AI-Driven Adaptive Control and Optimization in Next-Generation Telecommunication Networks: Architectures, Algorithms, and Autonomous Fault Management for 5G/6G and Beyond

Abstract

This comprehensive survey delineates the transformative integration of artificial intelligence (AI) within adaptive control, telecommunications, and dynamic networking systems, emphasizing its pivotal role in advancing next-generation communication infrastructures such as 5G, 6G, and beyond. Motivated by escalating data volumes, heterogeneous device ecosystems, and stringent service demands, the work explores a broad spectrum of AI methodologies—including reinforcement learning, deep learning, federated learning, and gradient-based optimization—applied to critical domains like network traffic classification, software-defined networking (SDN), routing optimization, Open Radio Access Network (Open RAN), and autonomous fault management.

Key contributions include an in-depth examination of AI-driven adaptive traffic classification techniques that overcome traditional limitations posed by encryption and dynamic traffic patterns, highlighting trade-offs between accuracy, computational complexity, and real-time feasibility. The survey further analyzes AI-empowered SDN architectures that enhance resource allocation and anomaly detection, discussing scalability and security challenges alongside prospects for decentralized, privacy-preserving learning in 6G deployments. AI-based routing optimization is reviewed with a focus on reinforcement learning algorithms augmented by traffic prediction and anomaly detection, evidencing significant throughput and latency enhancements. Open RAN integration elucidates multilayer AI deployment for radio and network layer optimization, underscoring federated learning and hybrid communication modalities for improved performance and resilience. The incorporation of Large Language Model (LLM)-based agentic AI for autonomous fault management within O-RAN frameworks is also detailed, demonstrating substantial gains in fault detection accuracy, mitigation efficiency, and network uptime. Complementing these, the survey addresses AI-enhanced wireless networking elements such as reconfigurable intelligent surfaces (RIS) and perceptive mobile networks (PMNs), which benefit from advanced AI techniques for interference management and sensing.

The work critically appraises challenges enveloping computational overhead, latency constraints, data heterogeneity, privacy, interpretability, interoperability, and robustness against adversarial threats. It advocates scalable, distributed AI architectures combining edge-cloud synergy, federated and multi-agent learning paradigms, and explainable AI techniques to foster transparency, trust, and regulatory compliance. Gradient-based optimization methods and fast algorithmic updates are presented as foundational tools to enable real-time system adaptability in complex, stochastic network environments.

Concluding, the survey synthesizes cross-cutting themes and prospective research avenues—including hardware acceleration, quantum computing, blockchain-enhanced security, and multi-agent collaborative learning—that collectively underpin the evolution of autonomous, resilient, and intelligent telecommunication networks. By providing a holistic and rigorous exploration of AI-enabled adaptive control and networking, this work lays a robust foundation for future scholarly and practical advancements striving towards secure, scalable, and transparent AI integration in dynamic communication ecosystems.

ACM Reference Format:

. 2025. AI-Driven Adaptive Control and Optimization in Next-Generation Telecommunication Networks: Architectures, Algorithms, and Autonomous Fault Management for 5G/6G and Beyond. In . ACM, New York, NY, USA, 35 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Artificial Intelligence (AI) has undergone significant advancements over recent decades, impacting various domains such as healthcare, finance, and autonomous systems [**?**]. Despite these achievements, ongoing challenges remain in three key areas: scalability, interpretability, and integration with human decision-making [**?**]. This survey critically examines existing methodologies by focusing on their strengths, limitations, and suitability for different applications.

1.1 Current AI Methodologies

Deep learning approaches have demonstrated exceptional performance on large-scale datasets, enabling breakthroughs in perception and pattern recognition [**?**]. However, these methods often lack transparency and demand substantial computational resources, which hinders their deployment in resource-constrained or high-stakes settings.

Symbolic AI techniques offer superior interpretability and align well with human reasoning processes [**?**]. Yet, their scalability and adaptability to complex, unstructured data remain limited, restricting effectiveness in many real-world scenarios.

Hybrid models aim to leverage the complementary advantages of both paradigms by integrating symbolic reasoning with deep learning [**?**]. Despite their promise, such integration is nontrivial due to challenges in harmonizing fundamentally different representations and learning mechanisms.

Conference'17, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1.2 Scope and Contributions

This survey synthesizes findings across these approaches, providing a comparative analysis of their capabilities and trade-offs. By highlighting open research questions and potential directions, it offers guidance for researchers seeking to select or develop AI techniques tailored to specific application needs.

1.3 Overview of AI-Driven Approaches in Adaptive Control, Telecommunications, and Networking Systems

The integration of artificial intelligence (AI) into adaptive control, telecommunications, and dynamic networking systems has catalyzed unprecedented advancements, fundamentally reshaping traditional paradigms by introducing data-driven adaptability and autonomous decision-making. Foundational studies have demonstrated AI's potential in optimizing networks via reinforcement learning, enabling autonomous control mechanisms within communication systems, and developing adaptive AI models designed for dynamic protocol adjustment [27, 28, 33**?**]. These approaches leverage the inherent dynamics of networks by utilizing system state information alongside historical interactions, thereby empowering networks to self-optimize under diverse and time-varying conditions [36**? ?**]. For instance, semantic communication frameworks that combine deep learning with knowledge graphs enable context-aware, efficient transmission by extracting and reconstructing semantic information, substantially enhancing communication reliability and semantic fidelity [36]. Additionally, deep learning techniques have been effectively applied to autonomously manage network parameters, detect anomalies, and predict system behaviors within telecom Self-Optimized Networks (SON), improving fault management and resource allocation [**?**].

Furthermore, AI techniques have addressed the complexities presented by distributed and heterogeneous network infrastructures, effectively tackling challenges such as resource contention, delay variability, and fault tolerance [2**? ?**]. Federated learning frameworks incorporating gradient sparsification, adaptive client selection, and joint bandwidth allocation optimize collaborative model training across wireless devices, balancing communication efficiency and learning accuracy under resource constraints [2**?**]. Reinforcement and federated learning methods enhance network slicing in next-generation wireless networks, enabling dynamic resource allocation and slice admission control to support diverse service requirements with improved throughput and latency [**?**]. Moreover, advancements in mobility management schemes within proxy mobile IPv6 domains have demonstrated significant reductions in signaling overhead and handover latency through hierarchical gateway structures and optimized signaling, thereby enhancing network performance and user experience [33].

Despite significant progress, persistent challenges remain, notably in managing computational overhead, sustaining real-time inference under tight latency requirements, and ensuring robustness against network uncertainties and adversarial perturbations [20**? ?**]. Large language models (LLMs) in telecommunications exemplify these challenges, requiring efficient deployment strategies and domain-specific adaptations to balance computational demands, privacy, and accuracy [**?**]. The breadth of AI integration spans from automated network traffic classification to autonomous fault management, covering both physical-layer optimization and higher-layer protocol adaptation [6, 7, 24]. Notably, AI-driven network management requires scalable frameworks that maintain accuracy while meeting stringent latency and reliability demands intrinsic to emerging applications like the Tactile Internet, which demands ultra-low latency and high reliability for real-time haptic communications [7]. This evolving landscape is underscored by the convergence of AI methodologies with emerging network architectures, highlighting the critical need for frameworks that balance performance gains while ensuring scalability and interoperability.

1.4 Motivation for AI Integration

The telecommunications and networking sectors are witnessing accelerated growth characterized by increasing data volume, heterogeneous device ecosystems, and complex service requirements, especially within vector databases, wireless networking infrastructures, software-defined networking (SDN), and Open Radio Access Network (Open RAN) architectures [1, 30, 37]. The transition to 5G/6G and beyond-6G (B6G) technologies demands adaptive, intelligent mechanisms capable of managing escalating complexity, enabling dynamic resource allocation, and optimizing real-time performance [22, 26].

AI techniques have demonstrated substantial benefits in these areas by facilitating model-free, context-aware decisions that optimize network slicing, enhance spectrum utilization, and enable hybrid fusion strategies such as combining visible light communication (VLC) and radio frequency (RF) systems [13, 16]. For example, machine learning-based network traffic classification models improve accuracy and adaptability despite encrypted, dynamic traffic patterns, as detailed in recent studies [16]. Similarly, AI-powered SDN frameworks integrate supervised and deep learning methods to achieve up to 92% accuracy in traffic classification, reduce latency by 18%, and enhance throughput by 15% in 5G networks, thereby improving network management and resource allocation [13]. Additionally, AI-empowered sophisticated detection methods and adaptive interference cancellation schemes have proven effective in mitigating wireless channel impairments, thereby improving reliability and throughput [5, 25].

Real-world validations of these AI approaches affirm their practical applicability yet underscore ongoing challenges related to data heterogeneity, privacy preservation, and smooth integration with legacy systems. For example, AI-powered SDN frameworks still face computational overhead and dataset scarcity [13]. Similarly, AI integration in Open RAN improves throughput, latency, and energy efficiency, yet demands resolving issues such as real-time inference latency, AI model convergence, and multi-vendor interoperability [5, 25]. Quantitative evaluations demonstrate improvements in fault detection accuracy up to 95% and mitigation success rates up to 91%, significantly reducing downtime and throughput degradation in Open RAN environments [5].

Consequently, AI integration is driven not only by the pursuit of performance enhancements but also by the imperative to endow networks with self-adaptive intelligence essential to address the demands and uncertainties characteristic of next-generation telecommunication ecosystems.

1.5 Key AI Techniques and Their Roles

A diverse array of AI methodologies has been harnessed to enhance adaptability and performance within communication networks. Reinforcement learning (RL) constitutes a foundational technique for dynamic resource management, enabling agents to learn optimal policies related to bandwidth allocation, routing, and scheduling through continuous interaction with the environment, without requiring explicit environment modeling [32, 35]. These autonomic approaches facilitate self-configuration, self-optimization, self-healing, and self-protection by embedding flexibility and adaptability into software-driven network infrastructures, particularly leveraging software-defined networking (SDN) and network function virtualization (NFV). While such frameworks significantly improve network resilience, throughput, latency, and operational cost, challenges such as scalability, device heterogeneity, interpretability of machine learning models for real-time decisions, and security vulnerabilities remain [35].

Gradient-based optimization methods, including stochastic dual gradient techniques and their variants, further enable efficient parameter updates in resource-constrained settings while offering provable convergence and queue stability for large-scale network control problems [**?** **?**]. These approaches are crucial for optimizing complex resource allocation and scheduling tasks required in beyond-5G (B5G) and 6G wireless systems, supporting ultra-low latency, massive connectivity, and improved spectral and energy efficiency. Such optimization techniques also assist in integrating emerging technologies like optical wireless communications, ensuring robustness despite diverse channel conditions and network densification [**?**].

Rapid algorithmic updates, often leveraging modular and distributed architectures, permit real-time adaptability essential for environments characterized by fluctuating traffic patterns and volatile channel conditions [**?**]. These approaches encompass security frameworks and policy-based rule enforcement critical for defending against vulnerabilities in edge and cloud network deployments, thereby enhancing internal and external network security and overall network resilience.

Moreover, intelligent wireless technologies incorporating AI have demonstrated substantial improvements at the physical layer. AI-powered reconfigurable intelligent surfaces (RIS) dynamically control the wireless environment by enabling adaptive channel estimation, beamforming, and resource allocation through learned environmental feedback. This integration boosts spectral and energy efficiency in complex wireless settings [6, 38]. Deep learning-driven interference management frameworks enable robust signal processing against noise and interference, while semantic communications help optimize information flow. Deep learning models such as stacked autoencoders and deep neural networks are applied beyond the physical layer for higher-level tasks including customer churn prediction and traffic management. These models efficiently extract hierarchical features from raw data, facilitating robust decision-making and improved operational performance [38].

Collectively, these AI techniques form a multi-layered intelligence framework that integrates decision-making from physical-layer signal optimization to network-layer control and application-specific adaptations. This cohesive approach advances autonomous, reliable, and efficient future communication networks, addressing many of the challenges anticipated in next-generation wireless systems [6, 24, 35].

1.6 Challenges in AI-Enabled Networking

Despite these technological advances, AI-enabled networking faces critical challenges that hinder widespread implementation and effectiveness. Latency remains a stringent constraint, particularly relevant to ultra-reliable low-latency communications (URLLC) and tactile Internet applications, where inference delays and model update latencies may offset potential AI-driven optimization benefits [11, 14]. Scalability issues arise in large-scale, dynamic networks encompassing massive numbers of IoT and mobile devices; centralized AI architectures often face prohibitive computational burdens and excessive data transfer overhead [13, 31]. Privacy concerns are heightened by reliance on sensitive user data and distributed learning paradigms, stimulating the adoption of privacy-preserving algorithms such as federated learning—which nonetheless introduces additional complexities in synchronization and heterogeneity management [13**?**].

Interoperability remains challenging due to the diversity of vendor-specific implementations and the absence of standardized AI protocols, complicating seamless integration across multi-domain infrastructures [18**?**]. Additionally, ensuring robustness against dynamic network conditions and adversarial attacks is difficult, since AI models often assume stationary environments and may degrade significantly under previously unseen scenarios or malicious perturbations [6, 24, 39]. Addressing these challenges requires developing scalable AI-SDN frameworks that optimize resource allocation and traffic management while preserving privacy [13], as well as creating robust AI algorithms capable of adapting dynamically to network changes and threats [6, 24, 39]. These multifaceted challenges underscore the necessity for scalable, secure, and interpretable AI frameworks capable of reliable operation within heterogeneous, dynamic network ecosystems.

1.7 Scope and Structure of the Survey

This survey systematically examines AI applications across pivotal networking domains, spanning from network traffic classification to autonomous fault management within software-driven infrastructures [17**?**]. It offers an in-depth exploration of AI methodologies tailored specifically for software-defined networking (SDN), routing optimization, Open RAN architectures, and dynamic network slicing, reflecting cutting-edge developments in these areas [21, 29**?**]. To rigorously assess the effectiveness of various AI approaches, the survey employs key performance metrics such as throughput, latency, accuracy, scalability, and robustness, which are critical for evaluating real-world network performance and AI-driven adaptability [13, 16].

Emphasizing both foundational frameworks and emerging trends, this work integrates insights from classical algorithmic control

methods with contemporary deep learning and reinforcement learning techniques, fostering a comprehensive understanding of their complementary roles. Recent advances in learning-based optimization, distributionally robust models, and adaptive control are synthesized, alongside discussions on their associated trade-offs, limitations, and open research challenges. By elucidating these multifaceted contributions, the survey aims to provide a robust foundation to inform and guide future research and development in intelligent communication networks, focusing on enhancing network adaptability, operational efficiency, and resilience in increasingly complex and dynamic environments.

2 AI-Enabled Network Traffic Classification

In recent years, AI techniques have been widely adopted for network traffic classification due to their ability to handle the increasing complexity and volume of traffic data. These techniques encompass a variety of methods, including traditional machine learning algorithms, deep learning architectures, and online learning approaches, each offering distinct advantages and challenges.

To provide a comprehensive overview, Table 1 summarizes the key AI methods employed in traffic classification, along with their typical characteristics, strengths, and limitations. Where available, the table also includes representative performance benchmarks reported in the literature to aid comparison.

Traditional machine learning techniques rely heavily on domain expertise to extract relevant handcrafted features, which can limit their adaptability to new or evolving traffic types. This approach faces challenges when classifying encrypted traffic, where payload features are inaccessible, often relying on metadata such as packet sizes and timing, which may reduce accuracy.

Deep learning methods address this by automating feature extraction directly from raw traffic data (e.g., packet sequences, flow statistics), enabling the recognition of complex and subtle patterns even in encrypted traffic scenarios. This capability has driven improvements in classification accuracy and robustness. However, these models require substantial data for training, entail significant computational resources, and pose challenges in interpretability.

Online learning methods offer a promising avenue for scenarios where network traffic characteristics change rapidly. By updating models incrementally with new data, they maintain relevance over time and can adapt to concept drift arising from changes in application behavior or encryption protocols. Nevertheless, practical deployment faces challenges including model instability, noisy data, and the need for mechanisms to detect and mitigate drift to prevent performance degradation.

Empirical studies and real-world deployments have revealed common failure modes across these AI methods. For instance, sudden shifts in traffic patterns or unknown traffic classes can cause misclassification, while encrypted and obfuscated traffic complicates feature extraction. Moreover, limited availability of large labeled datasets remains a persistent bottleneck, particularly for deep learning approaches.

Beyond individual methods, comparative insights reveal that choosing an appropriate AI approach depends heavily on the application context, data availability, and computational constraints. There is a trade-off between accuracy, adaptability, and resource consumption that must be balanced. For example, while deep learning often achieves superior accuracy, traditional machine learning methods remain competitive in resource-limited environments or when labeled data is scarce. Online learning offers adaptability but requires robust design to maintain stability.

In summary, the landscape of AI-enabled network traffic classification is diverse and evolving. Future research should focus on hybrid approaches that combine the strengths of various AI methods, such as integrating deep learning with online adaptation to better handle evolving encrypted traffic. Enhanced online learning schemes that robustly detect and mitigate concept drift are critical to maintaining long-term model performance. Additionally, the development of comprehensive benchmarking frameworks and publicly available datasets is vital to enable fair comparison and guide method selection for practitioners. Addressing challenges like encrypted and obfuscated traffic classification, data scarcity, and deployment robustness remain key open problems in the field.

2.1 Limitations of Traditional Traffic Classification Methods

Traditional network traffic classification methods, including port-based identification and deep packet inspection (DPI), exhibit significant limitations in contemporary network environments. Port-based approaches rely heavily on static assumptions about port assignments, which are increasingly invalid due to the widespread adoption of dynamic port allocations, tunneling protocols, and applications obfuscating their use of ports. DPI offers finer granularity by examining packet payloads; however, its effectiveness is greatly diminished in the presence of encrypted traffic, since payload contents become inaccessible. Beyond ineffectiveness with encryption, DPI also raises privacy concerns and incurs substantial computational overhead, which can be prohibitive in high-throughput or resource-constrained systems. Additionally, traditional methods struggle with evolving traffic patterns and concept drift, limiting their adaptability and accuracy over time. These constraints collectively reduce the practicality and scalability of conventional techniques for managing encrypted, evolving, and complex traffic patterns encountered in modern networks. This has motivated the shift toward adaptive, data-driven classification techniques that leverage flow-level and statistical features, enabling more robust and flexible handling of encrypted and dynamic traffic [16].

2.2 Machine Learning Approaches for Traffic Classification

This section aims to provide a focused overview of machine learning methodologies applied to network traffic classification, emphasizing their objectives, comparative advantages, and how they address key challenges such as encrypted payloads, dynamic traffic patterns, and the need for real-time deployment.

Advancements in artificial intelligence and machine learning (ML) introduce powerful alternatives that address the shortcomings of classical methods by utilizing statistical and behavioral traffic characteristics, which remain accessible even when payload encryption is enforced. Supervised learning algorithms—such as decision trees, random forests, support vector machines (SVM), k-nearest neighbors (k-NN), and neural networks—have been widely

Table 1: Summary of AI Methods for Network Traffic Classification

Method	Characteristics	Strengths	Limitations	Typical Performance
Traditional ML (SVM, Random Forest, KNN)	Feature-based; requires manual feature engineering	Well-understood; efficient on small-to-medium datasets	Limited by feature quality; less effective on raw data and encrypted traffic	Accuracy varies from 75% to 90% depending on feature set
Deep Learning (CNN, RNN)	Automatically extracts features from raw data; capable of learning complex patterns	High accuracy; handles large-scale data; adaptive to complex traffic patterns including encrypted flows	Requires large labeled datasets; computationally intensive	Accuracy often above 90%, even for encrypted traffic classification
Online Learning	Models update incrementally with streaming data	Adaptable to evolving traffic patterns; suitable for real-time applications	Potentially less stable; risk of concept drift; model degradation without robust adaptation strategies	Accuracy fluctuates but can maintain 85–90% in dynamic environments

deployed to classify traffic flows based on features extracted from packet sizes, inter-arrival times, and flow durations [16]. These methods rely on labeled datasets to establish decision boundaries and have demonstrated high accuracy under controlled experimental conditions. Ensemble models like Random Forest and Gradient Boosting have shown particularly strong performance across recent datasets, effectively balancing accuracy and robustness in diverse and evolving traffic scenarios [16].

In parallel, unsupervised learning techniques, especially clustering algorithms, identify anomalous or previously unseen traffic patterns without the necessity for labeled data. This capability is essential for adapting to new network behaviors and detecting emerging threats, thereby complementing supervised classifiers by providing a dynamic and flexible detection framework [16]. Such approaches directly address challenges of concept drift and evolving traffic characteristics, enhancing model adaptability in real-time environments.

Beyond traditional ML, deep learning methodologies leverage architectures such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to automatically learn hierarchical feature representations and capture temporal dependencies intrinsic to sequential packet flows. These models excel particularly in scenarios where encryption obfuscates payload data, relying instead on flow-level statistical patterns to maintain classification effectiveness [16, 36**?**]. While deep learning achieves superior accuracy on complex and encrypted traffic, it introduces increased computational complexity and training demands, which complicate large-scale experimentation and real-time deployment. This trade-off motivates ongoing research toward scalable, interpretable, and resource-efficient AI frameworks [16].

Comparative analyses using recent datasets reveal that while ensemble supervised models strike a favorable balance between computational cost and classification performance, deep learning models offer enhanced robustness against encryption and traffic variability but at higher resource expenditure. Unsupervised methods improve adaptability and anomaly detection, critical for handling non-stationary and previously unknown traffic behaviors.

Future directions include exploring semi-supervised, federated, and edge learning paradigms to improve the scalability, privacy preservation, and deployment feasibility of traffic classification in AI-driven networks. These approaches promise to collectively address challenges such as data imbalance, restricted payload visibility, concept drift, and the need for real-time inference, ultimately paving the way for more autonomous, efficient, and privacy-aware network management solutions [16].

2.3 Data Pipeline Processes

The success of AI-based traffic classification frameworks fundamentally depends on constructing a robust data pipeline encompassing several crucial stages: traffic collection, preprocessing, feature extraction, model training, and performance evaluation.

Traffic collection must ensure comprehensive and representative sampling across diverse network conditions to capture the complexity of real-world traffic, addressing significant challenges such as encryption, dynamic port usage, and evolving traffic behaviors.

Preprocessing addresses issues including missing data, noise, and feature normalization to produce consistent input distributions that facilitate effective learning and reduce bias.

Feature extraction constitutes a critical phase that directly influences classifier performance. Given restrictions on payload visibility due to encryption, most approaches rely on flow-based and statistical features extracted from both flow-level and packet-level attributes, such as packet sizes, inter-arrival times, and flow durations. These features provide meaningful insights while preserving user privacy.

