International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 12-No.3 (2026) pp. 254-261
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

Autonomous Radio Network Intelligence: Architectural Principles of Large Language Model-Driven Closed-Loop 5G Management Systems

Venkata Raghavendra[1]*, Vaishnav Vinjamuri[2]

[1]Verizon, USA
* **Corresponding Author Email:** venkat2a@gmail.com- **ORCID:** 0000-0002-5007-7220

[2]Verizon, USA
Email: vaishna2y@gmail.com- **ORCID:** 0000-0002-5007-1150

Article Info:

DOI: 10.22399/ijcesen.5491
Received : 05 May 2026
Revised : 15 July 2026
Accepted : 20 July 2026

Keywords

Autonomous Network Intelligence,
Large Language Models,
Spectral Efficiency,
Closed-Loop Systems,
5G Radio Access Network Management

Abstract:

Large language models (LLMs) are emerging as cognitive engines within telecommunications infrastructure, introducing a new paradigm for radio access network (RAN) design and management. Traditional methodologies for spectral efficiency and diagnostics have relied on manual workflows that are increasingly insufficient for the complexity of fifth-generation (5G) and evolving sixth-generation (6G) systems. This article presents architectural principles for autonomous network intelligence, focusing on closed-loop orchestration through Reasoning and Acting (ReAct) frameworks, deterministic tool invocation, and stateful memory management. Core concepts, like thread-aware state machines, pipeline toolchains, and geospatial infrastructure for data, facilitate the performance of diagnostics by intelligent agents and the execution of physical layer optimizations at expert levels. In the context of the enterprise architecture, factors are brought into consideration, particularly how a flexible design that supports LLMs' reasoning capabilities can be balanced with deterministic safety requirements. The resulting architecture provides a scalable and implementation-oriented foundation for autonomous closed-loop management in next-generation telecommunications networks. The proposed architecture further incorporates policy-driven automation layers, KPI-aware monitoring pipelines, and telemetry-assisted orchestration aligned with Open RAN and near-real-time radio intelligent controller (Near-RT RIC) environments. Network indicators, including Signal-to-Interference-plus-Noise Ratio (SINR), Block Error Rate (BLER), Channel Quality Indicator (CQI), and Physical Resource Block (PRB) utilization, are continuously correlated to support anomaly detection and adaptive remediation procedures. Deterministic authorization controls, rollback safeguards, and distributed telemetry processing are integrated to maintain operational reliability during autonomous optimization procedures.

1. Introduction

The unprecedented densification of fifth-generation Radio Access Networks (RANs) has exposed fundamental scalability limitations in traditional network management paradigms. As mobile operators deploy millions of macro sites, small cells, and distributed antenna systems to meet surging data demand, the cognitive bandwidth required to monitor, diagnose, and optimize network performance has exceeded human capacity. Signal-to-Interference-plus-Noise Ratio (SINR) degradation, Block Error Rate (BLER) spikes, and spectral efficiency drops, which were once manageable through manual RF engineering workflows, now emerge at frequencies and scales that demand autonomous, intelligent response systems capable of operating at machine speed. A single national 5G operator may manage hundreds of thousands of sector-band groups simultaneously, each producing continuous performance telemetry that must be evaluated against multi-dimensional quality thresholds in near real time.

The architectural challenge inherent in this transformation is not merely computational but fundamentally cognitive. RF diagnostics require contextual reasoning across multiple data dimensions simultaneously. including sector-level performance metrics, carrier band configurations, historical utilization patterns, interference topology, and real-time traffic distributions. Large language models (LLMs), with their demonstrated capacity for multi-step reasoning and contextual inference, offer a novel solution simultaneously. encoding this diagnostic intelligence within automated systems. However, deploying LLMs in mission-critical telecommunications environments introduces significant engineering challenges, particularly around hallucination mitigation, deterministic decision-making, and stateful context management across continuous monitoring sessions.

