CogNexus

Multilingual, International, Peer-Reviewed, Open Access Journal

Vol. 2(1) 2026, 192-244

Measuring AI-Driven Autonomy in 5G and FTTH Networks: A Benchmarking Framework for Saudi Vision 2030 Digital Infrastructure

[1]Mohammad Mustafa

[2]Mohammad Firoz

[1]Jawaharlal Nehru Technological University; Contact Center Company (ccc by stc), Riyadh, Saudi Arabia.
ORCID: 0009-0002-7535-3321

[2]Jawaharlal Nehru Technological University; Contact Center Company (ccc by stc), Riyadh, Saudi Arabia.
ORCID: 0009-0006-4529-9263

DOI: https://doi.org/10.63084/cognexus.v2i1.311

Abstract

Measuring autonomous network management requires evidence of authorized decisions, executed actions, and verified service outcomes, rather than improvements in throughput or prediction accuracy alone. This article proposes a technology-aware benchmarking framework for assessing operational autonomy in fifth generation (5G) and fiber-to-the-home (FTTH) networks, with Saudi Vision 2030 as the policy application context. Four normalized dimensions, verified closed-loop completion, human independence, disturbance robustness, and audit completeness, form a provisional Autonomy Assessment Score (AAS). The framework also reports network-specific service indicators and treats operational safety as a separate qualification condition. Its calculation workflow is illustrated with 4,320 synthetic management records spanning six scenarios, two network domains, and three notional management configurations. The illustrative AI-style configuration is generated through assumed probabilities, not a trained model or implemented controller; accordingly, the examples do not validate AI effectiveness or Saudi network performance. Reproducibility is supported by a fixed random seed, documented assumptions, event-level records, scoring scripts, and sensitivity analyses. Alternative score weights change numerical values, and completion overlaps substantially with robustness, underscoring the need for dimension-level reporting. The contribution is a reproducible methodological specification for comparing operational autonomy while separating autonomy, safety, and service performance. Independent simulator, testbed, and operator validation remains necessary before empirical or national-level conclusions can be drawn.

Keywords: Autonomous networks; 5G; Fiber-to-the-home; Closed-loop management; Autonomy assessment; Synthetic demonstration; Saudi Vision 2030

Introduction

Background and Research Context

The growing reliance on digital connectivity has placed greater demands on telecommunications networks, particularly in terms of service reliability, capacity, and operational efficiency. Fifth-generation (5G) mobile networks and fiber-to-the-home (FTTH) infrastructure are central to this development, providing the connectivity needed for cloud services, industrial applications, smart cities, and data-intensive communication. However, increasing network complexity has made conventional management approaches less effective in addressing rapidly changing traffic conditions, service requirements, and infrastructure failures. Boutaba et al. (2018) identified machine learning as an important means of improving network management through traffic prediction, resource allocation, fault diagnosis, and performance optimization.

The introduction of software-defined networking (SDN), network function virtualization (NFV), and intelligent management systems has gradually changed how telecommunications infrastructure is operated. Rather than depending entirely on manual configuration and reactive troubleshooting, operators can increasingly use network analytics and automated control to anticipate service disruptions and adjust network resources. In cellular networks, these developments have supported self-configuration, self-optimization, and self-healing capabilities (Klaine et al., 2017). Similar developments in optical networks have improved the potential for automated fault management, resource optimization, and service monitoring (Musumeci et al., 2019).

Nevertheless, automation and autonomy are not interchangeable. A network may execute predefined instructions without independently assessing changing conditions or verifying the consequences of its actions. Genuine operational autonomy requires coordinated observation, analysis, decision-making, execution, and feedback within established management policies. Liyanage et al. (2022) emphasized the importance of these capabilities in zero-touch network and service management, particularly as telecommunications services become increasingly heterogeneous.

Measuring such capabilities presents a substantial challenge. Conventional indicators, including throughput, latency, availability, and resource utilization, describe network performance but do not directly establish how independently a network operates. A reliable

assessment must therefore consider autonomous decision-making, successful closed-loop execution, human intervention, and operational outcomes.

Saudi Vision 2030 and Digital Infrastructure Development

Saudi Arabia's Vision 2030 places digital transformation among the foundations of economic diversification, public service modernization, and technological development. Reliable mobile and fixed broadband networks support these objectives by enabling digital government services, connected industries, business innovation, and broader access to information and communication technologies. As national dependence on digital services increases, network reliability and operational resilience become important considerations for infrastructure development (Saudi Vision 2030, 2025).

The Kingdom has made substantial progress in telecommunications connectivity. According to the Communications, Space and Technology Commission (CST, 2025), Saudi Arabia recorded an internet penetration rate of 99% in 2024, while median 5G download speeds reached 318.6 Mbps and median fixed broadband download speeds reached 120.4 Mbps. These figures indicate considerable development in network access and service capacity. However, they do not reveal the extent to which the underlying infrastructure is managed autonomously.

The distinction is important because high network performance does not necessarily imply intelligent or independent operations. Networks may achieve strong service indicators while relying heavily on manual monitoring, configuration, and fault recovery. Conversely, autonomous management systems may improve operational responsiveness without producing immediate increases in headline connection speeds. A structured autonomy assessment could therefore complement established telecommunications performance indicators by examining how effectively network management systems respond to operational conditions.

For Saudi Vision 2030, the immediate application is a measurement proposal, not evidence of improvements in Saudi networks. A future operator pilot could use the framework to complement national connectivity statistics with event-level indicators of service continuity, recovery, accountability, and bounded human intervention. Candidate inputs include authorized radio performance counters, OLT/ONU telemetry, network trouble tickets, change logs, and operational energy records, subject to access and data-protection restrictions.

Problem Statement

Although artificial intelligence has been applied extensively to telecommunications optimization, measuring its contribution to autonomous network operation remains methodologically challenging. Existing research has examined individual applications such as traffic prediction, network slicing, fault classification, and resource allocation (Bega et al., 2020; Gu et al., 2020). These studies provide valuable evidence concerning specific technical functions but do not necessarily establish the overall autonomy of the network management process.

The challenge becomes more pronounced when comparing 5G and FTTH infrastructure. Mobile networks must address radio interference, mobility, dynamic traffic distribution, and heterogeneous service requirements. FTTH networks operate under different physical and management constraints, including optical signal conditions, shared PON capacity, subscriber provisioning, and access equipment failures. Consequently, identical performance metrics cannot adequately represent every aspect of both technologies.

Existing autonomy assessment methodologies provide an important foundation, particularly the TM Forum Autonomous Network Levels Evaluation Methodology (TM Forum, 2023). However, applying such approaches alongside quantitative operational evidence across different access technologies requires careful consideration of measurement definitions, verification criteria, and technology-specific conditions.

The central problem addressed in this research is therefore the need for a reproducible, multidimensional benchmarking approach that can evaluate AI-driven operational autonomy across 5G and FTTH networks while preserving the technical distinctions between them.

Research Objectives

The principal objective is to specify and demonstrate a reproducible framework for assessing AI-driven autonomy in 5G and FTTH network management. This paper does not experimentally validate AI controllers or benchmark operational Saudi networks.