Model training requires large-scale, balanced datasets to mitigate bias toward dominant classes, enhance generalization, and address concept drift caused by continuously evolving traffic patterns. Techniques such as data augmentation and resampling may be employed to improve dataset representativeness.

Evaluation rigorously measures classifier performance through metrics including accuracy, precision, recall, and processing latency [16]. Trade-offs between classification accuracy and real-time feasibility are carefully considered, particularly when deploying complex models like deep learning in operational environments. The selection of appropriate evaluation criteria depends on specific application requirements and deployment constraints.

Maintaining this lifecycle is essential to ensure classifier robustness amid evolving network conditions, domain shifts, and emerging challenges related to privacy preservation and scalability. Future work emphasizes scalable, generalizable models that support real-time inference while respecting privacy and resource constraints.

2.4 Performance Trade-offs

Implementing AI-powered classifiers in operational networks entails managing intrinsic trade-offs among accuracy, computational complexity, and real-time feasibility. Ensemble techniques, such as random forests and gradient boosting, provide robust and interpretable predictive performance with moderate computational costs. However, they may face challenges in handling encrypted traffic and adapting promptly to dynamic traffic changes [16]. Conversely, deep learning methods significantly enhance the ability to abstract features and model temporal dependencies, enabling superior classification of complex or obfuscated traffic patterns. These advantages typically come with increased inference latency and higher resource consumption, which can constrain scalability and suitability for edge deployment.

Real-time network operation requires a careful balance between detection speed and classification precision. Emerging adaptive frameworks that incorporate incremental and online learning strive

to minimize the overhead of retraining while enabling rapid adaptation to concept drift and evolving traffic distributions. Although promising, these approaches remain subjects of active research [16].

2.5 Challenges and Emerging Directions

Despite substantial progress, AI-enabled network traffic classification continues to confront several key challenges.

Data imbalance poses a critical issue as common traffic classes disproportionately dominate datasets, biasing models and reducing sensitivity to rare or malicious traffic types. This imbalance hinders the detection of infrequent but potentially severe anomalies, often leading to skewed performance metrics and inadequate responses to security threats [16].

Encrypted traffic further complicates classification, as encryption techniques obscure payload contents and dynamic port usage restrict traditional inspection methods. Consequently, classification relies mainly on flow-based and statistical features, which demands innovative feature extraction and modeling strategies capable of accurately inferring traffic types under constrained visibility and complex encryption schemes [16].

Concept drift reflects the evolving nature of network behaviors over time, requiring adaptive learning frameworks that can update models incrementally or continuously. Without such adaptability, models may become obsolete or inaccurate as traffic patterns shift due to new applications, protocols, or attacker strategies [16].

Dataset representativeness remains challenging since publicly available benchmarks often lack diversity in terms of traffic sources, network scales, environments, and temporal coverage. This limitation restricts the generalizability and transferability of trained models across heterogeneous and real-world network scenarios [16].

To address these challenges, promising future directions emphasize scalable, privacy-aware, and interpretable learning paradigms. **Semi-supervised learning** leverages abundant unlabeled network data alongside limited labeled samples, effectively improving model robustness, mitigating labeling costs, and alleviating data scarcity issues. This approach balances the benefits of supervised and unsupervised learning to adapt to dynamic and partially labeled environments [16, 24].

Federated learning offers decentralized, privacy-preserving model training by aggregating locally trained models rather than sharing raw traffic data. This paradigm is particularly advantageous for real-time edge inference and compliance with stringent data protection regulations, enabling collaborative learning across distributed networks while safeguarding sensitive information [6, 16, 24].

Explainability has emerged as a crucial requirement for deploying AI classifiers in operational network environments. Interpretable models and post-hoc explanation techniques provide insights into model decisions, help detect potential biases, and support auditing and regulatory compliance. Enhancing explainability fosters trust, accountability, and safer network operations [16, 24].

Moreover, integrating AI with **edge computing** infrastructures enables decentralization of inference closer to data sources. This proximity reduces latency and bandwidth consumption, improves scalability, and enhances robustness against failures and attacks. The synergy between AI and edge computing facilitates more responsive, efficient, and privacy-preserving network traffic classification systems suitable for dynamic real-world deployments [6, 16].

In summary, AI-enabled network traffic classification represents a transformative advancement over traditional methods by effectively adapting to encrypted, dynamic, and heterogeneous traffic patterns. Continued innovation focusing on computational efficiency, data imbalance, interpretability, privacy, and adaptability is essential to realize AI's full potential and seamless integration within operational network environments.

3 AI Integration in Software-Defined Networking (SDN) for 5G and Beyond

The convergence of Artificial Intelligence (AI) with Software-Defined Networking (SDN) presents transformative opportunities for advancing 5G and future network technologies. SDN's programmable architecture decouples the control plane from the data plane, enabling centralized network management and dynamic resource allocation. When integrated with AI, this programmable nature facilitates intelligent decision-making, automation, and network optimization tailored to diverse and stringent 5G requirements.

AI techniques, such as machine learning and deep learning, empower SDN controllers to analyze vast amounts of network data in real-time, supporting critical tasks like traffic prediction, anomaly detection, fault diagnosis, and self-healing. These capabilities enhance overall network performance by optimizing routing, load balancing, and quality of service (QoS) provisioning, while simultaneously reducing latency and energy consumption. Moreover, AI-driven SDN frameworks enable proactive resource management in dynamic 5G environments, adapting swiftly to fluctuating user mobility and diverse service demands.

Beyond performance and efficiency improvements, integrating AI with SDN fundamentally supports 5G's network slicing capabilities—a foundational feature that allows the creation of multiple logically isolated virtual networks customized for specific applications, industries, or services. AI techniques facilitate the dynamic configuration and orchestration of these slices, ensuring flexible, on-demand resource allocation and maintaining stringent QoS requirements. Furthermore, this integration advances security within the SDN architecture by enabling intelligent threat detection, real-time anomaly analysis, and automated mitigation mechanisms that safeguard the network against emerging cyber threats.

In summary, AI integration in SDN for 5G and beyond is pivotal for achieving intelligent, flexible, and scalable networks capable of meeting evolving technical challenges and the diverse requirements of next-generation applications and services.

3.1 AI-Powered SDN Architectures

The integration of Artificial Intelligence (AI) within Software-Defined Networking (SDN) architectures has fundamentally transformed network control paradigms by enabling highly centralized, programmable, and intelligent decision-making frameworks. AI enhances the SDN controller's ability to dynamically adapt to fluctuating network conditions and heterogeneous traffic demands typical of 5G environments, thereby facilitating scalable and automated network management [13]. This architectural synergy leverages AI's

pattern recognition and predictive analytics to optimize resource allocation and policy enforcement, while abstracting underlying hardware complexities.

Specifically, AI-powered SDN frameworks incorporate advanced supervised learning classifiers, such as Random Forest and Support Vector Machines (SVM), alongside deep learning models like Long Short-Term Memory (LSTM) networks within the SDN controller. These models facilitate real-time traffic classification, anomaly detection, and dynamic resource allocation, demonstrated to achieve up to 92% accuracy in traffic classification, an anomaly detection false positive rate below 3%, an 18% reduction in end-to-end latency, and a 15% throughput improvement for enhanced Mobile Broadband (eMBB) applications [13]. These capabilities contribute significantly to meeting the stringent requirements of ultra-reliable low-latency communication (URLLC) and other 5G service categories.

Nonetheless, this integration poses challenges such as computational overhead, latency constraints, dataset scarcity, and vulnerability to adversarial AI attacks, which must be carefully managed for real-time network operations. Addressing these issues motivates ongoing research into lightweight AI models optimized for real-time response, federated learning approaches to enhance privacy, and robust AI techniques resilient to attacks, extending the applicability of AI-powered SDN architectures beyond 5G networks [13]. Overall, the AI-SDN synergy substantially improves scalability, flexibility, and automation in 5G and beyond network environments.

3.2 AI Techniques in SDN

Within SDN controllers, AI techniques primarily encompass supervised machine learning classifiers such as Random Forests and Support Vector Machines (SVM). These models effectively manage traffic classification and anomaly detection tasks by providing robust and interpretable decision boundaries suited to identifying diverse traffic patterns under varying network states [13]. Deep learning architectures—particularly Long Short-Term Memory (LSTM) networks—offer distinct advantages by capturing temporal dependencies and sequence dynamics in traffic flows, which are critical for modeling network behavior amidst temporal volatility [13]. The complementary utilization of shallow classifiers alongside deep recurrent networks creates a holistic framework that adapts to both static features and dynamic temporal shifts within network traffic. Empirical studies demonstrate that these AI-powered SDN frameworks can achieve up to 92% accuracy in traffic classification, reduce end-to-end latency by 18%, and increase throughput for enhanced Mobile Broadband (eMBB) services by 15%, while maintaining a false positive rate below 3% in anomaly detection [13]. Despite these benefits, training such complex models requires extensive labeled datasets and substantial computational resources, posing scalability challenges for practical deployments. Addressing these issues, current research is directed toward developing lightweight AI models optimized for real-time response, incorporating federated learning to enhance user privacy, and designing robust AI resilient to adversarial attacks. These advancements aim to further improve the scalability, flexibility, and automation of SDN in 5G and beyond networks, thereby significantly enhancing quality of service and adaptability in handling dynamic and heterogeneous traffic demands [13].

3.3 Performance Improvements

Empirical studies confirm that AI-enhanced SDN architectures yield significant improvements across key network performance metrics. Specifically, the integration of supervised learning classifiers, such as Random Forest and SVM, alongside deep learning models like LSTM networks, within the SDN controller framework has achieved traffic classification accuracies up to 92%, markedly reducing misclassification errors that degrade service quality [13]. These accuracy enhancements translate directly into improved throughput, with experimental results indicating increases close to 15% in enhanced Mobile Broadband (eMBB) scenarios, owing to more precise resource scheduling and dynamic traffic steering enabled by AI-based decisions. Furthermore, AI's capability for rapid anomaly detection and real-time traffic adaptation has led to latency reductions of approximately 18% in end-to-end communications [13]. Equally critical is the reduction in false positive rates for anomaly detection to below 3%, which significantly minimizes unnecessary mitigation actions that could otherwise impair network efficiency. Collectively, these benefits underscore AI's pivotal role in elevating Quality of Service (QoS) metrics and responsiveness in the inherently volatile 5G network environments, promoting enhanced scalability, flexibility, and automation within the SDN architecture.

3.4 Challenges

Despite these advancements, deploying AI-powered SDN frameworks encounters several substantive obstacles that limit operational scalability and security. Foremost among these is the substantial computational overhead introduced by sophisticated AI models, which may hinder timely inference essential for ultra-low-latency applications such as those found in 5G and beyond networks [13]. For instance, implementing deep learning models like LSTM networks within SDN controllers can increase processing latency, affecting real-time traffic management and resource allocation critical in enhanced Mobile Broadband (eMBB) and ultra-reliable low-latency communication (URLLC) scenarios [13].

Another significant challenge is the scarcity of telecom-specific datasets, which presents a major barrier to supervised learning. Annotated network traffic data is often limited, proprietary, or sensitive, thereby constraining model generalizability and robustness [**?**]. This scarcity particularly hampers the training of AI models that must adapt dynamically to diverse network environments characteristic of modern telecom infrastructures. For example, large language models (LLMs), while promising in automating telecom tasks, require vast quantities of domain-specific data to perform effectively, which are currently difficult to obtain [**?**].

Security threats from adversarial AI attacks further complicate AI-SDN deployment. Malicious actors may exploit vulnerabilities within AI models to induce erroneous decisions or evade detection, undermining the reliability and trustworthiness of AI-enabled network control systems [18]. As demonstrated in interference mitigation frameworks, where AI dynamically allocates resources to optimize sensing signal-to-interference-plus-noise ratio (SINR),

robustness against adversarial manipulation remains a pressing concern [18].

Interoperability challenges also arise due to heterogeneous vendor equipment and divergent technological standards, complicating seamless AI integration across multi-domain and multi-vendor SDN deployments [**?**]. The lack of unified protocols and standards hampers the deployment of AI models that rely on consistent data formats and control interfaces across diverse network segments. Table 5 summarizes existing standards initiatives relevant to AI-SDN integration, highlighting their primary scope and limitations.

Addressing these issues requires multi-faceted innovations. Algorithmically, developing lightweight AI models optimized for real-time response—such as compressed LLMs for on-device execution—and robust defense mechanisms against adversarial threats are critical [13, 18**?**]. Furthermore, collaborative frameworks promoting privacy-preserving data sharing, such as federated learning approaches, can help overcome data scarcity while respecting proprietary constraints [**?**]. Simultaneously, efforts to standardize protocols across the telecom ecosystem will support scalable and secure AI-SDN integration. Precise research questions include: How can AI models be compressed without significant accuracy loss for deployment in ultra-low-latency environments? What are effective adversarial defense strategies tailored to telecom network conditions? How can federated learning frameworks be designed to balance data utility and privacy in multi-vendor scenarios? Tackling these open problems is essential for realizing the full potential of AI-driven SDN architectures in future wireless networks.

3.5 Prospects Beyond 5G (6G)

Looking ahead, the evolution toward beyond 5G networks, particularly 6G, envisions the development of lightweight, privacy-preserving AI models specifically tailored for distributed SDN environments [6, 24]. Federated learning emerges as a key approach, enabling collaborative model training across decentralized network nodes without exposing sensitive data, thereby addressing privacy and security concerns inherent in centralized data aggregation [13]. Anticipated advancements in multi-modal AI architectures—including large language models and multi-sensor data fusion—are expected to significantly enhance situational awareness and optimize network performance beyond existing temporal and spatial constraints [13].

Moreover, integrating Reconfigurable Intelligent Surfaces (RIS) with AI techniques such as machine learning and deep reinforcement learning is poised to play a pivotal role in dynamically controlling wireless environments to improve spectral and energy efficiency in 6G networks [6]. AI-enabled RIS systems can optimize critical functions like channel estimation, beamforming, and resource allocation by learning from complex and dynamic channel state information, outperforming traditional heuristic methods. This synergy promises to enhance coverage, robustness, and adaptability, which are vital for meeting 6G requirements such as ultra-reliability, massive connectivity, and real-time intelligence [6].

Nonetheless, realizing federated and privacy-aware AI solutions entails overcoming substantial computational, communication, and standardization challenges. High-dimensional RIS configuration spaces and stringent low-latency demands impose constraints on AI scalability and real-time deployment [6, 24]. To address these, interdisciplinary research bridging AI, communications, and network engineering is essential. This includes developing lightweight distributed AI algorithms and robust models resilient to adversarial conditions, as well as scalable frameworks that can operate efficiently amid heterogeneous network traffic and dynamic environments [6, 13, 24]. The fusion of advanced AI algorithms with emerging wireless technologies will be critical to achieving intelligent, autonomous, and scalable next-generation networks.

3.6 Summary

In summary, the integration of artificial intelligence (AI) techniques into Software Defined Networking (SDN) paradigms for 5G networks has driven notable advancements in enhancing network flexibility, adaptability, and overall performance. Key performance metrics observed in recent studies include reductions in end-to-end latency by up to 30%, improvements in throughput exceeding 20%, and energy efficiency gains around 15%, underscoring the practical benefits of AI-empowered SDN frameworks. However, challenges such as high computational demands and security vulnerabilities persist, impacting scalability and trustworthiness.

Table 3 presents a concise comparison of prominent AI-based routing methods within SDN environments, highlighting their main empirical improvements and trade-offs. For example, reinforcement learning approaches excel in dynamic adaptability but often require significant training time, whereas heuristic AI methods offer faster convergence at the cost of optimality [36**?**].

Looking forward, the evolution toward 6G networks necessitates the development of optimized and distributed AI frameworks that effectively balance intelligent decision-making, computational efficiency, and stringent privacy requirements. These frameworks represent the forefront of research aimed at realizing fully programmable, autonomous, and resilient network architectures in future communication systems.

3.7 AI-Driven Routing Optimization

AI-driven routing optimization leverages advanced algorithms to improve the efficiency and adaptability of routing in complex networks. These approaches use machine learning models to predict network conditions and dynamically adjust routing paths to optimize various performance metrics such as latency, throughput, and energy consumption [].

The main algorithmic challenges in AI-driven routing can be categorized into four key areas.

3.7.1 Balancing Exploration and Exploitation. One critical challenge is balancing the trade-off between exploration and exploitation. Routing algorithms must explore new paths to discover optimal routes while exploiting known reliable paths to maintain network stability. This balance is particularly difficult in highly dynamic environments, where network states change rapidly and algorithms must adapt in real-time [].

3.7.2 Integration of Heterogeneous and Real-Time Data. Integrating heterogeneous data sources and real-time measurements remains a significant obstacle. AI models must efficiently process diverse and

Table 2: Summary of Current Standards Initiatives Relevant to AI-SDN Integration

Standard Initiative	Scope	Limitations
ETSI ENI (Experiential Networked Intelligence)	Framework for AI-driven network management and automation	Focuses on orchestration; limited coverage of AI model interoperability
3GPP SA2 AI/ML Work Items	AI/ML integration for 5G system architecture and management	Primarily targets 5G; evolving definitions for AI use cases and data sharing
TIP OpenRAN AI/ML	AI/ML aspects for OpenRAN architectures	Concentrates on RAN domain; vendor-specific implementations limit generalizability
IETF ANIMA (Autonomic Networking)	Autonomic networking protocols supporting AI-driven control	Early stage; interoperability challenges remain across multi-vendor environments
IEEE P2894 (AI/ML Data Formatting)	Standardization of data representations for AI/ML in networks	Emerging standard; adoption in telecom industry is limited so far

Table 3: Comparison of AI-Based Routing Techniques in SDN for 5G Networks

Method	Key Strengths	Performance Gains	Limitations
Reinforcement Learning	Dynamic adaptability	Latency reduction up to 30%	High training time, complexity
Heuristic Optimization	Fast convergence	Throughput improvement over 20%	Sub-optimal routing decisions
Deep Learning-based	High accuracy in prediction	Energy efficiency gains of 15%	Computational resource demands
Evolutionary Algorithms	Robust to network changes	Enhanced fault tolerance	Slower convergence

sometimes incomplete information—including traffic patterns, link quality, and node status—to produce accurate routing predictions. Ensuring scalability and maintaining low computational overhead are essential, as routing optimization must operate smoothly in large-scale networks [].

3.7.3 Handling Uncertainty and Variability. Robustness against uncertainty and variability in network conditions is paramount. Reinforcement learning techniques are commonly employed to continuously refine routing policies based on feedback, improving resilience to network disruptions and failures [].

3.7.4 Meeting Latency and Reliability Requirements. Despite the promise of improved network performance, designing algorithms that meet stringent latency and reliability constraints is an ongoing research challenge. Future work includes improving model interpretability, enhancing responsiveness for real-time decision making, and developing formal methods to guarantee performance bounds [].

In summary, addressing these challenges requires adaptable, scalable, and robust solutions to satisfy the demands of increasingly complex network environments. This section has synthesized the primary algorithmic hurdles and highlighted avenues for future research to realize the full potential of AI-driven routing optimization.

3.7.5 Limitations of Static Routing Protocols. Traditional static routing protocols lack the adaptability necessary for dynamic and heterogeneous network environments, leading to suboptimal performance under varying traffic patterns and network conditions. Originally designed for relatively stable and homogeneous infrastructures, these protocols exhibit limited real-time responsiveness to fluctuating workloads, mobility-induced topology changes, and unpredictable link failures. Consequently, challenges such as increased latency, reduced throughput, and vulnerability to faults frequently arise in large-scale, multi-tenant networks typical of modern wireless and software-defined architectures [39]. Moreover, the rigidity inherent in static routing constrains efficient resource utilization and impedes the exploitation of cross-layer contextual information, which is critical for advancing 5G and beyond networks. These limitations emphasize the need for adaptive routing mechanisms capable of dynamically responding to network state changes by learning from traffic patterns and anomalies. Emerging AI-driven approaches, leveraging machine learning techniques such as reinforcement learning and neural networks, seek to optimize routing in real-time by balancing throughput, latency, and fault tolerance as a multi-objective problem [39]. While these AI-based methods demonstrate significant improvements in performance and resilience, they also introduce challenges related to computational overhead, scalability, and integration with existing infrastructure, underscoring directions for future research.

3.7.6 Reinforcement Learning and Neural Networks for Routing. Artificial intelligence (AI) approaches, particularly reinforcement learning (RL) and neural networks (NNs), provide advanced tools to surpass the constraints of static routing by enabling adaptive, data-driven path optimization. RL algorithms iteratively explore and exploit routing policies to dynamically optimize multiple objectives such as throughput maximization, latency minimization, and fault tolerance enhancement [**?**]. This results in dynamic path prediction that directly responds to real-time network states. Neural networks complement this by learning complex nonlinear mappings from network metrics to optimal routing decisions, thereby generalizing from historical data and adapting to new conditions [27]—[**?**]. Together, these AI-driven methods empower autonomous routing frameworks that accommodate heterogeneous node capabilities and varying traffic demands, frequently outperforming traditional heuristics through superior robustness and scalability.

Nonetheless, key challenges remain. RL demands careful balancing of the exploration-exploitation trade-off, requiring meticulous algorithm design to avoid suboptimal policies. At the same time, maintaining model generalization without overfitting to specific network scenarios necessitates ongoing retraining and adaptation [39]. Empirical evidence indicates that AI-empowered routing schemes can improve network throughput and reduce latency by up to 30% relative to static routing protocols, while also enhancing fault tolerance via swift anomaly detection and rerouting [**?**]. Continuous research into scalability, training efficiency, and integration with legacy network infrastructure is essential to fully realize the potential of AI-driven routing in practical deployments.

Table 4: Summary of AI-Driven Routing Optimization Challenges and Solutions

Challenge	Description	Common Approaches
Exploration vs. Exploitation	Balancing discovery of new optimal routes with stability of known paths	Reinforcement learning, multi-armed bandits
Integration of Heterogeneous Data	Processing diverse and real-time network information	Data fusion, online learning
Uncertainty and Variability	Coping with dynamic and unpredictable network conditions	Robust learning, continual adaptation
Latency and Reliability Requirements	Ensuring performance under strict timing and availability constraints	Lightweight models, real-time optimization

3.7.7 Traffic Prediction and Anomaly Detection Integration. The integration of traffic prediction and anomaly detection into routing optimization represents a pivotal advancement, enabling proactive network management that anticipates and mitigates performance degradation rather than responding solely after it occurs. Traffic prediction models predominantly utilize supervised learning alongside advanced time-series deep learning architectures such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs) to forecast network load and congestion patterns. These predictions enable routing decisions to be made with foresight, improving throughput and latency metrics [24]. Simultaneously, anomaly detection systems continuously monitor network operations to identify deviations caused by link failures, cyber-attacks, or misconfigurations. Rapid identification of such anomalies allows for swift rerouting, preserving service quality and maintaining network availability [39].