The theoretical foundations for agentic AI systems capable of addressing these challenges have been advanced by recent research in language model orchestration. Yao et al. introduced the Reasoning and Acting (ReAct) framework as a methodology for enabling LLMs to interleave natural language reasoning with structured external tool invocations [1]. Their evaluations demonstrated that ReAct agents outperformed chain-of-thought approaches on complex multi-step tasks, establishing the viability of tool-augmented LLMs for domain-specific problem solving. The implications for telecommunications management are direct: a language model capable of reasoning about network conditions while simultaneously querying metric databases, executing diagnostic tools, and initiating reconfiguration commands represents a qualitatively new operational capability.

The International Telecommunication Union (ITU) has formalized the progression toward autonomous network management through its framework for autonomous networks, which defines five levels of autonomy ranging from fully manual operations at Level Zero to fully self-evolving systems at Level Five [2]. The architectural principles articulated in this article target Levels Three and Four autonomy, wherein AI systems can autonomously diagnose and remediate a broad class of anomalies while maintaining human oversight for high-impact decisions. Achieving this level requires not only deploying LLMs but also engineering the surrounding architecture, including tool-calling frameworks, memory systems, safety constraints, and physical network integration.

The economic necessity for automatic network management cannot be understated either. Cost studies for telecommunications have continually emphasized that the operations and maintenance segments account for a significant portion of total expenses, with a growing share of costs being incurred through manual troubleshooting and repair activities [3]. As network complexity increases with 5G densification, encompassing multiple frequency bands, heterogeneous node types, and dynamic beamforming configurations, these costs are projected to rise without intelligent automation. Systems that absorb diagnostic workloads previously requiring teams of senior RF engineers represent not only operational efficiencies but also strategic requirements for sustainable 5G operations at a national scale.

The complexities associated with ORAN systems are due to the implementation of disaggregated baseband systems, interoperability issues among multiple vendors, and near-real-time optimization needs. The challenges arise from the generation of constant telemetry data for distributed network functions, which need to be processed using correlation algorithms for policy, radio-frequency metrics, and infrastructure. Integration of large language model orchestration with near-RT RIC controllers, xApps, and telemetry-aware service pipelines enables adaptive coordination between diagnostic intelligence and radio resource management functions. Such integration supports scalable automation across dense urban deployments, heterogeneous spectrum layers, and dynamically varying traffic conditions.

This article addresses three interconnected architectural challenges in building autonomous RAN intelligence systems: designing constrained, deterministic tool-calling frameworks that harness LLM reasoning while preventing hallucination-driven errors in production environments; engineering stateful memory architectures that maintain contextual awareness across extended monitoring sessions; and structuring closed-loop integration between AI diagnostic systems and physical network controllers to enable safe, autonomous remediation. These principles are examined through design patterns applied in large-scale 5G deployments, with supporting evidence drawn from production performance measurements and industry research.

2. Core Architectural Principles of Autonomous Network Intelligence

2.1 Constrained Tool-Calling Frameworks and Hallucination Mitigation

The primary operational risk of deploying large language models (LLMs) in telecommunications environments is the hallucination problem, defined as the tendency of models to generate plausible but factually incorrect outputs when reasoning beyond their training distribution [13], [16]. In consumer applications, hallucinations may cause inconvenience; in autonomous network management systems, they can trigger erroneous remediation actions that affect service quality for millions of subscribers [6], [7]. Addressing this risk requires architectural mechanisms that constrain LLM outputs to verified, database-grounded measurements rather than inferred estimates. The structured tool orchestration mechanism addresses this issue by implementing a well-defined and strictly typed interface between the reasoning layer and the data infrastructure underpinning the networks [1]. Every tool takes only inputs that have been validated, such as sector ID, carrier frequency, and a time period, returning outputs that are valid responses to a query from an authoritative source in the form of production database records along with their provenance metadata.

