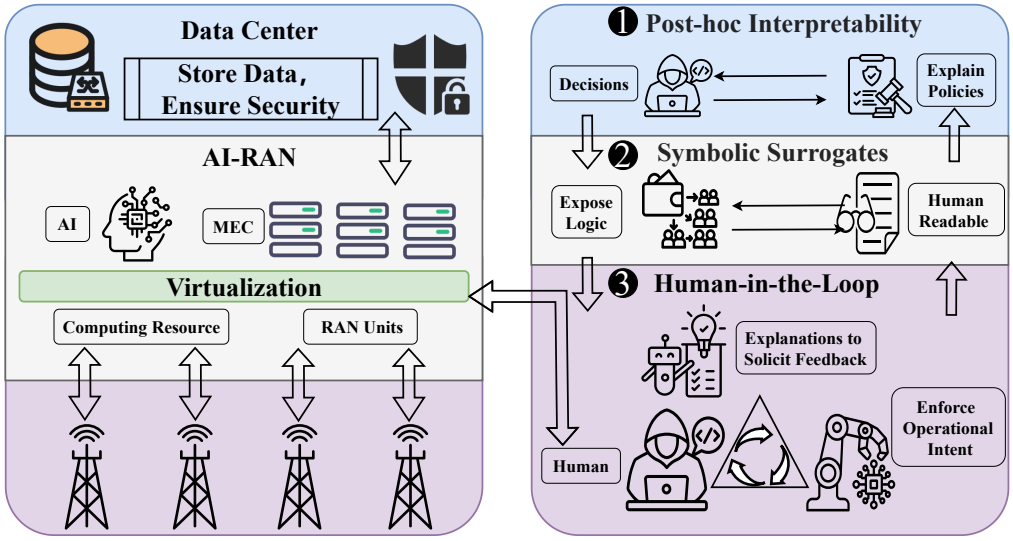
Graphical Abstract

Towards Transparent 6G AI-RAN: A Survey on Explainable Deep Reinforcement Learning for Intelligent Network Slicing

Shuaishuai Guo, Yutong Zhong, Zhenyu Feng, Shengqi Kang, Jichao Chen



System Architecture **XDRL Framework**

Highlights

Towards Transparent 6G AI-RAN: A Survey on Explainable Deep Reinforcement Learning for Intelligent Network Slicing

Shuaishuai Guo, Yutong Zhong, Zhenyu Feng, Shengqi Kang, Jichao Chen

- This paper provides a comprehensive survey of explainable deep reinforcement learning (XDRL) techniques for intelligent network slicing in 6G AI-driven Radio Access Networks (AI-RAN).
- It identifies key challenges and outlines promising research directions to achieve transparent, trustworthy, and human-centric AI-RAN systems for future 6G networks.

Towards Transparent 6G AI-RAN: A Survey on Explainable Deep Reinforcement Learning for Intelligent Network Slicing

Shuaishuai Guo[a,d], Yutong Zhong[b], Zhenyu Feng[c], Shengqi Kang[d], Jichao Chen[e,*]

[a] *Purple Mountain Laboratories, Nanjing, 210023, China*
[b] *Cognicore Artificial Intelligence Co., Ltd., Jinan, 250062, China*
[c] *Deep NeuroRAN Co., Ltd., Guangzhou, 510525, China*
[d] *School of Control Science and Engineering, Shandong University, Jinan, 250061, China*
[e] *Communication Systems Department, EURECOM, Sophia Antipolis, 06410, France*

Abstract

The advent of the sixth generation (6G) wireless networks envisions an artificial intelligence (AI)-native radio access network (AI-RAN), where deep reinforcement learning (DRL) emerges as a key enabler for intelligent and autonomous network slicing. Despite the demonstrated performance gains of DRL-based solutions in dynamic resource allocation and slice orchestration, their opaque decision-making nature raises critical concerns regarding trust, accountability, and operational deployment. To bridge this gap, explainable deep reinforcement learning (XDRL) has recently attracted significant attention as a means to enhance transparency, interpretability, and controllability of AI-RAN slicing policies. This survey provides a comprehensive overview of the state of the art in explainable DRL for intelligent network slicing. We first review the fundamental principles of DRL in the context of RAN slicing and identify the unique explainability challenges posed by high-dimensional, multi-slice environments. We then categorize existing XDRL approaches into post-hoc explanation, symbolic abstraction, and human-in-the-loop steering, analyzing their methodologies, strengths, and limitations. Furthermore, we

*Corresponding author.

E-mail addresses: shuaiguosdu@gmail.com (S. Guo), flsxzhong0124@gmail.com (Y. Zhong), zainfung@gmail.com (Z. Feng), kangsq@mail.sdu.edu.cn (S. Kang), jichao.chen@eurecom.fr (J. Chen).

highlight benchmark environments and experimental testbeds that have been employed to evaluate XDRL in realistic network scenarios. Finally, we outline key open challenges, including scalability, generalization across traffic patterns, integration with large language models (LLMs), and alignment with intent-based networking, and discuss promising research directions toward achieving transparent, trustworthy, and human-centric AI-RAN in 6G.

Keywords: Explainable deep reinforcement learning (XDRL), 6G AI-RAN, intelligent network slicing, transparency and interpretability, intent-based networking

1. Introduction

As the demand for increasingly diverse and sophisticated services continues to grow, network slicing has emerged as a pivotal technique for resource partitioning and optimization across multiple virtual networks [1, 2, 3, 4]. This approach enables the deployment of customized, efficient, and service-specific virtualized networks, each tailored to accommodate distinct traffic types, such as enhanced mobile broadband communications (eMBB), ultra-reliable low-latency communications (URLLC), and massive machine-type communications (mMTC) within fifth-generation (5G) wireless systems [5, 6]. Looking ahead, the vision for sixth-generation (6G) networks is to provide unprecedented connectivity characterized by ultra-high data rates, ultra-low latency, and the capability to support an enormous number of devices [7, 8]. In addition to simply enhancing throughput and coverage, 6G networks are expected to serve as the technological foundation for a wide variety of emerging applications, including autonomous vehicles, industrial automation, smart cities, holographic communications, and immersive virtual/augmented reality experiences [9, 10]. To achieve these ambitious goals, 6G networks will inevitably rely on the integration of advanced technologies such as AI, machine learning, and automation, which enable real-time adaptation and optimization in highly dynamic communication environments [11, 12, 13]. In particular, AI-driven radio access networks (AI-RAN) are envisioned as a cornerstone of 6G, leveraging AI techniques to enable self-organization, autonomous decision-making, and intelligent resource allocation [14, 15].

In 6G networks, it is expected that each network slice can be customized and optimized to meet specific service requirements, thereby providing a flexible and efficient way to address the diverse demands of future appli-

cations. For instance, a holographic-type communication (HTC) slice may prioritize ultra-high throughput and extremely low latency to enable real-time holographic conferencing and immersive virtual education [16]. A tactile Internet with digital twins (TI-DT) slice would instead emphasize ultra-reliable, deterministic latency to ensure seamless interaction between physical systems and their digital counterparts, supporting mission-critical applications such as industrial automation and remote robotic control [17, 18]. Similarly, AI-native service slices may allocate resources for distributed training and inference of large-scale models at the network edge, while ubiquitous sensing and communication (USC) slices integrate sensing and communication functions to support vehicular networks, smart transportation, and environment monitoring [19]. In addition, space-air-ground-sea integrated network (SAGSIN) slices can provide global, seamless connectivity by orchestrating resources across satellites, unmanned aerial vehicles (UAVs), terrestrial, and maritime nodes [20, 21]. The ability to intelligently manage and optimize these heterogeneous slices in real time is therefore critical for ensuring that each service achieves its required performance levels and complies with service-level agreements (SLAs). In the context of 6G, where service heterogeneity is expected to be more complex than ever before, network slicing becomes not only useful but indispensable [22].

To effectively address the resource management challenges in network slicing, traditional approaches primarily rely on mathematical optimization or heuristic algorithms. For instance, rigorous optimization frameworks typically model resource allocation as mixed-integer problems solved via approximations [23], or utilize advanced techniques such as Lyapunov optimization to maximize objectives under dynamic constraints [24]. Due to the NP-hard nature of these problems, heuristic approaches have been widely adopted to reduce computational complexity. Typical examples include adaptive Hungarian algorithms for bandwidth allocation [25], heuristics for soft slicing that balance utilization and dissatisfaction [26], and online placement algorithms based on the "power of two choices" for large-scale networks [27]. While these methods improve acceptance ratios and solving speed compared to brute-force baselines, they struggle to deliver optimal performance in the highly complex and non-stationary environments inherent to 6G [28]. This limitation arises primarily from their reliance on static, manually engineered rules that lack the adaptability to handle rapidly varying channel conditions and multi-objective trade-offs. Consequently, in the context of 6G AI-RAN, heuristic algorithms are often unable to meet the required performance guarantees, necessitating

the shift towards learning-based approaches [3].

To meet these challenges, deep reinforcement learning (DRL) has been reconginized as a powerful technique for optimizing complex and dynamic decision-making processes in communication networks [29, 30]. DRL, a subset of machine learning, enables an agent to learn optimal policies through continuous interaction with the environment, receiving feedback in the form of rewards [31]. In the context of AI-RAN, DRL has shown great promise in optimizing resource allocation [32], traffic routing [33], scheduling [34], and, most importantly, dynamic slice management [35]. Its ability to handle uncertainty, adapt to traffic fluctuations, and discover sophisticated policies beyond human-designed heuristics makes it particularly attractive for 6G network slicing [36]. By leveraging DRL, the network can autonomously learn to allocate radio resources, bandwidth, and power across multiple slices, thereby improving overall efficiency and user experience [37].

Despite these advantages, DRL suffers from a significant drawback: its black-box nature [38, 39]. While DRL-based solutions have demonstrated impressive performance in optimizing resource management tasks, the underlying decision-making process remains opaque and difficult to interpret. In traditional network management systems, operators rely on well-understood algorithms and protocols, which allow them to verify and trust the system's behavior. In contrast, the opacity of DRL models raises serious concerns about transparency, trustworthiness, and accountability [40]. Network operators and users alike may be reluctant to rely on AI systems that make decisions in ways that cannot be explained or validated, particularly in safety-critical or regulation-sensitive applications. This lack of interpretability could therefore become a major barrier to the practical deployment of DRL in AI-RAN systems [41, 42].

To deal with this, explainable deep reinforcement learning (XDRL) has been introduced as a promising paradigm aimed at enhancing the interpretability of DRL models [43]. XDRL seeks to make the decision-making process of DRL transparent and understandable, allowing network operators to gain insights into the rationale behind the actions chosen by the agent. Techniques such as symbolic abstraction [44], rule extraction [45], policy distillation [46], and post-hoc explanations [47] are employed to either approximate or directly reveal the internal logic of DRL models. These explainability mechanisms can improve operator trust, support human oversight, and ensure compliance with service requirements and regulatory standards. In this way, XDRL serves as a crucial bridge between the high performance of DRL and the necessity of

human-centric, transparent, and trustworthy AI systems [38].