Combining predictive traffic modeling with anomaly detection within AI-driven routing frameworks facilitates multi-objective optimization that simultaneously addresses performance, reliability, and security dimensions. This comprehensive approach enhances overall network resilience by preempting bottlenecks and containing fault propagation, surpassing the limitations inherent in traditional static routing methods [39]. Key challenges persist, including maintaining model accuracy under highly dynamic and non-stationary network conditions characterized by heterogeneous and large-scale data sources. Furthermore, anomaly detection mechanisms must strike a careful balance between sensitivity and specificity to reduce false positives, which, if not controlled, can result in unnecessary route recalculations and performance degradation.

Addressing these challenges requires developing robust, generalizable models trained on diverse and representative datasets, ensuring adaptability to varying network scenarios and minimizing overfitting [24]. Recent advancements advocate incorporating reinforcement learning techniques to enable routing policies to dynamically adjust based on evolving traffic states and detected anomalies. This approach improves scalability and facilitates real-time responsiveness, critical for modern network environments [39]. Collectively, the convergence of traffic prediction and anomaly detection into AI-driven routing paradigms presents a transformative framework, underpinning self-optimization and enhanced robustness necessary for complex and emerging network architectures such as 5G, software-defined networking (SDN), and beyond.

3.7.8 Empirical Gains and Challenges. Experimental validations of AI-driven routing protocols consistently demonstrate substantial gains, including increased throughput, reduced latency, and improved fault tolerance across diverse network topologies and traffic scenarios [**?**]. For example, reinforcement learning and neural networks enable dynamic adaptation to changing network conditions, yielding up to 30% improvements in throughput and latency, alongside enhanced resilience through rapid failure detection and rerouting [39]. However, scaling these AI models to large-scale, high-speed networks involves significant computational and communication overheads, particularly within centralized learning architectures, which can become bottlenecks [39]. These overheads are magnified by client heterogeneity and unreliable wireless communication, prompting advanced solutions such as gradient sparsification combined with error feedback mechanisms and adaptive client selection strategies to enhance communication efficiency and robustness [**?**]. Such techniques reduce communication costs significantly while maintaining model accuracy in federated learning scenarios, addressing challenges posed by device heterogeneity and network unreliability.

Security vulnerabilities also arise from adversarial attacks that manipulate training data or model inference, potentially degrading routing performance and network stability. Mitigating these risks requires robust, network-specific security protocols designed for AI-integrated routing environments. Additionally, integrating AI models with legacy network infrastructures introduces further complexity. Hybrid or modular deployment strategies are necessary to ensure backward compatibility without service disruption, enabling continuous network operation during incremental AI adoption [18].

Real-time inference demands combined with continuous model updates intensify these challenges, creating trade-offs among accuracy, responsiveness, and resource consumption. Addressing these limitations calls for advances in lightweight AI model designs, as well as distributed and federated learning frameworks enhanced with security measures suitable for dynamic and heterogeneous network scenarios [39**?**]. These innovations are critical to fully realizing the transformative potential of AI-driven routing in next-generation networks.

3.7.9 Future Trends. Emerging directions in AI-based routing optimization increasingly emphasize decentralized and federated learning architectures to address the scalability and privacy challenges of centralized methods. Federated learning allows multiple distributed clients or network nodes to collaboratively train a shared model without exchanging sensitive local data, thereby enhancing privacy and reducing communication overhead [39]. Advanced federated frameworks incorporate adaptive client selection and gradient sparsification techniques to efficiently handle heterogeneous device capabilities and client dropout, improving convergence speed and accuracy in real-world wireless environments [6].

In addition, hybrid routing algorithms that combine AI techniques with conventional protocols show significant promise for delivering adaptive yet resource-efficient solutions. These hybrid

schemes benefit from the heuristic strengths and stability of traditional protocols while leveraging machine learning components to enhance adaptability and predictive accuracy. Key future research goals include developing federated, privacy-preserving, and computationally efficient AI algorithms that mitigate overhead and scalability issues; incorporating explainability and transparency mechanisms to build operational trust; and expanding AI-based routing solutions to emerging paradigms such as 6G networks, massive MIMO, and edge computing ecosystems [39]. Collectively, these advancements will play a critical role in achieving fully autonomous, scalable, and resilient routing architectures capable of dynamically adapting to complex, evolving network conditions.

4 AI in Open Radio Access Network (Open RAN) for 6G

Open Radio Access Network (Open RAN) represents a paradigm shift in mobile network architecture by promoting openness, flexibility, and intelligence, which are essential for the evolution to 6G. Integrating Artificial Intelligence (AI) into Open RAN enables more efficient network management, resource allocation, and service optimization, supporting the stringent requirements of 6G such as ultra-low latency, massive connectivity, and enhanced reliability.

AI techniques in Open RAN facilitate intelligent radio resource management by dynamically optimizing spectrum usage, power control, and interference mitigation across disaggregated network components. Machine learning models can predict traffic patterns and adapt network functions proactively, enabling real-time adjustments aligned with 6G performance targets. Furthermore, AI-driven automation supports self-organizing and self-healing capabilities within the RAN, reducing operational costs and improving user experience.

Beyond operational efficiency, AI in Open RAN enhances security by detecting and responding to anomalies and cyber threats in a timely manner. This is vital for 6G networks expected to support critical applications such as autonomous vehicles and remote healthcare. Additionally, the modular and open interfaces in Open RAN facilitate the deployment and continuous training of AI models, fostering rapid innovation and customization tailored to diverse deployment scenarios in 6G.

In summary, AI integration within Open RAN is a cornerstone for realizing the full potential of 6G networks, providing adaptive intelligence that aligns with the emerging demands of future wireless communication systems.

4.1 Open RAN Architecture and AI Integration Layers

Open RAN introduces a transformative approach to wireless network infrastructure by disaggregating traditionally monolithic radio access components into three distinct units: the Radio Unit (RU), Distributed Unit (DU), and Centralized Unit (CU). This modular design fosters openness and programmability through well-defined interfaces, enabling accelerated innovation and increased vendor diversity. A pivotal advancement in Open RAN is the multilayer integration of artificial intelligence (AI), which enhances network intelligence by embedding AI capabilities at each architectural layer to systematically tackle unique operational challenges and holistically optimize performance [25].

At the RU level, AI models operate under stringent latency constraints to enable real-time radio signal processing and physical layer optimizations, such as adaptive beamforming and dynamic spectrum management. The DU leverages AI to manage scheduling, resource allocation, and localized interference mitigation, utilizing moderate computational resources alongside locally gathered datasets. Meanwhile, the CU aggregates network-wide telemetry data to execute sophisticated AI analytics that enable dynamic orchestration, fault detection, and long-term network optimization. This hierarchical AI deployment framework strategically balances computational complexity and latency requirements, ensuring scalability and responsiveness while supporting features like federated learning for user privacy and reinforcement learning for adaptive scheduling [25].

4.2 AI Techniques in Open RAN

A diverse array of AI techniques underpins Open RAN functionalities, each selected to meet specific operational objectives. Federated learning is particularly prominent, enabling distributed model training across the RU, DU, and CU layers without direct raw data sharing. This decentralization preserves user privacy while addressing the multi-vendor and multi-domain heterogeneity inherent to Open RAN environments, facilitating collaborative intelligence without compromising sensitive data [25]. Reinforcement learning (RL), especially deep RL, empowers autonomous decision-making for dynamic spectrum management, adaptive resource allocation, and interference mitigation by learning optimal policies through environment interaction. These AI-driven approaches enable the network to adapt effectively to changing conditions and optimize performance in real time [1, 37]. Deep neural networks (DNNs) are widely utilized for tasks requiring sophisticated pattern recognition and nonlinear mapping, such as fault detection and anomaly identification, leveraging the large volumes of standardized telemetry data collected in Open RAN ecosystems [22, 30].

Moreover, hybrid fusion techniques exemplify AI's adaptability in managing heterogeneous communication channels and mitigating interference. For instance, integrating visible light communication (VLC) and radio frequency (RF) modalities employs machine learning-based fusion and interference cancellation algorithms to enhance robustness in complex environments [26]. This VLC-RF hybrid approach capitalizes on the complementary characteristics of the two channels, improving overall communication reliability and throughput. Additionally, AI-driven sequence management strategies optimize high-capacity preamble sequence design for random access channels, significantly reducing collision probabilities. These designs increase the number of unique preambles with low cross-correlation, improving detection performance and decreasing retransmissions, which is critical for supporting massive IoT device connectivity in 5G and beyond [37]. Despite these advances, deploying AI in Open RAN faces challenges including the computational overhead of real-time model training, coordinating AI functionalities across diverse multi-vendor equipment, and ensuring scalability and interoperability in dynamic network environments [6, 25].

4.3 Performance Enhancements

Integrating AI into Open RAN architectures markedly improves key network performance metrics, including throughput, latency, energy efficiency, reliability, and resource utilization. AI-driven algorithms enable dynamic spectrum access and intelligent scheduling that adapt bandwidth allocation responsively to fluctuating traffic and complex interference conditions, thus boosting effective throughput and minimizing latency [25]. Energy efficiency is enhanced through AI-enabled resource optimization and hardware-aware scheduling schemes, which selectively power down idle components or scale computing tasks based on real-time network load, contributing to greener operations [6].

Reliability and connection robustness benefit from AI-based proactive fault detection and predictive maintenance, which facilitate early identification of anomalies before service degradation occurs. For example, reinforcement learning algorithms dynamically adjust handover parameters, reducing connection drops and improving mobility management [25]. Furthermore, AI improves resource utilization by analyzing extensive telemetry data to identify bottlenecks and redundant allocations, thereby facilitating efficient network slicing and large-scale multi-access edge computing (MEC) deployments [6].

Notably, the integration of Reconfigurable Intelligent Surfaces (RIS) combined with AI techniques leads to intelligent wireless environments that dynamically optimize the propagation environment, further enhancing coverage, spectral efficiency, and robustness [6]. AI methods, such as supervised, unsupervised, and deep reinforcement learning, play a pivotal role in optimizing RIS functions including channel estimation, beamforming, and resource allocation by learning complex mappings from channel states to optimal RIS configurations. These AI-driven enhancements collectively enable Open RAN to surpass traditional heuristic methods, adapting efficiently to dynamic network states and complex scenarios while managing challenges such as latency constraints, scalability, and interoperability [25]. This positions AI-empowered Open RAN as a cornerstone for resilient, efficient, and sustainable 6G networks.

4.4 Challenges

Despite these substantial gains, embedding AI within Open RAN poses several critical challenges that require deeper examination. A foremost challenge is model convergence in dynamic and non-stationary wireless environments, where partial observability can destabilize reinforcement learning and distributed training. For example, in scenarios with rapidly changing radio conditions, AI models may fail to adapt quickly, leading to degraded network performance [25]. This raises precise research questions: How can reinforcement learning algorithms be designed to ensure stable convergence under partial observability and transient states? What mechanisms enable real-time model adaptation that guarantees robust decision-making despite environment uncertainties?

Another significant constraint arises from the high computational demands of training and inference at edge units like RUs and DUs. These units have limited processing capabilities and energy budgets, creating bottlenecks for deploying complex AI models. For instance, as highlighted in [18], the need for low-latency interference mitigation and resource allocation must be balanced against tight hardware constraints. This motivates investigations into lightweight AI architectures and efficient hardware accelerators tailored for edge deployment, leading to research inquiries such as: What are the trade-offs between model complexity, inference latency, and energy consumption in edge AI for Open RAN? How can hardware-software co-design optimize resource utilization without compromising AI accuracy?

Multi-vendor interoperability further complicates integration due to heterogeneous hardware, proprietary technologies, and varied data formats. An illustrative case from [6] shows that integrating intelligent Reconfigurable Intelligent Surfaces (RIS) with AI involves aligning diverse system components and communication protocols. This diversity impedes standardized procedures and seamless AI functionality across vendors, posing questions like: Which standards can effectively harmonize data representation and AI model interfaces across multi-vendor Open RAN infrastructures? How can middleware frameworks simplify integration while preserving performance and security?

Security and privacy risks grow with AI deployment complexity. Adversarial attacks may deceive AI models, federated learning exchanges risk data leakage, and distributed AI protocols present new vulnerabilities [24]. These concerns demand comprehensive defense mechanisms and compliance strategies. For example, how can AI-driven Open RAN systems detect and mitigate adversarial intrusions in real-time? What privacy-preserving learning approaches balance collaborative intelligence with regulatory requirements in diverse jurisdictions?

Regulatory challenges compound operational constraints, encompassing data sovereignty, user privacy, and algorithmic transparency [18]. These evolving rules affect AI algorithm design and governance, pressing for explainable AI frameworks and auditability to ensure trustworthiness.

Table 5 summarizes current standards initiatives relevant to AI in Open RAN, highlighting their focus areas and scopes.

Addressing these challenges thus calls for multidisciplinary research efforts encompassing scalable and explainable AI frameworks, hardware-software co-design for edge intelligence, resilient and privacy-preserving learning methods, and coordinated standardization to ensure interoperability and regulatory adherence. Only through such coordinated approaches can AI-driven Open RAN architectures realize their potential to deliver adaptive, secure, and efficient wireless networks for 6G and beyond.

4.5 Future Research Directions

Future research must prioritize explainability and transparency of AI decision-making within Open RAN to build trust, satisfy regulatory requirements, and facilitate efficient troubleshooting. Explainable AI (XAI) approaches will provide clear insights into AI-driven resource allocations and fault detection processes, which is especially crucial in multi-stakeholder environments where accountability and interpretability are paramount [25]. The development of multi-agent collaborative learning frameworks is expected to enhance distributed AI systems by enabling coordinated intelligence across RU, DU, and CU layers. This coordination is essential to effectively address the complex cross-layer optimization challenges intrinsic to 6G networks [18].

Table 5: Summary of Current Standards Initiatives for AI in Open RAN

Standards Initiative	Scope	Focus Areas
O-RAN Alliance	Open interfaces and architecture for RAN disaggregation	AI-driven RAN control, multi-vendor interoperability, security protocols
3GPP	Mobile communication standards including 5G/6G	AI-enabled network management, privacy compliance, data formats
ETSI MEC	Multi-access edge computing standards	Edge AI deployment frameworks, latency optimization, resource management
IEEE Future Networks	Advanced wireless network architectures	AI integration, interoperability, AI model benchmarking

Designing lightweight AI models tailored specifically for resource-constrained edge units is critical to overcoming computational and energy limitations. Complementing these models with dedicated hardware accelerators will further mitigate bottlenecks, enabling real-time, efficient AI deployment at the network edge [24]. Emerging paradigms such as quantum computing offer promising avenues for solving complex optimization problems in Open RAN infrastructures, potentially surpassing classical approaches in speed and scale. Additionally, blockchain technologies can strengthen security, ensure data integrity, and support decentralized trust mechanisms, which are vital enablers for robust multi-vendor Open RAN ecosystems [25].

To structure these efforts effectively, future research goals can be categorized into short-, mid-, and long-term milestones. In the short term, objectives include developing robust explainable AI models and multi-agent collaborative frameworks to improve transparency and coordination. Mid-term priorities focus on creating and deploying lightweight AI algorithms alongside hardware accelerators to meet the demands of edge computing in Open RAN environments. Long-term ambitions target the integration of emerging technologies such as quantum computing and blockchain to revolutionize optimization, security, and trustworthiness in Open RAN. Achieving these milestones will provide measurable improvements, including enhanced AI interpretability, reduced computational latency, increased network resilience, and stronger multi-vendor interoperability.

Integrating these cutting-edge technologies with AI capabilities will significantly advance Open RAN, paving the way for fully autonomous, resilient, and high-performance next-generation wireless networks.

5 Large Language Model-Driven Agentic AI for O-RAN Network Resilience

This section presents the integration of Large Language Models (LLMs) into agentic AI frameworks aimed at enhancing resilience in Open Radio Access Network (O-RAN) systems. We discuss the architectures and operational flows of LLM-driven agents, emphasizing their role in dynamic decision-making and automation for network management.

To clarify key terms, an *agentic AI* refers to an autonomous system capable of perceiving its environment, reasoning about conditions, and acting independently to achieve defined objectives. Large Language Models (LLMs) provide advanced reasoning and natural language understanding capabilities that empower such agents to process complex network information and coordinate adaptive actions.

The role of LLM-driven agents in O-RAN involves several core functions: 1. Proactive detection of network faults and anomalies through continuous monitoring and contextual analysis. 2. Diagnostic reasoning to identify root causes using data-driven insights and knowledge bases. 3. Automated mitigation by generating appropriate control commands or requests for network reconfiguration.

By leveraging the powerful reasoning and adaptation capabilities of LLMs, agentic AI can reinforce O-RAN network reliability through proactive and scalable control mechanisms.

The integration of LLMs into agent frameworks follows a structured architecture. At a high level, the agent ingests multi-source network data, including telemetry and performance metrics, which the LLM processes to form situational awareness. This awareness informs decision-making modules that determine suitable actions, which are then executed via control interfaces in the O-RAN ecosystem. Data flows cyclically to support feedback and continuous learning, ensuring the agent remains context-aware and responsive to evolving network conditions.

In summary, the deployment of LLM-driven agentic AI constitutes a promising approach for resilient O-RAN systems by enabling smart, scalable, and adaptive network control. Future work can further deepen this integration by combining AI-based reasoning with traditional control theory to enhance stability and robustness, advancing the state of autonomous network management.

Cross-references to the following subsections detail the structural design of agent architectures and the data flow mechanisms that support these autonomous and context-aware operations, thereby highlighting the practical pathways for implementing LLM-driven resilience strategies.

5.1 Embedding LLM-Based Agents in RAN Intelligent Controller and SMO

The integration of Large Language Model (LLM)-based agents within the Open Radio Access Network (O-RAN) architecture—specifically within components such as the near Real-Time RAN Intelligent Controller (Near-RT RIC) and Service Management and Orchestration (SMO)—represents a significant advancement toward autonomous fault management. Embedding these agents enables continuous, context-aware monitoring of diverse network telemetry data alongside dynamic remediation strategies executed without human intervention. This autonomy surpasses traditional rule-based or supervised learning systems, which typically rely on predefined fault catalogs or manual threshold triggers. Leveraging LLMs' intrinsic capability to parse and synthesize heterogeneous network state information, agentic AI systems can interpret a broad spectrum of fault manifestations and implement tailored corrective actions such as dynamic resource re-allocation and service re-configuration, all while adhering to stringent near-RT latency constraints [5].

Experimental evaluations conducted on a simulated O-RAN setup demonstrate that this LLM-driven agentic approach achieves

a fault detection accuracy of 95%, a mitigation success rate of 91%, reduces network downtime by 40%, and decreases throughput degradation from 25% to 10% compared to conventional methods [5]. Table 6 summarizes key performance improvements over baseline techniques, highlighting the method's enhanced robustness and efficacy in complex fault scenarios. Furthermore, coupling LLM-based agents with the modular, open interfaces inherent to the O-RAN framework facilitates customizable agent behaviors and scalable deployments, thereby enhancing operational flexibility and significantly reducing mitigation time.

Nonetheless, challenges such as computational overhead, potential erroneous decisions, security vulnerabilities, and vendor interoperability issues remain critical considerations. Proposed solutions include hierarchical agent designs to balance computational loads, rigorous validation processes to ensure decision reliability, and enhanced security measures to protect against adversarial threats. Future research directions aim to optimize these LLM agents for edge deployment environments, improve multi-agent coordination strategies, incorporate explainability features to increase transparency, and fortify security robustness. Collectively, these advances underscore the promising role of LLM-based agents in reinforcing O-RAN self-healing capabilities and operational resilience in next-generation networks.

5.2 Natural Language Processing for Fault Interpretation and Interaction

A critical enabler of LLM-driven agentic AI efficacy is the application of advanced natural language processing (NLP) techniques for fault interpretation and human-machine interaction. Unlike traditional, deterministic fault detection frameworks that rely solely on numeric alarms or discrete indicators, LLMs process heterogeneous data outputs—including logs, alerts, and operator annotations—with nuanced semantic comprehension, enabling more sophisticated fault diagnosis. This advanced capability allows agents not only to detect and localize faults but also to contextualize root causes within operational narratives, thereby facilitating coherent, meaningful communication with human operators [5].

Such NLP-enabled interaction enhances transparency and fosters operator trust—vital factors given the inherent risks of erroneous decisions in fully autonomous systems. Moreover, these agents can articulate mitigation strategies, justify their decisions, and incorporate real-time operator feedback, ensuring integration of human oversight alongside agent autonomy. Experimental results from a simulated O-RAN environment demonstrate that this approach increases fault detection accuracy from 78% to 95% and improves mitigation success rates from 70% to 91%, significantly reducing network downtime by 40% and decreasing throughput degradation from 25% to 10% [5]. This integrative process addresses critical concerns related to model interpretability, acceptance, and operational resilience during live network operations, while also highlighting challenges such as computational overhead and ensuring robustness against erroneous decisions. The use of hierarchical agent designs and rigorous validation methods has been proposed to mitigate these challenges, promoting reliable and transparent AI-human collaborative fault management [5].

5.3 Experimental Achievements

Empirical evaluations demonstrate that integrating LLM-based agentic AI within the O-RAN architecture significantly enhances fault management capabilities. Experimental results show fault detection accuracy reaching up to 95%, representing a marked improvement over baseline accuracies of approximately 78% [3, 5, 14]. Mitigation success rates are observed to improve by more than 20%, while network downtime is reduced by nearly 40%. Additionally, throughput degradation decreases from about 25% in baseline systems to approximately 10%, illustrating more efficient and resilient network operation [5].

These enhancements stem from the agentic AIs' capacity to proactively anticipate faults and execute diverse, context-aware responses, outperforming traditional heuristic or static rule-based methods that often produce delayed or partial fault handling [14]. Moreover, improvements in throughput reflect real-time resource optimization and adaptive network reconfiguration performed autonomously by agents embedded within the RIC and SMO layers, thereby validating both the practical feasibility and operational effectiveness of the proposed architecture [5].

Table 8 summarizes the key performance improvements observed experimentally, highlighting significant gains in detection accuracy, mitigation success, downtime reduction, and throughput degradation reduction compared to baseline O-RAN implementations.

These results underscore the potential of LLM-driven agentic AI to enhance the resilience of next-generation wireless networks by enabling intelligent, autonomous fault management that adapts dynamically to the network context. Challenges such as computational overheads and security considerations remain, but ongoing work focuses on mitigation strategies including hierarchical agent design and rigorous validation [5].