The fundamental architectural principle here is the separation of duties, where the large language model is entrusted with conducting multi-stage reasoning, forming hypotheses, and planning remedial actions, and the tool invocations are tasked with retrieving data and validating measurements [16]. By doing so, it is ensured that each system component works in its area of competence. Moreover, the architecture allows for auditable diagnosis, as each statement made by the system can be directly linked to a certain invocation and query against the database [5].

2.2 Stateful Memory Architecture for Continuous Network Monitoring

Closed-loop network monitoring imposes requirements distinct from single-query diagnostic sessions. The intelligent system should be able to retain contextual awareness throughout the period of surveillance, which involves keeping tabs on changing circumstances, relating new occurrences to previous occurrences, and not repeating diagnoses of problems that have already been fixed [12], [14]. Standard LLM context windows, bounded by token limits, are insufficient without explicit memory management. Thread-aware state management addresses this challenge by assigning unique contextual identifiers to monitoring sessions for individual sectors or geographic clusters [2]. Each thread maintains a persistent state object containing the sector's current metric profile, a structured anomaly log, an inventory of active remediation actions, and a set of diagnostic hypotheses refined across monitoring intervals. When monitoring activity resumes, the agent retrieves the thread state to restore full contextual awareness without requiring the complete diagnostic history to fit within the active context window. This architecture has been validated in distributed deployment environments, where monitoring agents must maintain continuity despite interruptions or service restarts [8], [9]. Thread-aware memory eliminates repeated calls to diagnostic tools that can be made in a stateless manner, achieving efficiencies and faster response times [15]. Besides efficiencies, it establishes a permanent record of diagnostics and resolutions for post-event validation, testing of models, and continuous operational refinement [4].

The stateful monitoring architecture is also designed to facilitate coordination between O-RAN components based on their synchronization in terms of telemetry storage and execution of tasks based on policies. Memory objects can store telemetry about the history of KPI values, interference conditions at the level of sectors, the history of scheduler activities, and previous attempts to fix problems. This facilitates more efficient prioritization of measures required to correct issues while also minimizing redundant diagnostics within multiple sectors.

3. Spectral Efficiency Diagnostics: Cognitive Loop Design

3.1 Multi-Dimensional Signal Evaluation and Root Cause Correlation

Spectral efficiency degradation in 5G Radio Access Networks (RANs) rarely originates from a single isolated cause. Interference patterns, capacity saturation, hardware faults, and configuration errors often interact through non-linear mechanisms, producing similar observable symptom profiles across distinct causal pathways [7], [11], [12]. Effective autonomous remediation, therefore, requires a cognitive loop architecture capable of multi-dimensional correlation across carrier band configurations, sector geometries, temporal traffic patterns, and equipment-state information [14], [15].

The three-phase cognitive loop implemented in autonomous network intelligence systems operates sequentially through anomaly detection, evidence collection, and hypothesis formation. During the anomaly detection stage, constant monitoring of spectral efficiency measures detects sectors that have underperformed in terms of spectral efficiency below the operating threshold [2], [6]. On exceeding the operating threshold, the evidence-gathering stage is triggered, where the diagnostic software automatically engages structured telemetry acquisition procedures for collecting carrier-specific parameters, usage statistics, historical benchmarks, and neighboring sectors' interference statistics [11], [13].

The hypothesis formulation stage utilizes LLM's reasoning abilities to formulate hypotheses from the collected data. By comparing current profiles against libraries encoding known fault signatures such as inter-site interference from antenna downtilt, saturation from concentrated user load, hardware degradation, indicated elevated Error Vector Magnitude (EVM) measurements, and configuration drift from parameter mismanagement, the system generates ranked hypotheses with confidence levels [16]. This cognitive loop encodes expert-level diagnostic reasoning and aligns with ITU's framework for autonomous networks [2].

Additional orchestration layers may further correlate spectral efficiency degradation with radio scheduling imbalance, beamforming instability, uplink congestion behavior, and handover failure distribution across neighboring sectors. Correlation of Physical Resource Block (PRB) utilization, Channel Quality Indicator (CQI) variation, and retransmission statistics enables identification of congestion propagation patterns before large-scale quality degradation occurs. Such telemetry-driven reasoning mechanisms improve prioritization accuracy for remediation sequencing in dense urban and high-mobility deployment environments.