The integration of XDRL into intelligent network slicing is particularly relevant for 6G, as it combines the adaptability and efficiency of DRL with the interpretability required for real-world deployment [35]. By making DRL-based slicing decisions explainable, network operators are empowered to better monitor system behavior, diagnose potential issues, and intervene when necessary [48]. Moreover, XDRL opens the door to designing AI-RAN systems that are not only optimized for performance but also aligned with broader goals such as fairness, accountability, and sustainability [38]. This alignment is critical for ensuring that AI-driven 6G systems meet both technical performance targets and societal expectations [49]. Table 1 presents a detailed comparison between our work and existing literature, highlighting the distinct research gaps addressed herein. This paper aims to provide a comprehensive survey of state-of-the-art XDRL techniques and their applications to intelligent network slicing in 6G AI-RAN environments. We review the core concepts, methods, and evaluation frameworks for XDRL, analyze existing challenges and limitations, and identify key research directions that must be pursued to realize transparent, reliable, and human-centric AI solutions in next-generation networks.

Table 1: Comparison between our work and existing work.

Reference	Scenario	Focus
[2]	5G network	Role of slicing in meeting diverse 5G cases Service management and orchestration
[4]	5G network	Slicing solutions for RAN End-to-end network slicing
[22]	6G mobile and heterogeneous network	Resource management Power allocation, spectrum allocation
[29]	Decentralized and autonomous networks	DRL in communications and networking From basic RL to advanced DRL models
[30]	Communication in multi-agent systems	Classification and analysis of MADRL with communication
[50]	5G and beyond	Analyzes the role of XAI in a range of 5G enabling technologies
[51]	6G network	XAI for 6G wireless communications Vehicular networks and real-time/high-stakes scenarios
Our work	6G AI-RAN	Explainability in DRL for RAN slicing Evaluation benchmarks and testbeds for XDRL

The rest of the paper is organized as follows: In Section 2, we discuss the

challenges of applying DRL in AI-RAN and outline the need for explainability. In Section 3, we review the core concepts of XDRL and describe the different techniques used to explain DRL in intelligent network slicing. Section 4 highlights the evaluation frameworks and testbeds used to assess XDRL methods in real-world network scenarios. Section 5 identifies the open challenges and future research directions in XDRL for AI-RAN. Finally, Section 6 concludes the paper, summarizing the key findings and discussing the potential impact of explainable AI in the 6G era.

2. Challenges of DRL in AI-RAN

Deploying DRL in AI-RAN for 6G network slicing faces a broad spectrum of technical and operational hurdles that extend well beyond interpretability. First, the learning environment is inherently non-stationary: traffic demand, channel conditions, interference patterns, mobility, and even slice definitions evolve across time scales from milliseconds (PHY/MAC) to hours (diurnal load) and seasons (long-term planning) [3]. This violates the Markov and stationarity assumptions commonly used in DRL formulation and destabilizes policy learning [52]. Second, the partial observability of RAN states—due to delayed/aggregated key performance indicators (KPIs), measurement sparsity, and privacy constraints—forces agents to act under uncertainty and stale information, often requiring memory-based policies or belief tracking [35]. Third, the sample inefficiency of DRL clashes with stringent online safety and SLA guarantees: naive exploration in live networks can degrade quality of service (QoS), breach URLLC reliability, or waste spectrum/energy; purely simulated training, in turn, suffers from sim-to-real gaps (mismatched propagation models, unmodeled interference, hardware non-linearities) [53]. Fourth, reward design [54] is intrinsically multi-objective and hierarchical: throughput, latency/jitter, reliability, fairness, energy efficiency, carbon footprint, and cost must be optimized jointly across users, cells, and slices; poorly shaped rewards induce perverse incentives (e.g., favoring short flows, starving cell-edge users) and unstable training [55].

Scalability and coordination introduce additional barriers. In dense deployments (macro & small cells [56], UAVs [57, 58], and low earth orbit (LEO) satellites in SAGSIN [59]), decentralized or multi-agent DRL (MADRL) must address non-stationarity caused by other learning agents, credit assignment across layers, and communication overhead for coordination—while respecting fronthaul/backhaul constraints [60]. Safety and constraint satisfaction

are paramount: hard constraints (e.g., 10^{-5} block error rate (BLER) for URLLC, maximum transmit power, interference budgets, spectrum masks) must be enforced at all times [53, 61], not only in expectation; yet standard DRL optimizes long-term returns and offers no guarantees without additional mechanisms (shielding, Lyapunov- or constrained Markov decision process (CMDP)-based methods, or safe policy optimization) [62, 63]. Stability and convergence of training are fragile under high-dimensional continuous actions (power, physical resource block (PRB) allocation, beamforming, numerology) and delayed/rare-event rewards (e.g., outage, violation bursts), which impede temporal credit assignment [35, 64]. Exploration–exploitation [65] is constrained: aggressive exploration risks SLA breaches, whereas conservative exploration traps policies in locally optimal behaviors that fail to adapt to regime shifts (new services such as holographic-type communications or tactile Internet with digital twins). Data quality and availability are uneven across cells and time; logs can be biased (collected under legacy heuristics), noisy (measurement errors), or censored (privacy), hindering off-policy learning and evaluation [66, 67]. Finally, operationalization challenges—edge compute budgets, energy limits, model lifecycle management, online evaluation with counterfactual uncertainty, robustness to adversarial traffic or spoofed KPIs, and reproducible benchmarking—must be addressed before wide-scale deployment [15, 68, 69].

Black-box challenges occupy a central position because they entangle nearly all concerns above—safety, compliance, debugging, and operator trust [41, 51]. The opacity of DRL manifests at multiple levels: (i) Policy opacity (global). High-capacity function approximators (deep networks) encode policies that are difficult to summarize [70]. Operators cannot easily answer: Which features drive allocation decisions across slices? Under what traffic/channel state information (CSI) regimes will the policy throttle HTC in favor of TI-DT? How do beamforming and PRB assignments co-adapt with mobility and interference? Without global explanations, it is hard to validate that policies align with design intent (e.g., fairness for cell-edge users, carbon-aware scheduling). (ii) Decision opacity (local). For a specific action—say, reallocating PRBs from an AI-native inference slice to an URLLC TI-DT slice at the cell edge—operators need case-level rationales: What counterfactuals would have flipped the decision? Which KPIs (queue length, predicted latency violation risk, control channel utilization) were pivotal? Local explanations must be temporally aware, since current actions depend on long-term predictions and eligibility traces; static feature attributions

can mislead when action value stems from expected future congestion relief [71]. (iii) Temporal credit assignment and causality. DRL policies exploit long-horizon effects (e.g., proactively shaping traffic to prevent future queuing) [72]. Explanations must expose temporal causal chains—how a short-term power backoff reduces interference spillover, thereby lowering URLLC delay tails several frames later. Post-hoc saliency maps on instantaneous features miss these delayed effects; counterfactual and causal explanations (e.g., do-calculus-inspired or structural causal model (SCM)-backed analyses) are needed to justify sequences of actions [73, 74]. (iv) Constraint compliance and assurance. Operators require evidence that actions satisfy hard constraints before execution. Typical post-hoc explainable AI (XAI) provides plausibility but not guarantees [75]. DRL must pair with ante-hoc mechanisms—policy certificates [63], conservative value bounds [76], conformal risk control [77], or shielded action filters [78]—to deliver explanations that include formal compliance claims (e.g., predicted probability of violating 1 ms end-to-end (E2E) latency $< 10^{-5}$ under uncertainty sets). (v) Multi-agent explanations and scalability. In distributed AI-RAN, coordinated decisions (inter-cell interference coordination, handover orchestration, cooperative beamforming) arise from multiple agents' interactions [79]. Explanations must attribute outcomes to agents and communication messages—who influenced whom and why—while remaining compact and fast enough for near-real-time operations on edge hardware [80, 81]. (vi) Robustness, drift, and security. Explanations should surface sensitivity to distribution shifts (e.g., festival traffic spikes, new HTC codecs), adversarial perturbations (spoofed KPI injections), and model drift [82]. Operators need why a policy's reliance on a feature (e.g., predicted channel quality indicator (CQI)) has grown, and what safety fallbacks trigger when confidence drops.

Addressing these facets requires a blended toolbox that complements performance with interpretability and assurance. Policy distillation and abstraction can compress black-box policies into interpretable surrogates (decision trees, rule lists, symbolic automata, or linear-threshold ensembles) with fidelity tracking [44, 45]; hierarchical or options-based DRL exposes human-comprehensible sub-policies (e.g., "load-shed HTC," "protect TI-DT control plane"), improving simulatability [83]. Counterfactual and contrastive explanations articulate minimal state changes that alter actions (e.g., "if predicted URLLC queueing delay had been < 0.2 ms, the PRB reallocation would not occur"), which are actionable for operators [84, 85]. Causal explanations [74] and value decomposition [86] clarify how components (interference,

mobility, queue state) contribute to Q-values across time, aiding root-cause analysis. Uncertainty-aware outputs—calibrated value/policy confidence [87], risk measures, conformal prediction sets—allow explanations to include confidence and risk alongside rationales [77], enabling guardrails (shields) that veto unsafe actions [78]. Constraint-aware DRL (CMDPs, Lyapunov RL, Lagrangian methods) should externalize dual variables or certificates as part of the explanation payload, linking actions to explicit constraint satisfaction [76, 88, 89, 90]. Finally, human-in-the-loop (HITL) oversight benefits from explanation-aligned rewards and feedback channels: operators can critique rationales (not just outcomes), steering policies toward compliance, fairness, and sustainability objectives [91].

In summary, deploying DRL in 6G AI-RAN and network slicing faces profound challenges: highly non-stationary environments, lack of real-time safety guarantees, complex multi-objective reward design, and multi-agent coordination. Among these, the black-box nature of DRL policies remains the central barrier, as operators demand not only high performance but also trustworthy global understanding, local decision rationales, provable long-term constraint satisfaction, and robustness against distribution shifts and adversarial attacks. Therefore, a comprehensive, safety-first, and explainable framework is needed to overcome the aforementioned challenges.

3. From Black Box to Assured Control: Explainable DRL for 6G Network Slicing

XDRL targets the core limitation of DRL in AI-RANs—opacity by turning high-performance policies into auditable, simulatable, and assurance-ready control [40]. In 6G settings with heterogeneous slices, operators must understand why a policy reallocates PRBs, throttles an HTC slice, or raises modulation and coding scheme (MCS) under specific predicted CQI and queue states, and how such actions satisfy latency/reliability and interference constraints. XDRL therefore aims not only to make decisions understandable, but to couple each action with evidence about compliance, risk, and counterfactual behavior, thereby supporting trust, debugging, and policy governance.

At a conceptual level, XDRL enhances DRL pipelines by incorporating three key complementary axes that must be integrated, rather than treated as separate components, as illustrated in Fig. 1. These axes include: (i) post-hoc interpretability, which provides transparent explanations of trained policies

and the individual decisions made by the system; (ii) symbolic/structured surrogates, which translate complex policy logic into human-readable formats, making it easier for users to understand and trust the reasoning of system; (iii) HITL alignment and steering, where explanations are used not only to facilitate user feedback but also to ensure the system's actions align with human intentions and operational goals.