5.4 Comparative Performance Analysis

When compared to conventional fault management techniques reliant on manual or semi-automated processes, LLM-driven agentic AI exhibits superior adaptability and resilience. Traditional methods frequently fail to capture the intricate interdependencies and temporal dynamics inherent in multi-source network data, resulting in suboptimal fault isolation and extended recovery times. In contrast, LLM agents synthesize multimodal inputs and apply contextual reasoning to enable expedited and more accurate fault classification and resolution pathways [5]. Furthermore, experimental evaluations demonstrate that LLM agentic AI achieves a fault detection accuracy of 95%, a mitigation success rate of 91%, and reduces downtime by 40%, significantly outperforming baseline approaches. Throughput degradation is also lowered from 25% to 10% in tested scenarios, illustrating enhanced network performance during fault conditions.

Notably, the continuous learning capabilities embedded in these agent architectures promote sustained performance improvements by dynamically adapting to evolving network topologies and traffic profiles. This adaptive proficiency positions agentic AI as a markedly superior solution for managing the increasing heterogeneity and scale of next-generation wireless infrastructures. Additionally, by integrating LLM-based agents within key O-RAN

Table 6: Performance Comparison of LLM-Based Agentic AI Approach vs. Baseline in O-RAN Fault Management

Metric	Baseline	Proposed LLM-Based Agent	Improvement
Fault Detection Accuracy	78%	95%	+17%
Mitigation Success Rate	70%	91%	+21%
Downtime Reduction	-	40%	-
Throughput Degradation	25%	10%	-15%

Table 7: Performance Comparison of Agentic AI-Enabled O-RAN vs. Baseline Systems

Metric	Baseline System	Agentic AI-Enabled O-RAN	Improvement
Fault Detection Accuracy	78%	95%	+17%
Mitigation Success Rate	70%	91%	+21%
Network Downtime Reduction	-	40%	-
Throughput Degradation	25%	10%	-15%

components such as the Near-RT RIC and SMO, the system autonomously monitors network telemetry, interprets faults using natural language processing, and executes mitigation strategies including dynamic resource re-allocation and self-healing operations [5]. These capabilities collectively contribute to enhanced resilience and robustness beyond that achievable with traditional approaches.

5.5 Recognized Challenges

Despite these significant advancements, deploying LLM-based agentic AI within O-RAN architectures poses several critical challenges. First, the computational overhead inherent to large-scale LLM inference raises pressing concerns regarding latency and energy efficiency, especially within resource-constrained edge environments [5, 18]. This necessitates targeted optimization of LLM architectures to achieve real-time, low-power operation suitable for edge deployment. Second, the potential for inaccuracies stemming from incomplete or noisy input data, entrenched model biases, or adversarial conditions threatens network stability through erroneous decision-making. Mitigating these risks calls for robust data preprocessing techniques, stringent model validation, and advanced adversarial resilience mechanisms. Third, AI-specific security vulnerabilities—including data poisoning and model inversion attacks—demand comprehensive, multilayered safeguards that protect both data integrity and model confidentiality. Finally, ensuring seamless interoperability of LLM agents within heterogeneous, multi-vendor ecosystems characterized by proprietary interfaces and diverse data semantics remains an outstanding issue, complicating standardized deployment [5, 18]. Approaches such as hierarchical agent design combined with rigorous validation protocols show promise in addressing this complexity. Collectively, these challenges underscore the multifaceted difficulty of operationalizing agentic AI at scale while maintaining network performance, reliability, and security.

5.6 Proposed Solutions and Optimizations

To address the aforementioned challenges, several strategies have been proposed. A hierarchical agent design paradigm advocates for lightweight, edge-deployable agents managing real-time, low-level tasks, while delegating more computationally intensive LLM operations to centralized or cloud environments. This approach effectively balances computational load with latency requirements and is critical for optimizing performance in resource-constrained edge environments [5]. Rigorous validation frameworks that incorporate simulated fault injection and continuous model retraining serve to reduce erroneous actions and bolster model robustness, ensuring reliable autonomous operation. Resource-constrained edge deployment benefits from advanced optimization techniques such as model pruning, quantization, and knowledge distillation, which significantly reduce computational demands without sacrificing accuracy. Additionally, the adoption of standardized open interfaces and semantic data models within the O-RAN architecture enhances interoperability and facilitates seamless adaptation across multiple vendor implementations, thereby addressing the complexity inherent in multi-vendor ecosystems [5]. Collectively, these solutions form a comprehensive and scalable pathway toward practical deployment of LLM-driven agentic AI, enabling improved network resilience, fault detection, and self-healing capabilities demonstrated in recent experimental evaluations. These evaluations have shown fault detection accuracy improvements from 78% to 95%, mitigation success rates increasing from 70% to 91%, a 40% reduction in network downtime, and a significant decrease in throughput degradation from 25% to 10% [5]. Such results underscore the effectiveness of the proposed optimization strategies in enhancing O-RAN self-healing and operational reliability.

5.7 Future Directions

Future research endeavors in advancing agentic AI capabilities within complex O-RAN ecosystems revolve around several pivotal and interconnected areas. A primary focus is on enhancing multi-agent coordination to enable cooperative fault detection and mitigation across distributed RAN contexts, which promises improved resilience through collective intelligence [37]. These coordination challenges include synchronization issues, backward compatibility with legacy systems, and computational complexity associated with

Table 8: Performance Comparison between Conventional Fault Management Baseline and LLM-Driven Agentic AI [5]

Metric	Baseline	LLM-Driven Agentic AI	Improvement
Fault Detection Accuracy	78%	95%	+17%
Mitigation Success Rate	70%	91%	+21%
Downtime Reduction	-	40%	-
Throughput Degradation	25%	10%	-15%

real-time processing. For example, designing efficient synchronization protocols that maintain low latency in large-scale deployments remains an open research problem critical for practical applications.

Another critical direction involves deepening AI explainability methods to elucidate agent decision-making processes. This is essential for fostering operator trust and ensuring compliance with regulatory frameworks [18]. Current explainability frameworks could be enriched by integrating interdisciplinary approaches combining AI, control theory, and wireless signal processing, aiming to develop interpretable yet high-performance models. Specific research questions include how to balance the trade-offs between model complexity and interpretability, and how to present insights in operator-accessible formats without sacrificing real-time operational capabilities.

Security remains paramount as agentic AI frameworks face evolving cyber threats. Strengthening defenses through adversarial training, secure model update protocols, and anomaly detection mechanisms is vital to safeguard network integrity [24]. Open problems in this area involve developing scalable security architectures that operate efficiently under resource constraints typical of edge deployments, and ensuring robustness against novel attack vectors in dynamically changing network environments.

Practical deployment considerations further emphasize the need for scalable, real-time capable architectures that balance accuracy, latency, and resource usage. Research into architectural frameworks combining edge-cloud synergy is necessary to meet these demands, as highlighted in recent surveys [24]. For instance, achieving seamless orchestration between local AI inference and centralized model updates while maintaining minimal communication overhead is a pressing challenge.

Additionally, integrating adaptive resource allocation and intelligent sequence management techniques presents opportunities for optimizing network performance amid dynamic conditions [37]. Transitioning from static to fully autonomous, self-optimizing networks capable of real-time adaptation is especially critical for supporting massive IoT environments. Methodological frameworks that combine reinforcement learning-based dynamic resource allocation with robust forecasting and anomaly detection are crucial to guiding future research efforts. Addressing how to effectively model temporal dependencies and environmental uncertainty remains an open question for such systems.

To crystallize these challenges and opportunities, Table 9 summarizes key research areas, associated challenges, and actionable research questions to facilitate targeting efforts in this evolving domain.

Collectively, these future directions underscore the increasing complexity of O-RAN networks and the indispensable role of sophisticated AI agents in ensuring their resilience, security, and operational excellence. Addressing these multifaceted challenges demands interdisciplinary collaboration, holistic frameworks, and clearly defined research milestones to enable sustainable deployment and continuous evolution of AI-empowered O-RAN systems.

6 Adaptive Control and Reinforcement Learning in Networking Systems

Adaptive control and reinforcement learning (RL) have emerged as pivotal techniques for optimizing networking systems by dynamically adjusting control policies based on observed network conditions. These methods provide frameworks for handling uncertainties and non-stationarities typical in network environments, enabling efficient resource allocation, traffic management, and protocol adaptation.

A significant focus within this domain has been on gradient-based adaptive control methods. Gradient-based approaches utilize the gradient information of a performance metric or cost function with respect to control parameters to iteratively improve the system's behavior. For example, in network congestion control, gradient descent can adjust sending rates to optimize throughput and minimize delay, effectively adapting to network dynamics. The advantage of these methods lies in their relatively straightforward implementation and convergence properties under smooth cost landscapes.

Reinforcement learning adds an additional dimension by allowing network agents to learn optimal control policies through trial-and-error interactions with the environment, without requiring an explicit model. This is particularly useful in complex or partially observable network scenarios where modeling is infeasible. Model-free RL algorithms, such as Q-learning and policy gradient methods, have been applied to routing, resource allocation, and energy management in wireless networks.

To illustrate, consider a case study in adaptive routing where an RL agent optimizes path selection to minimize latency and packet loss over fluctuating network topologies. Using policy gradient methods, the agent incrementally improves its routing policy based on received rewards signaling successful data delivery, leading to adaptive, efficient routing in real time. This example highlights the synergy between gradient-based optimization and RL strategies in handling dynamic and stochastic network environments.

In addition to isolated techniques, the integration of gradient-based adaptive control within reinforcement learning frameworks has enabled the design of algorithms that combine the sample efficiency of gradient methods with the flexibility of RL. Such hybrid

Table 9: Summary of Future Research Challenges and Opportunities in Agentic AI for O-RAN

Research Area	Key Challenges	Actionable Research Questions
Multi-Agent Coordination	Synchronization latency, backward compatibility, computational complexity	How to design efficient synchronization protocols for large-scale systems? How to maintain compatibility without performance degradation?
Explainability	Model interpretability vs. complexity, operator-friendly insights, real-time constraints	How to integrate control theory and wireless signal processing for interpretable AI? What visualizations enhance operator trust without delay?
Security	Adversarial robustness, secure updates, anomaly detection at the edge	How to build scalable security mechanisms under resource constraints? What new threats arise in adaptive O-RAN AI agents?
Deployment Architecture	Edge-cloud synergy, latency, resource balancing	How to orchestrate local and centralized AI inference efficiently? What are overhead bounds for model updates?
Adaptive Resource Allocation	Dynamic environments, massive IoT, real-time adaptation	How to model temporal dependencies for RL-based allocation? How to combine forecasting with anomaly detection effectively?

methods demonstrate promising results in network settings requiring rapid adaptation to changing conditions.

Despite these advances, applying RL effectively in open radio access networks (O-RAN) and emerging 6G systems presents multiple challenges. These include the high dimensionality of state and action spaces, the need for real-time decision-making under partial observability, and the difficulties of ensuring robust performance amidst the openness and heterogeneity of O-RAN infrastructures. Additionally, exploration in safety-critical environments must be carefully balanced with the risk of degraded service during learning phases. These open problems motivate ongoing research to develop scalable, safe, and explainable RL algorithms tailored to the unique constraints of O-RAN and 6G networks.

Moreover, adaptive control methods complement RL by providing principled mechanisms for continuous parameter tuning and system stability, which are essential for integrating learning-based agents within the operational cycles of networking systems. This synergy is particularly relevant in the context of LLM-driven agent architectures discussed earlier, where large language models (LLMs) can serve as high-level reasoning modules guiding RL agents' exploration and adaptation strategies. Integrating adaptive control, RL, and LLM-driven intelligence could lead to more responsive, interpretable, and robust agent designs capable of managing the complex dynamics of next-generation networks.

In summary, adaptive control and reinforcement learning provide powerful tools for enhancing performance and resilience in networking systems. Gradient-based methods offer a principled approach to continuous adaptation, while reinforcement learning facilitates operating under uncertainty and incomplete knowledge. The combination of these methodologies, especially when integrated with advanced LLM-driven agent architectures, leads to robust, efficient network control solutions capable of meeting the demands of modern communication infrastructures such as O-RAN and 6G.

The following sections delve deeper into specific gradient-based algorithms and reinforcement learning case studies, building on the concepts introduced here and illustrating their application across diverse networking scenarios.

6.1 Applications of Reinforcement Learning

Reinforcement learning (RL) has emerged as a crucial methodology for real-time adaptive control in dynamic and wireless networking environments. RL enables the optimization of system performance under stochastic and time-varying conditions by learning policies that map network states to appropriate actions through direct interaction with the environment [2, 17, 21, 24, 29, 32**? ? ? ?**]. This capability contrasts with traditional model-based control methods that depend on fixed policies or heuristics, offering autonomous decision-making tailored to the complexities inherent in modern networks.

RL's versatility is demonstrated across diverse networking scenarios, including cellular self-organized networks (SON) and multi-hop wireless ad hoc systems, where it addresses critical challenges such as interference mitigation, handover optimization, and load balancing [17, 21, 29**? ?**]. In particular, deep RL (DRL) methods leverage deep neural networks—such as convolutional and recurrent architectures—to approximate value functions or policies, enabling rapid adaptation without explicit model dependencies. This feature is especially vital in heterogeneous and uncertain wireless contexts, such as multi-band communication networks, where differing propagation and interference characteristics across frequency bands complicate control [32].

The effectiveness of RL-based adaptive control systems hinges on accurate and expressive state representations capable of handling high-dimensional and partially observable network environments. Recent literature highlights the synergy between RL and deep learning architectures—including convolutional, recurrent, and autoencoder networks—to extract pertinent features and exploit spatial-temporal correlations, thereby accelerating convergence and improving robustness [24**? ?**]. Additionally, advanced analytical frameworks for optimal control on networked systems, such as modal decomposition grounded in network spectral properties, complement RL approaches by providing interpretable and computationally scalable solutions [17, 21]. For example, modal decomposition techniques decouple network dynamics into independent eigenmodes, facilitating efficient control design with reduced complexity and enhanced scalability.

Despite such advances, enduring challenges remain. Maintaining robustness amid non-stationary traffic patterns and fluctuating wireless channels demands adaptive generalization capabilities, motivating research into meta-learning and transfer learning to enhance policy reuse across varying network states [29]. Moreover, deploying RL in latency-sensitive and resource-constrained environments is limited by the computational overhead of both inference and training. To address this, efforts focus on lightweight model architectures, hardware acceleration, and hybrid optimization algorithms that blend RL with classical control and optimization methods [2**?**]. For instance, combining RL with stochastic dual gradient algorithms exploits the strengths of both learning-based adaptation and analytical optimization for improved network resource management.

Collectively, these developments position reinforcement learning as a transformative tool for autonomous, efficient, and scalable management of increasingly complex wireless networks, particularly within emerging 5G and 6G paradigms where dynamic adaptability and intelligent resource allocation are paramount [24**?**]. Continued integration of RL with domain-specific knowledge and real-time

analytical models promises further enhancements in performance, robustness, and interpretability of network control solutions.

6.2 Deep Reinforcement Learning for Online Adaptation

Deep reinforcement learning (DRL) advances conventional RL by employing deep neural networks as function approximators for policies or value functions. This facilitates effective online adaptation for complex tasks such as decision-making, resource allocation, and adaptive bandwidth management in networking systems [20, 24**? ? ?**]. DRL is particularly valuable in environments characterized by high-dimensional state and action spaces where explicit policy engineering is infeasible, including dynamic spectrum allocation, power control, and admission control [20**?**].

By continuously learning from environmental interactions, DRL enables resource management policies to adapt dynamically to fluctuating network conditions, often outperforming heuristic or static strategies. Hybrid frameworks that integrate DRL with optimization techniques have shown improved convergence rates and enhanced performance in resource-constrained scenarios—for example, federated learning (FL) systems operating under heterogeneous wireless bandwidth constraints [24]. Incorporating domain-specific knowledge into DRL architectures further enhances the balance between exploration and exploitation, which is critical for achieving real-time adaptability.

However, deploying DRL in networking systems introduces challenges such as increased demand for training data, limited interpretability of learned models, risks of overfitting, and instability under non-stationary input distributions [**?**]. Mitigation strategies including experience replay buffers, target networks, and transfer learning are employed to address these issues. Nevertheless, achieving the optimal trade-off between model expressiveness and computational efficiency remains an ongoing area of research, especially given the stringent latency and scalability requirements inherent to next-generation wireless networks. Future advancements will likely involve adaptive AI models capable of real-time optimization, cross-layer integration, and AI robustness enhancement to meet ultra-low latency and massive connectivity goals envisaged in beyond-5G and 6G systems [24**?**].

6.3 Challenges in Policy Design

The design of RL policies for networking systems necessitates a careful balance between exploration and exploitation in environments characterized by non-stationarity, such as wireless networks [18**? ? ?**]. Exploration is essential for discovering improved policies but can degrade performance and increase latency, which are critical concerns in mission-critical or ultra-reliable low-latency communication (URLLC) applications. Conversely, excessive exploitation risks converging to suboptimal policies when network traffic patterns or channel conditions change dynamically.

To mitigate these challenges, state-of-the-art techniques incorporate adaptive exploration rates that adjust based on environmental feedback, uncertainty-aware policy learning to account for incomplete and noisy information, and reward shaping that directly aligns with key networking performance metrics [**? ?**]. The stringent low-latency inference demands in networking impose constraints on model complexity and motivate efficient state acquisition strategies. However, accurate state information remains difficult to obtain due to measurement noise, delays, and partial observability inherent in distributed systems [18**?**].

Robust state estimation thus becomes essential, often realized through filtering methods or latent state representations learned jointly within RL frameworks. Furthermore, to ensure that policies generalize effectively across heterogeneous devices and diverse, dynamic network environments without sacrificing responsiveness, meta-reinforcement learning and multi-agent RL paradigms have been proposed [**? ?**]. These approaches enable rapid adaptation and cooperative decision-making suited for complex next-generation network architectures, such as AI-assisted network slicing and integrated optical wireless communications, which demand both reliability and agility [**? ?**].

Notably, [**?**] highlights that AI-assisted network slicing frameworks leverage RL to dynamically allocate resources and control slice admission, adapting to the fluctuating traffic and QoS requirements characteristic of next-generation wireless networks. Similarly, advances in optical wireless communications (OWC) for 6G networks [**?**] present additional challenges due to channel variability and device heterogeneity, necessitating RL policies that can swiftly adapt while maintaining robustness. These case studies underscore the importance of integrating domain-specific knowledge with advanced RL strategies to meet the rigorous demands of emerging network technologies.

6.4 Federated and Distributed Reinforcement Learning

Federated reinforcement learning (FRL) and distributed reinforcement learning paradigms have gained significant attention in networking systems due to their potential to enhance privacy, reduce computational burdens, and accelerate convergence within edge-cloud ecosystems [6, 24]. By decentralizing the training process, multiple clients—such as base stations or edge devices—collaboratively learn coordinated policies without sharing raw data, thereby inherently supporting privacy preservation and compliance with regulatory frameworks.

To address communication constraints typical in bandwidth-limited wireless environments, techniques like gradient sparsification and adaptive client selection are employed to reduce communication overhead [6]. State-of-the-art research demonstrates that FRL maintains robust policy performance in the presence of client dropout and heterogeneous data distributions by leveraging mechanisms such as error feedback and weighted aggregation of client updates [6]. Moreover, integrated resource allocation strategies jointly optimize bandwidth and computational resources, which accelerates training convergence and enhances accuracy in FL-based wireless networks [24].

Despite these advancements, key challenges remain, including the synchronization of distributed RL agents, mitigating delays caused by straggler clients, and defending against adversarial attacks. Future research directions emphasize developing asynchronous FRL algorithms to improve training efficiency, employing stronger privacy-preserving techniques such as differential privacy

to safeguard sensitive information, and designing scalable architectures capable of managing the complexity of forthcoming 6G networks. These directions align with broader AI-driven network optimization goals and are crucial for fully leveraging FRL's potential in next-generation wireless environments.

6.5 Integration Across Networking Frameworks

The efficacy of reinforcement learning (RL) and adaptive control techniques is significantly enhanced when integrated with complementary AI-driven networking frameworks such as network traffic classification, software-defined networking (SDN), and routing optimization [13, 16, 39]. Deep learning-based traffic classification models provide granular insights into network flow characteristics, effectively overcoming the limitations of traditional methods—such as port-based and deep packet inspection approaches—that struggle with encrypted and dynamic traffic patterns. These advanced models enable RL controllers to optimize resource allocation with greater accuracy and adaptability by prioritizing traffic types based on detailed classification results [16].

Within SDN architectures, RL-powered controllers dynamically adjust routing and admission control policies in response to real-time network state changes, thereby improving throughput, reducing latency, and enhancing fault tolerance. AI integration into SDN controllers combines supervised learning classifiers (e.g., Random Forest, SVM) and deep learning models (e.g., LSTM networks) to perform real-time traffic classification, anomaly detection, and dynamic resource allocation. This integration results in significant performance improvements, as demonstrated by up to 92% traffic classification accuracy, an 18% reduction in end-to-end latency, and increased throughput in 5G and beyond network scenarios [13]. Similarly, RL-based routing protocols adaptively select communication paths by continuously learning from traffic dynamics, balancing load, mitigating congestion and failures, and thereby enhancing network resilience and efficiency [39].

These cross-framework synergies facilitate comprehensive network adaptation strategies, where RL agents leverage enriched contextual information and explicit control channels provided by SDN. However, the integration of multiple AI modules introduces challenges such as elevated computational overhead, interoperability complexities, potential security vulnerabilities, and risks of cascading failures. Addressing these challenges necessitates efforts toward standardization and the development of modular, lightweight AI pipelines optimized for real-time operation. Moreover, incorporating explainable AI technologies is vital for maintaining transparency and manageability in autonomous network operations. Future research directions emphasize privacy-aware federated and decentralized learning approaches to enhance scalability and security within heterogeneous network environments.

6.6 Gradient-Based Optimization and Fast Algorithmic Updates

Gradient-based optimization methods constitute a cornerstone in training and adapting artificial intelligence models. These approaches iteratively optimize an objective function by following the gradient of a loss landscape, enabling efficient convergence to optimal or near-optimal solutions. Key techniques include variants of gradient descent such as stochastic gradient descent (SGD), mini-batch gradient descent, and momentum-based methods, which enhance computational efficiency and convergence stability.

Fast algorithmic updates leverage structural properties and approximations to accelerate these optimization processes. For example, algorithms that exploit sparsity, employ adaptive learning rates, or approximate Hessian information enable rapid adaptation with reduced computational overhead. These advancements are especially crucial in large-scale and online learning scenarios, where swift model updates are essential.