3.2 Predictive Load Modeling and Proactive Interference Mitigation

The most significant capability of autonomous network intelligence extends beyond reactive diagnostics to proactive intervention prediction and load rebalancing. This is possible through continuous analysis of past traffic trends, time-of-day patterns, events taking place within the venues, and the growing trends of the areas [12], [14]. The predictive nature of the AI system allows for preemptive actions to be taken where optimizations will be applied when signs of stress begin appearing and not after violations occur [7], [9]. Physical layer optimization actions initiated by predictive models may consist of adjusting the elevation angle of antennas, known as micro-downtilt, thereby minimizing interference between sectors in urban areas, as well as redistributing the user load across different frequencies [11], [15]. These interventions, previously requiring manual RF engineering work orders and maintenance windows, can now be executed autonomously within software-configurable parameters [3], [6]. In order to facilitate autonomous intervention actions, there is a certain hierarchy in terms of authorization based on the risk factor associated with the type of action. Low-risk software optimization actions can be undertaken without human approval, whereas medium-risk changes will result in alerts sent to operators, and high-risk physical actions would only occur after manual verification by humans [5], [8].Predictive orchestration models may also incorporate reinforcement learning-enabled policy adjustment for dynamic allocation of spectrum and routing of traffic in diverse frequency bands. Correlation of telemetry data regarding the distribution of users' mobility patterns, variability in scheduling loads, and sector-based throughput usage ensures effective optimization of resource allocations during high-demand periods. The integration with near-real-time control systems will ensure that local congestion incidents can be resolved automatically while ensuring deterministic protection measures for essential network services.

4. Enterprise Platform Architecture for Autonomous Intelligence

4.1 Geospatial Data Infrastructure and Hexagonal Spatial Indexing

Telecommunications intelligence systems operate on fundamentally spatial data: every network element occupies a physical location with specific coverage geometry, every customer qualification involves a geospatial proximity computation, and every planning decision is geographic in nature. Enterprise platform architecture for telecommunications AI must therefore treat geospatial computation as a primary concern. The Hexagonal Hierarchical Spatial Index (H3), originally developed for large-scale logistics optimization, provides a hierarchical spatial indexing framework that outperforms traditional latitude-longitude rectangular grid approaches for telecommunications applications [4]. The strength of the hexagonal indexing system is the constant-area feature. While rectangular systems generate cells that have a different effective area based on their latitude positions, hexagonal shapes created by H3 indexing remain roughly equal in terms of area throughout various regions, allowing for unbiased spatial evaluation on a global scale. The integration of the H3 system with column-oriented telemetry databases accelerates geospatial processes within seconds [11]. This capability is particularly critical for millimeter-wave deployments, where propagation physics impose strict line-of-sight requirements that necessitate precise geometric evaluation at building-level resolution [14].

Hierarchical spatial indexing additionally supports dynamic correlation between subscriber density, beam coverage overlap, and interference propagation boundaries across heterogeneous deployment environments. Integration of H3 indexing with distributed telemetry infrastructure enables rapid identification of congestion hotspots, coverage discontinuities, and mobility transition irregularities within large metropolitan regions.

Such spatial orchestration capabilities become increasingly important for ultra-dense 5G and emerging 6G infrastructures operating across millimeter-wave and sub-THz spectrum layers.