While the integration of post-hoc interpretability, symbolic surrogates, and HITL alignment offers a powerful framework for enhancing DRL systems, the balance between these axes involves inherent trade-offs that must be carefully managed. Post-hoc interpretability provides transparency and helps demystify the system's decisions; however, it can be limited by the complexity of the model—offering only high-level explanations that may lack the granularity needed for deep understanding in complex tasks. On the other hand, symbolic surrogates can offer more detailed and structured reasoning, but they may oversimplify certain aspects of the policy or fail to capture the full range of decision-making nuances, which could lead to misinterpretations. HITL alignment introduces a critical human element to guide the system's actions, ensuring that policies are aligned with human values and operational goals. However, it also relies on human feedback, which can be subjective, inconsistent, and slow, potentially creating a bottleneck in real-time applications. Therefore, achieving an optimal balance among these approaches by considering their specific strengths and limitations in particular scenarios represents a promising strategy. This approach ensures the system remains

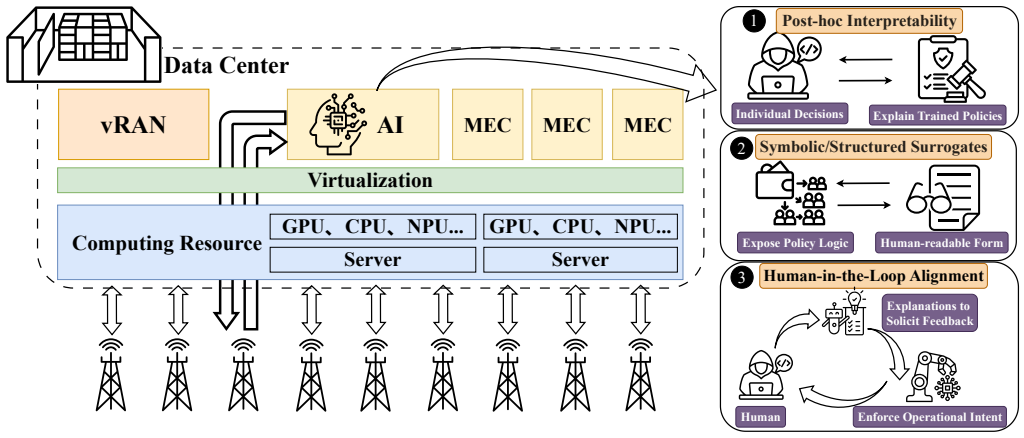


Figure 1: AI-RAN architecture integrating AI modules with conventional RAN components.

both transparent and adaptable while maintaining high performance and user trust.

In practice, successful implementations of XDRL integrate these axes into a unified, continuous feedback loop. Explanations serve to clarify the underlying rationales and highlight areas of uncertainty, while operators provide critical feedback and impose necessary constraints on the system's behavior. Based on this feedback, the policy is either updated to improve performance or protected from potentially harmful changes. At the same time, symbolic surrogates play a crucial role in tracking the reliability of the model and detecting drift in policy performance, ensuring that the system remains aligned with its intended objectives over time.

Concretely, post-hoc explanation provides both global and local insight [92, 93]. Global views quantify which features (e.g., predicted CQI, queue length, interference budgets, mobility indicators) drive action preferences across regimes, using attribution methods such as integrated gradients [94], Shapley additive explanations (SHAP) [92], or permutation tests [95], complemented by policy probes [96] (e.g., sensitivity to cell-edge load or URLLC risk). Local views deliver case-specific rationales and counterfactuals: "Had the predicted URLLC tail latency been $< 0.2\,\text{ms}$, the PRB reallocation from the AI-native inference slice would not have occurred." Saliency for high-dimensional inputs (e.g., radio maps, spectrum waterfalls) highlights which subbands or beams triggered a decision [97], while contrastive explanations articulate the minimal state change required to flip an action, making remediation actionable [98]. Post-hoc methods provide both global feature importance and local counterfactual rationales, delivering actionable transparency essential for trustworthy AI-RAN in 6G.

Structured surrogates translate black-box policies into compact control logic to aid simulation, verification, and auditing [44]. Distillation into decision trees [99], rule lists [100], linear-threshold ensembles [101], finite-state abstractions [102], or options/hierarchical graphs [103] yields simulatable controllers whose behavior can be stepped through under traffic/channel scenarios. Fidelity metrics (agreement rates, calibrated error bounds) [104] and coverage [105] (the share of states explained by simple rules) quantify when a surrogate suffices and when to fall back to the original policy. To capture delayed effects that are central to RAN control, SCM-backed analyses map state–action–KPI relations and support do-interventions and counterfactual queries across time, exposing temporal credit assignment: e.g., how a conservative power backoff now reduces interference spillover and URLLC

violations several frames later [74]. Structured surrogates distill black-box policies into compact, simulatable controllers, enabling formal verification, scenario stepping, and temporal credit-assignment analysis.

Because explainability without assurance is insufficient for safety-critical slices, XDRL should surface uncertainty and constraints alongside rationales. Uncertainty-aware heads (ensembles, distributional/value quantiles) provide calibrated confidence in Q-values and policies [87, 106]; risk measures (e.g., conditional value at risk (CVaR)) explain how tail performance is controlled [107]; conformal prediction sets report finite-sample guarantees for latency/outage risks [77]. Constraint-aware DRL can externalize dual variables or certificates as part of the explanation payload, linking an action to explicit compliance narratives—power, interference, BLER, and E2E latency budgets—before execution. In multi-agent AI-RAN, explanations must also attribute outcomes across agents and coordination messages ("who influenced whom and why") while remaining lightweight enough for edge inference [15, 81].

HITL integration closes the governance loop. Interactive explanations let operators critique not only outcomes but also rationales [108]; constraint-based steering transforms critiques into guardrails and shields that pre-filter unsafe actions [78]; reward shaping and preference learning align long-horizon objectives with fairness, carbon-awareness, and regulatory policies [91]. Over time, this process builds organization-specific explanation taxonomies (what evidence is needed for which slice and context) and playbooks (what corrective actions follow a given explanation pattern). XDRL ensures safety-critical trustworthiness with HITL governance that enables operators to steer via rewards learning, and evolve slice-specific explanation taxonomies and corrective playbooks for 6G AI-RAN.

Within intelligent network slicing, these capabilities yield tangible benefits. For HTC, explanations decompose how bandwidth, beam selection, and MCS jointly determine instantaneous throughput and jitter [111], with confidence indicators for immersive quality of experience (QoE) guarantees [112, 113]. For TI-DT, temporally grounded narratives justify proactive resource hardening that reduces future URLLC violation probabilities [114]. For AI-native slices, explanations quantify the latency–energy–accuracy trade-offs of edge–cloud offloading and how load shedding decisions are triggered [15]. In USC and SAGSIN, multi-agent attributions and compliance certificates enable cross-domain coordination under tight fronthaul/backhaul and spectrum constraints [115, 116]. To provide a synthesized view of the XDRL landscape, Table 2 categorizes existing approaches based on their mechanisms, objectives,

Table 2: Overview of XDRL axes in RAN slicing: techniques, objectives, applicability, and trade-offs.

Category	Core Techniques	Objectives	Applicability to RAN Slicing	Key Trade-offs (Pros / Cons)	Ref.
Post-hoc Interpretability	Integrated Gradients, SHAP, Permutation Tests, Policy Probes, Rationale & Counterfactual Explanations	Debugging & Actionable Transparency: Providing global and local insight into which features drive decisions.	Root Cause Analysis: Identifying interference sources or bottleneck features in multi-cell coordination; Offline auditing.	(+) Model-Agnostic; Transparency; High Fidelity. (−) High Computational Cost; Lacks Causality Guarantees.	[97] [109] [110]
Symbolic Surrogates	Decision Trees, Logical Rules, Linear-Threshold Ensembles; Finite-State Abstractions and Hierarchical Options/Graphs, SCMs	Verification & Auditing: Converting black-box policies into explicit logic or causal graphs.	Edge Inference: Lightweight rules suitable for resource-constrained RAN nodes; Verifiable admission control.	(+) Fast Inference; Formally Verifiable. (−) Accuracy-Complexity Trade-off; Limited Coverage in Complex Regimes.	[42] [44]
HITL & Assurance	Uncertainty-Aware DRL, Risk Measures, Shielding, Constrained DRL, Interactive RL	Safety & Intent Alignment: Ensuring safety and aligning policies with operator feedback and regulatory requirements.	Conflict Resolution: Managing competing objectives via safety shields in critical slices.	(+) Safety Guarantees; Intent Aligned. (−) Requires Rapid Human Feedback; Conservative Bounds.	[48] [107] [108]

and applicability to RAN slicing. As illustrated, post-hoc methods serve as the baseline for offline debugging and root-cause analysis, whereas symbolic surrogates are uniquely positioned for real-time edge inference due to their low latency. Furthermore, HITL and assurance mechanisms are highlighted as indispensable for safety-critical conflict resolution between slices.

Emerging alongside these techniques, the integration of large models, particularly large language models (LLMs), represents a paradigm shift in bridging the semantic gap between numerical DRL policies and human understanding. Recent frameworks have begun to leverage the reasoning and generative capabilities of these models for intelligent network slicing and open RAN (O-RAN) control. Notably, the pioneering SliceGPT framework [117] integrates a custom-trained GPT-3.5 agent with blockchain to enable dynamic slice brokerage, multi-stakeholder resource sharing, and LLM-driven optimization. In parallel, prompt-augmented multi-agent frameworks utilize domain-specific LLMs with learnable soft prompts to generate semantically structured state representations [118], achieving superior reward performance in dynamic O-RAN scenarios without expensive fine-tuning. Furthermore, recent LLM-based systems [119] demonstrate enhanced resource efficiency by dynamically allocating isolated slices and implementing permission-aware registration. From an orchestration perspective, visionary frameworks now employ LLMs for intent-to-requirement translation and cross-domain lifecycle management [120], effectively handling the complexity of 6G environments. Finally, to address the black-box nature of decision-making, approaches like composable XAI [110] use prompt engineering to convert opaque DRL decisions into human-readable textual explanations, directly addressing the transparency challenge in high-stakes optimization tasks.

To synthesize these methodologies into a unified system, we advocate a practical deployment blueprint "triad": during training, explanation-aligned rewards and uncertainty regularization; at inference, a lightweight explainability head coupled with a safety shield; during operations, human-feedback loops and drift monitoring with surrogate fidelity tracking.

In sum, XDRL reframes DRL for AI-RAN from a high-performance black box into a transparent, risk-aware, and policy-compliant controller. By unifying post-hoc interpretation, symbolic surrogates, causal/counterfactual reasoning, uncertainty and constraint reporting, and HITL governance, XDRL provides the evidential substrate required to deploy learning-based slicing at 6G scale—maintaining performance while earning operator trust and satisfying regulatory and SLA obligations.

