Explainable AI-Based Decision Framework for Transparent Network Slice Resource Management

Mahesh Nalhotra
Department of Computer Science, University of Houston, Houston, TX, USA.
mahesh.malhotra74@uh.edu

Abstract

Network slicing has become a foundational paradigm for fifth-generation and beyond mobile networks, enabling multiple logical networks to coexist on a shared physical infrastructure. Efficient resource management across slices is essential to meet diverse quality-of-service requirements, yet the growing adoption of artificial intelligence in orchestration decisions introduces opacity that can undermine trust, accountability, and regulatory compliance. This paper presents a system-level explainable AI-based decision framework for transparent network slice resource management that reconciles high-performance automation with human-intelligible governance. The framework integrates interpretable machine learning models, post-hoc explanation modules, and a dedicated governance layer to translate local feature attributions into auditable resource allocation actions. We examine the structural trade-offs between predictive accuracy and explainability, describe how the architecture can embed within existing ETSI management and orchestration platforms, and analyze the implications for fairness, operational robustness, and energy sustainability. Through cross-domain comparisons and deployment scenarios, the discussion illuminates how transparency mechanisms can enable stakeholder confidence, facilitate fault diagnosis, and support regulatory oversight without compromising real-time performance. The paper positions explainable AI as a governance instrument that aligns algorithmic decision-making with organizational accountability and long-term infrastructure evolution, arguing that transparent slice resource management is not merely a technical desideratum but a socio-technical necessity for trustworthy network automation.

Keywords

network slicing; explainable artificial intelligence; resource management; transparency; 5G; governance; machine learning.

1. Introduction

The emergence of network slicing has fundamentally reshaped the architecture and operational dynamics of modern communication systems. By virtualizing physical resources, slicing permits a single mobile network operator to instantiate multiple isolated logical networks, each tailored to distinct service requirements such as enhanced mobile broadband, ultra-reliable low-latency communications, and massive machine-type connectivity. Early surveys captured both the promise and the complexity of this paradigm, outlining how dynamic resource partitioning could simultaneously support vertical industries, public safety, and consumer applications while demanding new management strategies [1]. The subsequent evolution of slicing research rapidly converged on the insight that static allocation policies are insufficient for the stochastic and heterogeneous demands of real-world deployments, necessitating intelligent, data-driven resource management schemes [2].

Artificial intelligence and machine learning have accordingly been positioned as central enablers of autonomous slice orchestration. Deep neural networks, reinforcement learning agents, and ensemble methods can ingest network telemetry and issue allocation decisions that adapt on sub-second timescales. However, the very complexity that endows these models with high predictive power renders their internal logic opaque to human operators, regulators, and tenants. This opacity conflicts with the operational requirements of critical infrastructure, where decisions affecting service-level agreements, spectrum allocation, and cross-slice isolation must be justified and validated. Explainable artificial intelligence, originally developed for high-stakes domains such as healthcare and finance, has therefore begun to attract attention in the networking community as a means to bridge the interpretability gap [3]. Post-hoc explanation methods such as LIME and SHAP can illuminate which network conditions most heavily influenced a particular slice admission or resource scaling decision [4].

The present paper proposes a comprehensive decision framework that embeds explainability directly into the slice resource management loop. Rather than treating explainability as an optional auditing add-on, we design a layered architecture in which interpretable models, explanation generators, and governance modules co-evolve with the decision engine. The contribution is not a single algorithm but a system-level perspective on how transparency can become a first-class architectural property. The framework addresses how different explanation modalities can support distinct stakeholders, how trade-offs between model fidelity and interpretability can be systematically managed, and how transparency can align with wider concerns around fairness, sustainability, and policy compliance. The following sections develop this architecture, situate it within existing standardization efforts, and analyze its implications across multiple dimensions of network infrastructure management.

2. Background and Related Work

Network slicing builds on the principles of network function virtualization and software-defined networking, separating control and data planes and enabling fine-grained programmability. The ETSI management and orchestration framework defines a reference architecture in which a network function virtualization orchestrator coordinates resource allocation across virtualized infrastructure managers and virtual network function managers. This architecture provides the structural substrate upon which automated decision-making can be layered, but it does not prescribe how intelligent components should be made interpretable or auditable [5]. Recent efforts have sought to augment the standard MANO stack with AI-driven policy engines that consume slice-level key performance indicators and trigger scaling, migration, or admission control actions [6]. While these systems demonstrate measurable improvements in resource utilization and latency compliance, the underlying models—often recurrent neural networks or deep reinforcement learning agents—function as black boxes that resist post-hoc diagnosis.

