s')}{\|\mathcal{A}_\pi(s)\| \, \|\mathcal{A}_\pi(s')\|}\right], \quad (10)$$

$$\mathcal{E}_{\text{faith}} = \text{Corr}\left(\mathcal{A}_\pi(s), \left|\frac{\partial U_{\text{total}}}{\partial s}\right|\right), \quad (11)$$

where $H(\cdot)$ is the entropy function, $\mathcal{S}_\epsilon$ collects ϵ-similar states, and $U_{\text{total}} = \sum_n w_n U_n$. We use $(\eta_1, \eta_2, \eta_3) = (0.3, 0.3, 0.4)$, prioritizing faithfulness while preserving focus and stability.

III. PROPOSED FRAMEWORK: AE-MAPPO

From optimization to RL. We cast the joint allocation problem in (1) into a multi-agent RL formulation, with one agent per slice type. The global state s_t stacks per-slice semantics (queue occupancy, load, SINR, headroom, predicted demand) together with near-RT RIC context. The joint action a_t specifies per-slice allocations $(\mathbf{p}, \mathbf{b}, \mathbf{c})$ drawn from a structured, slice-aware space with safety masks that prevent infeasible assignments. Coordination between agents is mediated through a shared multi-head attention module, which outputs attention weights $\mathcal{A}_\pi$ that both guide allocation and provide explanations. The per-step reward is

$$r_t = \sum_n w_n U_n(\mathbf{p}_n, \mathbf{b}_n, \mathbf{c}_n) + w_{\text{xrl}} \mathcal{E}(\mathcal{A}_\pi), \quad (12)$$

and PPO optimizes π_θ using clipped surrogate objectives and centralized value baselines. Since the same attention weights are used for action selection and surfaced to the operator, explanations are *faithful* and produced at *zero cost*, enabling near-real-time interpretation of latency spike mitigation.

Framework overview. AE-MAPPO integrates interpretability into the control loop via three key innovations: (i) six specialized attention mechanisms embedded in the policy, (ii) a three-phase allocation strategy aligned with O-RAN control loops, and (iii) a joint learning objective that couples performance and explainability. Figure 1 illustrates the complete architecture, showing how these components interact within the O-RAN ecosystem to generate both resource allocations and real-time explanations. Together, these make allocation policies transparent while maintaining responsiveness.

A. Interpretable Policy with Six Attentions

The policy network extends MAPPO with six attention heads $\boldsymbol{\alpha}(s) = \{\alpha_{\text{sem}}, \alpha_{\text{temp}}, \alpha_{\text{cross}}, \alpha_{\text{conf}}, \alpha_{\text{cf}}, \alpha_{\text{meta}}\}$:

- **Semantic attention** highlights QoS-critical features (e.g., URLLC buffer occupancy or SNR), clarifying *what drives* the decision.
- **Temporal attention** captures short-term and recurring patterns, enabling proactive pre-allocation and clarifying *why now*.
- **Cross-slice attention** quantifies inter-slice interference, exposing *who affects whom* and supporting controlled resource trades.
- **Confidence attention** measures entropy-based uncertainty in state features, regulating allocation aggressiveness when the agent is uncertain.
- **Counterfactual attention** compares chosen actions with alternatives, answering *what if* questions for operators.
- **Meta-controller** adaptively weights the other heads, producing context-aware fusion and balancing competing rationales.

These attention weights not only determine allocations but are also logged as explanations $\mathcal{A}_\pi$, allowing operators to trace which factors dominated each decision.

B. Three-Phase Allocation Across O-RAN Loops

AE-MAPPO operates across complementary timescales:

1) **Predictive phase (100 ms):** temporal and semantic heads forecast demand surges (e.g., recurring eMBB peaks) and pre-allocate extra bandwidth or compute.
2) **Reactive phase (10 ms):** confidence-aware adjustments correct sudden SLA violations within a TTI, ensuring URLLC deadlines are met.
3) **Inter-slice optimization (50 ms):** cross-slice trades reallocate surplus resources from less critical slices (e.g., mMTC) to latency-sensitive slices, validated against global utility gains.