4.2 Microservices Architecture and High-Throughput Data Pipeline Design

Enterprise telecommunications AI platforms must concurrently serve multiple stakeholder groups, including RF engineers, capacity planners, operations personnel, and business intelligence teams, whose data access patterns and latency requirements differ substantially [7], [9]. It is impossible for a monolithic design to fulfill all these functions effectively without any contention or degradation during simultaneous use. A specialized microservices design provides a solution to this problem through service segmentation with independent scaling capabilities. These include geospatial computation, time-series metric storage, LLM orchestration, tool routing, natural language query processing, and visualization rendering [15]. High-throughput ingestion pipelines capable of processing terabytes of telemetry per hour require careful engineering attention to backpressure management, exactly-once semantics, and schema evolution as equipment vendors release updates [6]. These concerns are handled at the infrastructure layer through specialized ingestion frameworks, freeing the AI reasoning layer to focus on semantic intelligence extraction. Such a hierarchical separation makes it possible for reliability of infrastructure and semantic reasoning to develop independently from each other, thus ensuring sustainable, long-term infrastructure scalability over several generations of networks [12]. In addition to this, orchestrating pipelines in a distributed manner will enable the possibility of synchronization scalability for all layers involved in data intake, inference, policy control, and visualization, which operate in geographical areas separated by cloud architectures. Event processing technology, along with containerization technology for orchestrations, will provide quick deployment of network intelligence systems without compromising operational continuity.

5. Broader Societal and Industry Implications

The emergence of autonomous network intelligence brings about not only operational benefits for operators but also social benefits for society at large. With the emergence of 5G networks, which are expected to play the key role in providing public safety communications, emergency response capabilities, autonomous vehicles, remote surgeries, and industrial monitoring, reliable management systems mean much more than just business performance; indeed, they translate into social achievements [2], [5]. Systems capable of preventing any failures are qualitatively superior when compared to those that cannot do that [8], [10]. Just as important is the democratization of RF engineering skills via natural language processing. Through the abstraction of telecommunications-specific knowledge into an AI system, firms can empower their staff with advanced managerial skills without spending years training them [13]. This shift will allow the firm to move senior engineers away from repetitive diagnostic tasks towards more valuable innovation work while increasing the number of people who can manage complex network operations.

Workforce transformation industry research in relation to the technical nature of the field shows that automation solutions relying on the augmentation approach, where AI executes operational tasks while humans guide decision-making processes, provide more operational efficiency compared to completely manual and automated execution settings [5]. In the context of telecom infrastructure management, this model allows RF engineers and network operation experts to move from operational troubleshooting to optimization policy development, cross-domain collaboration, spectrum management, and infrastructure transformation. Systems analyzing data streams that help in identifying any deviations in network functioning and guide how to fix such deviations could be quite helpful in reducing the complexity involved in setting up wide-scale 5G networks. However, at the same time, humans have to confirm the remedial measures, ensure compliance with various guidelines, and control the communication infrastructure. Human-assisted autonomous operation provides a balance between both approaches.

Table 1: *Tool-Calling Frameworks and Hallucination Mitigation [1, 6, 7, 13]*

Dimension	Description	Benefits	Risks if Ignored
Hallucination Risk	LLMs may generate plausible but incorrect outputs in telecom environments	Identifies need for deterministic safeguards	Erroneous remediation actions affecting millions

Structured Interfaces	Strict, typed interfaces between reasoning and data infrastructure	Ensures validated inputs and verified outputs	Unverified data leading to compliance issues
Separation of Duties	LLMs handle reasoning; tools handle retrieval and verification	Each component operates in its domain of reliability	Overlap of responsibilities increases error rates
Auditability	Every diagnostic claim linked to tool invocation and database query	Transparent compliance and traceability	Lack of accountability in decision-making

Table 2: *Stateful Memory Architecture for Continuous Monitoring [2, 8, 9, 15]*

Component	Function	Operational Impact	Example Use Case
Contextual Identifiers	Assign unique IDs to monitoring sessions	Maintains continuity across long-running diagnostics	Sector-specific monitoring threads
Persistent State	Store metrics, anomaly logs, remediation actions, hypotheses	Provides full contextual awareness	Tracking evolving interference patterns
Distributed Continuity	Maintains monitoring despite interruptions or restarts	Ensures resilience in geographically distributed deployments	Multi-data center monitoring