The specific objectives are to define observable autonomy indicators and eligibility rules; formulate a technology-aware scoring procedure with explicit safety qualifications; demonstrate its calculations using transparent, assumption-driven synthetic events; examine sensitivity to indicator weights and scenario composition; and specify a future validation protocol for manual, rule-based, and trained AI-enabled controllers.

The study also identifies the operational data and governance conditions needed to use the framework in representative Saudi deployments. The relationship between autonomy, network performance, and energy efficiency remains a testable question rather than an established empirical result.

Research Questions

The study addresses five methodological questions. First, which event-level indicators distinguish autonomous operations from prediction accuracy and service quality? Second, how can comparable management functions be measured without conflating 5G and FTTH technical KPIs? Third, how does the proposed scoring procedure behave across notional manual, rule-based, and AI-style profiles?

Fourth, how sensitive are score interpretation and ranking to weighting, overlapping indicators, safety rules, and scenario mix? Fifth, what telemetry, implementation, and governance arrangements would be required to evaluate this framework with operational data in Saudi Arabia?

These questions guide the framework design, selection of benchmarking indicators, experimental comparisons, and interpretation of the resulting evidence.

Expected Research Contributions

This article contributes a task-level autonomy definition, an explicit AAS calculation and qualification procedure, a transparent synthetic event generator with sensitivity analyses, and a technology-aware protocol for later validation. Its principal contribution is a reproducible measurement design rather than a demonstrated improvement attributable to AI.

Literature Review and Research Gaps

Evolution of Intelligent and Autonomous Network Management

Telecommunications network management has evolved from equipment-centered supervision toward increasingly programmable and data-driven operations. Traditional management systems relied primarily on predefined configurations, periodic performance monitoring, and human intervention following service disruptions. Although these methods remain important, their limitations become more apparent in networks with rapidly changing traffic patterns, distributed infrastructure, and diverse service requirements.

SDN and NFV introduced greater flexibility by separating control functions from underlying infrastructure and allowing network services to be configured through software. These developments established important technical foundations for network slicing and dynamic service management in 5G systems (Ordonez-Lucena et al., 2017). Machine learning subsequently extended network management capabilities by supporting prediction, classification, anomaly detection, and adaptive resource allocation. Boutaba et al. (2018) demonstrated the broad applicability of these techniques across network management functions.

However, the ability to predict network conditions does not automatically establish autonomous operation. A predictive model may generate useful recommendations while leaving all operational decisions to human engineers. Similarly, deterministic automation can execute predefined actions without evaluating unfamiliar conditions. Autonomous management extends beyond these capabilities by integrating monitoring, decision-making, policy validation, execution, and outcome verification.

Liyanage et al. (2022) described zero-touch management as an architectural direction toward automated service lifecycle management and coordinated operations across network domains. This progression provides the conceptual basis for examining autonomy as an observable operational capability rather than a characteristic inferred solely from the presence of AI algorithms.

AI-Driven Autonomy in 5G Networks

Research on intelligent cellular networks has concentrated on improving resource utilization, service quality, mobility management, and network responsiveness. Klaine et al. (2017) reviewed machine learning applications in self-organizing cellular networks, identifying their relevance to automated configuration, optimization, and fault recovery. Zhang et al. (2019) further examined deep learning methods for mobile and wireless networking, demonstrating their potential in traffic analysis, resource management, and wireless system optimization.

The introduction of network slicing created additional management requirements because multiple services with different performance objectives must share common infrastructure. Foukas et al. (2017) highlighted the architectural and operational challenges associated with network slicing, while Afolabi et al. (2018) examined the enabling role of network softwarization. These developments increased the importance of adaptive resource allocation and service-level management.

Bega et al. (2020) proposed DeepCog, an AI-based capacity forecasting method for network slicing. Their evaluation demonstrated how learning-based resource prediction could reduce the costs associated with inappropriate capacity allocation. Nevertheless, the effectiveness of a forecasting model does not independently establish whether the complete resource management process operates autonomously. Such an assessment also requires evidence that decisions are authorized, executed, and verified without unnecessary human intervention.

Open Radio Access Network (O-RAN) architectures provide further opportunities for programmable and intelligent radio resource management. Arnaz et al. (2022) examined the integration of intelligence and programmability in open radio access networks, while Polese et al. (2023) analyzed O-RAN architecture, interfaces, control algorithms, and security challenges. These studies demonstrate the importance of standardized interfaces and programmable control in implementing intelligent RAN functions.

At the 5G core network level, 3GPP TS 23.288 specifies network data analytics capabilities that can support network functions through analytical information. However, the Network Data Analytics Function (NWDAF) should not be interpreted as a complete autonomous management system. Operational autonomy also depends on how analytics are incorporated into decisions, how corrective actions are executed, and whether the intended service conditions are subsequently restored.

Taken together, existing research establishes the technical feasibility of increasingly intelligent 5G management. A further benchmarking concern is how to distinguish the performance of individual AI components from the verified autonomy of the complete operational process.

AI-Driven Autonomy in FTTH Networks

FTTH networks provide high-capacity fixed broadband connectivity through optical access infrastructure. Passive optical network architectures commonly connect an optical line terminal (OLT) to multiple optical network units (ONUs) through a shared optical distribution network. Their operational requirements include subscriber service provisioning, optical signal monitoring, bandwidth allocation, fault localization, and service restoration.

Unlike mobile networks, FTTH systems are not primarily concerned with radio mobility or wireless interference. Their management challenges arise from optical transmission characteristics, shared access capacity, subscriber demand variations, and equipment or distribution network failures. Thangappan et al. (2020) reviewed dynamic bandwidth allocation

mechanisms in GPON and EPON systems, demonstrating the importance of allocating upstream transmission capacity according to traffic demand and service requirements.

Research on machine learning in optical networks has addressed several functions relevant to intelligent management. Musumeci et al. (2019) examined applications involving performance monitoring, fault management, and network optimization. Rafique and Velasco (2018) discussed learning-based network automation architectures, while Gu et al. (2020) reviewed intelligent optical networking approaches and their application to network control and survivability.

Fault identification represents a particularly important area of investigation. Shariati et al. (2019) demonstrated how information obtained from optical spectrum measurements could support failure detection and identification. However, techniques developed for optical transport networks cannot automatically be transferred to FTTH access networks without accounting for differences in equipment, telemetry availability, and physical architecture.

Recent publications have further examined the relationship between telemetry and autonomous optical networking. Cruzes (2026) discussed how real-time telemetry, machine learning, digital twins, and coordinated decision systems can support optical network automation. Rizzelli and Gaudino (2026) reviewed developments in high-speed passive optical networks, highlighting the continuing evolution of optical access technologies and associated capacity requirements.

These contributions provide a technical foundation for intelligent FTTH management but also reveal an important measurement distinction. Accurate optical fault prediction, for example, does not necessarily mean a network can independently restore service. Physical fiber damage may require field intervention, whereas certain configuration errors or bandwidth allocation problems can be addressed remotely. Consequently, autonomy assessment must distinguish recoverable events from situations requiring physical maintenance.