4. Evaluation and Testbeds

Rigorous evaluation of explainable DRL for 6G AI-RAN must jointly assess control performance, explanation quality, and assurance value (uncertainty, constraint compliance, safety), and it must do so under realistic traffic, channel, and multi-agent coordination conditions. Rather than isolating metrics and environments into separate silos, we advocate an integrated protocol that: (i) quantifies how well policies optimize slicing objectives; (ii) verifies that explanations are faithful, stable, and operationally useful; (iii) reports evidence and guarantees for risk and constraints; and (iv) validates sim-to-real transfer on progressively more realistic testbeds.

What to measure (metrics). Performance should be captured at user-, cell-, and slice-level: throughput/spectral efficiency, latency and tail latency (e.g., 95/99/99.9th percentiles) [121], reliability/BLER [122], outage/packet loss, SLA satisfaction rate, energy efficiency and carbon footprint, as well as convergence/stability and sample efficiency during training. Fairness across users and slices (e.g., Jain's index [123], minimum-rate guarantees, cell-edge protection) and cross-layer efficiency (PRB utilization, power budgets, interference spillover) are equally central. Explainability requires multi-faceted criteria: fidelity (agreement between explanations/surrogates and the underlying policy), coverage (fraction of states/actions for which a concise explanation applies) [124], simplicity (size/depth of trees, rule count, symbol complexity), stability/robustness (consistency of explanations under small input perturbations and across time), counterfactual validity (do predicted minimal changes actually flip actions) [125], and temporal adequacy (can explanations account for delayed effects relevant to queueing and interference dynamics) [73]. Because explainability without guarantees is insufficient for safety-critical slices, assurance metrics should accompany every report: uncertainty calibration (expected calibration error for value/policy heads) [126], risk control (CVaR or tail-risk at target quantiles) [107], constraint satisfaction (power, interference, latency, BLER budgets met per-step, not only on average) [88], and certificate quality (tightness of conformal bounds or Lagrange multipliers exposed by CMDP/Lyapunov methods). For learning from logs or shadow-mode operation, include off-policy evaluation diagnostics (e.g., importance sampling variants, doubly robust estimators, effective sample size) and sensitivity analyses to logging-policy drift [127, 128]. Finally, human factors—operator trust, task completion time, error detection uplift, and perceived usefulness—should

Table 3: Evaluation framework for AI-RAN.

Dimension	Core Metrics	Key Insights
Network Performance	Throughput/Spectral Efficiency Latency (mean, P95/P99/P99.9) Reliability (BLER) Outage/Packet Loss SLA Satisfaction Rate Energy & Carbon Efficiency	Track performance across multiple granularities (user, cell, slice) to ensure comprehensive QoS assessment and identify bottlenecks
Fairness & Resource Efficiency	Jain's Fairness Index Minimum-rate Guarantees Cell-edge Protection PRB Utilization Power Budget Adherence Interference Spillover	Ensure balanced allocation across users/slices and cross-layer efficiency
Explainability Quality	Fidelity (explanation-policy agreement) Coverage (fraction of explainable states) Simplicity (tree depth, rule count) Stability (robustness to perturbations) Counterfactual Validity Temporal Adequacy	Multi-faceted validation of explanation utility, not just interpretability
Safety Assurance	Uncertainty Calibration (ECE) Risk Control (CVaR at target quantiles) Per-step Constraint Satisfaction Certificate Tightness (conformal bounds)	Quantified guarantees mandatory for safety-critical slices
Offline Learning Validity	Off-policy Evaluation (IS, DR estimators) Effective Sample Size Sensitivity to Logging-policy Drift	Essential diagnostics before shadow-mode or live deployment
Human Factors	Operator Trust Task Completion Time Error Detection Uplift Perceived Usefulness	Measured via controlled studies or field logs; explanations must improve operations, not just scores

be measured via controlled studies or field logs to verify that explanations improve operations, not just scores. Table 3 summarizes our evaluation framework spanning network performance, fairness, explainability, safety, offline learning validity, and human factors—each essential for validating AI-RAN systems deployments. Additionally, the computational complexity and inference latency overhead associated with explainability mechanisms represent an important yet under-examined research direction. Although existing XDRL methods typically report manageable computational costs within their evaluation frameworks, standardized cross-method comparisons against vanilla DRL baselines on large-scale real-world platforms remain scarce. Notably, emerging evidence suggests that interpretability and efficiency may not necessarily be mutually exclusive: well-architected XDRL methods have shown potential to maintain or improve E2E latency by producing policies that converge faster or require fewer online environment interactions [48].

How to evaluate (protocols). Start with controlled simulation to stress distinct regimes (stationary vs. bursty HTC, URLLC with deterministic latency bounds, mobility gradients, interference-limited cells). Use train/validation/test splits in the space of traffic/channel seeds to prevent overfitting explanations to familiar regimes. Progress to hardware-in-the-loop (HIL) [129] and over-the-air emulation where the policy runs on edge hardware with realistic timing, and continue into shadow deployment [130] in live networks: the policy makes "ghost" decisions while the legacy controller acts, enabling counterfactual auditing and safe A/B trials with canary cells and automatic rollback [131]. Throughout, log explanations with their confidence and constraint evidence, and audit drift [82] by tracking surrogate fidelity and explanation stability over weeks.

Where to evaluate (testbeds and simulators). A diversified stack of environments reduces the sim-to-real gap. Packet-level simulators such as NS-3 [132] and OMNeT++ families [133] (e.g., SimuLTE/Simu5G [134]) enable detailed PHY/MAC modeling, realistic scheduler hooks, and trace-driven traffic. Software RANs like OpenAirInterface [135] or srsRAN (4G/5G) [136] allow real-time experiments with software-defined radios (SDRs) and commodity hardware, exposing practical fronthaul/backhaul and timing constraints. Large-scale wireless emulators (e.g., the Colosseum-class platforms [137, 138]) provide multi-node, interferer-rich, mobility-aware channels with real-time control, ideal for testing robustness and multi-agent coordination.

Table 4: Testbed and simulation stack for AI-RAN.

Category	Representative Platforms	Validation Capabilities	Realism Trade-off
Packet-Level Simulators	NS-3 [132], OMNeT++ [133], SimuLTE/Simu5G [134]	Detailed PHY/MAC modeling, scheduler hooks, trace-driven traffic, large-scale parameter sweeps	High control, low hardware cost; limited channel/timing realism
Software RANs	OpenAirInterface [135], srsRAN(4G/5G) [136]	Real-time execution with SDRs, fronthaul/backhaul constraints, practical timing effects	Medium scale, exposes implementation bugs; requires RF expertise
Large-scale Wireless Emulators	Colosseum [137, 138]	Multi-node RF emulation, mobility, interference, multi-agent coordination at scale	Controlled repeatability with realistic propagation; costly infrastructure
City-Scale Testbeds	POWDER [139], COSMOS [140], Arena-type [141] deployments	Over-the-air trials, heterogeneous radios, edge compute, cross-domain (space/air/ground) integration	Full realism, uncontrolled interference; limited reproducibility
O-RAN-Compliant Environments	Near-RT RIC with xApps/rApps, E2 interface testbeds [142]	Policy governance, explainability artifact integration, SLA auditing via control messages	Standards-aligned deployment; API maturity varies
DRL Toolkits	Gym-like APIs [143]	Unified observation/action spaces, reproducible logging, hyperparameter tuning	Abstraction layer; must preserve domain fidelity

City-scale testbeds (e.g., POWDER [139], COSMOS [140], Arena-type [141] deployments) support over-the-air trials with heterogeneous radios and edge compute, essential for space–air–ground–sea integration scenarios and cross-domain coordination. O-RAN–compliant environments (near-real-time RAN intelligent controller (Near-RT RIC) with xApps (Near-RT RIC applications), rApps (non-real-time RIC (Non-RT RIC) applications), and E2 interfaces) [142] are crucial to evaluate explainability artifacts in place: explanations and certificates can be attached to control messages, enabling policy governance and SLA audits. Finally, DRL toolkits (Gym-like APIs [143]) are useful as scaffolding to wrap the above environments with standardized observation/action spaces, logging, and reproducible evaluation harnesses. Table 4 presents an overview of representative testbed and simulation platforms across different types. Each platform type offers distinct validation capabilities and presents different trade-offs between control, scalability, and realism.

What to report (results and evidence). Beyond headline throughput/latency, results should present: (i) Pareto frontiers [144] showing trade-offs among performance, fairness, and energy; (ii) risk curves (tail latency vs. reliability target) and constraint dashboards with per-step violation rates; (iii)

fidelity/coverage–complexity plots for surrogates to demonstrate when simple explanations suffice; (iv) counterfactual audits quantifying the fraction of explanations whose predicted minimal changes actually flip actions in replay or HIL, with confidence intervals; (v) temporal attributions that expose multi-slot causal chains (e.g., how proactive power backoff reduces URLLC violations several frames later); and (vi) HITL gains (operator task time reduction, incident detection lift) from user studies or field logs. Robustness studies should include distribution shifts (festival spikes, codec upgrades for HTC, mobility surges), ablations of measurement noise and KPI spoofing, and multi-agent stress where coordination messages are delayed or dropped.

Putting it together. We recommend a three-stage pipeline: (1) Simulation stage to iterate on algorithms with full observability, rich counterfactual logging, and cheap ablations; (2) Emulation/HIL stage to test timing, SDR-in-the-loop PHY impacts, and explanation latency/overhead; (3) Shadow/limited rollout stage with canary cells under an O-RAN framework, where every action is paired with an explanation, uncertainty, and constraint certificate, and automatic shields veto unsafe actions. Success criteria span performance (SLA compliance and fairness), explainability (fidelity, coverage, stability, and counterfactual validity), and assurance (calibration, risk, and zero-violation guarantees in hard-constrained slices). This E2E approach ensures that learning-based slicing is not only fast and efficient but also transparent, risk-aware, and policy-compliant in the operational realities of 6G AI-RAN.

5. Open Challenges and Future Directions

Despite rapid progress, deploying XDRL at 6G scale remains an open endeavor that spans algorithms, systems, and governance. We highlight key gaps and outline concrete research directions, integrating scalability, generalization, language-mediated explanations, intent alignment, and ethics/ regulation into a single roadmap tailored to AI-RAN and intelligent network slicing.

5.1. Scalability of XDRL at 6G Scale.

As slices, cells, and control knobs proliferate (PRB assignment, power, beamforming, numerology, edge offloading), both policies and explanations must scale in space (high-dimensional state/action), time (multi-slot effects), and agents (distributed control) [145]. Lightweight, streaming explanations

that operate within edge compute budgets are needed: on-device attribution heads with bounded latency/overhead [146]; policy distillation into compact surrogates whose fidelity–coverage–complexity frontier is explicitly tracked [147]; and hierarchical/option-based abstractions that expose human-comprehensible sub-policies (e.g., "protect TI-DT control plane," "shed HTC under interference surge") [103]. Future work should formalize anytime explainability—producing progressively refined rationales under tight deadlines—and investigate efficient multi-agent attributions that identify "who influenced whom and why" without saturating fronthaul/backhaul. A promising direction is budgeted XDRL [148], where explanations are optimized jointly with control under compute/energy/time constraints and come with accuracy–latency trade-off certificates.

