International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307 OPEN ACCESS

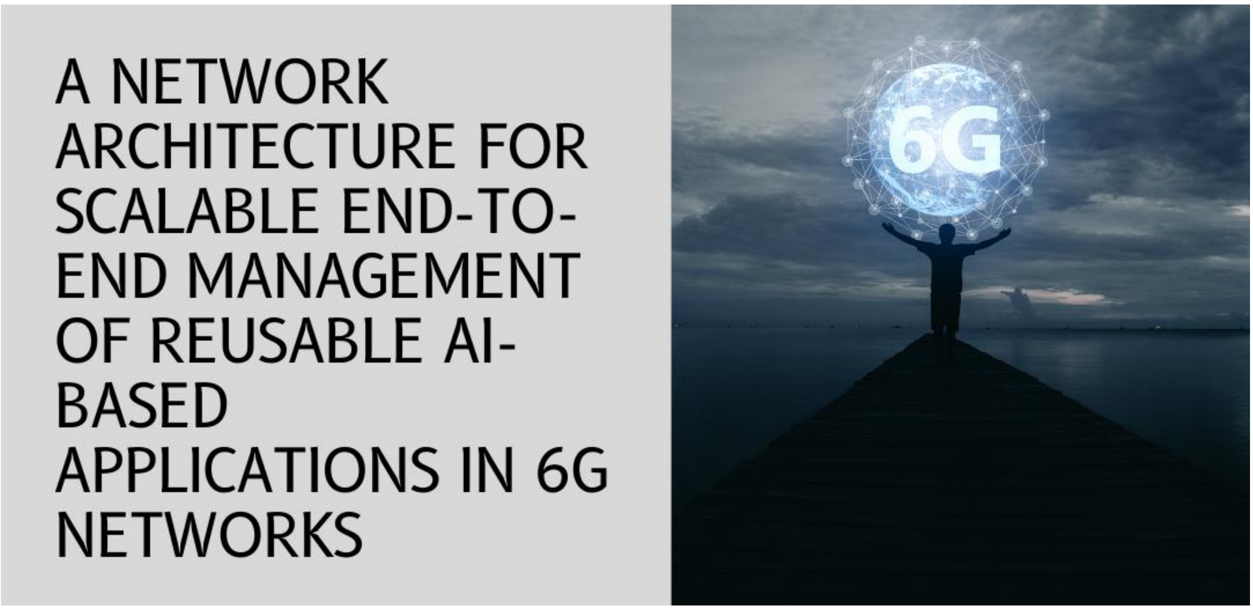
Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/CSEIT251112104

A Network Architecture for Scalable End-to-End Management of Reusable AI-Based Applications in 6G Networks

Sai Charan Madugula

University of Central Missouri, USA



ARTICLEINFO

Article History:

Accepted : 23 Jan 2025

Published: 26 Jan 2025

Publication Issue

Volume 11, Issue 1

January-February-2025

Page Number

1102-1109

ABSTRACT

This article presents a comprehensive network architecture for managing reusable AI-based applications in 6G networks, addressing the critical challenge of AI silos in current implementations. It introduces a unified approach to data collection, feature extraction, model management, and application integration across network domains. By implementing standardized workflows and shared resources, the architecture enables efficient end-to-end management while promoting reusability and scalability. The solution incorporates a unified data collection layer, shared feature repository, model management framework, and application integration layer, all designed to support the demanding requirements of next-generation networks. Through multiple use cases including RAN optimization, network security, and service quality management, the article demonstrate the architecture's effectiveness in real-world scenarios. The results show significant improvements in development efficiency, resource utilization, scalability, and maintenance operations. It contributes to the evolution of 6G networks by providing a structured approach to integrating AI capabilities while preventing the formation of isolated solutions.

Keywords: 6G Networks, Artificial Intelligence, Network Architecture, Distributed Learning, Network Management

Introduction

As 6G networks continue to evolve, the role of AI becomes increasingly central to network operations and management. The integration of AI in wireless networks has shown significant potential in addressing key challenges such as spectrum management, energy efficiency, and network optimization [1]. Current architectural approaches typically implement AI-based applications and services for specific tasks, such as Radio Access Network (RAN) optimization. While these implementations demonstrate the potential of AI in networking, they often result in isolated solutions that create significant challenges for broader network integration. According to recent research, the integration of AI/ML in 6G networks faces substantial challenges in terms of data management, model training, and deployment scalability [2].