Table 3: *Spectral Efficiency Diagnostics and Cognitive Loop [11, 12, 14]*

Phase	Activities	Outcomes	Example Fault Signature
Anomaly Detection	Continuous monitoring of spectral efficiency metrics	Identifies sectors below operational thresholds	P90 efficiency drop below 3.0 bits/Hz
Evidence Collection	Structured tool invocations to gather carrier data, utilization, and benchmarks	Builds multidimensional evidence-based	Adjacent sector interference profiles
Hypothesis Formation	LLM reasoning synthesizes evidence into ranked hypotheses	Produces prioritized root cause analysis	Antenna downtilt, saturation, and hardware degradation
Validation	Comparison against libraries of known fault signatures	Confirms expert-level diagnostic reasoning	Configuration drift from parameter errors

Table 4: *Enterprise Platform Architecture and Societal Implications [3, 4, 5, 14]*

Dimension	Description	Benefits	Broader Implications
Geospatial Infrastructure	Hexagonal Hierarchical Spatial Indexing (H3) for uniform spatial analysis	Consistent global coverage analysis, reduced distortion	Critical for millimeter-wave deployments
Column-Oriented Databases	Optimized for geospatial acceleration	Sub-second processing of billions of relationships	Enables national-scale planning

6. Conclusions

Autonomous radio network intelligence reflects the convergence of large language model orchestration, constrained tool-calling architectures, stateful memory management, and closed-loop integration with physical network layers. Together, these elements enable a new generation of telecommunications management that operates with speed, precision, and expert-level quality. The architectural principles outlined in this article, including structured tool frameworks for hallucination mitigation, thread-aware state management for monitoring continuity, predictive load modeling for interference prevention, and hexagonal-indexed geospatial infrastructure for coverage analysis, provide a replicable foundation for implementing advanced autonomous operations at a national scale. The article proves that these architectures offer tangible gains in terms of increased diagnostic agility, improved reliability of communications networks, and reduced consumption of resources necessary to ensure reliable operation within increasingly sophisticated architectures. Further, network autonomy in general serves to improve the reliability of future 5G infrastructures used for carrying out critical societal tasks. As networks progress towards the 6G architecture and native AI functionality, the designs developed at this stage will play a vital role in ensuring that telecommunication infrastructures can continue to serve intelligence needs in the future. Continued investments in the overlap of AI and telecommunications engineering are imperative, as technical expertise and social responsibility cannot be divorced here.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.

- **Author contributions:** The authors declare that they have equal right on this paper.

- **Funding information:** The authors declare that there is no funding to be acknowledged.

- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.
- **Use of AI Tools:** The author(s) declare that no generative AI or AI-assisted technologies were used in the writing process of this manuscript.

References

[1] Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., et al. (2023). ReAct: Synergizing reasoning and acting in language models. *arXiv*. https://arxiv.org/abs/2210.03629

[2] Benson, F., Rogers, R., Wilsome, D., & Eladawi, S. (2026). A framework for autonomous network reconfiguration in intelligent city networks. *ResearchGate Preprint*. https://www.researchgate.net/publication/400387252

[3] Mehmood, K., Kralevska, K., & Palma, D. (2023). Intent-driven autonomous network and service management in future cellular networks: A structured literature review. *Computer Networks, 220*, 109477. https://www.sciencedirect.com/science/article/pii/S1389128622005114

[4] De Anda-Suárez, J., López-Ramírez, J. L., Jimenez-Mendoza, D., Benitez-Quintero, J. M., Avina-Bravo, E. G., Gutierrez-Hernandez, D. A., et al. (2025). An integrated blockchain framework for secure autonomous vehicle communication systems. *Information, 16*(7), 557. https://www.mdpi.com/2078-2489/16/7/557

[5] Repetto, M., Striccoli, D., Piro, G., Carrega, A., Boggia, G., & Bolla, R. (2021). An autonomous cybersecurity framework for next-generation digital service chains. *Journal of Network and Systems Management, 29*, 37. https://link.springer.com/article/10.1007/s10922-021-09607-7