5.2. Generalization Under Traffic, Mobility, and Interference Shifts.

Real networks exhibit bursty loads, codec upgrades, handover storms, and emergent slices. Explanations that overfit to training regimes become misleading when distributions shift [149]. Beyond standard data augmentation, future XDRL should incorporate invariance discovery (e.g., causal representation learning that factors out spurious correlations) [150, 151], risk-sensitive objectives (CVaR-aware policies with tail-focused explanations) [152], and uncertainty-calibrated [126] rationales that degrade gracefully ("low-confidence explanation — falling back to conservative control"). Off-policy evaluation with doubly robust estimators [127] and sensitivity analyses must be extended to the explanation layer, quantifying how robust rationales remain when replayed under new traffic/channel seeds. We also advocate explanation-time domain randomization [153]: stress-testing explanations by injecting controlled perturbations to KPIs and measurements to assess stability.

5.3. Language-enabled Explanations via LLMs—Grounded and Verifiable

LLMs can translate symbolic traces, counterfactuals, and certificates into operator-facing narratives, but must be prevented from hallucinating [154]. Research is needed on grounded natural language explanations (NLEs) that are compiled from structured evidence [155]: feature attributions with confidence, SCM-backed do-intervention outcomes, constraint duals from CMDPs/ Lyapunov methods, and conformal risk bounds. Tool-augmented LLMs [156] should be restricted to read-only access to these artifacts and required to emit traceable claims (each sentence linked to an evidence token), enabling automated explanation linting for consistency. Additional directions include

operator preference learning from language feedback; bilingual/controlled-vocabulary NLEs [157] for global operations centers; and safety prompts [158] that force disclosure of uncertainty, alternatives, and expected impact on SLAs.

5.4. Alignment with Intent-Based Networking (IBN).

Operators express high-level intents ("keep URLLC tail latency < 1 ms at 99.999%," "ensure HTC smoothness above mean opinion score (MOS) threshold while capping energy") that must be compiled into constraints, rewards, and monitors [159]. XDRL should provide a bidirectional mapping: (i) intent $\rightarrow$ policy & shield, translating intents into CMDP constraints [90], risk targets, and reward shaping; (ii) policy $\rightarrow$ intent compliance report, generating per-action explanations that reference the intents, including counterfactuals ("if the HTC demand had been 15% lower, no PRB reallocation would be triggered") and pre-execution assurance ("predicted violation probability $< 10^{-5}$"). Future work includes intent refinement (suggesting minimally invasive edits to intents when infeasible), conflict arbitration among competing intents, and lifecycle management where explanation drift flags when a deployed policy no longer realizes the original intent.

5.5. Ethical, Regulatory, and Auditing Requirements

With learning-based control affecting service access and quality, explanations must support fairness, privacy, and accountability [160]. Research priorities include: (i) fairness-aware XDRL that surfaces how actions impact protected or disadvantaged cohorts (e.g., cell-edge users) and quantifies trade-offs on Pareto frontiers; (ii) privacy-preserving explanations (e.g., differential privacy (DP) aware attribution that releases aggregate rationales without leaking user-level data) [161]; (iii) audit-ready artifacts [162] that bind each action to evidence—uncertainty calibration metrics, constraint certificates, and SCM-backed causal justifications—signed and stored for post-incident analysis; and (iv) red-teaming explanations [163] against adversarial KPI spoofing or social engineering, with detectors that cross-check narratives against machine-verifiable traces. Coordination with O-RAN ecosystems is key so that explanations and certificates can travel alongside E2 control messages for real-time governance.

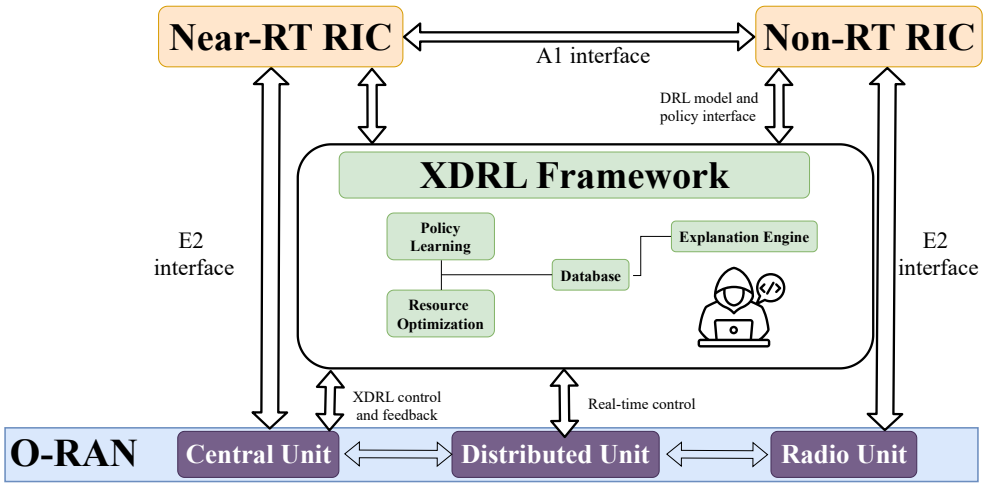


Figure 2: The XDRL-O-RAN integration architecture, featuring the RIC layer, XDRL framework layer, and O-RAN component layer, interconnected through standardized interfaces for data exchange and intelligent decision optimization.

5.6. Bridging Sim-to-Real and Standardization

To ensure transferability, both explanations and policies should be evaluated jointly across a ladder of environments, ranging from packet-level simulators (NS-3/Simu5G) to software-defined RANs (srsRAN /OpenAirInterface), wireless emulators (Colosseum-class), and large-scale city testbeds. Throughout this process, it is crucial to maintain consistent logging APIs for attributions, counterfactuals, and certification purposes. Fig. 2 illustrates the complete integration architecture of XDRL with O-RAN. The architecture consists of three primary layers: (1) The top-tier RIC layer comprises the Non-RT RIC and Near-RT RIC, interconnected via the A1 interface for policy management and model distribution; (2) The XDRL framework layer integrates multiple reinforcement learning agents and a policy engine responsible for intelligent decision optimization; (3) The O-RAN component layer includes the central unit, distributed unit, and radio unit, coordinated through the interfaces. The Near-RT RIC exchanges real-time control and monitoring data with O-CU/O-DU via the E2 interface. This E2E architecture supports intelligent management from user equipment to intelligent control.

5.7. Putting the Roadmap Into Practice

We envision a deployable stack where: (1) training integrates explanation-guided rewards [164] and uncertainty regularization; (2) inference includes a lightweight explanation head and a safety shield that vetoes actions lacking sufficient confidence or violating certificates; (3) operations run HITL critique with drift monitors on both the policy and explanations; and (4) LLM-based NLEs are strictly grounded in machine-checkable artifacts. Success will be measured by consistent SLA attainment across slices, low tail-risk under shifts, high surrogate fidelity at bounded complexity, valid counterfactual rates, and demonstrated operator trust gains in field studies. Advancing along this path will turn learning-based slicing from a high-performance black box into transparent, risk-aware, and policy-compliant control suitable for the operational realities of 6G AI-RAN.

6. Conclusion

This paper positioned XDRL as the linchpin for trustworthy, large-scale automation in 6G AI-RAN, where intelligent network slicing must satisfy stringent SLAs under non-stationary traffic, partial observability, and tight safety constraints. We synthesized a holistic toolkit that unifies post-hoc interpretation, symbolic surrogates, causal/counterfactual reasoning, uncertainty calibration, and constraint-aware control—embedded within a HITL governance loop—to transform high-performance black-box policies into transparent, risk-aware, and policy-compliant controllers. We proposed evaluation protocols and testbed pathways that couple performance with explanation fidelity, coverage, stability, counterfactual validity, and assurance metrics, facilitating sim-to-real transfer and auditability. Finally, we outlined a forward-looking roadmap on scalability, generalization, language-grounded explanations, intent alignment, and ethics/regulation. Advancing along this path will enable learning-based slicing to deliver not only peak efficiency but also verifiable reliability, fairness, and accountability in operational 6G networks.

References

[1] S. Zhang, "An overview of network slicing for 5G," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 111–117, 2019.

[2] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5G: Survey and challenges," *IEEE communications magazine*, vol. 55, no. 5, pp. 94–100, 2017.

[3] W. Wu, C. Zhou, M. Li, H. Wu, H. Zhou, N. Zhang, X. S. Shen, and W. Zhuang, "AI-native network slicing for 6G networks," *IEEE Wireless Communications*, vol. 29, no. 1, pp. 96–103, 2022.

[4] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini, and H. Flinck, "Network slicing and softwarization: A survey on principles, enabling technologies, and solutions," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2429–2453, 2018.

[5] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, "5G wireless network slicing for eMBB, URLLC, and mMTC: A communication-theoretic view," *Ieee Access*, vol. 6, pp. 55 765–55 779, 2018.

[6] E. J. dos Santos, R. D. Souza, J. L. Rebelatto, and H. Alves, "Network slicing for URLLC and eMBB with max-matching diversity channel allocation," *IEEE communications letters*, vol. 24, no. 3, pp. 658–661, 2019.

[7] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What should 6G be?" *Nature Electronics*, vol. 3, no. 1, pp. 20–29, 2020.

[8] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6G: A comprehensive survey," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 334–366, 2021.

[9] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjöland, and F. Tufvesson, "6G wireless systems: Vision, requirements, challenges, insights, and opportunities," *Proceedings of the IEEE*, vol. 109, no. 7, pp. 1166–1199, 2021.

[10] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE network*, vol. 34, no. 3, pp. 134–142, 2019.

[11] C.-X. Wang, X. You, X. Gao, X. Zhu, Z. Li, C. Zhang, H. Wang, Y. Huang, Y. Chen, H. Haas *et al.*, "On the road to 6G: Visions,

requirements, key technologies, and testbeds," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 905–974, 2023.

[12] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE communications magazine*, vol. 57, no. 8, pp. 84–90, 2019.

[13] Y. Siriwardhana, P. Porambage, M. Liyanage, and M. Ylianttila, "AI and 6G security: Opportunities and challenges," in *2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*. IEEE, 2021, pp. 616–621.

[14] N. A. Khan and S. Schmid, "AI-RAN in 6G networks: State-of-the-art and challenges," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 294–311, 2023.

[15] K. B. Letaief, Y. Shi, J. Lu, and J. Lu, "Edge artificial intelligence for 6G: Vision, enabling technologies, and applications," *IEEE journal on selected areas in communications*, vol. 40, no. 1, pp. 5–36, 2021.

[16] Y. Lu and X. Zheng, "6G: A survey on technologies, scenarios, challenges, and the related issues," *Journal of Industrial Information Integration*, vol. 19, p. 100158, 2020.