For FTTH networks, relevant benchmarking indicators should therefore include successful automated fault diagnosis, verified restoration of software-recoverable failures, human intervention requirements, bandwidth allocation efficiency, and service continuity. These indicators provide a more defensible basis for autonomy evaluation than prediction accuracy or optical transmission performance alone.

CogNexus
Multilingual, International, Peer-Reviewed, Open Access Journal
Vol. 2(1) 2026, 192-244

Existing Autonomous Network Standards and Benchmarking Approaches

International telecommunications standards provide an established foundation for evaluating autonomous management capabilities. The TM Forum Autonomous Network Levels Evaluation Methodology, IG1252, defines a structured approach to assessing autonomy through operational processes, tasks, and associated evaluation criteria (TM Forum, 2023). Its importance lies in assessing operational capabilities rather than assigning autonomy levels solely according to the technologies deployed.

The European Telecommunications Standards Institute (ETSI, 2019) established a reference architecture for zero-touch network and service management, incorporating management services and closed-loop operational principles. In parallel, 3GPP specifications define relevant 5G analytics functions, while ITU-T recommendations provide technical requirements for optical access systems. Industry specifications from the Broadband Forum address access network management and the associated data models.

Table 1. Relevant telecommunications standards and their roles in the proposed autonomy benchmarking framework

Standard or framework	Primary relevance	Application in the proposed study
TM Forum IG1252	Evaluation of autonomous network maturity through operational processes and tasks	Reference framework for assessing autonomy levels independently of the proposed numerical score
ETSI GS ZSM 002	Zero-touch service management and closed-loop operational architecture	Guidance for monitoring, decision-making, execution, and verification processes
3GPP TS 23.288	Network data analytics within the 5G System	Definition of relevant 5G analytics capabilities and information sources
O-RAN architecture	Programmable radio access management and intelligent control	Evaluation of AI-supported RAN optimization where applicable interfaces are available
ITU-T G.9807.1	Technical specifications for 10-gigabit-capable symmetric passive optical networks	Reference for XGS-PON system characteristics and optical access constraints

Broadband Forum TR-383 and TR-369	Access network data models and user services management	Support for standardized telemetry and management interfaces where implemented

The standards summarized in Table 1 offer substantial guidance for designing and evaluating automated telecommunications systems. However, their purposes differ. TM Forum IG1252 addresses autonomy maturity, ETSI ZSM supports management architecture, and technology-specific specifications describe functions and interfaces within particular network environments.

Consequently, a cross-technology benchmarking framework must preserve these distinctions while establishing consistent operational measurement rules. It should complement existing standards by linking autonomy classifications to measurable evidence, including successful closed-loop execution, human intervention, service recovery, and policy compliance.

Critical Analysis and Research Gaps

The reviewed literature demonstrates considerable progress in intelligent telecommunications management. Learning-based approaches have been investigated for cellular resource allocation, network slicing, optical fault identification, and traffic optimization (Bega et al., 2020; Gu et al., 2020; Shariati et al., 2019). Standards development has also strengthened the architectural foundations for autonomous operations. Nevertheless, several methodological issues remain important when applying these advances to a common assessment of 5G and FTTH networks.

First, much of the technical literature focuses on individual algorithms, network functions, or technology-specific management problems. Although these studies establish the effectiveness of particular approaches, their metrics do not necessarily support direct comparison of operational autonomy across different access technologies.

Second, model performance and autonomous operation are distinct measurement concepts. High prediction accuracy, reduced latency, or improved bandwidth utilization may demonstrate technical effectiveness without proving that a management system can independently complete a verified operational task.

Third, the incremental contribution of AI requires careful separation from the benefits already provided by conventional automation. Comparisons restricted to manual management and AI-enabled systems may attribute improvements to AI that could also be achieved through deterministic rules.

Fourth, autonomy introduces operational considerations beyond efficiency, including system robustness, decision traceability, policy compliance, and appropriate human oversight. These factors are particularly important when automated actions could affect service continuity or network security.

Finally, national telecommunications indicators primarily describe connectivity and service performance. They provide limited evidence about the autonomy of the underlying management systems. This creates an opportunity to examine how operational autonomy indicators could complement established infrastructure performance measurements.

Table 2. Identified research gaps and corresponding methodological responses

Research gap	Issue requiring investigation	Proposed methodological response
G1: Cross-technology comparability	Technology-specific evaluations may not support consistent comparison of autonomy across 5G and FTTH networks.	Develop common operational autonomy dimensions supported by separate technology-specific performance indicators.
G2: Distinction between AI performance and autonomy	Prediction accuracy and network performance improvements do not establish independent operational capability.	Measure verified closed-loop completion, decision independence, execution success, and operational outcomes.
G3: Attribution of AI benefits	Improvements attributed to AI may partly result from conventional automation.	Compare manual management, rule-based automation, and AI-enabled closed-loop management under controlled conditions.
G4: Operational safety and accountability	Autonomy assessments may not adequately represent robustness, human oversight, policy compliance, and decision traceability.	Incorporate explicit safety requirements, auditability indicators, robustness evaluation, and human escalation criteria.

G5: National infrastructure relevance	Connectivity and service performance indicators do not directly measure autonomous management capability.	Establish a defined relationship between autonomy benchmarking outcomes and relevant Saudi digital infrastructure priorities.

The gaps presented in Table 2 establish the rationale for a common evaluation methodology. The framework measures task-level autonomy through observable events while preserving technology-specific service outcomes, eligibility criteria, and operational safeguards.

Importantly, the framework must be assessed against existing methods rather than assume superiority. Its scientific value will depend on the validity of its indicators, the consistency of its measurements, and the extent to which controlled evaluations demonstrate meaningful differences between management approaches.

Research Positioning

The reviewed studies establish three complementary areas of knowledge: AI-based network optimization, standards-driven autonomous management, and technology-specific performance evaluation. However, integrating these perspectives into a reproducible cross-technology assessment remains a relevant research objective.

This study addresses that objective by proposing a benchmarking framework that combines standards-aligned autonomy assessment, measurable closed-loop operational evidence, and network-specific performance indicators. The framework considers 5G and FTTH as distinct technical environments while applying common definitions of autonomous task completion, human intervention, robustness, and operational accountability.

Its intended contribution is not to introduce another isolated network optimization algorithm, but to establish a consistent method for evaluating how effectively AI-enabled management systems operate under defined conditions. The resulting measurements may also provide a basis for examining operational efficiency, service resilience, and infrastructure management priorities associated with Saudi Vision 2030.

Proposed AI-Driven Autonomy Benchmarking Framework

Framework Design Principles

The proposed benchmarking framework establishes a structured method for measuring how effectively artificial intelligence supports autonomous operations in 5G and fiber-to-the-home (FTTH) networks. Its primary purpose is to assess operational autonomy through measurable evidence rather than the presence of AI algorithms or conventional network performance indicators alone. The framework combines network monitoring, intelligent decision-making, controlled execution, service verification, and performance assessment within a common evaluation structure. This approach builds on established research concerning machine learning for network automation and zero-touch service management (Rafique & Velasco, 2018; Liyanage et al., 2022).