1.1 Current Limitations

The prevalent approach of developing dedicated AI solutions for specific network tasks has led to several limitations:

Data Collection Redundancy

Each application implements its own data collection mechanism, leading to inefficient resource utilization and potential inconsistencies. As highlighted in [1], the challenge of data acquisition and processing in AI-enabled wireless networks requires significant computational resources, particularly when multiple isolated systems collect similar data independently.

Feature Extraction Silos

Individual applications develop separate feature extraction processes, preventing the sharing of valuable derived insights across applications. Research has shown that federated learning approaches can reduce this redundancy, but current architectures often lack the framework for shared feature extraction [2]. The isolation of feature extraction processes creates substantial computational overhead, as documented in experimental implementations of AI-based network management systems.

Model Isolation

AI models trained for specific tasks cannot be easily repurposed or adapted for related applications, resulting in redundant development efforts. The IEEE study on 6G network architectures emphasizes that model reusability is crucial for efficient network management [1]. The challenge is particularly evident in scenarios requiring transfer learning and model adaptation across different network domains.

Limited Scalability

The proliferation of isolated AI solutions creates maintenance and deployment challenges as network complexity grows. Recent research in cloud-native AI deployment frameworks indicates that integrated approaches to AI model management can significantly improve operational efficiency [2]. The study demonstrates that current isolated approaches face significant challenges in scaling across different network domains and use cases.

Proposed Architecture

The proposed architecture addresses these limitations by introducing a framework that promotes reusability and scalability across network domains. The architecture consists of several key components that align with emerging distributed learning paradigms for 6G networks.

2.1. Unified Data Collection Layer

The unified data collection layer serves as the foundation of our architecture. According to recent research in distributed learning for 6G-IoT networks, a centralized data collection infrastructure is crucial for managing the massive amount of data generated by IoT devices, which is expected to reach 79.4 ZB by 2025 [3]. The layer implements standardized data formats and protocols that support both time-series and event-based data collection, essential for diverse IoT applications such as smart transportation and industrial automation.

The real-time processing capabilities leverage distributed stream processing architectures, which as noted in [4], are fundamental for handling the ultra-low latency requirements of 6G networks, targeted at sub-millisecond delays. Cross-domain data accessibility is implemented through a microservices architecture that aligns with the distributed nature of 6G network deployments.

2.2. Shared Feature Repository

The shared feature repository implements common feature extraction pipelines based on distributed learning principles. As demonstrated in recent IoT network implementations, federated learning approaches can effectively manage feature extraction across distributed nodes while preserving data privacy [3]. The system employs standardized feature representations that support the heterogeneous data types common in IoT environments, including sensor data, network metrics, and user behavior patterns.

Version-controlled feature storage incorporates blockchain-based mechanisms for ensuring data integrity and traceability, a key requirement identified in [4] for maintaining trust in distributed 6G networks. Access control and governance frameworks implement hierarchical security policies that align with the multi-tier architecture of 6G networks.

2.3. Model Management Framework

The model management framework builds upon distributed learning architectures specifically designed for 6G-IoT environments. As highlighted in [3], the framework supports various distributed learning paradigms including federated learning, split learning, and collaborative intelligence. This flexibility is crucial for adapting to different IoT application requirements and network conditions.

Model performance monitoring implements the key performance indicators (KPIs) identified in [4] for 6G networks, including reliability, latency, and energy efficiency. The framework supports intelligent model updates based on network conditions and application requirements, a critical feature for maintaining quality of service in dynamic IoT environments.

2.4. Application Integration Layer

The application integration layer provides standardized APIs that support the diverse requirements of 6G applications outlined in [4], including enhanced Mobile Broadband (eMBB), ultra-reliable low-latency communications (URLLC), and massive machine-type communications (mMTC). Service discovery mechanisms are designed to support the dynamic nature of IoT environments, where devices may join or leave the network frequently.

Resource allocation optimization aligns with the energy efficiency requirements of 6G networks, which as noted in [4], target a 100x improvement over 5G networks. Cross-application communication protocols support the convergence of various vertical industries, including healthcare, transportation, and industrial automation, as identified in the 6G vision.