[6]Polese, M., Bonati, L., D'Oro, S., Basagni, S., & Melodia, T. (2023). Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges. *IEEE Communications Surveys & Tutorials, 25*(2), 1376–1411. https://dl.acm.org/doi/10.1109/COMST.2023.3239220

[7] Liyanage, M., Braeken, A., Shahabuddin, S., Ranaweera, P., et al. (2023). Open RAN security: Challenges and opportunities. *Journal of Network and Computer Applications, 214*, 103621. https://www.sciencedirect.com/science/article/pii/S1084804523000401

[8] Boutaba, R., Tornatore, M., & Rak, J. (2021). AI-driven closed-loop automation in 5G and beyond mobile networks. In *Proceedings of the 4th FlexNets Workshop on Flexible Networks: Artificial Intelligence Supported Network Flexibility and Agility*. https://dl.acm.org/doi/epdf/10.1145/3472735.3474458

[9] Wypiór, D., Klinkowski, M., & Michalski, I. (2022). Open RAN—Radio access network evolution, benefits and market trends. *Applied Sciences, 12*(1), 408. https://www.mdpi.com/2076-3417/12/1/408

[10] Wani, M. S., Kretschmer, M., Schröder, B., Grebe, A., & Rademacher, M. (2025). Open RAN: A concise overview. *IEEE Open Journal of the Communications Society, 6*. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10601697

[11] Cho, H., Mukherjee, S., Kim, D., Noh, T., & Lee, J. (2023). Facing wireless network densification in 6G: Challenges and opportunities. *ICT Express, 9*(3), 517–524. https://www.sciencedirect.com/science/article/pii/S2405959522001448

[12] Sufyan, A., Khan, K. B., Khashan, O. A., Mir, T., & Mir, U. (2023). From 5G to beyond 5G: A comprehensive survey of wireless network evolution, challenges, and promising technologies. *Electronics, 12*(10), 2200. https://www.mdpi.com/2079-9292/12/10/2200

[13] Zhang, C., Patras, P., & Haddadi, H. (2019). Deep learning in mobile and wireless networking: A survey. *IEEE Communications Surveys & Tutorials, 21*(3), 2224–2287. https://homepages.inf.ed.ac.uk/ppatras/pub/comst19.pdf

[14] Sinha, S., Reddy, G. P., Kim, S.-M., & Park, S.-H. (2026). Native AI-based hybrid deep learning for wireless link quality prediction in NTN waterside scenarios. *ICT Express, 12*(2), 311–318. https://www.sciencedirect.com/science/article/pii/S2405959525001559
[15] Doha, S. R., & Abdelhadi, A. (2025). Deep learning in wireless communication receivers: A survey. *IEEE Access, 13*. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=11053759
[16] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language models are few-shot learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems (NeurIPS '20)* (pp. 1877–1901). https://dl.acm.org/doi/abs/10.5555/3495724.3495883
[17] European Telecommunications Standards Institute (ETSI). (2019). *Zero-touch network and service management (ZSM); Reference architecture* (ETSI GS ZSM 002 V1.1.1). https://www.etsi.org/deliver/etsi_gs/ZSM/001_099/002/01.01.01_60/gs_ZSM002v010101p.pdf
[18] Bega, D., Gramaglia, M., Pérez, R., Fiore, M., Banchs, A., & Costa-Pérez, X. (2020). AI-based autonomous control, management, and orchestration in 5G: From standards to algorithms. *IEEE Network, 34*(6), 14–20. https://www.it.uc3m.es/banchs/papers/ieeenet20.pdf
[19] Wen, W., Zhou, C., Li, M., Wu, H., Zhou, H., Zhang, N., et al. (2022). AI-native network slicing for 6G networks. *IEEE Wireless Communications, 29*(1), 96–103. https://www.semanticscholar.org/reader/51eaaebdb856096ea5d8fceccd9bf1d40d0ac002