Five principles guide the framework's development. Interoperability ensures that autonomy can be evaluated across different network technologies, management systems, and equipment vendors. Measurability requires that operational decisions and their consequences be supported by identifiable events, timestamps, intervention records, and verification results. Comparability establishes common definitions for autonomy while maintaining separate technical performance indicators for mobile and optical access networks. Safety and accountability require autonomous actions to comply with operational policies, authorization requirements, and service protection mechanisms. Finally, reproducibility requires consistent measurement procedures, documented assumptions, and clearly defined evaluation conditions.

These principles address a central difficulty in autonomous network assessment: different technologies may perform similar management functions through substantially different mechanisms. For example, a 5G controller may respond to congestion by adjusting radio resource allocation, while an FTTH controller may modify bandwidth allocation policies at an optical line terminal (OLT). Although the technical actions differ, both can be evaluated according to their ability to detect the problem, select an authorized response, execute the change, and verify its effectiveness. The proposed framework therefore separates autonomy capabilities from the technology-specific outcomes used to establish their operational value.

Conceptual and Technical Architecture

The proposed architecture comprises six interconnected layers that organize network information, intelligent management functions, autonomous control, and benchmarking activities. Its design draws on the principles of closed-loop management described by the European Telecommunications Standards Institute (ETSI, 2019) and the use of programmable

network interfaces discussed by Polese et al. (2023). The architecture is intended to support coordinated assessment without requiring identical management implementations across the two network domains.

The network infrastructure layer contains the physical and virtual resources being evaluated. In the 5G domain, these include radio access network components, relevant core network functions, and transport connectivity. The FTTH domain includes OLTs, optical network units (ONUs), and the associated passive optical distribution infrastructure. Each domain retains its own management interfaces and operational constraints.

The data acquisition layer collects network measurements, configuration changes, equipment alarms, and operational event records. Relevant 5G information includes radio resource utilization, traffic conditions, connection failures, latency, and service performance. FTTH measurements include optical signal power, loss-of-signal alarms, PON utilization, subscriber traffic, and equipment status. Depending on the implementation, data may be obtained through network management systems, standardized telemetry interfaces, and authorized application programming interfaces. The 5G Network Data Analytics Function (NWDAF), defined in 3GPP TS 23.288, can provide relevant analytical information where supported by the deployed architecture.

The data processing layer prepares collected information for analysis by validating records, aligning timestamps, resolving duplicated events, correlating alarms, and extracting operational features. This layer also records missing or unreliable measurements, since incorrect telemetry can compromise autonomous decision-making. Musumeci et al. (2019) emphasized the importance of accurate monitoring information in machine learning applications for optical networks.

The AI and decision management layer evaluates network conditions and identifies possible management responses. Depending on the operational task, its functions may include traffic forecasting, fault classification, anomaly detection, and resource optimization. The selected algorithms must be documented and evaluated separately from the autonomy assessment. Decisions are checked against operating policies, resource constraints, and predefined service requirements before execution. This distinction is necessary because an accurate model prediction does not guarantee that the resulting network action is appropriate.

The autonomous control layer executes authorized management actions through the available control interfaces. It records the selected action, execution status, and subsequent network

response. Where an action fails, violates a predefined safeguard, or produces an unacceptable service condition, the system must support rollback or escalation to a human operator. Physical faults that cannot be corrected remotely are directed to appropriate maintenance procedures rather than classified as successful autonomous restoration.

The benchmarking and governance layer evaluates the operational evidence generated by the preceding layers. It calculates autonomy indicators, measures service outcomes, monitors policy compliance, and maintains records for independent assessment. Autonomy scores are reported alongside their individual components and relevant network performance indicators. Figure 1 presents the proposed architectural arrangement.

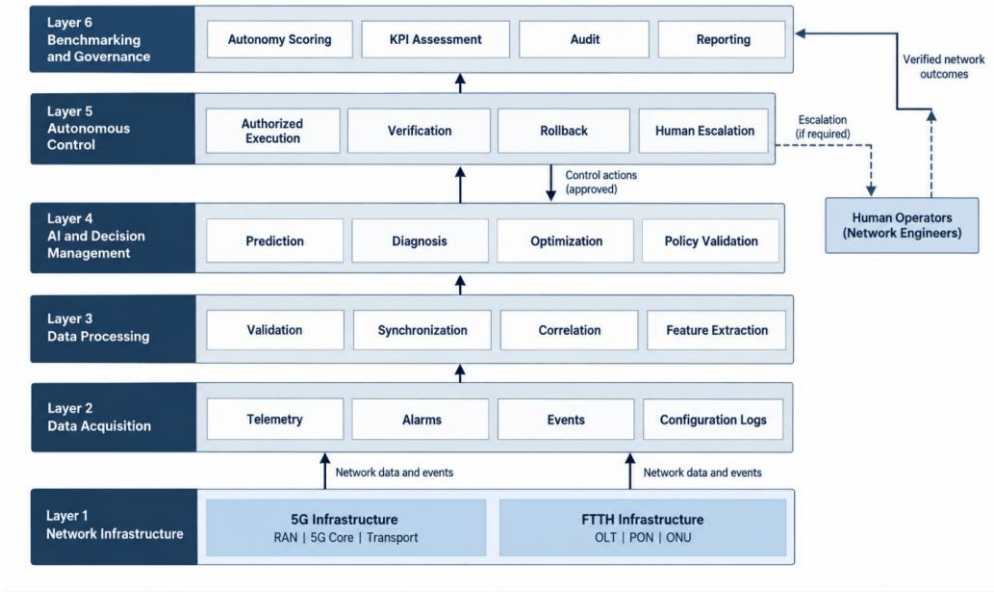


Figure 1. Proposed layered architecture for AI-driven autonomy benchmarking in 5G and FTTH networks (conceptual design)

Operational Benchmarking Scenarios

The framework evaluates autonomy using operational scenarios that represent common network management responsibilities. Six functions are selected: fault management, resource optimization, service assurance, predictive maintenance, recovery management, and service provisioning. These functions allow comparison of management capabilities across 5G and FTTH networks while preserving their different technical requirements.

In the 5G environment, the evaluation addresses radio resource allocation, traffic congestion, mobility-related failures, and service management. In the FTTH environment, it addresses optical access faults, bandwidth allocation, subscriber service quality, and equipment

management. Such distinctions are consistent with research on intelligent radio access networks and optical network automation (Arnaz et al., 2022; Gu et al., 2020).