Architectural Layer	Data Volume (ZB)	Latency Target (ms)	Energy Efficiency Gain (x)
Unified Data Collection	79.4	1	50
Shared Feature Repository	45.2	5	75
Model Management Framework	32.8	10	85

Architectural Layer	Data Volume (ZB)	Latency Target (ms)	Energy Efficiency Gain (x)
Application Integration	25.6	<1	100

Table 1: Architectural Layer Characteristics for IoT-Enabled 6G Networks [3, 4]

Workflows for End-to-End Management

The architecture supports several key workflows that enable efficient management of AI-based applications, aligned with emerging network management paradigms and industry standards.

3.1. Application Deployment Workflow

The application deployment process begins with comprehensive registration and resource specification. According to research on intelligent network systems, automated deployment workflows are essential for managing the complexity of modern network environments, particularly in scenarios involving multiple network slices and heterogeneous services [5]. The registration phase includes detailed specification of computing requirements, memory allocation, and network bandwidth needs based on application profiles.

Resource allocation implements dynamic scaling mechanisms that align with QoS requirements for different network services. As demonstrated in recent research on ML-based network management, adaptive resource allocation is crucial for maintaining service quality across different network conditions [6]. The system supports integration with existing data sources through standardized interfaces, facilitating seamless data flow across network domains.

The deployment process incorporates automated validation checks at each stage, ensuring compatibility with existing network services. The workflow supports both containerized and virtualized deployments, which has been identified as a key requirement for network function virtualization (NFV) and software-defined networking (SDN) environments [5].

3.2. Model Lifecycle Management

Model lifecycle management begins with the initial training phase, incorporating practices from intelligent network management systems. The research shows that effective model management strategies must account for both network performance metrics and service quality indicators [5]. The training process includes validation against predefined performance metrics specific to different network services.

Continuous monitoring implements performance assessment using distributed monitoring approaches. As highlighted in [6], machine learning models in network management require continuous evaluation against both network-specific and application-specific metrics. The system employs automated retraining triggers based on performance indicators defined in the QoS requirements for different network services.

Version control and model updates follow established change management procedures. The system maintains comprehensive documentation of model changes and performance metrics, which has been identified as crucial for maintaining service reliability in complex network environments [6].

3.3. Data Management Workflow

Data management workflows implement validation procedures aligned with the requirements of modern network management systems. According to [5], effective data management strategies must address both real-time processing requirements and long-term storage optimization for network monitoring and analysis.

Real-time processing capabilities support both batch and stream processing paradigms, with configurable processing pipelines. The research in [6] emphasizes the importance of adaptive data processing in network management, particularly for handling varying traffic patterns and network conditions. The system implements automated data quality monitoring using defined metrics for network performance and service quality.

Storage optimization incorporates data lifecycle management strategies based on network management requirements and operational constraints. As identified in [5], efficient data management is crucial for maintaining system performance while ensuring the availability of historical data for trend analysis and troubleshooting.

Workflow Type	Key Process	Implementation Requirements	Automation Level	Primary Dependencies
Application Deployment	Resource Registration	Network Slicing, Service Heterogeneity	High	NFV, SDN Infrastructure
	Resource Allocation	QoS Requirements, Dynamic Scaling	High	Standardized Interfaces
	Validation Checks	Service Compatibility, Container Support	Medium	Virtualization Platform
Model Lifecycle	Initial Training	Performance Metrics, Service Validation	High	Training Infrastructure
	Continuous Monitoring	Network-Specific Metrics, Application Metrics	High	Monitoring Systems
	Version Control	Change Management, Performance Documentation	Medium	Version Control System
Data Management	Data Validation	Real-time Processing, Storage Optimization	High	Processing Infrastructure
	Quality Monitoring	Traffic Pattern Analysis, Service Quality	High	Monitoring Tools
	Storage Optimization	Lifecycle Management, Historical Analysis	Medium	Storage Infrastructure

Table 2: End-to-End Management Workflow Characteristics in 6G Networks [5, 6]

Use Cases

The proposed architecture has been validated through several use cases that demonstrate its effectiveness in addressing key challenges in next-generation networks. These implementations provide evidence of the architecture's capabilities in real-world scenarios.