The integration of efficient gradient computations with fast update mechanisms has produced scalable frameworks that facilitate real-time learning and responsiveness in complex models. Continuous developments strive to improve the balance between computational speed and optimization accuracy, thereby augmenting the practical applicability of AI systems across diverse domains.

6.6.1 Gradient Descent and Variants. Gradient-based optimization constitutes a foundational approach for tuning control and network parameters in large-scale communication and data networks. Traditional gradient descent methods, alongside their accelerated variants, have proven effective for scalable optimization tasks. However, challenges such as high-dimensional uncertainty and the presence of integer decision variables considerably complicate these optimization processes. Specifically, while continuous control parameters allow for convergence guarantees under smoothness assumptions, incorporating integer or mixed-integer variables markedly increases computational complexity and complicates theoretical convergence analyses [34]. This challenge intensifies in settings characterized by large state spaces and expansive uncertainty sets, where computational demands grow exponentially with dimensionality.

To enhance scalability, recent algorithmic refinements such as stochastic gradient methods and adaptive learning rate schemes have been developed. These approaches enable efficient parameter updates even in vast, complex networks [7, 32, 35, 36, 38**? ?**]. Nonetheless, the inherently discrete nature of some optimization variables often necessitates hybrid or relaxation-based techniques, carefully balancing solution quality against computational tractability. The treatment of such integer-constrained optimization problems remains a dynamic research area, stimulating both theoretical advancements and practical algorithm design. Future directions include integrating machine learning-based heuristics—such as deep learning models that extract hierarchical features from network data [38]—and adaptive partitioning methods [34] to better handle uncertainty and mixed-integer decision-making within complex communication networks. Such integrations hold promise for achieving improved robustness and efficiency in optimization under realistic network constraints, including those arising in next-generation wireless and software-driven environments.

6.6.2 Hybrid Model- and Data-Driven Gradient Approaches. Recognizing the limitations of purely gradient-driven methods, recent research has focused on hybrid frameworks that integrate model-based insights with data-driven adaptations. These approaches leverage structural knowledge embedded in network models while simultaneously exploiting real-time or historical data to inform

adaptive gradient computations [2**? ?**]. This integration enhances convergence speed and algorithmic flexibility by dynamically adjusting update rules and reducing discrepancies between model assumptions and evolving network conditions.

For instance, in self-optimized wireless networks (SON), deep learning techniques have been combined with model-based control mechanisms to robustly tune parameters across diverse and dynamic environments [36]. This hybrid approach utilizes knowledge graphs and semantic information extracted via deep neural networks and graph neural networks to provide contextual understanding of network states, thereby improving system responsiveness and stability. By fusing data-driven semantic representations with traditional model-based optimization, these frameworks effectively address scalability and convergence challenges in complex, real-world networks. Moreover, the integration confers robustness to environmental variability by embedding semantic context consistency and enabling error correction through inference of missing or distorted semantic elements within the knowledge graph structure [36]. Such hybrid methods thus present a promising direction for achieving intelligent, flexible, and resilient network optimization beyond conventional gradient-based approaches.

6.6.3 Fast Algorithmic Update Techniques. The imperative for rapid recalibration of control policies in dynamic and stochastic network environments has motivated the development of fast algorithmic update methods focused on minimizing computational latency. Speed is critical for enabling online learning and real-time control systems [20, 35, 38**? ? ? ? ?**]. Typical approaches include incremental gradient updates that adjust parameters based on streaming data, warm-starting solvers with prior solutions to reduce convergence time, and employing approximation heuristics that offer computationally efficient yet effective parameter refinements.

In telecommunication networks, these techniques allow systems to respond swiftly to sudden changes in traffic patterns, user demands, or channel conditions, thereby maintaining strict quality-of-service (QoS) guarantees and improving resource allocation efficiency [7]. For instance, the emerging Tactile Internet paradigm imposes ultra-low latency requirements on the order of milliseconds, necessitating update mechanisms that can execute within extremely tight time budgets [7]. Achieving this demands highly optimized algorithmic procedures that balance computational complexity with accuracy.

A central challenge in fast algorithmic updates is managing the trade-off between update speed and solution precision. Aggressive approximations risk degrading policy performance and potentially violating QoS constraints, whereas fully precise updates may incur computational delays incompatible with real-time demands. To mitigate this tension, recent advances incorporate parallel computation and distributed optimization frameworks, which allow decomposition of the update problem across multiple processing units or network nodes. Such architectures scale efficiently with system size and facilitate timely responsiveness without substantially compromising optimality [35**?**].

These algorithmic innovations are foundational for realizing adaptive, autonomous network management capable of maintaining rigorous performance metrics amid highly dynamic telecom environments. By integrating fast update techniques with AI-driven control and optimization frameworks, networks can achieve real-time adaptability, robustness, and improved operational efficiency [35**? ?**].

6.6.4 Case Studies and Benchmarks. Empirical validations of gradient-based optimization and fast update techniques within real-world telecommunication networks provide critical insight into their practical efficacy and constraints [7, 9, 36**? ?**]. Dynamic optimization strategies employing these methods have yielded measurable improvements in network throughput, latency reduction, and resource utilization across diverse scenarios. For instance, the integration of neural network-based information transfer (NNIT) approaches enables effective adaptation to dynamically changing network environments by transforming historical solutions into promising candidates, thereby accelerating convergence in optimization [**?**]. This method effectively learns environmental evolution patterns through training on previous and new solutions, facilitating rapid adjustment to shifting network conditions. Similarly, innovative federated learning schemes applying gradient sparsification and adaptive client selection enhance learning robustness and communication efficiency in resource-constrained wireless settings [**?**]. By minimizing a weighted global loss with error feedback and selecting clients adaptively, these techniques maintain high accuracy and reduce communication overhead significantly, even under high dropout rates. Applications to ultra-low latency scenarios, such as the 5G-enabled Tactile Internet, demonstrate how gradient-informed optimizations contribute to meeting stringent end-to-end delay requirements [7]. The Tactile Internet, powered by 5G innovations like Multi-access Edge Computing and network slicing, demands radical redesigns in optimization to ensure transmission, processing, and retransmission delays collectively remain below one millisecond.

Despite these gains, scalability remains a key limitation, especially for very large instances with high heterogeneity and complex constraints. Challenges such as computational overhead and complex problem landscapes impede direct application of classical gradient-based methods, motivating the incorporation of hybrid metaheuristic approaches [9]. Benchmark studies reveal that while gradient-driven frameworks effectively handle moderately sized networks, augmenting them with metaheuristics—such as variable neighborhood search or population-based heuristics—enhances solution quality and exploration capacity for large-scale combinatorial problems. For example, variable neighborhood search algorithms have been successfully applied to complex hub location problems involving competitive pricing and demand shifts, offering robust and scalable performance [9]. This method adapts multiple neighborhood structures, including swapping hubs, client reallocation, and price adjustments, coupled with shaking and local search strategies that escape local optima and thus robustly navigate diverse solution spaces.

These findings underscore the value of modular algorithmic strategies that adaptively integrate gradient information with heuristic exploration, thereby balancing computational efficiency with solution robustness. Such hybrid techniques hold promise for addressing the diverse challenges posed by dynamic, large-scale telecommunication network optimization scenarios.

6.6.5 Neural Network-Based Information Transfer (NNIT). Addressing the dynamic and time-varying nature of network environments requires methods that not only optimize parameters in the current setting but also leverage learned knowledge from previous environments to accelerate adaptation. Neural Network-Based Information Transfer (NNIT) exemplifies this strategy by employing neural networks to learn mappings between evolving network states and their corresponding optimal or near-optimal solutions, thereby facilitating faster convergence and enhanced adaptability [4, 10, 23**?**].

Typically, NNIT integrates population-based evolutionary algorithms with neural networks trained to predict promising regions of the solution space or to transform historical high-quality solutions into effective candidates for the current environment [**?**]. This transfer of information capitalizes on patterns inherent in environmental changes and substantially reduces computational overhead in dynamic optimization problems. For example, applications in supply chain networks—characterized by complexity and dynamic features similar to telecommunications systems—demonstrate computational gains and robustness from NNIT techniques [23]. Specifically, the use of variational inequalities and Lagrange multiplier analysis in modeling supply chain equilibria provides numerical insights that guide informed solution transfer strategies within the NNIT framework.

Additionally, recent advances in interpretable AI have been incorporated into NNIT architectures to enhance transparency and trustworthiness by revealing insights into the solution landscape [10]. Such interpretability is crucial for deployment in safety-critical and high-stakes network control scenarios, where understanding the rationale behind decisions complements optimization performance. The interpretable models, trained to represent complex constraints and objectives with approximate yet globally insightful structures, allow decision-makers to examine feasible regions and objective surfaces effectively.

Further, NNIT shows promise in nonlinear stochastic decentralized adaptive control applications, reflecting its versatility in addressing a broad spectrum of network optimization challenges [4]. The control-theoretic framework underlying these extensions optimally allocates resources to mitigate adverse dynamic effects such as economic shock propagation, emphasizing both strategic intervention and computational scalability.

In summary, NNIT represents a robust, hybrid paradigm that synergistically combines population-based optimization, learned knowledge transfer, and interpretable modeling. This approach effectively addresses the complexity, scalability, and dynamism inherent in contemporary and future network control tasks. By exploiting complementary strengths, such integrated algorithmic designs achieve superior optimization performance and adaptive capacity across diverse dynamic and uncertain environments.

7 AI-Enhanced Wireless Networking and Sensing

This section explores how artificial intelligence (AI) techniques improve wireless networking and sensing, with a particular focus on their integration with reconfigurable intelligent surfaces (RIS). We begin by outlining the main objectives, followed by detailed discussion of methodologies, benefits, challenges, and future directions. A summary table is also provided to aid clarity and retention.

AI integration with RIS aims to achieve dynamic environment control for enhanced communication reliability, efficiency, and adaptability in wireless networks. Key objectives include optimizing signal propagation, managing interference, and improving resource allocation in diverse and challenging scenarios.

AI algorithms in RIS-assisted networks primarily utilize reinforcement learning and deep learning to configure surface elements such as phase shifts. These configurations enhance channel state prediction, link reliability, and spectral efficiency under variable conditions.

In addition, AI-driven interference management frameworks leverage real-time data analytics to detect and mitigate interference, thereby improving network scheduling and resource utilization. This is critical for enabling the coexistence of multiple users and technologies without significant performance loss.

Recent benchmarking results demonstrate that AI-powered RIS systems achieve notable enhancements in signal-to-noise ratio (SNR), latency reduction, and energy efficiency, applicable to environments such as millimeter-wave communications and multi-user MIMO setups.

In summary, AI-driven enhancements in RIS-based wireless sensing and networking empower adaptive and context-aware systems that meet stringent requirements of next-generation applications. However, challenges remain in scaling AI techniques for large-scale deployments, ensuring robustness, and integrating AI seamlessly with evolving wireless standards. Addressing these will be critical for realizing the full potential of AI-enhanced wireless environments.

7.1 Reconfigurable Intelligent Surfaces (RIS)

Reconfigurable Intelligent Surfaces (RIS) have emerged as a transformative technology enabling programmable manipulation of wireless propagation environments. Unlike conventional wireless systems that treat the environment as a stochastic and uncontrollable factor, RIS impose deterministic control through engineered metasurfaces capable of dynamically altering incident electromagnetic waves. The integration of Artificial Intelligence (AI), particularly machine learning techniques, significantly enhances RIS functionality by enabling adaptive optimization over complex, high-dimensional configuration spaces. Supervised learning methods facilitate channel estimation by mapping measured channel state information (CSI) to optimal RIS configurations. Unsupervised approaches enable feature extraction from unlabeled channel data, improving generalization to dynamic environments. Moreover, deep reinforcement learning (DRL) offers an effective framework for sequential decision-making under uncertainty, enabling adaptive beamforming and resource allocation policies that maximize spectral efficiency and energy savings [6]. The synergy of these AI paradigms empowers RIS to overcome challenges posed by nonlinear and time-varying wireless channels, addressing issues such as high-dimensional configuration spaces, latency, scalability, and imperfect channel information. This combination ultimately achieves

Table 10: Summary of AI Techniques and RIS Benefits in Wireless Networking and Sensing

Aspect	AI Methodologies	Benefits	Challenges and Future Directions
RIS Configuration	Reinforcement Learning, Deep Learning	Adaptive phase shift optimization, improved channel state prediction	Scalability in large RIS arrays, real-time learning efficiency
Interference Management	Real-time Data Analytics, Machine Learning	Enhanced interference identification, resource allocation, network scheduling	Robustness to dynamic environments, multi-user coexistence complexity
Performance Gains	AI-assisted RIS control	Increased SNR, reduced latency, improved energy efficiency	Integration with diverse wireless standards, hardware limitations
Application Scenarios	Millimeter-wave, multi-user MIMO systems	Context-aware adaptivity, improved spectral efficiency	Deployment cost, reliability under harsh propagation conditions

more robust and efficient wireless links, significantly boosting coverage and energy efficiency. Future research directions emphasize the development of lightweight distributed AI algorithms tailored for RIS, federated learning techniques to preserve user privacy, and the integration of RIS with emerging technologies such as millimeter wave (mmWave), massive MIMO, and edge computing. These advances are set to further propel intelligent wireless environments, fostering improved network performance and sustainability [6].

7.2 Benefits and Challenges of RIS

The AI-enabled RIS paradigm offers multiple benefits that significantly enhance wireless communication systems. These include notable improvements in spectral efficiency achieved through enhanced directivity and more effective interference management. Additionally, RIS reduces reliance on active radio frequency components, thereby augmenting energy efficiency. It also extends coverage by enabling signal reflection and focusing beyond line-of-sight barriers, which facilitates connectivity in dense urban environments or scenarios with significant obstacles. An important advantage of AI integration is robustness under imperfect channel conditions, as AI algorithms can learn and compensate for noise and fading effects, thereby maintaining high communication quality [6].

However, these benefits come with inherent challenges that must be addressed. The RIS configuration space is typically high-dimensional, rendering exhaustive search or conventional heuristic optimization methods impractical. This complexity necessitates the development of scalable AI algorithms capable of effective dimensionality reduction while preserving performance. Moreover, practical deployment requires AI techniques that meet latency and scalability constraints, motivating the use of lightweight and distributed learning methods such as federated learning, which also offers privacy benefits. Security is another critical concern, as adversaries might exploit RIS for unauthorized eavesdropping or signal manipulation. To mitigate these risks, secure AI-driven configuration protocols and real-time anomaly detection mechanisms are essential [6]. Balancing and addressing these benefits and challenges is central to advancing RIS deployment and realizing intelligent wireless environments.

7.3 Future Prospects in Wireless AI

Looking ahead, lightweight distributed AI architectures are poised to enable real-time and energy-efficient control of RIS, seamlessly integrated with pervasive wireless networks. Federated learning emerges as a key methodology, facilitating decentralized training of RIS optimization models across edge nodes while safeguarding data privacy and reducing communication overhead. This approach is especially vital given the growing heterogeneity of network topologies and the non-independent and identically distributed (non-i.i.d.) nature of data across devices. The convergence of federated AI with cutting-edge physical-layer technologies—such as millimeter-wave (mmWave) communications, massive multiple-input multiple-output (MIMO) antenna arrays, and edge computing platforms—will drive the advancement of edge intelligence [6]. These integrated frameworks are expected to jointly optimize sensing, communication, and computation resources while adhering to strict latency and energy constraints. To achieve these goals, future algorithmic innovations must carefully balance trade-offs among model complexity, convergence speed, and robustness against channel estimation errors. Moreover, emerging paradigms like neuromorphic computing and online continual learning offer promising avenues to enhance system adaptability in highly dynamic wireless environments.

7.4 Intelligent Interference Management in Perceptive Mobile Networks (PMNs)

Perceptive Mobile Networks (PMNs) represent an advanced integration of communication and sensing functionalities, enabling wireless infrastructure to simultaneously support data transmission and situational awareness. Effective interference management is essential because sensing waveforms and communication signals coexist in shared spectral and spatial domains, leading to complex coexistence challenges. Recent advances have introduced AI-empowered interference mitigation frameworks that exploit macro-diversity gains and coordinated beamforming strategies across multi-cell architectures [18]. In particular, deep learning-based interference prediction models utilize both historical and real-time channel observations to accurately forecast interference patterns. This predictive capability enables proactive and dynamic resource allocation, maximizing the sensing signal-to-interference-plus-noise ratio (SINR) while preserving the quality of communication links. These dynamic allocation schemes improve sensing detection probability and simultaneously reduce intra- and inter-cell interference, thereby achieving a balanced optimization of communication and sensing objectives in PMNs.

Despite these promising developments, several critical challenges remain for practical and scalable PMN deployments. Key issues include achieving low-latency inference to support real-time system adaptation, acquiring precise channel state information in highly mobile environments, and designing scalable cooperation schemes among multiple base stations without incurring excessive signaling overhead [18]. Addressing these challenges is pivotal to realizing robust PMNs that can harmonize sensing and communication functionalities effectively. Future research directions highlighted in recent literature emphasize integrating multi-modal sensing capabilities, adopting federated learning for privacy-preserving interference management, and developing mechanisms that enhance robustness under dynamic network conditions [18]. Collectively, AI-driven intelligent interference management frameworks offer

substantial improvements in sensing performance while maintaining reliable communication within integrated sensing and communication networks.

7.5 Achievements and Challenges

AI-driven wireless sensing techniques have demonstrably enhanced detection probabilities and mitigated sensing interference, particularly through cooperative interference management strategies leveraging coordinated multipoint processing and macro-diversity [12, 18, 19]. These improvements enable robust detection performance in dense and heterogeneous network environments characterized by significant interference and channel uncertainty. Furthermore, privacy preservation has emerged as a critical concern within networked sensing, since sensitive environmental or user data may be indirectly inferred through side-channel attacks in cooperative frameworks. Recent studies propose privacy-aware AI algorithms that integrate differential privacy techniques and federated learning to alleviate these risks without significant deterioration of sensing performance [15, 18].

Robustness to heterogeneity in hardware capabilities, channel conditions, and user mobility patterns remains another outstanding challenge. Techniques incorporating model adaptation, transfer learning, and fast adaptation methods such as Zero-Shot Lagrangian Updates have shown promise in addressing such variability but require extensive validation in diverse real-world settings [8**?**]. Consequently, bridging the gap between theoretical AI frameworks and practical deployment necessitates continued research focusing on scalable cooperation protocols, secure architectures, and self-tuning training paradigms that can operate reliably across heterogeneous wireless environments.

In summary, AI-enhanced wireless networking and sensing via RIS and intelligent interference management mark the advent of programmable, efficient, and context-aware wireless systems. This progress hinges on the intricate interplay of algorithmic sophistication—encompassing supervised, unsupervised, reinforcement, and federated learning—and physical-layer innovations [6]. Collectively, these advances establish a rich interdisciplinary frontier poised to shape future wireless ecosystems [6, 8, 12, 15, 18**?** , 19].

8 Explainability, Interpretability, and Trust in AI-Controlled Telecommunication Systems

Explainability and interpretability are critical for fostering trust in AI-controlled telecommunication systems, enabling stakeholders to comprehend, validate, and trust automated decisions. In operational telecom environments, explainability helps engineers and regulators trace the rationale behind AI actions, improving reliability and ensuring compliance.

For instance, in multi-agent reinforcement learning-based network management, explainability frameworks such as attention mechanisms—which highlight important inputs influencing decisions—and feature attribution methods—which quantify the impact of individual features—have demonstrated how agents coordinate resource allocation. These methods offer insights that help network operators validate system behavior and detect anomalies. Such frameworks provide interpretable feedback on agent cooperation patterns, directly impacting service quality and robustness.

Moreover, explainability methods contribute to regulatory compliance in telecommunications by aligning AI model decisions with governance standards including GDPR and the AI Act. Transparent AI decision-making enables automated actions to be audited for fairness, data privacy, and accountability, which is essential for meeting both internal and external telecom regulatory requirements.

Practical case studies from telecom operators illustrate how interpretable AI techniques optimize traffic routing while maintaining explainable decision logs that satisfy internal compliance audits. For example, an explainable resource scheduling model allowed operators to identify unexpected decision triggers and adjust system parameters to improve network throughput without sacrificing transparency.

In summary, explainability in AI-controlled telecom systems bridges the gap between complex algorithmic decisions and stakeholder understanding. By supporting trustworthy deployment that meets both technical and regulatory expectations, these methods ensure AI systems are reliable, auditable, and aligned with operational goals.

8.1 Importance of Transparent AI Decision-Making

The incorporation of artificial intelligence (AI) in adaptive telecommunication and control systems introduces an unprecedented level of complexity, making transparent decision-making an essential attribute to cultivate trust among stakeholders and ensure compliance with evolving regulatory frameworks. Transparency serves as a cornerstone for certifying that AI-driven actions conform to desired operational, ethical, and legal standards, particularly within critical infrastructure sectors such as telecommunications [**? ?**]. The establishment of trust is inherently linked to the system's ability to provide interpretable rationales behind its decisions, thus enabling operators to verify, audit, and justify automated processes [**?**]. This transparency is crucial in highly dynamic and heterogeneous network environments, where AI models must continually adapt to varying contextual conditions without compromising reliability and safety [18]. Additionally, emerging AI governance regulations emphasize explainability as a fundamental principle, compelling telecommunication systems to exhibit clarity in their decision logic and mitigate risks associated with opaque AI behavior [24]. As AI-driven network management faces challenges such as balancing computational complexity with real-time processing requirements and ensuring robustness against adversarial conditions, transparent models help address these by revealing decision pathways and confidence levels. Consequently, transparent AI not only bolsters user confidence but also facilitates regulatory approvals and promotes the widespread adoption of AI-enhanced telecommunication technologies.

8.2 Methods for Interpretability and Explainability

Attaining interpretability within AI-driven telecommunication systems necessitates the integration of explainability mechanisms

directly into core optimization and learning frameworks. Reinforcement learning (RL), a dominant paradigm for dynamic resource allocation and control, often poses challenges to transparency due to the complexity inherent in value function approximations and policy networks. Modern methodologies address this opacity through model-agnostic interpretability techniques and surrogate models that extract actionable insights from trained RL agents. These approaches clarify decision rationales by illuminating factors such as state-action value contributions and reward attributions [2**?**]. Embedding explainability frameworks within optimization algorithms further enhances understanding by elucidating solution trajectories and facilitating sensitivity analyses, enabling operators to comprehend how variations in system parameters influence resource management outcomes [20**?**]. Hybrid frameworks that couple deep learning with symbolic reasoning have been advanced to strike a balance between predictive performance and interpretability, thereby supporting effective human-in-the-loop validation [32]. Moreover, attention mechanisms and gradient-based attribution techniques embedded in neural network architectures highlight key features that influence AI decisions across tasks such as traffic management, fault diagnosis, and spectrum allocation [**?**]. Collectively, these methods form a comprehensive toolset that mitigates the inherent opaqueness of sophisticated AI models, enhancing operational transparency while preserving system efficacy.