Table 3. Proposed operational scenarios for cross-technology autonomy benchmarking

Operational function	5G evaluation scenario	FTTH evaluation scenario
Fault management	Detection and diagnosis of radio link failures, cell degradation, and network connectivity faults	Detection and localization of optical signal degradation, OLT/ONU faults, and loss-of-signal events
Resource optimization	Adjustment of radio resources and network capacity in response to changing traffic demand	Optimization of PON bandwidth allocation and access capacity utilization
Service assurance	Identification and correction of service-level degradation affecting latency, throughput, or network quality	Identification and remediation of subscriber service degradation, congestion, and access network performance problems
Predictive maintenance	Prediction of equipment failures and abnormal operating conditions requiring preventive action	Prediction of OLT/ONU degradation and potential optical access equipment failures
Recovery management	Automated recovery from software-recoverable radio or network service failures	Automated restoration of recoverable optical access service or configuration failures
Service provisioning	Policy-driven configuration and activation of supported network services or slices	Automated subscriber service activation, profile configuration, and provisioning verification

Each experimental scenario must identify the initial network condition, the expected management response, and the evidence required to establish successful completion. For example, an automated congestion-management task is considered successful only when the controller executes a permitted configuration change and subsequent measurements confirm that the specified service condition has been restored. Issuing a command without verifying its outcome does not constitute successful autonomous completion.

The benchmarking procedure should distinguish eligible autonomous tasks from incidents requiring physical intervention or actions outside the controller's authority. These events remain relevant to overall network reliability, but they must not distort the denominator used to measure the completion of technically feasible autonomous tasks. All exclusions must be documented to prevent selective reporting.

Closed-Loop Autonomy Evaluation

Closed-loop management provides the operational basis for the proposed autonomy assessment. The framework follows six connected stages: observation, analysis, decision, execution, verification, and learning or adjustment. These stages extend established network automation principles by requiring evidence of both autonomous control and successful operational outcomes. ETSI (2019) identifies closed-loop operation as an essential mechanism for coordinated, automated network management.

During observation, the system collects information describing current network conditions. The analysis stage identifies service degradation, abnormal behavior, or opportunities for optimization. During decision-making, the controller determines an appropriate response based on available information, predicted outcomes, and operational policies. The execution stage applies the authorized action through supported management interfaces. Verification examines whether the action achieved the specified objective without introducing unacceptable service effects. Finally, learning or adjustment updates the controller's operational knowledge or revises its response according to the capabilities of the implemented system.

An important requirement is that every autonomous action pass a policy-validation process before execution. Decisions that fall outside approved operational boundaries must be rejected or referred to a human operator. Similarly, unsuccessful actions should trigger predefined recovery procedures, rollback where technically supported, or escalation.

The framework records each stage using a unique event identifier, timestamps, decision provenance, action status, and verification evidence. These records establish whether the management task was completed independently, required human involvement, or failed to achieve its objective. The same assessment procedure applies to both network domains, although the permissible actions and service verification conditions remain technology-specific.

CogNexus
Multilingual, International, Peer-Reviewed, Open Access Journal
Vol. 2(1) 2026, 192-244

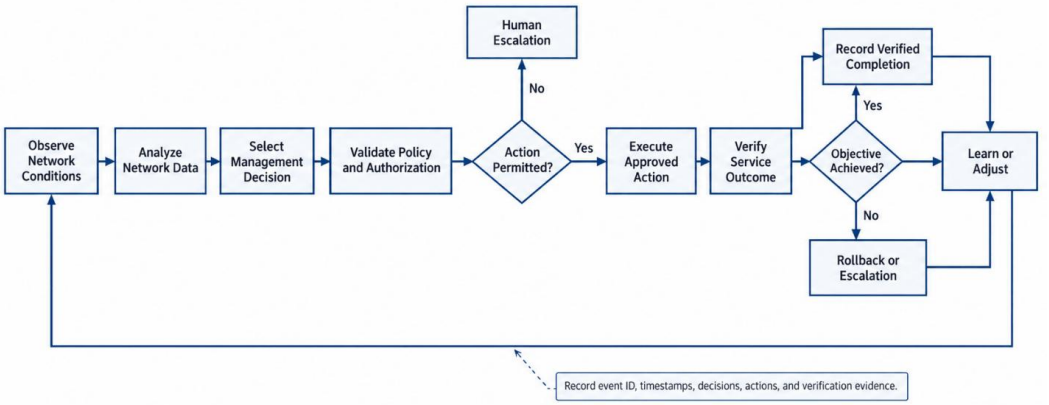


Figure 2. Proposed closed-loop autonomy evaluation and safeguard mechanism (conceptual design).

Mathematical Modelling and Autonomy Measurement

Formal Definition of Network Autonomy

The proposed framework defines autonomy as the ability to observe conditions, choose authorized operational actions, execute them, and verify outcomes without a human operational decision. It distinguishes task completion from prediction accuracy and network service performance (Liyanage et al., 2022).

Let d denote the network domain (5G or FTTH), s an operational scenario, and N(eligible; d,s) the count of events meeting prespecified technical and authorization criteria. Physical-repair cases remain in the incident registry but are excluded from the autonomy completion denominator only when a remote remedy is genuinely infeasible. Eligibility counts and exclusions must be disclosed.

The framework evaluates autonomy through four principal dimensions: verified closed-loop completion, independence from human operational decisions, robustness under disturbances, and audit completeness. These dimensions are measured separately before being combined into a proposed autonomy assessment score. Service quality and operational efficiency are evaluated alongside the score rather than treated as direct evidence of autonomy.

This distinction is important because a management system may operate independently while producing poor service outcomes. Conversely, a network may maintain excellent service performance through extensive human supervision. The proposed model therefore considers

both autonomy and operational effectiveness without assuming that improvements in one necessarily establish improvements in the other.

Closed-Loop Completion Rate

Closed-loop completion measures the proportion of eligible management events successfully resolved through independent and verified autonomous actions. It captures whether a management system completes the entire operational sequence rather than merely generating predictions or issuing configuration commands.

The closed-loop completion rate is expressed as:

$$C_{d,s} = \frac{N_{d,s}^{verified}}{N_{d,s}^{eligible}} \quad \textbf{(1)}$$

Here N(verified; d,s) is the count of eligible events with a permitted autonomous action, no human operational decision, and independent outcome verification; N(eligible; d,s) is the denominator defined before observing controller outcomes. A missing or unverifiable outcome cannot be recorded as a successful completion.

An event qualifies as successfully completed only when the required action is executed, the intended network condition is independently verified, and applicable safety requirements are satisfied. Failed actions, unverifiable outcomes, and events requiring human intervention are not counted as successful autonomous completions.

The indicator ranges from zero to one, with higher values representing a greater proportion of verified autonomous tasks. Its interpretation must consider scenario complexity, event eligibility, and service verification requirements. Physical faults requiring field repairs are recorded in the broader incident dataset but excluded from the autonomous completion denominator when they cannot be resolved through the available control mechanisms.

Human Independence Ratio

Human independence measures the extent to which eligible network management events proceed without human operational decisions. Although human oversight remains necessary for safety and accountability, frequent manual involvement indicates that the management process has not achieved substantial operational independence.

The human independence ratio is defined as:

$$I_{d,s} = 1 - \frac{N_{d,s}^{human}}{N_{d,s}^{eligible}} \quad (\mathbf{2})$$

Here N(human; d,s) is the number of eligible events requiring at least one human operational decision and N(eligible; d,s) is the number of eligible events. Mandatory monitoring without a decision should be classified separately from an intervention that changes the operational response.

A value approaching one indicates fewer events requiring human intervention. However, independence does not necessarily establish successful task completion. A controller may operate without human assistance but fail to restore the required service condition. Consequently, the independence ratio must be interpreted alongside verified closed-loop completion and operational reliability.