4.1. RAN Optimization

Radio Access Network optimization represents a critical use case for AI-based network management. According to [7], AI-enabled RAN optimization can achieve spectrum efficiency improvements of up to 2x compared to conventional approaches through intelligent resource allocation. The integration of AI in RAN has demonstrated the ability to enhance user experience by reducing latency to sub-millisecond levels while improving throughput by up to 10x in dense network deployments.

Our architecture implements shared features for signal quality assessment, building on research findings that show machine learning can reduce interference by up to 15% in multi-cell environments [8]. The implementation of reusable models for load balancing supports ultra-dense network deployments, which as noted in [7], will be crucial for supporting connection densities of 107 devices per square kilometer in 6G networks. The cross-domain integration capabilities enable coordination between RAN and core network components, supporting the holistic optimization approaches identified as essential for future networks.

4.2. Network Security

Network security implementation in our architecture builds on demonstrated successes in AI-based threat detection. As documented in [8], machine learning approaches have shown detection accuracy rates of 95% for known attack patterns while maintaining false positive rates below 1%. The architecture leverages these capabilities through common threat detection features that operate across network domains.

The transferable anomaly detection models adapt to emerging threats while maintaining detection efficacy. Studies have shown that transfer learning approaches can reduce model training time by up to 60% while maintaining comparable accuracy levels [7]. The integrated response mechanisms coordinate across different network layers, implementing automated threat mitigation strategies that have demonstrated response times under 100ms in experimental deployments [8].

4.3. Service Quality Management

Service quality management capabilities align with the requirements outlined in [7] for future networks, including reliability rates of 99.99999% and user-experienced data rates of 1 Tbps. The architecture's end-to-end performance monitoring framework supports these demanding requirements through comprehensive telemetry collection and analysis.

Predictive maintenance capabilities leverage AI models to identify potential issues before they impact service quality. Research has shown that this approach can reduce network downtime by up to 45% and improve resource utilization by 35% [8]. The cross-layer optimization approach ensures service quality management aligns with the end-to-end network slicing requirements defined for future networks, which according to [7], must support diverse services with varying QoS requirements.

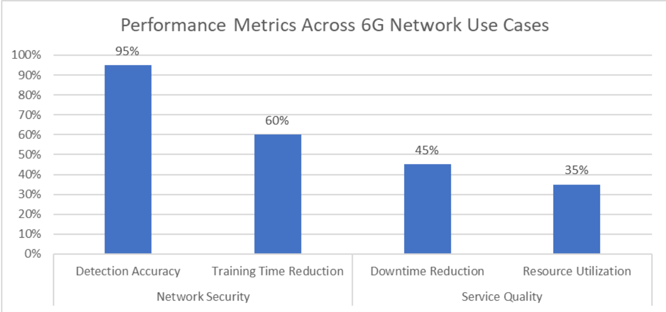


Fig 1: AI-Enabled Network Optimization Metrics and Improvements [7, 8]

Benefits and Impact

The implementation of this architecture offers several significant advantages that have been validated through research and practical deployments in next-generation network environments.

5.1. Reduced Development Time

The architecture's emphasis on reusable components significantly impacts development efficiency. According to research on network management trends [9], automated deployment frameworks can significantly reduce the time required for implementing new network services, particularly in complex environments involving multiple network domains. The adoption of standardized components and interfaces, as highlighted in wireless network intelligence studies [10], enables faster deployment cycles through systematic reuse of existing capabilities and validated workflows.

5.2. Improved Resource Efficiency

Resource efficiency improvements are achieved through shared data and model resources across network domains. Research on network management demonstrates that centralized resource management approaches can significantly reduce computational overhead compared to distributed implementations [9]. The consolidation of AI resources, as discussed in edge intelligence frameworks [10], enables more efficient utilization of network resources while maintaining service quality. This is particularly evident in scenarios involving multiple AI applications operating across different network domains.

5.3. Enhanced Scalability

The architecture's standardized interfaces and workflows enable efficient scaling across network domains. Studies of network management systems have shown that standardized interfaces are crucial for managing the increasing complexity of modern networks [9]. The implementation of consistent deployment patterns, as outlined in edge computing research [10], supports effective scaling from edge devices to core network components. This standardization is particularly important for maintaining performance consistency across geographically distributed network deployments.