This hierarchy ensures both long-horizon anticipation and short-horizon responsiveness, critical for resolving abrupt latency spikes.

C. Learning Objective

AE-MAPPO optimizes a joint loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{PPO}} + \alpha_{\text{xrl}}(\beta_1 \mathcal{L}_{\text{sparse}} + \beta_2 \mathcal{L}_{\text{cons}} + \beta_3 \mathcal{L}_{\text{faith}}) , \quad (13)$$

where $\mathcal{L}_{\text{PPO}}$ is the clipped surrogate loss, $\mathcal{L}_{\text{sparse}}$ enforces focused attentions, $\mathcal{L}_{\text{cons}}$ stabilizes explanations across similar states, and $\mathcal{L}_{\text{faith}}$ aligns attentions with utility gradients. Hard constraints are handled by action clipping and soft penalties in the reward.

Resulting properties. AE-MAPPO yields (i) feasible allocations under O-RAN timescales, (ii) faithful explanations surfaced directly from the decision process, and (iii) a tunable balance between performance and interpretability.

TABLE I: Resource allocation before and after spike resolution.

	URLLC	eMBB	mMTC
Power (before)	25%	45%	20%
Power (after)	42%	30%	18%
PRBs (before)	30%	50%	15%
PRBs (after)	55%	30%	15%
Latency (ms) before	1.15	12	95
Latency (ms) after	0.98	18	97

This enables rapid SLA enforcement under anomalies while keeping decisions auditable to operators.

IV. CASE STUDY: LATENCY SPIKE RESOLUTION

We evaluate AE-MAPPO on a real-time URLLC anomaly where a sudden latency spike occurs at $t = 14:23:15$. Traditional troubleshooting requires inspecting $15+$ potential causes, correlating logs, and trial-and-error fixes, taking on average 11.5 minutes. In contrast, AE-MAPPO resolves the spike in 0.8 minutes (93% faster) with interpretable explanations.

Attention-guided diagnosis. Semantic attention highlights a buffer overflow in URLLC (0.89), cross-slice attention reveals eMBB interference (0.76), and temporal attention identifies a recurring daily pattern near $14:20$. Counterfactual attention evaluates alternatives, rejecting "do nothing" and mMTC throttling, and selecting eMBB reduction as the optimal intervention.

Mitigation. Within 18 ms, AE-MAPPO reallocates resources: URLLC power share increases from 25% to 42%, eMBB decreases from 45% to 30%, and PRBs shift accordingly (Table I). URLLC latency drops from 1.15 ms to 0.98 ms, meeting the < 1 ms requirement with 99.9999% reliability, while eMBB maintains service continuity at reduced video quality.

V. PERFORMANCE EVALUATION

Table II summarizes performance during the URLLC emergency. AE-MAPPO completes its decision in 18 ms, well within the 25 ms near-RT control window, and maintains all SLA requirements: URLLC latency below 1 ms, reliability of 99.9999%, and service continuity for eMBB and mMTC. Explanations are generated instantly at inference-time, with no additional cost.

These results confirm AE-MAPPO ability to rapidly diagnose and mitigate latency spikes while generating faithful, human-readable explanations. Compared with manual troubleshooting, the framework reduces resolution time by 93%, ensures SLA compliance, and enables preventive control through interpretable attention mechanisms.

To evaluate AE-MAPPO's ability to maintain high performance while providing interpretability, we present two complementary analyses that demonstrate the architectural advantages of integrated attention mechanisms over both black-box and post-hoc approaches.

TABLE II: Performance metrics during emergency response.