Human intervention records should distinguish routine supervision, mandatory authorization, corrective decisions, and emergency escalation. Only intervention categories included in the predefined measurement protocol should affect the calculated independence ratio. This distinction prevents legitimate safety oversight from being misinterpreted as unnecessary operational dependence.

Composite Autonomy Assessment Score

Individual autonomy indicators provide useful information, but no single measure fully represents autonomous network management. The proposed framework therefore introduces an Autonomy Assessment Score (AAS) that combines four normalized dimensions into a common quantitative index.

In addition to completion and independence, the model incorporates robustness and audit completeness.

Operational robustness measures the ability of the management system to complete eligible tasks under predefined disturbances, such as changing traffic conditions, degraded telemetry, or unexpected network behavior. It is expressed as:

$$R_{d,s} = \frac{N_{d,s}^{verified,dist}}{N_{d,s}^{eligible,dist}} \quad (3)$$

Here N(verified,dist; d,s) counts verified independent completions in predesignated disturbed events, and N(eligible,dist; d,s) counts all eligible disturbed events. Because robustness is a subset-specific completion rate, it is mathematically related to the overall completion indicator and must not be treated as independent evidence.

Audit completeness measures whether operational events contain the required information for independent evaluation:

$$Q_{d,s} = \frac{N_{d,s}^{audit}}{N_{d,s}^{eligible}} \quad (\mathbf{4})$$

Here N(audit; d,s) is the number of eligible events with all required decision, authorization, action, timestamp, and outcome fields. Audit completeness covers successful and unsuccessful outcomes and is not itself proof of autonomy.

The composite score is defined as:

$$AAS_{d,s} = 100\left(w_C C_{d,s} + w_I I_{d,s} + w_R R_{d,s} + w_Q Q_{d,s}\right) \quad (5)$$

subject to:

$$w_C + w_I + w_R + w_Q = 1$$
$$w_k \geq 0, \qquad 0 \leq AAS_{d,s} \leq 100$$

The nonnegative weights w_C, w_I, w_R, and w_Q apply to completion, independence, robustness, and audit completeness, respectively; their sum equals one. The AAS is a proposed index for method demonstration, not a standardized autonomy level.

For the base illustration, the weights are fixed before generating any result:

$$AAS_{d,s} = 100\left(0.40C_{d,s} + 0.25I_{d,s} + 0.20R_{d,s} + 0.15Q_{d,s}\right)$$

The base allocation gives verified task completion the largest share (0.40), followed by absence of human decisions (0.25), completion under disturbed conditions (0.20), and record completeness (0.15). This ordering encodes a design preference for operational accomplishment over documentation, not an empirically estimated importance structure. Two prespecified alternatives are examined: completion-priority (0.55, 0.20, 0.15, 0.10) and governance-priority (0.25, 0.20, 0.15, 0.40). Because R and C overlap, the reporting protocol

must also show both dimensions, assess their correlation, and consider a score that excludes R in later construct validation.

The raw AAS is always reported as a diagnostic index, but a qualified autonomy classification requires C > 0, I > 0, sufficient eligible events for all included dimensions, complete reporting of excluded cases, and no executed critical policy violation within the assessed unit. A single executed critical safety violation disqualifies that unit from receiving a favorable autonomy classification, regardless of raw AAS. Less severe violations, attempted policy blocks, missing data, and human escalation remain separately reportable. These minimum conditions are methodological safeguards, not externally validated standards; a zero-completion manual configuration is labeled "not autonomous" even when audit points produce a nonzero raw score.

The proposed AAS is distinct from the autonomous network maturity levels defined by the TM Forum (2023). While TM Forum IG1252 assesses autonomy through operational processes and associated task criteria, AAS provides a continuous numerical measure based on observed events. The score is therefore intended to complement existing autonomy assessments and must not be converted directly into standardized maturity levels without independent validation.

Operational Performance and Service Quality Metrics

Autonomy measurement must be accompanied by operational performance assessment to establish whether independent management decisions produce acceptable network outcomes. Relevant indicators include fault recovery time, service availability, latency, throughput, resource utilization, and service-level agreement (SLA) compliance. Previous studies have demonstrated the importance of these outcomes in intelligent network management and optimization (Boutaba et al., 2018; Bega et al., 2020).

One important reliability indicator is mean time to repair or restore service (MTTR), defined for successfully restored incidents as:

$$MTTR = \frac{1}{N_r}\sum_{i=1}^{N_r}(t_i^{restore} - t_i^{onset}) \quad (6)$$

Here N_r is the number of restored incidents, t_i(onset) is the time the ith incident begins, and t_i(restore) is its independently verified restoration time. MTTR refers only to restored

incidents and must be accompanied by unresolved-event counts and right-censoring information.

Lower MTTR values generally indicate faster recovery. However, reporting MTTR only for restored incidents can conceal failures that remain unresolved. The evaluation must therefore report unresolved incident rates and use time-to-event analysis where observations are censored at the end of the measurement period.

Service availability is calculated as:

$$A_d = \frac{T_d^{operational}}{T_d^{observed}} \times 100 \ (\mathbf{7})$$

Here T_d(operational) denotes time meeting the specified service availability conditions in domain d, while T_d(observed) denotes total observation time over the same measurement interval.

Network latency and throughput are assessed using independent service measurements. For 5G, additional indicators include handover failures, radio link failures, and radio resource utilization. FTTH-specific measurements include optical received power, loss-of-signal events, OLT/ONU availability, and PON bandwidth utilization. These technology-specific indicators reflect the different operating characteristics of cellular and optical access infrastructure (Polese et al., 2023; Thangappan et al., 2020).

Operational efficiency is evaluated through human effort, management overhead, and energy consumption. Where infrastructure energy and delivered traffic can be measured reliably, traffic-normalized energy intensity may be calculated as:

$$EI_d = \frac{E_d^{network}}{V_d^{traffic}} \ (\mathbf{8})$$

Here E_d(network) is independently metered network energy in kWh and V_d(traffic) is delivered traffic in TB during the same measurement period. The illustrative dataset contains only assumed controller energy in Wh/event, so network energy intensity cannot be estimated from it.

Energy consumed by the management controller should be reported separately from total network infrastructure energy. This prevents an apparent improvement in management

efficiency from being incorrectly interpreted as a reduction in overall telecommunications energy consumption.

Benchmarking Indicators

The proposed framework combines common autonomy indicators with operational metrics specific to each network domain. The common indicators establish a basis for comparing autonomous management capabilities, while the technology-specific measures determine whether those capabilities deliver acceptable outcomes under relevant service conditions.