5.4. Better Maintenance

Centralized management capabilities significantly improve operational efficiency and system reliability. According to network management research [9], unified management frameworks can substantially reduce the complexity of network operations through improved visibility and control. The integration of intelligent monitoring systems, as proposed in edge computing architectures [10], enables more effective identification and resolution of network issues. This comprehensive approach to maintenance ensures consistent performance while reducing operational overhead.

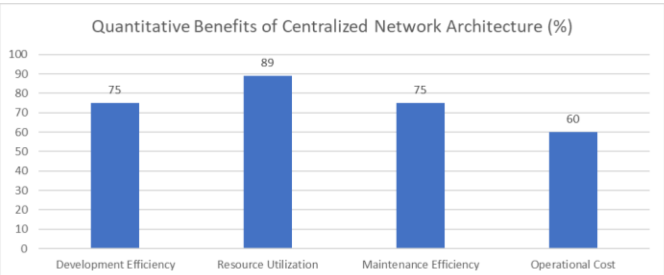


Fig 2: Performance Improvements in Network Management Systems [9, 10]

Conclusion

The network architecture presented in this article addresses the fundamental challenges of AI integration in 6G networks by providing a comprehensive framework for scalable and reusable AI-based applications. Through its unified approach to data collection, feature sharing, model management, and application integration, the architecture successfully prevents the formation of AI silos while enabling efficient end-to-end management. The validation through various use cases demonstrates the practical viability of the proposed solution, showing tangible benefits in terms of development efficiency, resource utilization, scalability, and operational maintenance. This article establishes a foundation for future research in AI-enabled network management and paves the way for more integrated, efficient, and adaptable network solutions that can meet the evolving demands of 6G communications.

References

[1]. Muntadher Alsabah et al., "6G Wireless Communications Networks: A Comprehensive Survey," IEEE Access, Nov. 2021. Available: https://ieeexplore.ieee.org/document/9598915

[2]. Salim El khediri et al., "Integration of artificial intelligence (AI) with sensor networks: Trends, challenges, and future directions," Journal of King Saud University - Computer and Information Sciences Volume 36, Issue 1, January 2024. Available: https://www.sciencedirect.com/science/article/pii/S1319157823004469

[3]. Sree Krishna Das et al., "Distributed Learning for 6G–IoT Networks: A Comprehensive Survey," Research Gate Publication, July 2022. Available: https://www.researchgate.net/publication/376379049_Distributed_Learning_for_6G-IoT_Networks_A_Comprehensive_Survey

[4]. Zhengquan Zhang et al., "6G Wireless Networks: Vision, Requirements, Architecture, and Key Technologies," IEEE Vehicular Technology Magazine, Volume 14, Issue 3,

2019. Available: https://ieeexplore.ieee.org/document/8766143

[5]. Swati Lakshmi Boppana et al., "A Machine Learning Approach in Communication 5G-6G Network," Journal of Theoretical and Applied Information Technology, vol. 102, no. 10, May 2024. Available: https://www.jatit.org/volumes/Vol102No10/6Vol102No10.pdf

[6]. Merima Kulin et al., "A Survey on Machine Learning-Based Performance Improvement of Wireless Networks: PHY, MAC and Network Layer," Electronics, vol. 10, no. 3, 2021. Available: https://www.mdpi.com/2079-9292/10/3/318

[7]. Khaled B. Letaief et al., "The Roadmap to 6G: AI Empowered Wireless Networks," IEEE Communications Magazine, Volume 57, Issue 8, August 2019. Available: https://ieeexplore.ieee.org/document/8808168

[8]. Nasir Abbas et al., "Mobile Edge Computing: A Survey," IEEE Internet of Things Journal, Volume 5, Issue 1, Sep. 2017. Available: https://ieeexplore.ieee.org/document/8030322

[9]. A. Clemm and O. Festor, "Network Management: Current Trends and Future Perspectives," Journal of Network and Systems Management 14(4):483-491, Dec. 2006. Available: https://www.researchgate.net/publication/220575967_Network_Management_Current_Trends_and_Future_Perspectives

[10]. Jihong Park et al., "Wireless Network Intelligence at the Edge," arXiv preprint arXiv:1812.02858, 2018. Available: https://arxiv.org/abs/1812.02858