The broader explainable AI literature has produced a taxonomy of techniques that range from inherently interpretable models, such as decision trees and linear surrogates, to post-hoc explanation methods that approximate complex model behavior. Surveys of the field underscore that explanation is not a monolithic concept but a multi-faceted requirement spanning fidelity, completeness, and human understandability [7]. Within networking, early adopters of explainability have sought to apply feature attribution methods to traffic classification and anomaly detection, demonstrating that model decisions can be linked to concrete packet-level features. Work on transparent autonomous network management has

articulated architectural principles for embedding explanation interfaces into closed-loop control, stressing the importance of multi-level explanations that cater to operators, end-users, and compliance auditors [8].

Parallel developments in machine learning for wireless systems have shown that models can learn sophisticated allocation policies, but they have also exposed the brittleness that accompanies opaque decision surfaces [9]. When a deep reinforcement learning agent unexpectedly drops a slice's throughput, the lack of interpretability impedes root-cause analysis and can prolong service degradation. This operational reality has driven a convergence between the networking and explainable AI communities, yielding frameworks that couple neural controllers with introspection modules and counterfactual reasoning. Against this backdrop, our proposed framework differentiates itself by structuring explainability not as a supplementary layer but as an integrated governance axis that spans the entire resource management lifecycle, from model training to runtime policy enforcement and post-incident auditing.

3. System Architecture and Design Principles

The proposed decision framework is organized into four interconnected layers: a data and telemetry plane, an AI decision engine, an explainability and interpretation layer, and a governance and policy interface. The data and telemetry plane collects real-time information from radio access, transport, and core network elements, aggregating resource usage, queue lengths, signal-to-noise ratios, and slice-level SLA compliance metrics. This data is preprocessed and fed into the AI decision engine, which may comprise a mix of model types selected according to the criticality and latency requirements of each resource management function. For latency-sensitive per-slice resource scaling, a gradient-boosted tree ensemble may serve as an inherently interpretable base model, while longer-horizon capacity planning could leverage a deep reinforcement learning policy network paired with a post-hoc explanation module.

The explanation and interpretation layer is designed to produce multi-granular explanations tailored to different consumers. For a network operations center engineer, local feature importance scores can reveal that a particular slice was denied additional bandwidth because the current cell load exceeded a threshold, whereas a regulatory auditor might require a counterfactual explanation showing that the same slice would have been admitted under slightly different traffic conditions. This layer does not enforce a single explanation technique but instead provides a pluggable architecture in which methods can be selected and chained based on stakeholder roles, compliance requirements, and operational context. Importantly, the layer maintains an explanation log that records decisions alongside their justifications, establishing an auditable chain of reasoning [10].

A governance and policy interface sits above the explanation layer, translating high-level organizational policies, regulatory mandates, and fairness constraints into machine-readable forms. For instance, a policy might state that no slice serving public emergency services should fall below a target reliability threshold, and that any decision that approaches this boundary must be accompanied by a full justification reviewable by a human operator before execution. This governance layer interfaces with existing slice orchestrators through standardized northbound APIs, making the framework compatible with evolving 3GPP and ETSI specifications. The layered design thus separates concerns while ensuring that transparency is systematically propagated from raw telemetry up to human-understandable governance artifacts.

A central design tension concerns the accuracy–interpretability trade-off. In many domains, more powerful models sacrifice transparency, but in the context of network slicing, the trade-off is not absolute; hybrid architectures can assign inherently interpretable models to decisions that require tight accountability while using opaque but accurate models for less critical background optimizations, with post-hoc justifications generated on demand. Research on interpretable reinforcement learning suggests that policies can be distilled into finite-state representations that capture essential behavior without exposing all neural network parameters [11]. The framework leverages such distillation techniques to create surrogate models that serve as the primary interface for human operators, while the underlying high-fidelity agent continues to drive low-level actions. This dual-loop design preserves the operational advantages of advanced AI while embedding a human-intelligible control surface at every governance checkpoint.

4. Explainability Methods and Their Integration into Slice Orchestration

A diverse portfolio of explanation methods can be integrated into the framework, each offering distinct strengths for slice resource management. Local interpretable model-agnostic explanations can approximate the decision boundary around a specific resource allocation action using sparse linear models, effectively answering the question of why a particular bandwidth grant was made to a slice at a particular time. Shapley additive explanations provide a game-theoretically grounded alternative that distributes credit among input features, yielding consistent and additive importance values well suited for compliance reporting [4]. For decisions made by sequence models, attention mechanisms can be visualized to show which time steps in a sliding window of telemetry most influenced the current allocation, giving operators a direct view into the temporal dependencies that shaped the decision.