Table 4. Proposed core autonomy and operational performance metrics for 5G and FTTH benchmarking

Measurement category	Benchmarking indicator	Unit or reporting measure	Preferred measurement evidence
Autonomy	Verified closed-loop completion	Percentage of eligible events (%)	Controller execution logs and independent verification records
Autonomy	Human independence	Percentage of eligible events (%)	Operator intervention and escalation records
Autonomy	Disturbance robustness	Verified completion rate under predefined disturbances (%)	Controlled disturbance and scenario execution logs
Governance	Audit completeness	Percentage of events containing required records (%)	Decision histories, authorization records, execution logs, and verification evidence
Governance	Policy compliance	Number and severity of policy violations	Policy enforcement and security audit logs
Reliability	Mean time to repair or restore service	Seconds or minutes	Fault onset and verified restoration timestamps

Reliability	Service availability and unresolved incident rate	Percentage (%)	Service monitoring and incident records
Service quality	Network latency and throughput	Milliseconds (ms); Mbps or Gbps	Independent traffic and service performance measurements
5G-specific	Handover failure rate, radio link failure rate, and resource utilization	Percentage (%), events, and utilization ratio	RAN performance counters and network management logs
FTTH-specific	Optical received power, loss-of-signal events, and PON utilization	dBm, event count, and utilization ratio	OLT/ONU monitoring records and optical access telemetry
Operational efficiency	Human operational effort per eligible event	Minutes/event	Operational activity and intervention records
Sustainability	Network energy intensity and management energy overhead	kWh/TB; Wh/event	Energy meters, traffic measurements, and controller power records
AI reliability	Prediction performance, uncertainty calibration, and decision latency	Task-specific metrics; milliseconds	Held-out model evaluation and decision execution logs

The indicators in Table 4 serve different analytical purposes. Autonomy and governance metrics describe the management system's operational capabilities, whereas reliability and service quality metrics establish the consequences of its decisions. AI reliability measures assess the underlying models but do not independently demonstrate network autonomy.

For cross-technology comparison, the framework prioritizes common dimensionless indicators such as verified completion, independence, and audit completeness. Raw performance values, including latency and optical signal power, are interpreted within their respective network

environments. The benchmarking process must also account for differences in scenario difficulty, event frequency, and service requirements to prevent misleading comparisons.

Framework Validation Criteria

The scientific value of the proposed framework depends on whether its indicators accurately represent network autonomy and produce consistent results across different operating conditions. Validation will therefore address construct validity, content validity, repeatability, inter-rater reliability, sensitivity, and operational validity.

Construct validity will examine whether the selected indicators measure autonomous operational capability rather than unrelated network performance characteristics. Content validity will assess whether the framework adequately represents the principal functions involved in autonomous network management, including independent decision-making, execution, verification, robustness, and accountability. These assessments may be supported by expert review and comparison with established autonomous network evaluation criteria.

Repeatability will be examined through repeated benchmarking trials conducted under equivalent operating conditions. The resulting scores should demonstrate acceptable consistency after accounting for expected network variability. Inter-rater reliability will assess whether independent evaluators applying the same criteria produce comparable autonomy classifications and measurements.

Sensitivity analysis will investigate how changes in indicator weights, event eligibility rules, scenario distributions, and measurement uncertainty affect the composite autonomy score. Because the four dimensions are not completely independent, the analysis will also examine their relationships and identify cases in which the composite index conceals important differences between management configurations.

Finally, operational validity will examine whether successful autonomous decisions correspond to independently verified improvements or acceptable maintenance of network service conditions. This assessment is essential because an autonomy score should not be interpreted as evidence of effective network management unless the underlying actions satisfy the defined service and safety requirements.

The validation procedure will therefore report individual autonomy indicators alongside the composite score and measured network outcomes. This provides a basis for determining

whether the proposed framework can produce meaningful, reproducible assessments across 5G and FTTH environments.

Research Methodology and Experimental Design

Research Design

The study develops a measurement specification and demonstrates its calculations using scripted synthetic events. It does not implement a 5G or FTTH network controller. The research design combines a standards-informed conceptual framework, event-level autonomy definitions, numerical sensitivity assessment, and a prespecified protocol for subsequent experimental validation. Rafique and Velasco (2018) and Liyanage et al. (2022) provide relevant technical foundations for closed-loop management.

Five methodological steps are distinguished: requirements and literature synthesis; definition of event eligibility, safeguards, and AAS dimensions; reproducible synthetic calculation; scoring and scenario-mix sensitivity analysis; and specification of independent simulator or operator validation. Autonomy indicators are separated from service performance, so the paper makes no claim that generated scenario differences prove an AI benefit.

The methodology comprises five stages. The first identifies benchmarking requirements through a critical review of relevant telecommunications standards, autonomous management architectures, and existing evaluation methods. The second develops the proposed framework, including its technical architecture, autonomy indicators, mathematical models, and operational assessment criteria. The third involves implementing the framework within suitable 5G and FTTH experimental environments. The fourth applies controlled operating scenarios to compare manual management, deterministic automation, and AI-enabled closed-loop control. The final stage evaluates the reliability, validity, and statistical significance of the resulting measurements.

The investigation distinguishes two related outcomes: the degree of operational autonomy achieved and the consequences of that autonomy for network performance. Autonomous completion, human independence, robustness, and audit completeness will be assessed using the indicators established in Section 4. Service availability, restoration time, resource utilization, and operational efficiency will be measured separately. This distinction prevents

improvements in conventional network performance from being incorrectly interpreted as evidence of autonomous operation.

Experimental Environment

The proposed experimental environment consists of separate 5G and FTTH network configurations connected to a common benchmarking and data collection system. Both environments will support controlled traffic generation, fault injection, network monitoring, and management action logging. Their configurations will be documented to permit independent reproduction of the experiments.

For a first independent validation campaign, the selected 5G implementation would be ns-3 with the 5G-LENA NR module. A Python control process would consume simulator trace callbacks for load, radio link state, packet latency, and traffic counters, then request permitted scheduler or service-configuration changes through an experiment-specific actuation interface. Both telemetry and actuation connections require integration testing; neither is implemented in the current study.

The selected FTTH implementation would be a discrete-event XGS-PON service model with explicit OLT/ONU states, subscriber queues, upstream allocation, and optical-degradation events. The controller would use simulated OLT/ONU counters to request supported DBA or service-profile adjustments and log independently verified outcomes. This is a proposed access-management abstraction, not a physical-layer-accurate optical testbed.

Each environment would export the same event, decision, authorization, execution, and verification schema. The present 4,320 records instead come from the documented probability generator; no simulator platform, trained AI controller, or access-network equipment produced them.

5G network environment. The 5G configuration will represent a standalone mobile network incorporating relevant radio access, core network, and management functions. Depending on available resources, the implementation may use a 5G-compatible ns-3 simulation environment or a laboratory configuration based on suitable components from srsRAN, OpenAirInterface, and free5GC. The selected components must support the interfaces and management actions required by the experimental scenarios.

The environment will reproduce variations in user demand, traffic congestion, mobility conditions, radio resource availability, and recoverable network failures. Relevant

measurements will include network throughput, latency, radio resource utilization, connection failures, and service restoration time. Polese et al. (2023) demonstrated the importance of programmable control interfaces and intelligent management functions in open radio access networks, providing a technical foundation for the proposed 5G evaluation.

FTTH network environment. The FTTH configuration will represent a passive optical access network consisting of an optical line terminal (OLT), multiple optical network units (ONUs), and shared optical access capacity. A GPON or XGS-PON implementation will be selected according to the available simulation tools or laboratory equipment. The experimental configuration will reproduce subscriber traffic variation, PON congestion, optical signal degradation, equipment alarms, and recoverable service configuration failures.