[17] M. Yaqoob, R. Trestian, M. Tatipamula, and H. X. Nguyen, "Digital-twin-driven end-to-end network slicing toward 6G," *IEEE Internet Computing*, vol. 28, no. 2, pp. 47–55, 2023.

[18] K. Liang, W. Guo, Z. Li, C. Li, C. Ma, K.-K. Wong, and C.-B. Chae, "Customizable and robust internet of robots based on network slicing and digital twin," *IEEE Network*, vol. 38, no. 3, pp. 17–24, 2024.

[19] Y. Cui, F. Liu, X. Jing, and J. Mu, "Integrating sensing and communications for ubiquitous IoT: Applications, trends, and challenges," *IEEE network*, vol. 35, no. 5, pp. 158–167, 2021.

[20] H. Guo, J. Li, J. Liu, N. Tian, and N. Kato, "A survey on space-air-ground-sea integrated network security in 6G," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 53–87, 2021.

[21] N. Cheng, H. Jingchao, Y. Zhisheng, Z. Conghao, W. Huaqing, L. Feng, Z. Haibo, and S. Xuemin, "6G service-oriented space-air-ground integrated network: A survey," *Chinese Journal of Aeronautics*, vol. 35, no. 9, pp. 1–18, 2022.

[22] H. F. Alhashimi, M. N. Hindia, K. Dimyati, E. B. Hanafi, N. Safie, F. Qamar, K. Azrin, and Q. N. Nguyen, "A survey on resource management for 6G heterogeneous networks: current research, future trends, and challenges," *Electronics*, vol. 12, no. 3, p. 647, 2023.

[23] S. Jošilo and G. Dán, "Joint wireless and edge computing resource management with dynamic network slice selection," *IEEE/ACM Transactions on Networking*, vol. 30, no. 4, pp. 1865–1878, 2022.

[24] J. Feng, Q. Pei, F. R. Yu, X. Chu, J. Du, and L. Zhu, "Dynamic network slicing and resource allocation in mobile edge computing systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 7, pp. 7863–7878, 2020.

[25] Y.-H. Chen, "An adaptive heuristic algorithm to solve the network slicing resource management problem," *International Journal of Communication Systems*, vol. 36, no. 8, p. e5463, 2023.

[26] Z. Sasan and S. Khorsandi, "Balancing resource utilization and slice dissatisfaction through dynamic soft slicing for 6G wireless networks," *Scientific Reports*, vol. 15, no. 1, p. 22987, 2025.

[27] J. J. A. Esteves, A. Boubendir, F. Guillemin, and P. Sens, "Heuristic for edge-enabled network slicing optimization using the "power of two choices"," in *2020 16th International Conference on Network and Service Management (CNSM)*. IEEE, 2020, pp. 1–9.

[28] N. Sen *et al.*, "Intelligent admission and placement of O-RAN slices using deep reinforcement learning," in *2022 IEEE 8th International Conference on Network Softwarization (NetSoft)*. IEEE, 2022, pp. 307–311.

[29] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y.-C. Liang, and D. I. Kim, "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE communications surveys & tutorials*, vol. 21, no. 4, pp. 3133–3174, 2019.

[30] C. Zhu, M. Dastani, and S. Wang, "A survey of multi-agent deep reinforcement learning with communication," *Autonomous Agents and Multi-Agent Systems*, vol. 38, no. 1, p. 4, 2024.

[31] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE signal processing magazine*, vol. 34, no. 6, pp. 26–38, 2017.

[32] Y. Wei, F. R. Yu, M. Song, and Z. Han, "Joint optimization of caching, computing, and radio resources for fog-enabled IoT using natural actor–critic deep reinforcement learning," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2061–2073, 2018.

[33] C. Liu, M. Xu, Y. Yang, and N. Geng, "DRL-OR: Deep reinforcement learning-based online routing for multi-type service requirements," in *IEEE INFOCOM 2021-IEEE Conference on Computer Communications*. IEEE, 2021, pp. 1–10.

[34] S. Chinchali, P. Hu, T. Chu, M. Sharma, M. Bansal, R. Misra, M. Pavone, and S. Katti, "Cellular network traffic scheduling with deep reinforcement learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[35] J. A. Hurtado Sanchez, K. Casilimas, and O. M. Caicedo Rendon, "Deep reinforcement learning for resource management on network slicing: A survey," *Sensors*, vol. 22, no. 8, p. 3031, 2022.

[36] A. Gharehgoli, A. Nouruzi, N. Mokari, P. Azmi, M. R. Javan, and E. A. Jorswieck, "AI-based resource allocation in end-to-end network slicing under demand and CSI uncertainties," *IEEE Transactions on Network and Service Management*, vol. 20, no. 3, pp. 3630–3651, 2023.

[37] Y. Cai, P. Cheng, Z. Chen, M. Ding, B. Vucetic, and Y. Li, "Deep reinforcement learning for online resource allocation in network slicing," *IEEE Transactions on Mobile Computing*, vol. 23, no. 6, pp. 7099–7116, 2023.

[38] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins *et al.*,

"Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information fusion*, vol. 58, pp. 82–115, 2020.

[39] M. M. Morovati, F. Tambon, M. Taraghi, A. Nikanjam, and F. Khomh, "Common challenges of deep reinforcement learning applications development: an empirical study," *Empirical Software Engineering*, vol. 29, no. 4, p. 95, 2024.

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

[41] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature machine intelligence*, vol. 1, no. 5, pp. 206–215, 2019.

[42] A. Duttagupta, M. Jabbari, C. Fiandrino, M. Fiore, and J. Widmer, "SymbXRL: symbolic explainable deep reinforcement learning for mobile networks," in *IEEE INFOCOM 2025-IEEE Conference on Computer Communications*. IEEE, 2025, pp. 1–10.

[43] G. A. Vouros, "Explainable deep reinforcement learning: state of the art and challenges," *ACM Computing Surveys*, vol. 55, no. 5, pp. 1–39, 2022.

[44] A. Verma, V. Murali, R. Singh, P. Kohli, and S. Chaudhuri, "Programmatically interpretable reinforcement learning," in *International conference on machine learning*. PMLR, 2018, pp. 5045–5054.

[45] O. Bastani, Y. Pu, and A. Solar-Lezama, "Verifiable reinforcement learning via policy extraction," *Advances in neural information processing systems*, vol. 31, 2018.

[46] Y. Coppens, K. Efthymiadis, T. Lenaerts, A. Nowé, T. Miller, R. Weber, and D. Magazzeni, "Distilling deep reinforcement learning policies in soft decision trees," in *Proceedings of the IJCAI 2019 workshop on explainable artificial intelligence*, 2019, pp. 1–6.

[47] A. Heuillet, F. Couthouis, and N. Díaz-Rodríguez, "Explainability in deep reinforcement learning," *Knowledge-Based Systems*, vol. 214, p. 106685, 2021.

[48] F. Rezazadeh, H. Chergui, and J. Mangues-Bafalluy, "Explanation-guided deep reinforcement learning for trustworthy 6G RAN slicing," in *2023 ieee international conference on communications workshops (icc workshops)*. IEEE, 2023, pp. 1026–1031.

[49] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, "6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 957–975, 2020.

[50] T. Senevirathna, V. H. La, S. Marcha, B. Siniarski, M. Liyanage, and S. Wang, "A survey on XAI for 5G and beyond security: Technical aspects, challenges and research directions," *IEEE Communications Surveys & Tutorials*, vol. 27, no. 2, pp. 941–973, 2025.

[51] H. Sun, Y. Liu, A. Al-Tahmeesschi, A. Nag, M. Soleimanpour-Moghadam, B. Canberk, H. Arslan, and H. Ahmadi, "Advancing 6G: Survey for explainable AI on communications and network slicing," *IEEE Open Journal of the Communications Society*, 2025.

[52] S. Padakandla, P. KJ, and S. Bhatnagar, "Reinforcement learning algorithm for non-stationary environments," *Applied Intelligence*, vol. 50, no. 11, pp. 3590–3606, 2020.

[53] M. Bennis, M. Debbah, and H. V. Poor, "Ultrareliable and low-latency wireless communication: Tail, risk, and scale," *Proceedings of the IEEE*, vol. 106, no. 10, pp. 1834–1853, 2018.

[54] J. Eschmann, "Reward function design in reinforcement learning," *Reinforcement learning algorithms: Analysis and Applications*, pp. 25–33, 2021.

[55] Y. Wang, L. Zhao, X. Chu, S. Song, Y. Deng, A. Nallanathan, and K. Liang, "Deep reinforcement learning-based optimization for end-to-end network slicing with control- and user-plane separation," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 11, pp. 12 179–12 194, 2022.

[56] X. Huang, S. Leng, S. Maharjan, and Y. Zhang, "Multi-agent deep reinforcement learning for computation offloading and interference coordination in small cell networks," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 9, pp. 9282–9293, 2021.

[57] J. Chen, O. Esrafilian, H. Bayerlein, D. Gesbert, and M. Caccamo, "Model-aided federated reinforcement learning for multi-UAV trajectory planning in IoT networks," in *2023 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2023, pp. 818–823.

[58] A. M. Seid, G. O. Boateng, B. Mareri, G. Sun, and W. Jiang, "Multi-agent DRL for task offloading and resource allocation in multi-UAV enabled IoT edge network," *IEEE Transactions on Network and Service Management*, vol. 18, no. 4, pp. 4531–4547, 2021.

[59] X. Li, H. Zhang, H. Zhou, N. Wang, K. Long, S. Al-Rubaye, and G. K. Karagiannidis, "Multi-agent DRL for resource allocation and cache design in terrestrial-satellite networks," *IEEE Transactions on Wireless Communications*, vol. 22, no. 8, pp. 5031–5042, 2022.

[60] A. Feriani and E. Hossain, "Single and multi-agent deep reinforcement learning for AI-enabled wireless networks: A tutorial," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 1226–1252, 2021.

[61] M. U. A. Siddiqui, H. Abumarshoud, L. Bariah, S. Muhaidat, M. A. Imran, and L. Mohjazi, "URLLC in beyond 5G and 6G networks: An interference management perspective," *IEEe Access*, vol. 11, pp. 54 639–54 663, 2023.

[62] J. Garcıa and F. Fernández, "A comprehensive survey on safe reinforcement learning," *Journal of Machine Learning Research*, vol. 16, no. 1, pp. 1437–1480, 2015.

[63] S. Gu, L. Yang, Y. Du, G. Chen, F. Walter, J. Wang, and A. Knoll, "A review of safe reinforcement learning: Methods, theories and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[64] B. Han, Z. Ren, Z. Wu, Y. Zhou, and J. Peng, "Off-policy reinforcement learning with delayed rewards," in *International conference on machine learning*. PMLR, 2022, pp. 8280–8303.

[65] R. S. Sutton, A. G. Barto *et al.*, *Reinforcement learning: An introduction*. MIT press Cambridge, 1998, vol. 1, no. 1.