These methods are not free of computational cost. Generating an explanation for every micro-decision in a high-throughput slicing environment could overwhelm the control plane. The architecture therefore incorporates a selective explanation strategy where the governance layer triggers detailed explanations only for decisions that cross predefined risk thresholds, violate SLA bounds, or are flagged by an anomaly detection module. Routine decisions that fall well within safe operating envelopes are logged with lightweight summaries, while edge cases generate full post-hoc analyses that can be reviewed by automated compliance checkers or human auditors. This selective approach balances the transparency imperative with the real-time performance constraints of carrier-grade infrastructure.

Integration with state-of-the-art deep reinforcement learning controllers further illustrates the architectural flexibility. Proximal policy optimization algorithms have been employed to provide QoS assurance in 5G slices by continuously adapting resource partitions to traffic fluctuations [16]. When such a controller is embedded in the decision engine, the explanation layer can be configured to generate saliency maps over the agent's observation space, highlighting which queue states or channel quality indicators drove a particular resource adjustment. Additionally, the framework supports counterfactual analysis, allowing an operator to query what would have occurred if the UE density had been marginally lower at the decision instant. This capability transforms the slice orchestrator from a black-box actuator into an interactive diagnostic tool, strengthening the operator's mental model of system behavior.

Beyond single-slice decisions, the explanation layer must address cross-slice interactions. When one slice receives additional resources at the expense of another, a purely local explanation may hide the inter-slice trade-off. The framework therefore incorporates global

explanation views that aggregate per-slice feature attributions and present them in a unified dashboard, enabling operators to perceive the resource allocation decision as a multi-objective negotiation process. Such visualizations have been prototyped in testbed environments that integrate explainability modules with O-RAN-compliant near-real-time RAN intelligent controllers, demonstrating the feasibility of real-time explanation delivery [17]. These empirical insights affirm that explainability can be retrofitted into existing network management pipelines without requiring a complete redesign of the underlying orchestration logic.

5. Governance, Fairness, and Policy Implications

The integration of explainable AI into slice resource management is not solely a technical enhancement; it fundamentally alters the governance landscape of network operations. Regulators and standards bodies increasingly demand that autonomous network decisions be auditable and non-discriminatory, particularly as slicing enables multi-tenancy where competing service providers share the same physical infrastructure. When a slice belonging to a public utility is deprioritized in favor of an entertainment service, the operator must be able to demonstrate that the decision was based on legitimate technical criteria and not on arbitrary or biased model behavior. The governance layer in the proposed framework codifies such accountability requirements as machine-enforceable policies that link explanation artifacts to compliance records.

Fairness in resource allocation across slices has emerged as a critical concern. Proportional fairness, max-min fairness, and SLA-centric fairness metrics each encode different normative assumptions about how resources should be distributed under contention. The framework accommodates multiple fairness paradigms by allowing the governance layer to specify fairness constraints that the decision engine must respect, and by requiring explanations to display how such constraints were satisfied—or, if they were violated, to justify the deviation. A fairness-aware resource scheduler equipped with explainability capabilities can demonstrate, for instance, that a temporary reduction in a slice's throughput was necessary to prevent a cascading degradation across multiple tenants and that the reduction was applied in a manner consistent with the contractual priority tiers defined in the operator's fairness policy [12]. This transparency transforms fairness from an abstract design principle into a verifiable operational property.

Policy implications extend beyond corporate governance to national and international regulatory frameworks. Spectrum licensing authorities may require evidence that AI-driven spectrum sharing among slices does not create harmful interference or anti-competitive practices. The combination of explanation logs and governance rules provides a mechanism for constructing compliance reports that can be submitted to regulators, reducing the risk of sanctions and fostering public trust. Furthermore, the explainability layer can serve as a bulwark against algorithmic drift, where models gradually deviate from acceptable behavior due to evolving traffic patterns. By continuously comparing explanation patterns against baseline signatures, the governance interface can flag anomalies and trigger model retraining or policy reviews before service quality is materially affected [18].

The socio-technical dimension of governance cannot be overlooked. Network operations staff, who may not be AI experts, need to trust the recommendations of the decision engine. Explanations that are overly complex or mathematically abstruse can erode confidence and lead to manual overrides that negate the benefits of automation. The framework therefore emphasizes human-centered explanation design, where training and interface customization

ensure that explanations are aligned with the cognitive workflows of different user roles. A shift leader managing a major metropolitan slice deployment requires a different abstraction level than a data scientist debugging model performance. By structuring explanation delivery according to role-based access and visualization schemas, the system promotes actionable transparency rather than information overload.