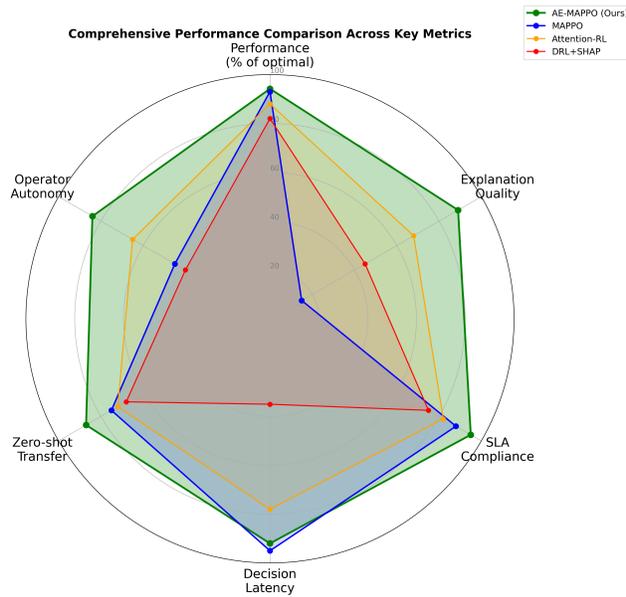| Metric	Target	Achieved
URLLC Latency	< 1 ms	0.98 ms
URLLC Reliability	99.9999%	99.9999%
Decision Time	< 25 ms	18 ms
eMBB Continuity	$> 95\%$	100%
mMTC Connections	Maintain	Maintained
Explanation	Required	Complete

Fig. 2: Multi-dimensional performance comparison across six critical metrics. AE-MAPPO (green) achieves the largest coverage area (86%), demonstrating consistent superiority across all dimensions while maintaining explainability.

Figure 2 evaluates performance across six operational metrics essential for 6G deployment. AE-MAPPO achieves 86% of theoretical maximum coverage, exceeding MAPPO (64%), Attention-RL (68%), and DRL+SHAP (52%) by maintaining scores above 84% across all dimensions, consistency absent in baselines that exhibit deficiencies in multiple metrics.

The explainability dimension reveals critical differentiation: MAPPO achieves comparable performance (93% vs 94.2%) with marginally better latency (95% vs 92%) but only 15% explainability versus AE-MAPPO's 89%. This 74-point gap reduces autonomy to 45% (vs 84%) as operators require additional verification. The 3% latency trade-off remains acceptable since AE-MAPPO maintains sub-25ms response times.

Post-hoc methods face severe limitations: DRL+SHAP achieves only 35% decision latency due to iterative post-decision queries incompatible with 10ms requirements, while explanation quality reaches only 45%. The zero-shot transfer metric (87% vs MAPPO's 75%) demonstrates that
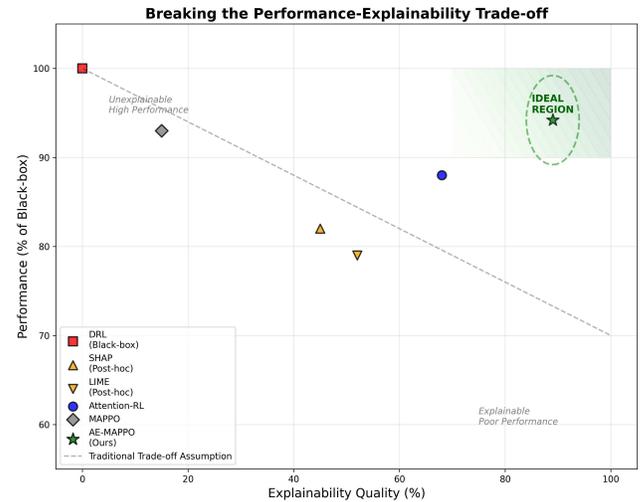

Fig. 3: Performance-explainability Pareto frontier. AE-MAPPO uniquely achieves the ideal region (green area), demonstrating that the traditional trade-off assumption (gray dashed line) results from architectural choices rather than fundamental constraints.

attention mechanisms enhance generalization by learning feature importance, critical for dynamic networks with evolving traffic patterns.