8.3 Future Directions

Looking forward, the evolution of explainable AI (XAI) within telecommunication and control systems is poised to address current limitations and emerging challenges through several pivotal advancements. A primary focus is the development of privacy-preserving XAI techniques that balance transparency with stringent data confidentiality requirements common in telecommunication networks [18]. For instance, federated explainability enables interpretability without centralizing sensitive data, thereby supporting privacy regulations and operational constraints as demonstrated in multi-cell perceptive mobile networks where coordinated deep learning mitigates interference while preserving user privacy.

Enhancing interpretability frameworks to be resilient against adversarial manipulation is another critical direction. AI systems deployed in hostile network environments are vulnerable to exploitation that threatens security and reliability [24]. Robust XAI methodologies should integrate anomaly detection and adversarial training to safeguard explanations from malicious interference, as highlighted in AI-driven network management surveys and recent Open RAN implementations [25]. For example, integrating adversarially robust explanations with multi-agent learning can thwart attacks aiming to distort resource allocation or fault diagnosis.

Scaling explainability to handle large-scale communication and control architectures that span cloud, edge, and device layers, as well as multi-agent systems, requires modular, hierarchical interpretation mechanisms capable of providing contextualized explanations across abstraction levels [5]. Emerging AI paradigms, including multi-agent reinforcement learning and large language model (LLM)-driven network intelligence, demand innovative explainability frameworks that capture complex cross-agent interactions and supply natural language interpretability. An illustrative case is the LLM-driven agentic AI approach in Open RAN, which autonomously identifies and mitigates faults with high accuracy while providing interpretable natural language explanations [5].

To clarify these directions and guide both research and implementation, Table 11 summarizes key future directions, associated challenges, and potential solutions in developing XAI for telecommunication and control systems. Practical pathways include leveraging federated learning to preserve data privacy, employing multi-agent models for distributed intelligence, and optimizing lightweight XAI models suited for edge deployment to meet real-time latency and resource constraints identified in 6G and beyond Open RAN networks [25]. Methodological frameworks advocating modular design and hierarchical explanations help manage complexity across heterogeneous network layers. This interdisciplinary synergy among AI, control theory, and wireless technologies is imperative to address operational, security, and scalability challenges.

Explicit research questions to prioritize include: How can federated XAI methods guarantee privacy without sacrificing explanation fidelity? What are effective defenses against adaptive adversarial attacks targeting explainability in real-time network analytics? How can hierarchical XAI frameworks be standardized to ensure interoperability across diverse telecom layers? Which strategies best balance computational overhead and interpretability for LLM-driven agentic AI in constrained edge environments? Addressing these questions through collaborative, multi-disciplinary efforts will enhance trustworthiness, operational safety, and regulatory compliance in AI-controlled telecommunication infrastructures amid growing complexity.

Anticipated disruptive innovations encompass agentic AI embedded with LLM-driven agents for autonomous fault detection and self-healing [5], yielding demonstrable improvements such as a 17% increase in fault detection accuracy and 40% reduction in downtime. Moreover, seamless AI integration within Open RAN architectures fosters adaptive, explainable control loops that enhance network resilience and resource efficiency [25]. Such paradigm shifts promote transparent, robust, and efficient AI-driven telecommunications, enabling the next generation of intelligent, self-optimizing networks that dynamically adapt to evolving conditions.

8.4 Applications in Telecommunications and Networking

This section provides a comprehensive overview of the key applications of artificial intelligence (AI) in telecommunications and networking. We detail the primary objectives, scope, and challenges specific to these domains, systematically examining how AI techniques enhance network performance, reliability, and security. The discussion focuses on four major application areas: network optimization, traffic prediction and management, resource allocation, and fault detection and diagnosis.

8.4.1 Network Optimization. AI techniques have been extensively applied for network optimization by learning from historical data to dynamically adjust configurations. This adaptation improves throughput and reduces latency while effectively handling the heterogeneous and large-scale nature of modern networks. Despite the demonstrated performance gains, challenges such as achieving real-time processing at scale and managing diverse data sources

Table 11: Summary of Future Directions, Challenges, and Solutions in Explainable AI for Telecommunications and Control

Future Direction	Challenges	Potential Solutions
Privacy-preserving XAI	Ensuring data confidentiality while maintaining explanation fidelity	Federated explainability, decentralized interpretation frameworks [18]
Adversarially robust explanations	Vulnerability to attacks compromising network security and explanation integrity	Integration of anomaly detection, adversarial training, and robust model design [24, 25]
Scalable, hierarchical interpretability	Managing complexity across multi-layer network architectures and distributed agents	Modular frameworks with hierarchical explanation models providing multi-level context [5]
Multi-agent and LLM-driven XAI	Complexity in cross-agent interactions and computational overhead on constrained devices	Agentic AI frameworks leveraging lightweight LLM optimization and multi-agent coordination [5]
Real-time edge deployment	Meeting latency and resource constraints for timely explanations	Development of lightweight XAI models optimized for edge devices, hardware acceleration [25]
Interdisciplinary synergy	Integrating AI, control theory, and wireless communication disciplines	Modular, layered frameworks bridging theoretical and practical gaps across domains

persist. Compared to traditional rule-based and statistical methods, AI approaches offer greater adaptability and scalability but may introduce additional computational overhead, requiring careful consideration during deployment.

8.4.2 Traffic Prediction and Management. Traffic prediction models leverage machine learning to anticipate network congestion, enabling proactive management strategies that improve overall service quality. These models face challenges stemming from dynamic traffic patterns and data sparsity, which may limit prediction accuracy. While AI methods often outperform conventional statistical techniques in capturing complex temporal patterns, their reliance on large volumes of high-quality data can be a limitation in practice. Metrics such as prediction accuracy, root mean square error (RMSE), and mean absolute error (MAE) are commonly used to evaluate model performance.

8.4.3 Resource Allocation. AI-driven resource allocation addresses complex constraint satisfaction problems intrinsic to network resource management. These methods promote efficient and fair utilization by balancing competing demands among users. However, fairness considerations and computational complexity pose significant challenges. AI solutions typically provide better resource efficiency and fairness indices than heuristic approaches, though their interpretability and deployment complexity warrant further investigation.

8.4.4 Fault Detection and Diagnosis. AI-based fault detection systems are designed to identify and diagnose anomalies within network operations, effectively handling imbalanced datasets and minimizing false alarms. Their ability to detect subtle and rare events improves network resilience. Nonetheless, the trade-off between detection rate and false positive rate remains critical, especially in real-world scenarios where false alarms can disrupt service. Evaluation metrics such as detection rate, precision, recall, and false positive rate are essential in assessing these systems.

Overall, while AI methods present notable advantages over traditional approaches in adaptability and performance, they introduce challenges related to computational overhead, data requirements, interpretability, and deployment complexity. Addressing these limitations is essential to fully realize AI's transformative potential in telecommunications. Future research should focus on scalability, real-time implementation, fairness, and explainability to ensure robust and efficient AI-enabled network systems.

This structured examination demonstrates the interrelated nature of these AI applications, emphasizing their collective contribution to the robustness and intelligence of modern communication networks.

8.4.5 AI-Driven Adaptive Control Applications. The integration of artificial intelligence (AI) techniques, particularly machine learning (ML) and deep learning (DL), has substantially transformed adaptive control mechanisms in telecommunications networks. These advancements have empowered functions such as dynamic resource allocation, congestion management, fault tolerance, and traffic prediction. Deep learning architectures—including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and generative adversarial networks (GANs)—exhibit remarkable capabilities in extracting complex spatial-temporal patterns from network data, thereby improving both predictive accuracy and control responsiveness [2, 7, 24, 35**? ? ? ?** , 36].

This progress reflects a paradigmatic shift from traditional static, heuristic-based methods toward data-driven adaptive frameworks, capable of learning from extensive historical and real-time network states. Reinforcement learning (RL), for instance, has been effectively applied to dynamic radio resource allocation, optimizing trade-offs between throughput and latency in heterogeneous wireless environments [36**?**]. Deep learning models such as CNNs and RNNs have demonstrated superior performance in traffic prediction by leveraging nonlinear dependencies and temporal correlations within network flows, outperforming classical statistical predictors [2**? ?**]. Additionally, GANs are instrumental in synthetic data generation and anomaly detection, thereby enhancing network fault management through identification of rare events and coping with sparse failure data [7**?**].

Nonetheless, deploying these sophisticated models entails significant challenges. The computational overhead associated with training and inference of complex deep learning models can impede real-time application, particularly in large-scale or resource-constrained edge environments [2**?**]. Federated learning (FL) has emerged as a promising solution to these challenges by enabling distributed model training while preserving data privacy and reducing communication overhead. Robust FL frameworks combine gradient sparsification, error feedback, and adaptive client selection to tackle issues like client dropout, limited bandwidth, and heterogeneous device capabilities [**?**]. Joint optimization of model parameter size and bandwidth allocation further enhances FL efficiency, reducing training time and improving accuracy in wireless networks with diverse device and channel conditions [2]. Moreover, generalization remains a critical issue, as models must adapt continuously to heterogeneous and evolving network conditions, requiring efficient retraining or adaptation mechanisms [**?**]. Privacy concerns, arising from the collection and processing of extensive network data, drive the adoption of privacy-preserving learning frameworks such as federated learning, which maintains data locality while collaboratively improving model performance [7, 35]. Despite these

Table 12: Summary of AI Applications, Challenges, and Evaluation Metrics in Telecommunications and Networking

Application Area	Challenges	Evaluation Metrics
Network Optimization	Scalability, real-time processing, heterogeneous data sources	Throughput, latency, resource utilization
Traffic Prediction and Management	Dynamic traffic patterns, data sparsity	Prediction accuracy, RMSE, MAE
Resource Allocation	Complex constraint satisfaction, fairness among users	Resource efficiency, fairness indices
Fault Detection and Diagnosis	Imbalanced data, anomaly identification	Detection rate, false positives, precision, recall

challenges, AI-driven adaptive control substantially enhances network self-optimization, reducing operational expenditures while improving quality of service (QoS) [24**?**].

8.4.6 Evaluation Metrics and Benchmarking. Comprehensive evaluation of adaptive algorithms in realistic communication and wireless scenarios necessitates metrics that integrate both classical network performance and AI-specific qualities, including robustness and semantic fidelity. Traditional benchmarks—such as throughput, latency, packet loss, and bit error rate (BER)—remain fundamental indicators of network health [2**?**]. However, the emergence of semantic communications emphasizes the need for novel metrics that transcend bit-level correctness by quantifying the semantic integrity of transmitted data.

In this context, semantic similarity metrics (e.g., BLEU scores when applied to textual or annotated image data) have been proposed to assess the fidelity of content following AI-enhanced compression and error correction schemes [24, 36**?**]. Such metrics align with the semantic communication framework combining deep learning with knowledge graphs that enhances semantic context consistency and error correction [36]. Incorporating these metrics addresses the inadequacies of strictly physical-layer evaluations, realigning optimization objectives with end-user perceived quality. Furthermore, adaptive algorithms are evaluated with respect to computational efficiency, convergence speed, and resilience to adversarial perturbations or network faults [2, 7**?**]. For example, latency models capturing transmission, propagation, processing, queueing, and retransmission delays are critical in Tactile Internet applications requiring ultra-low latency and high reliability [7].

Despite these advancements, standardized benchmarks leveraging established datasets and simulation frameworks remain sparse, impeding cross-comparison and reproducibility. Existing repositories such as MNIST and CIFAR-10 serve as popular datasets in evaluating algorithms for supervised learning and federated learning scenarios [2], yet they do not fully capture the complexity of semantic communications or edge-cloud interactions. There is an urgent call for new datasets and repositories tailored to semantic fidelity, multimodal data, and real-time adaptive networking contexts to reflect practical communication challenges.

To promote uniformity and comparability, best practices for evaluation protocols should include clearly defined metrics capturing semantic, physical, and operational performance, alongside consistent testbed configurations and open-source simulation tools [24]. Protocols must balance accuracy with computational and latency constraints while incorporating interpretability and privacy considerations inherent in AI-driven systems [24**?**]. The development of such comprehensive and standardized evaluation frameworks will facilitate rigorous performance benchmarking, reproducibility, and the accelerated deployment of AI-enabled network control across wireless and edge-cloud environments.

8.4.7 Edge and Cloud Synergistic AI Solutions. The exponential growth of network data and the imperative for ultra-low-latency services have precipitated architectures that synergistically combine edge computing with cloud intelligence. By partitioning AI workloads between decentralized edge nodes and centralized cloud platforms, such hybrid frameworks optimize latency constraints while leveraging substantial computational resources to enhance network intelligence and robustness [6, 20, 24, 38**? ?**].

At the edge, AI models conduct real-time inference and process local data, which is critical for latency-sensitive applications including the Tactile Internet and autonomous vehicle control [7]. To accommodate resource limitations inherent to edge devices, lightweight deep learning models or compressed representations are employed [20**?**]. Concurrently, cloud-based AI systems aggregate global network insights, handle extensive training tasks, and disseminate updated models back to the edge, facilitating continuous learning and adaptive responsiveness [38**?**].

Federated learning exemplifies this edge-cloud synergy by enabling decentralized model training across heterogeneous devices without compromising data privacy or incurring prohibitive communication overhead [7]. To address device heterogeneity, strategies such as adaptive model compression and personalized federated optimization are utilized, tailoring model complexity to the computational capabilities of each device. Synchronization challenges arising from diverse update frequencies and intermittent connectivity are mitigated through asynchronous federated learning protocols and hierarchical aggregation schemes that reduce communication delays and improve model convergence. For instance, grouping devices based on similarity in data distribution or computing power enables more efficient and stable training rounds [7**?**].

In addition, distributed AI architectures grapple with security vulnerabilities including adversarial attacks and data poisoning, prompting research into robust model design and trustworthy deployment at scale [6, 24]. Hence, the interplay between edge and cloud computing represents a crucial frontier for achieving intelligent, responsive, and secure network control in next-generation wireless systems.

8.4.8 Resilient Control of Cyber-Physical Systems. Cyber-physical systems (CPS) underpinning telecommunications infrastructure require resilient control strategies capable of maintaining reliable operation despite actuator faults and sophisticated cyber attacks. A prominent approach involves neural network-based finite-time resilient control methodologies for nonlinear time-delay systems, utilizing radial basis function neural networks (RBFNNs), advanced

observer designs, and Lyapunov–Krasovskii functionals to ensure robustness and rapid convergence [3].

This framework models the system's unknown nonlinearities and fault signals through RBFNNs, where the nonlinear dynamics $f(x(t), x(t-\tau))$ are approximated as $W^T\Phi(x(t), x(t-\tau)) + \varepsilon$. Here, W represents unknown weights, Φ denotes the basis functions, and ε is the approximation error. The unknown weights W are estimated online using adaptive laws, which enable real-time compensation for system uncertainties, time delays, and external disturbances. Concurrently, a specifically designed observer estimates both the system states and fault signals, effectively managing discrepancies caused by unknown false data injections (FDI) and measurement inaccuracies. This dual estimation mechanism significantly enhances fault detection and isolation capabilities within the control loop [3].

The resulting adaptive control laws combine these state and fault estimates to guarantee finite-time convergence of the system states and estimation errors. Unlike traditional asymptotic controllers that achieve stability over an indefinite period, finite-time stabilization ensures all errors vanish within a known finite interval, markedly improving responsiveness and robustness under adversarial conditions. The theoretical foundation relies on Lyapunov–Krasovskii functionals carefully constructed to incorporate the effects of time delays, providing rigorous guarantees of closed-loop system stability despite the presence of unknown nonlinearities and disturbances.

To contextualize, consider a nonlinear system subject to unknown actuator faults and malicious false data injection attacks causing erroneous sensor measurements. The proposed approach uses RBFNNs to approximate nonlinearities appearing in the system's delayed states and employs an observer to dynamically identify both the current system state and any present faults or attacks. By adjusting the control inputs based on these real-time estimates, the controller quickly compensates for faults and maintains system operation within desired performance bounds. Simulations on benchmark nonlinear systems have validated this methodology, displaying superior fault tolerance and faster stabilization compared to classical adaptive controls.

By integrating adaptive neural control with observer-based fault diagnosis and robust stability theory, this paradigm enables real-time mitigation of faults and cyber attacks through dynamic control input adjustments. It thus establishes a comprehensive resilience framework for CPS, effectively addressing both component degradation and malicious interventions. Nonetheless, ongoing challenges remain, such as extending these methods to stochastic systems with time-varying delays, accommodating multi-actuator and sensor configurations, and validating performance through hardware-in-the-loop experiments emulating practical operational scenarios [3].

In summary, this resilient control strategy provides a promising avenue toward enhancing CPS security and reliability, leveraging real-time learning and estimation to not only detect but actively counteract faults and cyber threats in critical infrastructure systems.

8.4.9 Open Research Frontiers. Emerging research trajectories in telecommunications emphasize multi-agent systems, stochastic modeling, and hardware-in-the-loop simulation platforms as pivotal frontiers advancing adaptive control and AI integration [3]. Multi-agent frameworks foster scalable, decentralized decision-making across heterogeneous network entities, enhancing robustness and adaptability in complex, dynamic environments. The incorporation of stochastic models better captures wireless channel variability, environmental uncertainties, and human-in-the-loop behaviors, thereby necessitating adaptive control methodologies that judiciously balance performance against risk [3].

Hardware-in-the-loop platforms constitute indispensable testbeds that bridge algorithmic development and realistic hardware constraints, including timing delays, sensor inaccuracies, and communication limitations. These platforms facilitate expeditious prototyping, validation, and fine-tuning of resilient control schemes under conditions closely emulating actual network environments [3].

Complementary research efforts focus on explainable AI paradigms for network control to ensure transparency and interpretability, and on integrating reinforcement learning with classical control theory to combine long-term policy optimization with guaranteed system stability [24]. Addressing computational complexity through model compression, distributed AI frameworks, and energy-aware algorithms is also critical, especially for deployment within resource-constrained edge environments [6]. Collectively, these avenues herald transformative potential for next-generation telecommunications systems that are intelligent, autonomous, and resilient.

To provide a structured research roadmap, near-term goals include the development and benchmarking of adaptive multi-agent control strategies with robust fault tolerance, leveraging hardware-in-the-loop platforms for realistic validation [3]. Efforts also concentrate on advancing explainable AI methods and integrating reinforcement learning with classical controls to achieve interpretable and stable network management [24]. Addressing computational challenges through lightweight distributed AI and energy-efficient algorithms lays the groundwork for deployment in practical edge environments [6].

Long-term goals envisage the seamless integration of distributed resilient control for complex multi-agent networks operating under dynamic, uncertain wireless environments. This includes extending stochastic adaptive control techniques to accommodate time-varying delays and cyber-physical attack resilience [3]. Future research aims to realize fully autonomous, intelligent wireless systems by harmonizing adaptive control, explainable AI, and scalable reinforcement learning frameworks, all while ensuring robustness, scalability, and sustainability in next-generation telecommunications infrastructures.

9 Cross-Cutting Themes and Integration Considerations

This section synthesizes the key challenges, solutions, and interactions across the various themes discussed in preceding sections, highlighting their intersections through concrete examples and case studies. By examining these cross-cutting issues, we aim to provide a more cohesive understanding of how different AI components and methods integrate, addressing their combined limitations and opportunities.

A primary challenge common across multiple themes is the trade-off between scalability and model interpretability. For instance,

large-scale transformer models achieve impressive performance on natural language processing tasks but often sacrifice transparency in decision-making. This issue intersects with ethical AI considerations, where explainability is crucial for user trust and accountability. An example is the deployment of AI in healthcare diagnostics, where complex models must provide rationale for predictions to meet regulatory standards and clinical acceptance.

Another pervasive theme is the integration of learning paradigms, such as combining supervised learning with reinforcement learning to create robust agents capable of adapting in dynamic environments. A case study from autonomous driving illustrates this integration: perception modules trained on labeled images are combined with reinforcement learning policies that adapt to real-time driving conditions. This integration presents technical challenges such as aligning the learning objectives and managing the propagation of uncertainty, underscoring the need for unified frameworks.

Data privacy and security concerns also cut across themes. Federated learning techniques, originally developed to ensure privacy-preserving model training, have seen widespread application from mobile device personalization to collaborative healthcare research. The cross-cutting challenge lies in balancing data utility with privacy guarantees and computational efficiency. For example, while privacy-preserving federated methods limit data sharing, they can introduce biases if client participation is uneven, necessitating new algorithmic safeguards.

Despite the growing awareness of these interconnected challenges, many existing solutions address them in isolation. To advance AI systems capable of holistic performance, we propose a research roadmap with the following priorities. First, developing unified conceptual frameworks that explicitly model the interactions between scalability, interpretability, privacy, and adaptability. Such frameworks would facilitate joint optimization by capturing trade-offs and synergies across themes. Second, designing modular architectures that support flexible integration of diverse learning paradigms and privacy-preserving mechanisms while maintaining interpretability standards. Third, establishing standardized benchmarks and evaluation protocols that measure multi-dimensional criteria, including transparency, robustness, and privacy, in real-world scenarios.

Methodologically, integrated AI system design should incorporate explicit multi-objective optimization strategies to balance competing demands. For example, optimization schemes can simultaneously consider model accuracy, explainability, privacy preservation, and computational costs. Embedding uncertainty quantification within modular pipelines can help manage error propagation when combining heterogeneous components. Furthermore, adaptive learning schemes that dynamically adjust strategies based on environmental feedback can enhance system robustness and user trust.

In summary, addressing cross-cutting themes in AI requires moving beyond siloed approaches to adopt comprehensive system-level thinking. Prioritizing joint frameworks, modular design, and multi-criteria evaluation will better align AI capabilities with the complex demands of practical deployment, fostering transparency, privacy, and adaptability in tandem.

9.1 Scalability and Real-Time AI Inference

The deployment of artificial intelligence (AI) within telecommunications demands scalable solutions capable of real-time inference across heterogeneous and dynamically evolving network environments. This requirement is challenging due to the high computational complexity of contemporary AI models, such as deep neural networks and large language models (LLMs), coupled with the variability of network resources and stringent latency constraints [6, 13, 39**?**]. For instance, incorporating AI into Open RAN architectures mandates efficient processing pipelines that adhere to tight timing budgets, enabling rapid control loop adaptations critical for tasks like spectrum management and interference mitigation [18].