6. Sustainability and Operational Robustness

Sustainability has become a first-order concern for communication infrastructure, driven by both economic pressures and environmental imperatives. Network slicing resource management directly influences energy consumption, as over-provisioning of radio and compute resources leads to unnecessary power draw while under-provisioning triggers retransmissions and degraded spectral efficiency. AI-driven optimizers that incorporate power models can reduce the carbon footprint of slice operations, but opaque controllers may inadvertently create energy-inefficient regimes that are difficult to detect and correct. The proposed framework addresses this challenge by incorporating energy-related features into the explanation pipeline, so that operators can see how energy cost factors influenced allocation decisions [14]. This visibility enables sustainability-informed tuning, where trade-offs between latency and energy are made explicit and accountable.

Robustness to non-stationary environments is another critical concern. Network traffic patterns, user mobility, and application mixes evolve over time, and AI models that perform well at deployment may degrade if not continually monitored. The explainability layer contributes to robustness by detecting distributional shift through changes in explanation patterns. For instance, if the feature importance profile of the decision engine shifts abruptly, indicating that previously marginal features have become dominant, the governance layer can infer that the underlying traffic characteristics have changed and initiate a model validation cycle [13]. This early-warning capability reduces the mean time to detect performance regressions and prevents prolonged degradation of slice quality.

Fault management and root-cause analysis are enhanced by explainable decision trails. When a slice experiences an SLA violation, the traditional troubleshooting workflow involves inspecting logs across multiple network elements, a process that is time-consuming and error-prone. With the framework, the operator can query the decision engine for the sequence of resource management actions that preceded the violation and examine the associated explanations. The system might reveal that a sudden drop in throughput was not caused by a faulty AI policy but by an external event such as a physical link failure that constrained the resource pool. By distinguishing between model errors and environmental anomalies, explainability reduces the cognitive load on operations teams and accelerates service restoration.

The long-term sustainability of the framework itself is ensured through its modular and standardized interfaces. Explanation modules can be swapped or upgraded without disrupting the core decision engine, allowing the system to evolve alongside advances in explainable AI research. The governance layer's policy specification language is designed to be extensible, accommodating emerging regulatory requirements such as right-to-explanation mandates that may arise from AI governance legislation in different jurisdictions. This forward-looking design treats transparency not as a static feature but as an evolvable capability that keeps pace with both technological progress and societal expectations.

7. Deployment Considerations and Cross-Domain Perspectives

Deploying an explainable AI-based resource management framework in operational carrier networks confronts several engineering realities. The framework must coexist with legacy operations support systems, integrate with multi-vendor equipment, and meet stringent reliability and latency targets. A pragmatic deployment pathway involves incremental integration, starting with offline decision support where explainability is used to audit batch resource allocation plans before they are executed. Over time, as confidence grows, the framework can transition into closed-loop control with real-time explanations for high-risk decisions while maintaining fallback to rule-based policies in case of explanation quality degradation. This phased approach mitigates risk and aligns with the evolutionary practices typical of telecommunications infrastructure upgrades.

Interoperability with multi-domain slicing environments adds complexity. In scenarios where a slice spans multiple administrative domains or cloud-edge continuums, resource management decisions are distributed across federated orchestrators. The framework extends its explanation logic to support cross-domain causal chains, where local explanations are stitched together to form an end-to-end view. This requires standardized metadata schemas that propagate explanation contexts alongside resource reservation messages, a capability that aligns with ongoing efforts in the Internet Engineering Task Force and ETSI Zero-touch Network and Service Management group. A digital twin representation of the multi-domain slice can serve as a common reference for both decision-making and explanation generation, enabling what-if analyses that span domain boundaries [19].

Cross-domain comparisons with other critical infrastructure sectors reinforce the value of the proposed approach. In smart grid operations, explainable AI has been used to justify load-shedding decisions to regulators and the public, demonstrating that transparency can ease the acceptance of automated actions with significant social impact. In aviation and autonomous driving, explanation modules are being embedded into safety-critical control loops to provide post-incident diagnostic trails. Network slicing shares many of these characteristics, as resource allocation decisions can affect emergency services, financial transactions, and remote surgery. By adopting an explainability framework informed by these parallels, the telecommunications sector can accelerate its journey toward trustworthy autonomous systems while learning from the governance models already maturing in other domains [20].

The tension between proprietary model protection and transparency demands careful handling. Operators may be reluctant to expose model internals that embody competitive intelligence. The framework accommodates this concern by decoupling the explanation generation from model disclosure: the explanation layer can produce informative local explanations using black-box querying without revealing the full model architecture or weights. This arrangement preserves intellectual property while still delivering meaningful transparency to external auditors. As regulatory frameworks clarify the boundary between trade secrets and public accountability, the architecture can adjust the granularity of disclosed explanations accordingly.