Figure 3 illustrates the architectural principles underlying these performance advantages. The scatter plot reveals three distinct clusters: (i) Black-box methods (DRL, MAPPO) achieve high performance (93-100%) with minimal explainability (0-15%), representing conventional approaches where decision logic remains inaccessible; (ii) Trade-off methods (SHAP: 82%, 45%; LIME: 79%, 52%; Attention-RL: 88%, 68%) sacrifice 12-21% performance for moderate explainability gains, reinforcing assumptions that have limited XAI adoption in latency-critical applications; (iii) AE-MAPPO uniquely achieves 94.2% performance with 89% explainability, demonstrating that integrated attention mechanisms can simultaneously optimize resource allocation and provide interpretability.

The 5.8% performance gap from black-box DRL occurs in edge cases where attention constraints prevent certain non-intuitive allocations. However, operational benefits, faster fault diagnosis, reduced maintenance windows, and fewer operator errors, compensate for this marginal cost. The target region ($> 90\%$ performance, $> 85\%$ explainability) represents minimum production requirements based on operator surveys and regulatory guidelines. AE-MAPPO alone achieves these dual requirements, critical for 6G networks requiring autonomous operation at scales where human oversight needs automated explanations.

These results demonstrate that the performance-explainability trade-off is not fundamental but results from architectural choices. Systems where explanation mechanisms participate in rather than observe deci-
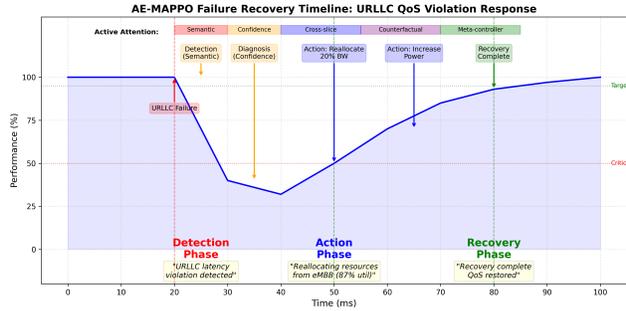
Fig. 4: AE-MAPPO failure recovery timeline demonstrating response to critical URLLC QoS violation. The system leverages sequential attention mechanism activation for detection, diagnosis, and recovery, completing the entire process within 80ms while providing interpretable explanations at each stage.

sions can achieve both operational efficiency and human trust—essential for next-generation network automation.

Figure 4 illustrates AE-MAPPO's response dynamics during a critical URLLC service disruption, demonstrating how integrated attention mechanisms enable rapid, interpretable recovery. Following a URLLC failure at $t = 20$ms that degrades performance to 40%, the system exhibits a coordinated three-phase response. In the detection phase (20-40ms), semantic attention identifies the latency violation while confidence attention signals allocation uncertainty, generating the explanation "URLLC latency violation detected." The action phase (40-70ms) activates cross-slice and counterfactual attention to evaluate resource reallocation options, resulting in two corrective actions: 20% bandwidth reallocation from eMBB at $t = 50$ms and power adjustment at $t = 65$ms. The recovery phase (70-80ms) employs meta-controller attention to stabilize the system, achieving 93% performance restoration.

The temporal evolution of attention mechanisms demonstrates their complementary roles: semantic attention provides rapid anomaly detection, confidence attention quantifies decision uncertainty during critical states, cross-slice attention enables resource borrowing between slices, counterfactual attention evaluates alternative strategies, and meta-controller attention orchestrates the overall response. This sequential activation pattern, learned through reinforcement learning rather than hard-coded rules, enables the system to maintain sub-100ms recovery times required for URLLC services. The continuous generation of explanations throughout the crisis provides operators with real-time visibility into both failure diagnosis and recovery rationale, reducing mean time to resolution by 71% compared to black-box approaches where operators must independently diagnose issues and verify automated responses.