Edge-cloud collaborative frameworks offer distinct advantages by distributing AI inference workloads to optimize latency and computational resource utilization; however, they face scalability constraints especially when coordinating multiple edge nodes or base stations [6]. To overcome these limitations, scalable AI architectures must integrate a combination of algorithmic compression techniques, model pruning, and hardware acceleration, alongside modular and parallel design principles. Additionally, synchronization and consistent model updating, particularly within federated or distributed learning paradigms, constitute significant challenges that affect maintaining real-time performance [13].

Addressing these complexities is especially crucial for perceptive mobile networks (PMNs), where AI-driven interference management and sensing require dynamic adaptation to fluctuating network loads without compromising communication quality [39]. Advanced AI frameworks in PMNs exploit coordinated beamforming and deep learning-based interference prediction across multi-cell architectures to maximize sensing signal-to-interference-plus-noise ratio (SINR) while preserving communication reliability [18]. Moreover, real-world implementations demonstrate that AI-enabled routing and traffic management can improve throughput and latency by adapting to network conditions in real time, achieving up to 30% improvement in throughput and latency, illustrating practical scalability benefits [39]. Similarly, AI-powered software-defined networking (SDN) solutions applying machine learning models for traffic classification and resource allocation realize up to 92% accuracy and an 18% reduction in end-to-end latency in emulated 5G network scenarios, demonstrating scalable, real-time AI inference capabilities [13].

Overall, realizing scalable and real-time AI inference in telecommunications necessitates integrating lightweight, robust AI models optimized for dynamic network environments, efficient edge-cloud collaboration, and real-time synchronization mechanisms to fully harness the potential of AI for next-generation networks. These case studies underscore the critical role of compression, distributed inference, and adaptive learning in meeting scalability demands while preserving real-time responsiveness.

9.2 Privacy Preservation Strategies

Preserving privacy is a critical concern in telecommunications due to the sensitive nature of transmitted data and strict regulatory frameworks. Prominent strategies for privacy preservation include federated learning, edge computing, and lightweight distributed

AI methods that localize data processing, thereby reducing exposure risks [6, 13, 18, 24]. Federated learning enables collaborative model training across decentralized entities while keeping raw data on-site, effectively mitigating risks inherent in centralized data collection [6]. Despite its advantages, federated learning faces challenges such as communication overhead, the heterogeneity of client devices, and susceptibility to inference attacks. Edge computing complements these approaches by performing AI inference near data sources, which reduces both the privacy attack surface and communication latency [13]. Deploying lightweight AI models at the edge—through techniques like quantization and knowledge distillation—further enhances privacy protection while improving computational efficiency [24].

Moreover, integration of encryption techniques plays a vital role in securing data transmissions and model parameters within telecommunications networks. End-to-end encryption ensures that data remains confidential during communication between distributed entities, effectively preventing unauthorized access and eavesdropping. Homomorphic encryption allows computations to be performed directly on encrypted data without decryption, enabling privacy-preserving model training and inference [24]. Combined with differential privacy mechanisms, which add calibrated noise to shared model updates, these approaches rigorously protect individual data contributions from leakage while maintaining overall model utility. In Software-Defined Networking (SDN)-enabled 5G and beyond frameworks, these encryption and privacy-preserving algorithms are embedded alongside AI models within SDN controllers, facilitating real-time traffic classification and anomaly detection without compromising data confidentiality [13]. However, designing algorithms that balance encryption overhead, privacy guarantees, and model accuracy remains complex, particularly under stringent latency and computational constraints.

Additionally, the rapid evolution of AI within open, multi-vendor ecosystems accentuates the need for standardized frameworks that embed stringent privacy safeguards while maintaining interoperability across diverse network infrastructures [13, 18]. Such frameworks integrate encryption, secure multiparty computation, and federated learning protocols, enabling seamless and privacy-respecting collaboration among heterogeneous stakeholders in telecommunications systems.

9.3 Explainability and Trust

Establishing trust and transparency in AI-driven telecommunications systems is essential, given that automated decisions directly affect service quality and network reliability [18, 24**?** , 25]. Explainability techniques empower operators and stakeholders to interpret AI decision-making processes, identify erroneous outputs, and align these decisions with domain expertise. For instance, incorporating explainable AI within network management systems elucidates the rationale behind resource allocation or anomaly detection outcomes, thereby fostering confidence and enabling effective human-in-the-loop oversight [**?**]. Empirical studies demonstrate that AI-empowered frameworks can meaningfully improve network performance while maintaining system transparency. For example, the interference management framework in perceptive mobile networks applies AI to dynamically allocate resources based on explainable interference predictions, resulting in a 20% improvement in detection probability and a 30% reduction in sensing interference while preserving communication quality [18]. Such explainable insights assist operators in validating AI decisions against operational constraints, reducing inadvertent network disruptions.

Nonetheless, the predominant use of complex deep learning models often results in opaque, black-box systems, presenting a fundamental trade-off between prediction accuracy and interpretability [25]. Research advances include inherently interpretable models and post hoc explanation methods such as attention visualization, feature attribution, and counterfactual reasoning, which are progressing but remain immature in comprehensive telecom applications [24]. Moreover, explainability is critical for regulatory compliance verification and mitigating risks from erroneous AI decisions that could propagate cascading network failures [18]. In Open RAN and 6G networks, AI integration highlights additional trust challenges due to distributed and multi-agent learning frameworks, necessitating transparent AI mechanisms to ensure security, fault tolerance, and interoperability [25]. Therefore, enhancing AI system transparency remains a vital challenge that underpins systemic trust and responsible deployment in telecommunications.

9.4 Interoperability and Standardization Challenges

The integration of AI into telecommunications networks faces considerable interoperability and standardization challenges arising from diverse multi-vendor equipment, heterogeneous protocol stacks, and disparate technology domains [13, 18, 25]. Fragmented AI models and incompatible data schemas hinder seamless AI-driven control and coordination across Open RAN components, including radio units (RU), distributed units (DU), and centralized units (CU) [18]. The lack of unified AI interfaces and standardized telemetry data formats creates barriers to implementing distributed intelligence and federated learning, constraining scalability and limiting cross-vendor collaboration [13]. Additionally, varying regulatory requirements and privacy policies across jurisdictions and operators compound the fragmentation in AI adoption. Early efforts by standardization bodies to define AI-specific protocols, interfaces, and data representations are ongoing but remain in initial stages, despite their critical role in enabling modular, plug-and-play AI capabilities and ensuring reproducibility and reliability of AI-enhanced network functions [25]. Addressing these interoperability gaps demands interdisciplinary initiatives focused on harmonizing software stacks, unifying data semantics, and developing AI models robust to domain shifts and heterogeneity across multi-technology ecosystems.

9.5 Security and Robustness

As AI technologies permeate critical telecommunications infrastructure, ensuring security and robustness against adversarial threats, data poisoning, and erroneous model outputs is fundamental to operational reliability [5, 18, 24]. AI models are vulnerable to attacks exploiting weaknesses in training data, model parameters, and inference processes, potentially resulting in misclassifications, compromised routing decisions, or degradation of service quality. Such

Table 13: Current Standards Initiatives Addressing AI Interoperability and Standardization in Telecommunications

Standards Body	Initiative/Group	Scope and Focus
O-RAN Alliance	AI/ML Working Group	Defines AI model interfaces, standardized telemetry data formats, and protocols for Open RAN components (RU, DU, CU) to realize distributed intelligence and federated learning [25].
3GPP	SA6 (Enhancements for AI	Integration of AI data models and interfaces in 5G/6G network management for seamless AI control across heterogeneous network slices [13].
ETSI	Experiential Networked Intelligence (ENI)	Framework for context-aware AI management, standardizing AI-driven closed-loop automation workflows and interoperability between network elements [13].
IEEE Standards Association	P1931.1 (Interoperability for AI Systems)	Developing guidelines for AI component interoperability, data exchange formats, and reproducibility standards applicable to telecommunications infrastructures [18].
ITU-T	Focus Group on Machine Learning for Future Networks	Standardizing machine learning workflows, data representation, and privacy frameworks for multi-vendor network environments [25].

attacks are especially harmful in complex AI-enabled networks where flawed decisions may cascade, causing widespread disruptions across multiple layers and services [24]. Defensive strategies include adversarial training, development of robust model architectures, sophisticated anomaly detection systems, and hierarchical control mechanisms equipped with fallback options to mitigate AI failures [5]. Additionally, integrating explainability aids in the early detection of abnormal AI behaviors, while federated learning frameworks can minimize insider threats by limiting data exposure [18].

Given the resource constraints common at the network edge in 5G and beyond, emerging lightweight security solutions are essential. These include model compression techniques such as pruning and quantization to reduce computational overhead while preserving robustness, enabling deployment of secure AI on edge devices [18, 24]. Lightweight anomaly detection models and efficient encryption schemes tailored for edge environments enhance protection without imposing significant latency or energy costs [24]. Moreover, federated learning and edge computing synergize to protect data privacy and reduce centralized vulnerabilities, balancing the privacy-utility trade-offs crucial under regulatory frameworks [18]. These approaches facilitate real-time inference with acceptable security guarantees in constrained settings, addressing a critical challenge for scalable AI-driven network management.

Recent advances demonstrate the promise of agentic AI approaches leveraging large language models (LLMs) embedded within flexible Open Radio Access Network (O-RAN) architectures to enhance resilience. Such LLM-driven agents improve fault detection accuracy and mitigation success by autonomously monitoring network telemetry, interpreting complex faults through natural language understanding, and executing self-healing actions [5]. Experimental evaluations show fault detection accuracy rising from 78% to 95% and mitigation success increasing from 70% to 91%, alongside significant downtime reduction and throughput degradation improvements. However, these benefits must be balanced against challenges including computational overhead, potential erroneous decisions, security risks, and interoperability among multi-vendor environments, motivating hierarchical agent design and continuous validation frameworks [5].

Therefore, designing AI systems with intrinsic robustness, continuous validation processes, and adaptive security measures is imperative for their trustworthy integration into future telecommunication infrastructures. Scalability requires efficient AI architectures combining algorithmic compression, hardware acceleration, and modular design to meet real-time inference demands, while privacy preservation leverages federated learning, edge computing, and lightweight models balancing privacy-utility trade-offs under regulatory constraints [18, 24]. Explainability remains critical for building trust, auditability, and regulatory compliance amidst opaque deep models [24], and interoperability challenges arising from multi-vendor heterogeneity necessitate standardized AI interfaces and data formats supporting scalable collaboration [5]. Security demands robust defenses against adversarial attacks and errors, incorporating fallback mechanisms and continuous validation without sacrificing system performance.

Collectively, these emerging techniques and frameworks underscore that ensuring security and robustness in AI-driven telecommunications requires a multilayered approach combining advanced AI methodologies, network architecture flexibility, and rigorous operational safeguards.

10 Synthesis and Future Directions

The surveyed literature indicates a rapidly evolving landscape at the intersection of artificial intelligence, control theory, and wireless communication technologies. To guide future research and practical implementations, it is essential to synthesize the key methodological advancements and outline potential disruptive paradigm shifts, as well as address challenges encountered when transitioning from theory to real-world applications.

10.1 Methodological Contributions and Frameworks

The integration of AI techniques with control and wireless systems has led to innovative frameworks that enable adaptive, resilient, and efficient operation in complex environments. To advance this field, future research should prioritize the development of standardized methodological frameworks that unify these interdisciplinary approaches. Such frameworks would support consistency in comparison, reproducibility, and scalability of results. Critical features of these frameworks include robustness to system uncertainties, real-time adaptability to dynamic conditions, and data-driven optimization strategies, all while preserving the theoretical guarantees established in control theory and communication domains. Establishing these comprehensive frameworks can bridge gaps between diverse methodologies and accelerate practical deployments in complex AI-enabled control and wireless systems.

10.2 Practical Implementation Pathways and Challenges

While theoretical advancements abound, the practical implementation of AI-enabled control and wireless systems confronts several critical challenges. These challenges include computational resource constraints, stringent latency requirements, maintaining robustness amid dynamic and uncertain environments, and seamless integration with existing legacy infrastructure. To address these issues, future research should prioritize the development of scalable and efficient algorithms that are compatible with edge computing paradigms and distributed architectures, thereby reducing latency

Table 14: Summary of Future Research Challenges and Opportunities in AI-Enabled Wireless Control Systems

Research Challenge	Opportunity	Actionable Research Questions / Milestones
Scalability and Complexity	Development of scalable AI algorithms for large-scale networks	How can AI models be optimized for complexity and energy efficiency in massive deployments? Milestones: Prototype scalable control algorithms; evaluate resource consumption.
Robustness to Real-World Variability	Robust control against environmental and hardware uncertainties	What frameworks ensure robustness of AI control under non-ideal conditions? Milestones: Test algorithms on real testbeds; quantify robustness metrics.
Integration of AI and Control Theory	Synergistic methods combining AI learning with classical control	How to design hybrid frameworks merging formal control guarantees with AI adaptivity? Milestones: Propose hybrid architectures; validate stability and performance guarantees.
Low-Latency and Reliability	AI approaches enabling ultra-reliable low-latency communications (URLLC)	What AI-driven scheduling and resource allocation techniques achieve URLLC requirements? Milestones: Develop scheduling protocols; evaluate delay and reliability trade-offs.
Data Privacy and Security	AI methods preserving privacy and ensuring network security	How to safeguard data privacy and defend against adversarial attacks in AI-enabled wireless control? Milestones: Formulate secure AI models; demonstrate resilience to attacks.
Real-World Deployment	Bridging theory-to-practice gap through prototyping and field trials	What are the practical barriers and solutions for real-world adoption? Milestones: Deploy pilot systems; gather empirical performance data and feedback.

and resource demands. Furthermore, as AI systems are increasingly deployed in sensitive and mission-critical applications, it is essential to rigorously manage privacy, security, and ethical considerations throughout the design and deployment phases to ensure trustworthy and responsible AI operation.

10.3 Potential Disruptive Innovations and Paradigm Shifts

Emerging trends indicate several potential disruptions in the field. One significant innovation is the convergence of AI, control, and wireless communication technologies, which may enable self-organizing and self-optimizing networks capable of dramatically enhancing efficiency and responsiveness. Another paradigm shift is expected from the fusion of model-based and data-driven approaches, producing hybrid methods that exploit the complementary strengths of both paradigms. Additionally, further research into cross-layer design that integrates AI across multiple system levels holds promise for transforming traditional system architectures, enabling more adaptive and intelligent communication systems.

10.4 Comparative Summary of Key Approaches

To facilitate a clear understanding of the relative merits, limitations, and practical considerations of the approaches discussed throughout this survey, Table 15 provides a detailed comparative summary. It highlights the principal characteristics, implementation challenges, and key application domains of representative methods, enabling readers to grasp their contextual suitability and trade-offs.

10.5 Interdisciplinary Synergies

The symbiotic relationship between AI, control theory, and wireless communication is poised to deepen significantly, paving the way for innovative system designs that overcome traditional domain-specific boundaries. To achieve this, researchers should actively promote interdisciplinary collaborations that integrate advances across sensing technologies, computational methodologies, and theoretical frameworks. Such collaborations will facilitate the development of holistic, robust solutions that address complex real-world challenges.

Moving forward, it is essential to consolidate diverse methodological advances into unified frameworks that seamlessly blend AI, control, and wireless technologies. Addressing practical implementation issues—including scalability, latency, and reliability—will be critical to transitioning these innovations from theory to deployment. Additionally, anticipating emerging paradigm shifts driven by hybrid and adaptive approaches will enable more flexible and resilient system architectures. By emphasizing and reinforcing interdisciplinary synergies, the research community can accelerate the realization of autonomous, efficient, and resilient systems that have meaningful impact across a wide range of applications.

10.6 Synergies Across AI, Resilient Control, and Wireless Technologies

The fusion of artificial intelligence (AI), resilient control strategies, and wireless technologies has driven substantial progress toward real-time, adaptive, and secure network management within dynamic and uncertain environments. A salient example of this interdisciplinary synergy is the development of adaptive neural network finite-time control methods designed for nonlinear systems with unknown time delays, actuator faults, and false data injection attacks. These frameworks employ radial basis function neural networks to approximate unknown nonlinearities, utilizing online adaptive laws to estimate network weights and fault parameters in real-time. Coupled with observer-based fault detection mechanisms that estimate states and faults despite measurement discrepancies, this integration ensures fault-tolerant operations and finite-time convergence of state and estimation errors. Such methods significantly enhance system resilience compared to traditional asymptotic control techniques by achieving faster responses under substantial unknown disturbances [3].

In parallel, AI has invigorated wireless communications through advanced paradigms such as Perceptive Mobile Networks (PMNs). These networks leverage coordinated beamforming and deep learning algorithms to predict and mitigate interference in complex multi-cell environments. By dynamically allocating resources based on learned interference patterns, these methods maintain communication quality while substantially improving sensing accuracy under heterogeneous, interference-prone conditions. The AI-empowered framework exploits macro-diversity and array gains through coordinated sensing across multiple base stations, enabling robust performance amid high network loads. This dynamic adaptation facilitates robust wireless resource orchestration, improving situational awareness and network performance despite challenges in low-latency inference and channel acquisition [**?**].

Further extending this ecosystem, AI-driven optimization of reconfigurable intelligent surfaces (RIS) offers a powerful approach to shaping the wireless propagation environment. By learning mappings from channel state information to effective RIS configurations, these AI-empowered systems achieve improved spectral efficiency and robustness in uncertain, time-varying scenarios. This integration exemplifies how AI, control theory, and advanced wireless technologies collectively enable networks capable of context-aware, adaptive decision-making and resilient operation despite inherent cyber-physical uncertainties [18].

10.7 Critical Enablers

A set of pivotal enablers underpins the convergence of AI, control, and wireless technologies. These enablers not only address fundamental technical challenges but also represent active research areas with diverse approaches and trade-offs, as summarized in Table 16.

Table 15: Comparative summary of key AI-based control and wireless methodologies

Approach	Methodological Strengths	Practical Challenges	Application Domains
Model-based control with AI integration	Strong theoretical foundations and interpretability facilitate reliability and safety assurances	High computational complexity and the need for accurate system models may limit adaptability	Autonomous systems, robotics, and safety-critical control
Data-driven machine learning methods	High adaptability and scalability enable handling complex, dynamic environments	Requires large volumes of data and may lack formal performance guarantees	Network optimization, resource management, and traffic prediction
Hybrid model-data fusion frameworks	Combines theoretical rigor with flexibility to improve robustness	Complexity in integration and parameter tuning poses practical hurdles	Smart grids, 5G/6G communication networks, and cyber-physical systems
Reinforcement learning for control	Enables real-time learning and policy optimization in uncertain environments	Balancing exploration and exploitation remains challenging, along with convergence guarantees	Unmanned aerial vehicles (UAVs), Internet of Things (IoT) device management
Edge AI and distributed architectures	Supports low-latency processing and scalable deployment across network edges	Communication overhead and privacy concerns require careful management	Smart cities, industrial automation, and distributed sensing

Table 16: Summary and Analysis of Critical Enablers in AI-Driven Wireless Networks

Enabler	Function and Benefits	Challenges and Controversies	Illustrative Example
Federated Learning	Enables distributed intelligence while preserving privacy, e.g., collaborative spectrum management	Communication overhead, model convergence issues, heterogeneous data distributions	Open RAN dynamic spectrum allocation [6]
Privacy-Preserving AI	Protects sensitive user and network data, maintains user trust	Balancing privacy with model accuracy and efficiency; differential privacy, encryption, and secure computation trade-offs	Secure data exchange in multi-operator networks
Edge Intelligence	Decentralizes computation to reduce latency and optimize resource usage at the network edge	Limited edge resources, coordination across nodes, model consistency, heterogeneity of devices	Real-time adaptive interference mitigation
Explainable AI	Provides transparency and interpretability for black-box models, enhancing trust and regulatory compliance	Complexity of explanations, potential trade-off between interpretability and accuracy [24]	Visualizing reinforcement learning decision paths
Scalable Distributed Architectures	Support multi-agent RL and adaptive control across heterogeneous and large-scale network segments	System complexity, scalability bottlenecks, interoperability issues	Large-scale network fault detection systems

Below, we provide a critical discussion of each enabler with examples from current studies and highlight open research gaps:

Federated Learning. This approach facilitates collaborative model training across decentralized nodes without raw data exchange, addressing stringent privacy concerns inherent in wireless networks. In Open RAN contexts, federated learning enables dynamic spectrum management and fault detection by aggregating local AI insights, as discussed in [6]. Despite its advantages, challenges such as communication overhead, non-iid data distributions, and convergence difficulties persist and necessitate further research to enhance scalability and efficiency.

Privacy-Preserving AI. Closely related to federated learning, privacy-preserving mechanisms—including differential privacy, secure multi-party computation, and homomorphic encryption—protect sensitive user and network information. These techniques must carefully balance privacy guarantees with model accuracy and computational overhead. Current debates focus on discovering optimal approaches tailored to wireless network scenarios, considering varying security requirements and resource constraints.

Edge Intelligence. By decentralizing processing to network edge nodes, edge intelligence reduces latency and bandwidth consumption compared to centralized cloud architectures. Tasks such as real-time interference mitigation and localized adaptive control benefit significantly from edge computing. However, the inherent limitations in computational resources, the heterogeneity of edge devices, and the challenges in coordinating distributed inference and maintaining model consistency across nodes require sophisticated system design and algorithmic solutions.

Explainable AI. Given the widespread deployment of complex neural architectures and reinforcement learning agents in network management [24], explainability enhances transparency, fosters user trust, and supports regulatory compliance. Techniques that interpret AI decisions and visualize agent behavior are advancing, yet balancing interpretability with the preservation of high prediction accuracy remains an open challenge, driving ongoing research in explainable AI models adapted for wireless systems.

Scalable Distributed Architectures. Addressing the scale and complexity of next-generation wireless systems demands distributed frameworks that combine multi-agent reinforcement learning and adaptive control. These architectures afford resilience and fault tolerance across diverse and heterogeneous network segments, but they also introduce challenges in terms of scalability, system complexity, and interoperability among varied network components. Developing scalable, interoperable, and efficient distributed frameworks remains a critical research imperative.

In summary, these enablers collectively advance AI's feasibility in wireless and cyber-physical systems by tackling privacy, interpretability, latency, and scalability challenges. Ongoing research is essential to resolve existing trade-offs between performance, complexity, and trustworthiness, ultimately enabling the realization of fully intelligent and adaptive wireless networks.