[66] M. Uehara, C. Shi, and N. Kallus, "A review of off-policy evaluation in reinforcement learning," *arXiv preprint arXiv:2212.06355*, 2022.

[67] C. Voloshin, H. M. Le, N. Jiang, and Y. Yue, "Empirical study of off-policy policy evaluation for reinforcement learning," *arXiv preprint arXiv:1911.06854*, 2019.

[68] I. Ilahi, M. Usama, J. Qadir, M. U. Janjua, A. Al-Fuqaha, D. T. Hoang, and D. Niyato, "Challenges and countermeasures for adversarial attacks on deep reinforcement learning," *IEEE Transactions on Artificial Intelligence*, vol. 3, no. 2, pp. 90–109, 2021.

[69] J. Fu, M. Norouzi, O. Nachum, G. Tucker, Z. Wang, A. Novikov, M. Yang, M. R. Zhang, Y. Chen, A. Kumar *et al.*, "Benchmarks for deep off-policy evaluation," *arXiv preprint arXiv:2103.16596*, 2021.

[70] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," in *International conference on machine learning*. Pmlr, 2014, pp. 387–395.

[71] S. Milani, N. Topin, M. Veloso, and F. Fang, "Explainable reinforcement learning: A survey and comparative review," *ACM Computing Surveys*, vol. 56, no. 7, pp. 1–36, 2024.

[72] G. Ma, X. Wang, M. Hu, W. Ouyang, X. Chen, and Y. Li, "DRL-based computation offloading with queue stability for vehicular-cloud-assisted mobile edge computing systems," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 4, pp. 2797–2809, 2022.

[73] E. Pignatelli, J. Ferret, M. Geist, T. Mesnard, H. van Hasselt, O. Pietquin, and L. Toni, "A survey of temporal credit assignment in deep reinforcement learning," *arXiv preprint arXiv:2312.01072*, 2023.

[74] P. Madumal, T. Miller, L. Sonenberg, and F. Vetere, "Explainable reinforcement learning through a causal lens," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 03, 2020, pp. 2493–2500.

[75] D. Vale, A. El-Sharif, and M. Ali, "Explainable artificial intelligence (XAI) post-hoc explainability methods: Risks and limitations in non-discrimination law," *AI and Ethics*, vol. 2, no. 4, pp. 815–826, 2022.

[76] Y. Chow, O. Nachum, E. Duenez-Guzman, and M. Ghavamzadeh, "A Lyapunov-based approach to safe reinforcement learning," *Advances in neural information processing systems*, vol. 31, 2018.

[77] A. N. Angelopoulos, S. Bates, A. Fisch, L. Lei, and T. Schuster, "Conformal risk control," *arXiv preprint arXiv:2208.02814*, 2022.

[78] M. Alshiekh, R. Bloem, R. Ehlers, B. Könighofer, S. Niekum, and U. Topcu, "Safe reinforcement learning via shielding," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.

[79] G. Ananthanarayanan, X. Foukas, B. Radunovic, and Y. Zhang, "Distributed AI platform for the 6G RAN," *arXiv preprint arXiv:2410.03747*, 2024.

[80] Z. Ding, T. Huang, and Z. Lu, "Learning individually inferred communication for multi-agent cooperation," *Advances in neural information processing systems*, vol. 33, pp. 22 069–22 079, 2020.

[81] X. Du, Y. Ye, P. Zhang, Y. Yang, M. Chen, and T. Wang, "Situation-dependent causal influence-based cooperative multi-agent reinforcement learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 16, 2024, pp. 17 362–17 370.

[82] F. Hinder, V. Vaquet, and B. Hammer, "One or two things we know about concept drift—a survey on monitoring in evolving environments. part a: detecting concept drift," *Frontiers in Artificial Intelligence*, vol. 7, p. 1330257, 2024.

[83] A. M. Rahimi, A. Ziaeddini, and S. Gonglee, "A novel approach to efficient resource allocation in load-balanced cellular networks using hierarchical DRL," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 5, pp. 2887–2901, 2022.

[84] Y. Cai, W. Li, X. Meng, W. Zheng, C. Chen, and Z. Liang, "Adaptive contrastive learning based network latency prediction in 5G URLLC scenarios," *Computer Networks*, vol. 240, p. 110185, 2024.

[85] S. Verma, V. Boonsanong, M. Hoang, K. Hines, J. Dickerson, and C. Shah, "Counterfactual explanations and algorithmic recourses for machine learning: A review," *ACM Computing Surveys*, vol. 56, no. 12, pp. 1–42, 2024.

[86] P. Sunehag, G. Lever, A. Gruslys, W. M. Czarnecki, V. Zambaldi, M. Jaderberg, M. Lanctot, N. Sonnerat, J. Z. Leibo, K. Tuyls *et al.*, "Value-decomposition networks for cooperative multi-agent learning," *arXiv preprint arXiv:1706.05296*, 2017.

[87] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," *Advances in neural information processing systems*, vol. 30, 2017.

[88] J. Achiam, D. Held, A. Tamar, and P. Abbeel, "Constrained policy optimization," in *International conference on machine learning*. PMLR, 2017, pp. 22–31.

[89] Q. Liu, N. Choi, and T. Han, "Constraint-aware deep reinforcement learning for end-to-end resource orchestration in mobile networks," in *2021 IEEE 29th International Conference on Network Protocols (ICNP)*. IEEE, 2021, pp. 1–11.

[90] E. Altman, *Constrained Markov decision processes*. Routledge, 2021.

[91] P. F. Christiano, J. Leike, T. Brown, M. Martic, S. Legg, and D. Amodei, "Deep reinforcement learning from human preferences," *Advances in neural information processing systems*, vol. 30, 2017.

[92] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. of NIPS*, 2017, p. 4768–4777.

[93] M. T. Ribeiro, S. Singh, and C. Guestrin, ""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*, 2016, pp. 1135–1144.

[94] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *International conference on machine learning*. PMLR, 2017, pp. 3319–3328.

[95] A. Altmann, L. Toloşi, O. Sander, and T. Lengauer, "Permutation importance: a corrected feature importance measure," *Bioinformatics*, vol. 26, no. 10, pp. 1340–1347, 2010.

[96] G. Alain and Y. Bengio, "Understanding intermediate layers using linear classifier probes," *arXiv preprint arXiv:1610.01644*, 2016.

[97] H. Zhou, J. Bai, Y. Wang, J. Ren, X. Yang, and L. Jiao, "Deep radio signal clustering with interpretability analysis based on saliency map," *Digital Communications and Networks*, vol. 10, no. 5, pp. 1448–1458, 2024.

[98] A.-H. Karimi, G. Barthe, B. Schölkopf, and I. Valera, "A survey of algorithmic recourse: contrastive explanations and consequential recommendations," *ACM Computing Surveys*, vol. 55, no. 5, pp. 1–29, 2022.

[99] N. Frosst and G. Hinton, "Distilling a neural network into a soft decision tree," *arXiv preprint arXiv:1711.09784*, 2017.

[100] W. M. Czarnecki, R. Pascanu, S. Osindero, S. Jayakumar, G. Swirszcz, and M. Jaderberg, "Distilling policy distillation," in *The 22nd international conference on artificial intelligence and statistics*. PMLR, 2019, pp. 1331–1340.

[101] A. Malinin, B. Mlodozeniec, and M. Gales, "Ensemble distribution distillation," *arXiv preprint arXiv:1905.00076*, 2019.

[102] K. Mallik, A.-K. Schmuck, S. Soudjani, and R. Majumdar, "Compositional synthesis of finite-state abstractions," *IEEE Transactions on Automatic Control*, vol. 64, no. 6, pp. 2629–2636, 2018.

[103] P.-L. Bacon, J. Harb, and D. Precup, "The option-critic architecture," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2017.

[104] N. Posocco and A. Bonnefoy, "Estimating expected calibration errors," in *International conference on artificial neural networks*. Springer, 2021, pp. 139–150.

[105] X. Xu, R. Beckett, K. Jayaraman, R. Mahajan, and D. Walker, "Test coverage metrics for the network," in *Proceedings of the 2021 ACM SIGCOMM 2021 Conference*, 2021, pp. 775–787.

[106] M. G. Bellemare, W. Dabney, and R. Munos, "A distributional perspective on reinforcement learning," in *International conference on machine learning*. PMLR, 2017, pp. 449–458.

[107] Y. Chow and M. Ghavamzadeh, "Algorithms for CVaR optimization in MDPs," *Advances in neural information processing systems*, vol. 27, 2014.

[108] W. B. Knox and P. Stone, "Interactively shaping agents via human reinforcement: The TAMER framework," in *Proceedings of the fifth international conference on Knowledge capture*, 2009, pp. 9–16.

[109] F. Rezazadeh, H. Chergui, L. Alonso, and C. Verikoukis, "SliceOps: Explainable MLOps for streamlined automation-native 6G networks," *IEEE Wireless Communications*, vol. 31, no. 5, pp. 224–230, 2024.

[110] M. Ameur, B. Brik, and A. Ksentini, "Leveraging LLMs to eXplain DRL decisions for transparent 6G network slicing," in *2024 IEEE 10th International Conference on Network Softwarization (NetSoft)*, 2024, pp. 204–212.

[111] D. d. S. Brilhante, J. C. Manjarres, R. Moreira, L. de Oliveira Veiga, J. F. de Rezende, F. Müller, A. Klautau, L. Leonel Mendes, and F. A. P. de Figueiredo, "A literature survey on AI-aided beamforming and beam management for 5G and 6G systems," *Sensors*, vol. 23, no. 9, p. 4359, 2023.

[112] J. Feng, L. Liu, X. Hou, Q. Pei, and C. Wu, "QoE fairness resource allocation in digital twin-enabled wireless virtual reality systems," *IEEE journal on selected areas in communications*, vol. 41, no. 11, pp. 3355–3368, 2023.

[113] Y. Jiang, J. Kang, X. Ge, D. Niyato, and Z. Xiong, "QoE analysis and resource allocation for wireless metaverse services," *IEEE Transactions on Communications*, vol. 71, no. 8, pp. 4735–4750, 2023.

[114] S. Hendaoui, F. Hendaoui, and N. Zangar, "Dynamic proactive–reactive scheduling for URLLC in 5G: Leveraging XGBoost and network virtualization," *Physical Communication*, vol. 68, p. 102553, 2025.

[115] F. A. Dicandia, N. J. Fonseca, M. Bacco, S. Mugnaini, and S. Genovesi, "Space-air-ground integrated 6G wireless communication networks: A review of antenna technologies and application scenarios," *Sensors*, vol. 22, no. 9, p. 3136, 2022.

[116] D. Van Huynh, S. R. Khosravirad, S. L. Cotton, H. Shin, and T. Q. Duong, "Multi-agent reinforcement learning for optimal resource allocation in space-air-ground integrated networks," *IEEE Internet of Things Journal*, 2025.