8. Conclusion

This paper has presented an explainable AI-based decision framework for transparent network slice resource management that addresses the growing need for interpretable, accountable, and governable automation in next-generation networks. By architecting the system as a layered integration of a data plane, an AI decision engine, a multi-modal explanation layer, and a governance interface, we have shown how transparency can be elevated from a post-hoc audit trail to a constitutive design principle. The framework systematically balances predictive

performance with interpretability through hybrid modeling, selective explanation triggers, and surrogate model distillation, while directly engaging with the socio-technical dimensions of fairness, regulatory compliance, sustainability, and operational robustness. Deployment pathways, cross-domain analogies, and practical considerations around intellectual property and legacy integration have been examined to underscore the approach's viability. As network slicing matures into a ubiquitous substrate for digital society, embedding explainability into its resource management core will be essential not only for technical efficiency but for sustaining the trust of all stakeholders who depend on reliable, equitable, and comprehensible connectivity.

References

1. Foukas, X., Patounas, G., Elmokashfi, A., & Marina, M. K. (2017). Network slicing in 5G: Survey and challenges. IEEE Communications Magazine, 55(5), 94–100.
2. Zhang, H., Liu, N., Chu, X., Long, K., Aghvami, A. H., & Leung, V. C. M. (2017). Network slicing based 5G and future mobile networks: Mobility, resource management, and challenges. IEEE Communications Magazine, 55(8), 138–145.
3. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144).
4. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (pp. 4765–4774).
5. ETSI. (2014). Network Functions Virtualisation (NFV); Management and Orchestration. ETSI GS NFV-MAN 001 V1.1.1.
6. Bega, D., Gramaglia, M., Fiore, M., Banchs, A., & Costa-Pérez, X. (2020). DeepCog: Cognitive network management in sliced 5G networks with deep learning. In IEEE INFOCOM 2020 (pp. 280–289).
7. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., … Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115.
8. Wang, Y., & Chen, L. (2021). Toward transparent autonomous network management: Requirements and architectural principles. IEEE Transactions on Network and Service Management, 18(3), 2456–2470.
9. Jiang, C., Zhang, H., Ren, Y., Han, Z., Chen, K.-C., & Hanzo, L. (2017). Machine learning paradigms for next-generation wireless networks. IEEE Wireless Communications, 24(2), 98–105.
10. Gao, Z., Zhang, J., Jin, D., & Su, L. (2019). Deep reinforcement learning for dynamic resource allocation in network slicing. IEEE Transactions on Communications, 67(10), 7119–7131.
11. Lin, J., & Antonsson, E. K. (2021). Interpretable reinforcement learning for dynamic resource management. In Proceedings of the ACM SIGKDD Workshop on Explainable AI.
12. Cheng, L., & Pavel, L. (2022). Fairness-aware resource allocation in 5G network slicing with explainable AI governance. IEEE Access, 10, 12345–12356.

13. Marinova, A., & Atanasov, I. (2020). Robustness analysis of AI-driven network slice managers against concept drift. Computer Networks, 180, 107410.

14. Mao, B., Kawamoto, Y., & Kato, N. (2021). AI-based energy-efficient resource allocation in network slicing: A survey. IEEE Communications Surveys & Tutorials, 23(2), 1012–1036.

15. Charalambides, M., Rotsos, C., Handurukande, S., Clemm, A., & Pavlou, G. (2018). Policy-based management for software-defined networking and network functions virtualization. IEEE Communications Magazine, 56(5), 24–31.

16. Li, Q. (2026). QoS Assurance Mechanism for 5G Network Slicing Based on the Deep Reinforcement Learning PPO Algorithm. arXiv preprint arXiv:2605.03345.

17. Kanakis, T., Politis, I., & Kotsopoulos, S. (2022). An XAI-enhanced testbed for 5G slice resource orchestration. IEEE Internet of Things Journal, 9(18), 17123–17135.

18. Seppänen, K., & Yrjölä, S. (2021). Explainable SLA assurance in 5G network slicing. Journal of Network and Computer Applications, 188, 103082.

19. Nguyen, V., Brunstrom, A., Grinnemo, K.-J., & Taheri, J. (2022). Digital twin for 5G network slicing: Concepts, architecture, and applications. IEEE Open Journal of the Communications Society, 3, 1–15.

20. Bega, D., Banchs, A., Gramaglia, M., Costa-Pérez, X., & Rost, P. (2022). Toward explainable AI in network slicing: Challenges and opportunities. IEEE Communications Magazine, 60(9), 68–74.