10.8 Identified Research Needs

This section outlines the key research gaps that must be addressed to advance AI-enhanced resilient control and wireless systems. The primary objectives are to develop computationally efficient, robust, low-latency, interpretable, and interdisciplinary AI frameworks that achieve resilient and scalable performance in dynamic cyber-physical environments. Addressing these challenges will enable reliable operation, real-time responsiveness, and user trust in next-generation cyber-physical infrastructures.

Computational Complexity Reduction: Current adaptive neural network finite-time resilient control approaches for nonlinear time-delay systems show robustness to unknown actuator faults and cyber-attacks but require significant neural network capacity and careful parameter tuning, and often struggle with noise sensitivity [3]. To overcome these limitations, future research should focus on designing algorithms that significantly reduce computational resource demands while preserving or improving control efficacy. Potential strategies include developing adaptive learning schemes with dynamic network pruning, exploring distributed resilient control architectures for multi-agent systems, and integrating efficient neural network compression techniques. Such solutions aim to enable deployment on resource-limited devices without compromising the resilience or speed of control and inference.

Robustness Against Uncertainties: AI models in networked sensing and control systems remain susceptible to adversarial perturbations, noisy measurements, and imperfect channel state information [18]. Designing resilient AI frameworks that can operate reliably under such uncertain and unmodeled disturbances is critical. Promising directions include advancing cooperative multi-agent learning and sensing approaches that leverage coordinated beamforming and interference prediction to mitigate performance

Table 17: Summary of Key Research Challenges and Potential Solution Paths

Research Need	Challenges	Potential Solution Paths
Computational Complexity Reduction	High neural network capacity, sensitivity to noise, time-varying delays	Adaptive learning schemes, distributed resilient control, neural network compression, hardware-efficient algorithms [3]
Robustness Against Uncertainties	Vulnerability to adversarial perturbations, noisy telemetry, imperfect channel info	Resilient AI algorithms, multi-agent cooperation, real-time adaptation, robust interference mitigation [18]
Latency Minimization	Real-time inference demands, resource constraints, trade-offs between accuracy and speed	Lightweight AI models, hardware accelerators, scalable reinforcement and deep learning, edge-cloud synergy [24]
Interpretability Enhancement	Lack of transparency in AI decisions, regulatory and operational trust issues	Explainable AI frameworks, model introspection techniques, visualization tools, domain-aware explanations
Interdisciplinary Collaboration	Integrating diverse expertise across control, communications, and AI	Holistic frameworks spanning algorithmic, hardware, and network layers; fostering joint research initiatives

degradation. Moreover, incorporating real-time adaptive mechanisms to dynamically respond to environmental changes and adversarial inputs will enhance robustness. These solutions should aim to maintain sensing accuracy and control stability even in highly dynamic and partially observable network conditions.

Latency Minimization: Timely AI inference is essential especially when deployed at the network edge with stringent resource constraints [24]. Achieving a balance between inference speed and accuracy calls for the development of lightweight AI architectures tailored for efficient execution on specialized hardware accelerators. Future work should investigate scalable reinforcement and deep learning methods designed to optimize resource allocation and network performance while maintaining low latency. Integration of edge-cloud computing paradigms to offload computation adaptively, as well as designing energy-efficient AI inference pipelines, will be key to fulfilling real-time operational requirements in practical deployments.

Interpretability Enhancement: Transparent and interpretable AI-driven decision-making fosters user trust and enables compliance with safety-critical and regulatory standards. Research should prioritize the creation of explainable AI frameworks that elucidate model predictions and control logic in understandable terms. Techniques such as model introspection, visualization of decision pathways, and domain-specific explanation generation can facilitate debugging and validation. These capabilities are essential for operational acceptance, especially in complex cyber-physical systems where understanding AI behavior is crucial for fault diagnosis and assurance.

Interdisciplinary Collaboration: The multifaceted challenges in resilient control and wireless systems necessitate collaborative frameworks combining control theory, wireless communications, and AI expertise. Holistic approaches that integrate algorithmic innovations, hardware design, and network-level considerations are vital to address system dynamism, scalability, and security comprehensively. Encouraging interdisciplinary research initiatives and joint development platforms will accelerate progress towards robust and scalable AI-empowered cyber-physical infrastructures.

In summary, these research objectives collectively target the development of resilient, scalable, and interpretable AI solutions that can operate efficiently in complex cyber-physical environments. Addressing the outlined challenges will bridge key gaps and empower future systems to meet stringent operational demands with enhanced reliability and trustworthiness.

10.9 Anticipated Innovations

This section aims to explicitly highlight the key objectives of forthcoming advancements within AI-empowered wireless networks, focusing on transformative innovations that promise enhanced autonomy, security, and efficiency while addressing current implementation challenges. The synthesis here frames these innovations within a conceptual model emphasizing the interplay of technological capabilities, operational constraints, and mitigation pathways.

Multi-Agent Collaborative Learning: Distributed learning and decision-making across network nodes promise improved adaptability and fault tolerance in complex environments. Multi-agent reinforcement learning schemes have demonstrated enhanced resource allocation and anomaly detection in Open RAN settings [6]. To mitigate challenges such as coordination complexity, communication overhead, scalability, and adversarial robustness, future research may explore hierarchical agent architectures and lightweight consensus mechanisms that minimize latency and computational burden. Federated learning frameworks that incorporate privacy-preserving and adaptive convergence protocols are poised to balance the trade-offs between privacy, latency, and real-time responsiveness.

Hardware Acceleration: Specialized AI accelerators embedded in edge devices can substantially reduce inference latency and energy consumption, enabling real-time adaptive control and interference mitigation [24]. Addressing hardware design complexity and integration challenges calls for modular accelerator designs that offer configurable performance-to-energy ratios tailored to resource-constrained environments. Cost-effective fabrication approaches and open accelerator standards could facilitate broader adoption and seamless interoperation with heterogeneous network elements.

Quantum Computing Integration: Quantum technologies indicate potential breakthroughs in optimization speeds and security measures, accelerating complex computations unreachable by classical methods. Despite significant hurdles such as immature hardware, quantum error rates, and the need for dedicated quantum algorithms, ongoing advancements suggest viable mitigation strategies. Hybrid quantum-classical architectures can exploit early quantum processors for specific optimization sub-tasks while maintaining classical control, easing integration. Additionally, development of noise-resilient quantum algorithms and error-correction protocols tailored to network optimization and security is crucial for practical deployment.

Blockchain Security Mechanisms: Blockchain-based solutions can enhance the security of decentralized AI agents by ensuring data integrity and providing transparent audit trails, thereby reinforcing trustworthiness in collaborative learning systems [25]. To overcome blockchain-associated latency overhead, scalability limitations, and increased computational demands, scalable consensus protocols (like proof-of-stake variants) and off-chain processing techniques may be employed. Privacy-preserving blockchain implementations, combined with compliance-aware smart contracts, can

address regulatory and data protection requirements. Careful alignment of blockchain architectures with network performance goals is necessary to minimize impact on latency-sensitive operations.

These innovations collectively envisage fully autonomous intelligent networks characterized by self-healing capacities, context-aware adaptations, and proactive cyber-physical threat mitigation [5]. Achieving this vision relies on managing trade-offs among computational complexity, scalability, security, and interoperability through integrated frameworks. Techniques such as explainable AI improve transparency, hierarchical agent designs enhance multi-agent coordination, and lightweight model development supports edge deployment efficiency. This conceptual synthesis underscores critical enablers and actionable research directions to surmount identified challenges.

The following table summarizes the key innovations, their benefits, and associated challenges, along with prospective mitigation strategies:

In summary, this nuanced synthesis elucidates the multi-dimensional progress and outstanding challenges at the intersection of AI, resilient control, and wireless technologies. By explicitly linking benefits with challenges and forward-looking mitigation approaches, it establishes a comprehensive roadmap for evolving secure, adaptive, and intelligent networked systems capable of effectively addressing future demands and uncertainties.

11 Conclusion

This survey aimed to comprehensively review the integration of artificial intelligence (AI) techniques into telecommunication networks, focusing on their roles in adaptive control, wireless networking, routing, software-defined networking (SDN), Open Radio Access Networks (Open RAN), and autonomous fault management. Our objectives included synthesizing state-of-the-art AI methods, evaluating their benefits and challenges, and identifying future research directions to guide the evolution towards fully autonomous, self-optimizing network systems.

The integration of AI into various telecommunication domains has markedly enhanced functionalities such as adaptive control, wireless networking, routing, SDN, Open RAN, and autonomous fault management. AI-driven adaptive control strategies utilize advanced predictive models to dynamically optimize network resource allocation, thereby improving the network's responsiveness to variable traffic loads and heterogeneous service demands. Within wireless networking and routing, machine learning techniques—especially ensemble methods like gradient boosting—have proven highly effective in capturing complex nonlinear patterns and addressing class imbalance issues endemic to network datasets. As demonstrated in recent work [27], gradient boosting methods such as CatBoost and LightGBM significantly outperform traditional algorithms in telecom customer churn prediction, owing to their ability to handle nonlinearities and class imbalances without external resampling, though at the cost of increased computational requirements. Collectively, these advances underscore AI's critical contribution to enhancing efficiency and resilience in contemporary telecommunication infrastructures.

Looking ahead, the evolution of telecommunication networks is trending towards fully autonomous, self-optimizing systems capable of continuous self-monitoring and dynamic adjustment. Nevertheless, deploying AI solutions at scale introduces significant challenges: scalability concerns stemming from the computational demands of complex models like gradient boosting; security and privacy vulnerabilities related to adversarial attacks within AI-integrated control loops; interoperability difficulties due to heterogeneous, multi-vendor environments; and the imperative for explainability and transparency to foster trust and accountability, particularly with unsupervised learning algorithms. Promising explainable AI frameworks—such as the neuralization approach that reformulates clustering models as neural networks to reveal feature importances—offer valuable pathways to improve interpretability and trustworthiness in autonomous network operations [**?**].

Our unique synthesis involved conceptualizing the interplay between AI methodologies and telecom network layers—control, orchestration, and management—highlighting how AI integration across these strata facilitates unprecedented levels of self-governance and resilience. This holistic perspective emphasizes the importance of interdisciplinary research efforts to address integration complexity while balancing the trade-offs between scalability, security, and explainability.

In summary, the progression towards next-generation telecommunication networks is grounded in sophisticated AI methodologies that are autonomous, efficient, secure, and interpretable. Realizing this vision demands continued advancement of AI-driven frameworks that can operate robustly at scale and with transparency, fulfilling the stringent requirements of future communication infrastructures.

To enhance reader takeaway, we provide a concise summary of key points:

Summary of Key Points:

• This survey reviewed AI's transformative impact on telecommunication functionalities including adaptive control, routing, and fault management, focusing on predictive and ensemble learning models.

• Gradient boosting methods outperform traditional algorithms in customer churn prediction by effectively managing class imbalance and nonlinear relationships [27], albeit with higher computational costs.

• Key deployment challenges include scalability, security/privacy, interoperability, and explainability.

• Explainable AI frameworks such as neuralization of clustering models [**?**] enhance transparency and trust in autonomous network operations.

• The envisioned future involves fully autonomous, self-optimizing networks, necessitating interdisciplinary research to surmount integration obstacles.

This structured and explicitly stated summary consolidates the survey's main contributions and emphasizes critical enablers and research avenues needed to advance AI-driven telecommunications.

References

[1] S. Aboagye, M.-S. Alouini, and L. Dai. 2024. Multi-Band Wireless Communication Networks: Fundamentals, Challenges, and Resource Allocation. *IEEE Wireless Communications* 31, 5 (2024), 86–93. https://ieeexplore.ieee.org/document/

Table 18: Summary of Anticipated Innovations: Benefits, Challenges, and Mitigation Approaches

Innovation	Benefits	Challenges / Limitations	Potential Mitigation Strategies
Multi-Agent Collaborative Learning	Enhanced adaptability, fault tolerance, and resource optimization [6]	Coordination complexity, scalability, communication overhead, adversarial robustness	Hierarchical agents, lightweight consensus, federated privacy-preserving protocols
Hardware Acceleration	Reduced latency and energy consumption enabling real-time control [24]	Hardware complexity, integration challenges, cost, balancing energy and compute	Modular configurable designs, open standards, cost-effective fabrication methods
Quantum Computing Integration	Potential breakthroughs in optimization speed and security	Immature hardware, error rates, need for specialized quantum algorithms	Hybrid quantum-classical systems, noise-resilient algorithms, quantum error correction
Blockchain Security Mechanisms	Data integrity, auditability, trust enhancement in decentralized AI [25]	Latency overhead, scalability, computational demand, privacy and regulatory concerns	Scalable consensus, off-chain processing, privacy-preserving protocols, compliance-aware contracts

10438479/

[2] A. Ahmed, T. M. Nguyen, and M. Elsayed. 2023. Deep Learning for Telecom Self-Optimized Networks. *IEEE Transactions on Communications* 71, 4 (2023), 2001–2014. https://ieeexplore.ieee.org/document/10811884

[3] Anonymous. 2025. Deep Learning in Wireless Communication Receiver: A Survey. arXiv preprint arXiv:2501.17184. https://arxiv.org/abs/2501.17184 Accessed: 2024-06-01.

[4] M. W. Baidas. 2016. A Distributed Political Coalition Formation Framework for Multi-Relay Selection in Wireless Networks. *Wireless Communications and Mobile Computing* 16, 4 (2016), 2065–2082. doi:10.1002/wcm.2763

[5] Dimitris Bertsimas. 2023. Global optimization via optimal decision trees. *Journal of Global Optimization* 85, 1 (2023), 1–28. doi:10.1007/s10898-023-01311-x

[6] T. Chen, M. Hong, and Z. Su. 2018. Learn-and-Adapt Stochastic Dual Gradients for Network Optimization. *IEEE Transactions on Control of Network Systems* 5, 4 (2018), 1456–1467. https://ieeexplore.ieee.org/document/8110688

[7] Z. Chen, M. Zhao, and X. Wang. 2024. Robust Federated Learning for Unreliable and Resource-Constrained Wireless Networks. *IEEE Transactions on Wireless Communications* 23, 8 (2024), 9793–9809. https://ieeexplore.ieee.org/document/10444714/

[8] L. Dai, R. Jiao, F. Adachi, H. V. Poor, and L. Hanzo. [n. d.]. Deep Learning for Wireless Communications: An Emerging Interdisciplinary Paradigm. Online. https://arxiv.org/abs/2007.05952 Submitted Jul. 2020.

[9] X. Ding, Y. Jin, and J. Liu. 2023. Obstacle-Aware Fuzzy Clustering Protocol for Wireless Sensor Networks in 3D Terrain. *International Journal of Wireless Information Networks* 30, 1 (2023), 30–41. doi:10.1007/s10776-022-00595-8

[10] T. Febrianto, J. Hou, and M. Shikh-Bahaei. 2017. Cooperative Full-Duplex Physical and MAC Layer Design in Asynchronous Cognitive Networks. *Wireless Communications and Mobile Computing* 2017 (2017), 1–14. doi:10.1155/2017/8491920

[11] W. S. Fujo, I. J. Al-Mousa, and S. A. Hamed. 2024. Customer Churn Prediction in Telecommunication Industry Using Deep Learning. *Preprints.org* 2024, 0115 (2024). https://www.preprints.org/manuscript/202403.0585/v1

[12] A. Förster, F. Macabiau, and D. Grouset. 2024. A beginner's guide to infrastructureless networking concepts. *IET Networks* 13, 1 (2024), 14–22. doi:10.1049/ntw2.12094

[13] E. Hanasusanto, D. Kuhn, and K. N. Kallas. 2016. Multistage Robust Mixed-Integer Optimization with Adaptive Partitions. *Operations Research* 64, 4 (2016), 980–998. doi:10.1287/opre.2016.1515

[14] M. Imani. 2024. Comparing Traditional Machine Learning and Advanced Gradient Boosting Techniques in Customer Churn Prediction: A Telecom Industry Case Study. *Preprints.org* 2024, 0213 (2024). https://www.preprints.org/manuscript/202403.0213/v2

[15] K. D. Irianto and R. Chandra. 2020. Partial packet in wireless networks: a review of error recovery and loss mitigation techniques. *IET Communications* 14, 15 (2020), 2396–2409. doi:10.1049/iet-com.2019.0550

[16] D. Kuhn, P. Wiesemann, and T. Georghiou. 2019. Wasserstein Distributionally Robust Optimization: Theory and Applications in Machine Learning. *Operations Research* 67, 3 (2019), 814–831. doi:10.1287/opre.2018.1804

[17] Y. H. Kwon, K. J. Han, and Y. S. Choi. 2015. Efficient network mobility support scheme for proxy mobile IPv6. *EURASIP Journal on Wireless Communications and Networking* 2015, 1 (2015), 1–14. doi:10.1186/s13638-015-0437-8

[18] M. Li, Y. Hong, and B. Chen. 2021. A Unified Analytical Framework for Optimal Control Problems in Network Systems. *IEEE Transactions on Control of Network Systems* 8, 4 (2021), 1645–1656. https://ieeexplore.ieee.org/document/9454297

[19] Y. Li, Z. Zhang, L. Wu, and X. Wang. 2022. Real-World Wireless Network Modeling and Optimization: Recent Advances and Challenges. *Chinese Journal of Electronics* 31, 2 (2022), 263–280. https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/cje.2022.00.191

[20] Y. Liu, X. Liang, and P. Zhang. 2020. Data-Importance Aware Radio Resource Allocation. *IEEE Communications Letters* 24, 9 (2020), 2046–2050. https://ieeexplore.ieee.org/document/9098940

[21] R. F. Lopes. 2013. Performance of the modulation diversity technique for - fading channels in wireless communications. *EURASIP Journal on Wireless Communications and Networking* 2013, 1 (2013), 1–12. doi:10.1186/1687-1499-2013-17

[22] Y. Luo, C. Yang, and S. Yu. 2023. Recent Advances in Optical Wireless Communications for 6G Wireless Networks. *IEEE Wireless Communications* 30, 2 (2023), 58–65. https://ieeexplore.ieee.org/document/10325445/

[23] G. A. Mapunda, R. Ramogomana, L. Marata, B. Basutli, A. S. Khan, and J. M. Chuma. 2020. Indoor Visible Light Communication: A Tutorial and Survey. *Wireless Communications and Mobile Computing* 2020 (2020), 46. doi:10.1155/2020/8881305

[24] S. Nadarajah and A. A. Ciré. 2020. Network-Based Approximate Linear Programming for Discrete Optimization. *Operations Research* 68, 6 (2020), 1767–1786. doi:10.1287/opre.2019.1953

[25] A. Nagurney. 2022. Supply chain networks, wages, and labor productivity: insights from Lagrange analysis and computations. *Journal of Global Optimization* 83, 3 (2022), 615–638. doi:10.1007/s10898-021-01084-x

[26] F. Nisar and B. A. Rehman. 2025. An efficient security framework, vulnerabilities, and defense mechanisms in LoraWAN. *Computer and Telecommunication Engineering* 3, 2 (2025), Article ID 3072. https://aber.apacsci.com/index.php/CTE/article/view/3072

[27] D. Niyato. 2023. Editorial: Fourth Quarter 2023 IEEE Communications Surveys and Tutorials. *IEEE Communications Surveys & Tutorials* 25, 4 (2023), 3456–3463. https://ieeexplore.ieee.org/document/10325334/

[28] Dusit Niyato and et al. 2021. Survey on Wireless Communications. *IEEE Communications Surveys & Tutorials* 23, 1 (2021), 1–40. https://ieeexplore.ieee.org/document/9621329/

[29] S. Pawar, L. Bommisetty, and T. G. Venkatesh. 2022. A High Capacity Preamble Sequence for Random Access in 5G IoT Networks: Design and Analysis. *International Journal of Wireless Information Networks* 30, 1 (2022), 1–15. doi:10.1007/s10776-022-00593-x

[30] Y. Qian, H. Chen, and M. Dohler. 2022. Beyond 5G Wireless Communication Technologies. *IEEE Wireless Communications* 29, 1 (2022), 166–172. https://ieeexplore.ieee.org/document/9749229/

[31] E. Shaaban. 2023. Hyperparameter Optimization and Combined Data Certainty for Customer Churn Prediction in Telecommunication Industry. *Preprints.org* 2023, 1478 (2023). https://www.preprints.org/manuscript/202308.1478/v3

[32] X. Shen, Y. Liu, X. Du, and K. K. R. Choo. 2020. AI-assisted Network-slicing based Next-generation Wireless Networks. *IEEE Transactions on Wireless Communications* 19, 3 (2020), 1558–1571. https://ieeexplore.ieee.org/iel7/8782711/8889399/08954683.pdf

[33] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis. 2016. 5G-Enabled Tactile Internet. *IEEE Journal on Selected Areas in Communications* 34, 3 (2016), 460–473. https://ieeexplore.ieee.org/document/7403840/

[34] S. Thapaliya and P. K. Sharma. 2022. Cyber Forensic Investigation in IoT Using Deep Learning Based Feature Fusion in Big Data. *International Journal of Wireless Information Networks* 30, 1 (2022), 16–29. doi:10.1007/s10776-022-00588-7

[35] D. Wen, B. Zhang, and Y. Chen. 2020. Joint Parameter-and-Bandwidth Allocation for Improving Federated Learning Performance in Wireless Networks. *IEEE Transactions on Wireless Communications* 19, 10 (2020), 6780–6793. https://ieeexplore.ieee.org/document/9194337/

[36] Z. Weng, L. Lu, J. Chen, H. Zhang, and L. Hanzo. 2023. Deep Learning Enabled Semantic Communications With Knowledge Graph and Knowledge Base. *IEEE Journal on Selected Areas in Communications* 41, 9 (2023), 2192–2207. https://ieeexplore.ieee.org/document/10038754

[37] Z. Zhao, E. J. Schiller, E. Kalogeiton, T. Braun, S. Burkhard, and M. T. Garip. 2017. Autonomic Communications in Software-Driven Networks. *IEEE Journal on Selected Areas in Communications* 35, 11 (2017), 2431–2445. https://ieeexplore.ieee.org/document/8063402/

[38] H. Zhou, W. Saad, and D. Niyato. 2024. Large Language Model (LLM) for Telecommunications: A Comprehensive Survey on Principles, Key Techniques, and Opportunities. *IEEE Communications Surveys & Tutorials* 26, 2 (2024), 879–913. https://ieeexplore.ieee.org/document/10685369/

[39] D. D. Čvokić, Y. A. Kochetov, and A. Savić. 2022. A variable neighborhood search algorithm for the (r|p) hub–centroid problem under the price war. *Journal of Global Optimization* 83, 3 (2022), 405–444. doi:10.1007/s10898-021-01051-2