[117] E. Bandara, P. Foytik, S. Shetty, R. Mukkamala, A. Rahman, X. Liang, N. W. Keong, and K. D. Zoysa, "SliceGPT – OpenAI GPT-3.5 LLM, blockchain and non-fungible token enabled intelligent 5G/6G network slice broker and marketplace," in *2024 IEEE 21st Consumer Communications & Networking Conference (CCNC)*, 2024, pp. 439–445.

[118] F. Lotfi, H. Rajoli, and F. Afghah, "Prompt-tuned LLM-augmented DRL for dynamic O-RAN network slicing," *arXiv preprint arXiv:2506.00574*, 2025.

[119] B. Liu, J. Tong, and J. Zhang, "LLM-Slice: Dedicated wireless network slicing for large language models," in *Proceedings of the 22nd ACM Conference on Embedded Networked Sensor Systems*, 2024, pp. 853–854.

[120] A. Dandoush, V. Kumarskandpriya, M. Uddin, and U. Khalil, "Large language models meet network slicing management and orchestration," *arXiv preprint arXiv:2403.13721*, 2024.

[121] J. Dean and L. A. Barroso, "The tail at scale," *Communications of the ACM*, vol. 56, no. 2, pp. 74–80, 2013.

[122] T. Fehrenbach, R. Datta, B. Göktepe, T. Wirth, and C. Hellge, "URLLC services in 5G low latency enhancements for LTE," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*. IEEE, 2018, pp. 1–6.

[123] R. K. Jain, D.-M. W. Chiu, W. R. Hawe *et al.*, "A quantitative measure of fairness and discrimination," *Eastern Research Laboratory, Digital*

Equipment Corporation, Hudson, MA, vol. 21, no. 1, pp. 2022–2023, 1984.

[124] M. T. Ribeiro, S. Singh, and C. Guestrin, "Anchors: High-precision model-agnostic explanations," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.

[125] R. Guidotti, "Counterfactual explanations and how to find them: literature review and benchmarking," *Data Mining and Knowledge Discovery*, vol. 38, no. 5, pp. 2770–2824, 2024.

[126] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *International conference on machine learning*. PMLR, 2017, pp. 1321–1330.

[127] N. Jiang and L. Li, "Doubly robust off-policy value evaluation for reinforcement learning," in *International conference on machine learning*. PMLR, 2016, pp. 652–661.

[128] D. Precup, R. S. Sutton, and S. Singh, "Eligibility traces for off-policy policy evaluation," 2000.

[129] F. Mihalič, M. Truntič, and A. Hren, "Hardware-in-the-loop simulations: A historical overview of engineering challenges," *Electronics*, vol. 11, no. 15, p. 2462, 2022.

[130] L. Yang, A. Zhou, X. Ma, Y. Zhang, Y. Li, and S. Wang, "Flexible shadow: Enhancing service reliability in resource-constrained edge computing," in *2024 IEEE International Conference on Web Services (ICWS)*. IEEE, 2024, pp. 767–777.

[131] P. Thomas and E. Brunskill, "Data-efficient off-policy policy evaluation for reinforcement learning," in *International conference on machine learning*. PMLR, 2016, pp. 2139–2148.

[132] G. F. Riley and T. R. Henderson, "The ns-3 network simulator," in *Modeling and tools for network simulation*. Springer, 2010, pp. 15–34.

[133] A. Varga, "OMNeT++," in *Modeling and tools for network simulation*. Springer, 2010, pp. 35–59.

[134] G. Nardini, D. Sabella, G. Stea, P. Thakkar, and A. Virdis, "Simu5G–an OMNeT++ library for end-to-end performance evaluation of 5G networks," *IEEE Access*, vol. 8, pp. 181 176–181 191, 2020.

[135] N. Nikaein, M. K. Marina, S. Manickam, A. Dawson, R. Knopp, and C. Bonnet, "OpenAirInterface: A flexible platform for 5G research," *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 5, pp. 33–38, 2014.

[136] I. Gomez-Miguelez, A. Garcia-Saavedra, P. D. Sutton, P. Serrano, C. Cano, and D. J. Leith, "srsLTE: An open-source platform for LTE evolution and experimentation," in *Proceedings of the Tenth ACM International Workshop on Wireless Network Testbeds, Experimental Evaluation, and Characterization*, 2016, pp. 25–32.

[137] L. Bonati, P. Johari, M. Polese, S. D'Oro, S. Mohanti, M. Tehrani-Moayyed, D. Villa, S. Shrivastava, C. Tassie, K. Yoder *et al.*, "Colosseum: Large-scale wireless experimentation through hardware-in-the-loop network emulation," in *2021 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*. IEEE, 2021, pp. 105–113.

[138] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "ColO-RAN: Developing machine learning-based xApps for open RAN closed-loop control on programmable experimental platforms," *IEEE Transactions on Mobile Computing*, vol. 22, no. 10, pp. 5787–5800, 2022.

[139] J. Breen, A. Buffmire, J. Duerig, K. Dutt, E. Eide, M. Hibler, D. Johnson, S. K. Kasera, E. Lewis, D. Maas *et al.*, "POWDER: Platform for open wireless data-driven experimental research," in *Proceedings of the 14th International Workshop on Wireless Network Testbeds, Experimental evaluation & Characterization*, 2020, pp. 17–24.

[140] D. Raychaudhuri, I. Seskar, G. Zussman, T. Korakis, D. Kilper, T. Chen, J. Kolodziejski, M. Sherman, Z. Kostic, X. Gu *et al.*, "Challenge: COSMOS: A city-scale programmable testbed for experimentation with advanced wireless," in *Proceedings of the 26th annual international conference on mobile computing and networking*, 2020, pp. 1–13.

[141] L. Bertizzolo, L. Bonati, E. Demirors, A. Al-Shawabka, S. D'Oro, F. Restuccia, and T. Melodia, "Arena: A 64-antenna SDR-based ceiling

grid testing platform for sub-6 GHz 5G-and-beyond radio spectrum research," *Computer Networks*, vol. 181, p. 107436, 2020.

[142] M. Polese, L. Bonati, S. D'oro, S. Basagni, and T. Melodia, "Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1376–1411, 2023.

[143] M. Towers, A. Kwiatkowski, J. Terry, J. U. Balis, G. De Cola, T. Deleu, M. Goulão, A. Kallinteris, M. Krimmel, A. KG *et al.*, "Gymnasium: A standard interface for reinforcement learning environments," *arXiv preprint arXiv:2407.17032*, 2024.

[144] C. F. Hayes, R. Rădulescu, E. Bargiacchi, J. Källström, M. Macfarlane, M. Reymond, T. Verstraeten, L. M. Zintgraf, R. Dazeley, F. Heintz *et al.*, "A practical guide to multi-objective reinforcement learning and planning," *arXiv preprint arXiv:2103.09568*, 2021.

[145] P. Rost, C. Mannweiler, D. S. Michalopoulos, C. Sartori, V. Sciancalepore, N. Sastry, O. Holland, S. Tayade, B. Han, D. Bega *et al.*, "Network slicing to enable scalability and flexibility in 5G mobile networks," *IEEE Communications magazine*, vol. 55, no. 5, pp. 72–79, 2017.

[146] L. L. Zhang, S. Han, J. Wei, N. Zheng, T. Cao, Y. Yang, and Y. Liu, "Nn-meter: Towards accurate latency prediction of deep-learning model inference on diverse edge devices," in *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*, 2021, pp. 81–93.

[147] R. Poyiadzi, X. Renard, T. Laugel, R. Santos-Rodriguez, and M. Detyniecki, "Understanding surrogate explanations: the interplay between complexity, fidelity and coverage," *arXiv preprint arXiv:2107.04309*, 2021.

[148] N. Carrara, E. Leurent, R. Laroche, T. Urvoy, O.-A. Maillard, and O. Pietquin, "Budgeted reinforcement learning in continuous state space," *Advances in neural information processing systems*, vol. 32, 2019.

[149] J. Miller, K. Krauth, B. Recht, and L. Schmidt, "The effect of natural distribution shift on question answering models," in *International conference on machine learning*. PMLR, 2020, pp. 6905–6916.

[150] M. Arjovsky, L. Bottou, I. Gulrajani, and D. Lopez-Paz, "Invariant risk minimization," *arXiv preprint arXiv:1907.02893*, 2019.

[151] B. Schölkopf, F. Locatello, S. Bauer, N. R. Ke, N. Kalchbrenner, A. Goyal, and Y. Bengio, "Toward causal representation learning," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 612–634, 2021.

[152] Y. Chow, M. Ghavamzadeh, L. Janson, and M. Pavone, "Risk-constrained reinforcement learning with percentile risk criteria," *Journal of Machine Learning Research*, vol. 18, no. 167, pp. 1–51, 2018.

[153] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2017, pp. 23–30.

[154] P. Manakul, A. Liusie, and M. J. Gales, "SelfCheckGPT: Zero-resource black-box hallucination detection for generative large language models," *arXiv preprint arXiv:2303.08896*, 2023.

[155] H. Rashkin, V. Nikolaev, M. Lamm, L. Aroyo, M. Collins, D. Das, S. Petrov, G. S. Tomar, I. Turc, and D. Reitter, "Measuring attribution in natural language generation models," *Computational Linguistics*, vol. 49, no. 4, pp. 777–840, 2023.

[156] A. Parisi, Y. Zhao, and N. Fiedel, "Talm: Tool augmented language models," *arXiv preprint arXiv:2205.12255*, 2022.

[157] L. Qin, Q. Chen, Y. Zhou, Z. Chen, Y. Li, L. Liao, M. Li, W. Che, and P. S. Yu, "A survey of multilingual large language models," *Patterns*, vol. 6, no. 1, 2025.

[158] A. Kumar, C. Agarwal, S. Srinivas, A. J. Li, S. Feizi, and H. Lakkaraju, "Certifying LLM safety against adversarial prompting," *arXiv preprint arXiv:2309.02705*, 2023.

[159] A. Clemm, L. Ciavaglia, L. Z. Granville, and J. Tantsura, "Rfc 9315: Intent-based networking-concepts and definitions," 2022.

[160] E. Tabassi, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," National Institute of Standards and Technology

(NIST), Tech. Rep. NIST AI 100-1, Jan. 2023, [Online]. Available: https://doi.org/10.6028/NIST.AI.100-1.

[161] T. T. Nguyen, T. T. Huynh, Z. Ren, T. T. Nguyen, P. L. Nguyen, H. Yin, and Q. V. H. Nguyen, "Privacy-preserving explainable AI: a survey," *Science China Information Sciences*, vol. 68, no. 1, p. 111101, 2025.

[162] I. Munoko, H. L. Brown-Liburd, and M. Vasarhelyi, "The ethical implications of using artificial intelligence in auditing," *Journal of business ethics*, pp. 209–234, 2020.

[163] M. Li, L. Li, Y. Yin, M. Ahmed, Z. Liu, and Q. Liu, "Red teaming visual language models," *arXiv preprint arXiv:2401.12915*, 2024.

[164] S. Mahmud, S. Saisubramanian, and S. Zilberstein, "Explanation-guided reward alignment." in *IJCAI*, 2023, pp. 473–482.