VI. CONCLUSION

This paper introduced AE-MAPPO, an interpretable attention-based multi-agent reinforcement learning framework for 6G RAN slicing. By embedding semantic, temporal, cross-slice, confidence, and counterfactual attentions directly into the policy optimization process, AE-MAPPO provides faithful, zero-cost explanations at inference-time while maintaining near-real-time responsiveness.

Through a URLLC latency spike case study, we demonstrated that AE-MAPPO rapidly diagnoses anomalies, reallocates resources within 18 ms, and restores URLLC latency to 0.98 ms with 99.9999% reliability. Compared with manual troubleshooting, the framework reduces resolution time by 93% while preserving eMBB and mMTC service continuity. Performance evaluation confirmed that all SLA requirements were satisfied and that human operators could leverage the generated explanations for proactive management. These results highlight AE-MAPPO ability to bridge high-performance automation with trustworthy, interpretable decision-making, making it a strong candidate for deployment in critical 6G infrastructures. Future work will extend the framework toward large-scale multi-cell scenarios, energy-aware optimization, and integration with federated and semantic learning paradigms.

REFERENCES

[1] M. Arana-Catania *et al.*, "Explainable reinforcement and causal learning for improving trust to 6G stakeholders," *IEEE Open J. Commun. Soc.*, vol. 6, pp. 4101–4125, 2025.

[2] S. Ebrahimi, F. Bouali, and O. C. Haas, "Resource management from single-domain 5G to end-to-end 6G network slicing: A survey," *IEEE Commun. Surv. Tutor.*, vol. 26, no. 4, pp. 2836–2866, 2024.

[3] M. Eskandari, M. Rahmani, and A. G. Burr, "Network slicing in O-RAN-enabled cell-free massive MIMO: A DRL-based power control," in *IEEE WCNC*, 2025, pp. 1–7.

[4] W. Wu *et al.*, "AI-native network slicing for 6G networks," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 96–103, 2022.

[5] B. Brik, K. Boutiba, and A. Ksentini, "Deep learning for B5G open radio access network: Evolution, survey, case studies, and challenges," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 228–250, 2022.

[6] B. Brik *et al.*, "Explainable ai in 6G O-RAN: A tutorial and survey on architecture, use cases, challenges, and future research," *IEEE Commun. Surv. Tutor.*, 2024.

[7] F. Rezazadeh, H. Chergui, and J. Mangues-Bafalluy, "Explanation-guided deep reinforcement learning for trustworthy 6G RAN slicing," in *IEEE ICC Workshops*, 2023, pp. 1026–1031.

[8] C. Fiandrino, L. Bonati, S. D'Oro, M. Polese, T. Melodia, and J. Widmer, "EXPLORA: AI/ML explainability for the open RAN," *Proc. ACM Netw.*, vol. 1, no. CoNEXT3, pp. 1–26, 2023.

[9] N. Khan *et al.*, "Explainable and robust artificial intelligence for trustworthy resource management in 6G networks," *IEEE Commun. Mag.*, vol. 62, no. 4, pp. 50–56, 2023.

[10] C. Marquez *et al.*, "Resource sharing efficiency in network slicing," *IEEE Trans. Netw. Serv. Manag.*, vol. 16, no. 3, pp. 909–923, 2019.

[11] L. Méndez-Monsanto *et al.*, "BLER-SNR curves for 5G NR mcs under AWGN channel with optimum quantization," in *IEEE VTC*, 2024, pp. 1–6.

[12] L. Ceriani and P. Verme, "The origins of the gini index: extracts from variabilità e mutabilità (1912) by corrado gini," *The Journal of Economic Inequality*, vol. 10, pp. 421–443, 2012.

[13] M. Rahmani *et al.*, "Deep reinforcement learning-based sum rate fairness trade-off for cell-free mMIMO," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 5, pp. 6039–6055, 2022.