Measurements will include optical received power, access capacity utilization, subscriber throughput, latency, equipment status, and service restoration time. These indicators reflect established optical access management requirements, particularly those concerning bandwidth allocation and fault monitoring (Musumeci et al., 2019; Thangappan et al., 2020).

Although the two environments differ technically, both will follow equivalent event recording and performance evaluation procedures. The benchmarking system will collect operational evidence independently of the management controllers to reduce the possibility that a controller incorrectly reports its own actions as successful.

Data Requirements and Collection

The study will collect operational data describing network conditions, management decisions, and service outcomes. Required information includes performance measurements, equipment alarms, traffic statistics, configuration changes, AI predictions, automated execution records, human interventions, and service verification results. The importance of reliable telemetry for learning-based network management has been established in previous research on intelligent optical networks (Gu et al., 2020).

Each experimental event will receive a unique identifier linking its initial network condition to the corresponding management response. The event record will contain the network domain, operational scenario, management configuration, event severity, traffic conditions, and relevant timestamps. It will also indicate whether the selected action was authorized, successfully executed, independently verified, or referred to a human operator.

A common data structure will be used for both network domains, while additional fields will capture technology-specific characteristics. In the 5G environment, these may include radio utilization, handover failures, and service latency. FTTH records will include optical power measurements, loss-of-signal events, PON utilization, and OLT/ONU operating conditions.

Before analysis, the collected records will undergo quality checks to identify duplicate events, missing measurements, inconsistent timestamps, and incomplete verification evidence. Events with insufficient evidence will not be classified as successful autonomous completions. Missing-data patterns and event exclusions will be reported to ensure that the resulting autonomy scores are not artificially inflated.

Where authorized operational datasets become available, they may be used to assess the framework under commercial network conditions. However, contextual telecommunications statistics cannot substitute for the event-level records required to evaluate autonomous management.

Comparative Experimental Configurations

Three management configurations will be evaluated to distinguish the effects of manual operation, conventional automation, and AI-enabled autonomous control. The comparison is designed to identify whether AI provides additional operational benefits beyond those achievable through predefined automation rules.

Table 5. Experimental management configurations for comparative benchmarking

Configuration	Management approach	Evaluation purpose
B0: Manual management	Human operators monitor network conditions, identify faults, select corrective actions, and perform or authorize management operations.	Establish a reference for service recovery, human effort, operational efficiency, and network performance.
B1: Rule-based automation	Predefined conditions trigger deterministic management actions. Events outside supported rules are escalated to human operators.	Evaluate the performance of conventional automation without AI-based decision-making.
B2: AI-enabled closed-loop management	Trained AI models support fault diagnosis, prediction,	Evaluate whether AI improves verified

	or optimization. Actions are executed within predefined policy constraints and followed by outcome verification.	autonomous completion, operational resilience, and service performance beyond B1.

B0 represents human-issued operational decisions; B1 uses an explicit deterministic ruleset with the same action permissions; B2 is reserved for an implemented trained AI controller in future experiments. In the current synthetic demonstration, B2_AI_proxy is not B2 as an implemented AI system. The proxy draws decisions and outcomes from preset probabilities, and no model parameters are fitted from the demonstration data.

In a future comparison, all three configurations would receive matched traffic and faults, the same service thresholds, and equivalent permissible actions. Run order and scenario seeds should be randomized or counterbalanced; initial network state and controller-version records should be archived. Differences from B1 would then be interpreted as effects of the tested AI controller only within those controlled settings.

To maintain experimental fairness, the three configurations will use equivalent network resources, event definitions, and permissible corrective actions wherever feasible. Management policies and verification requirements will remain consistent. The experimental order will be randomized when practical, and the network will be restored to an equivalent initial state before subsequent trials.

Experimental Scenarios and Procedures

The experiments will cover the six operational functions identified in Section 3.3: fault management, resource optimization, service assurance, predictive maintenance, recovery management, and service provisioning. Each function will be evaluated under normal operating conditions and predefined disturbances.

An implementable pilot protocol would use 10 minutes of warm-up followed by a 180-minute observation window per independent simulation run, with each fault injected at a logged time and followed to restoration or censoring. Three predefined load regimes (35%, 65%, and 90% of nominal capacity) and two stress levels (20% load increase or 10% missing telemetry under controlled injection) would supply a starting design. These are proposed trial settings, not measurements or validated deployment limits. Scenario-specific acceptance thresholds, including verification duration and permissible actions, must be fixed before the run. Physical-

repair events would remain in the incident inventory but be excluded only from automation-eligible denominators.

A pilot study would estimate between-run variation and determine the replication count needed for the primary completion-rate contrast. A preliminary planning target of 30 independent runs per domain-scenario-configuration is not a power result and must be revised after pilot estimates. Each run would retain fault injection details, random seed, controller version, event eligibility, verification traces, intervention records, and safety outcomes. The present five batches of 24 toy events are solely generator groups, not independent network trials.

The implementation plan also requires negative-control events for which action is prohibited, missing or conflicting telemetry, and physical faults requiring maintenance. These scenarios assess safe refusal, fallback behavior, and policy compliance rather than rewarding unnecessary intervention.

AI Model Development and Validation

For an actual validation trial, B2 would use a versioned gradient-boosted classifier for fault type and a gradient-boosted regression model for short-horizon demand, combined with a deterministic policy layer that restricts available actions. The model and controller must be trained, calibrated, and tested with domain-specific telemetry before any operational comparison. This intended controller is distinct from the current B2_AI_proxy, which is not a trained model.

Training, validation, and testing would be separated by run and, where possible, site or temporal block. Inference results would be compared against engineering-confirmed labels and measured service outcomes. Decision confidence, latency, abstentions, and fallback would be recorded at the same event granularity as executed actions. The synthetic file contains train/validation/holdout labels by generated trial, but no model was trained on any split.

The classifier would be assessed with precision, recall, confusion matrices, and probability calibration; the regressor with held-out error and residual analyses. An autonomous completion would be counted only if a supported action passed policy checks and independent telemetry verified the resulting service outcome. A model prediction alone cannot earn autonomous-completion credit.

Classification models will be assessed using precision, recall, and false-positive rates. Forecasting models will be evaluated using appropriate prediction errors, while decision-

making systems will also be assessed for execution latency, uncertainty, and operational consequences. Where predictions influence network control, the model must support rejection or escalation when its confidence or available evidence is inadequate.

Importantly, model accuracy will not be treated as evidence of completed network autonomy. A successful prediction must lead to an authorized action and a verified operational outcome before the corresponding event contributes to the autonomous completion indicator.

Ethical, Security, and Data Governance Considerations

The research will incorporate data protection, network security, and operational safety throughout experimentation. Access to operator infrastructure or subscriber-related information will require appropriate authorization. Sensitive records will be minimized, anonymized or pseudonymized where necessary, and stored under controlled access conditions.

Autonomous control will be restricted to approved management interfaces and predefined operational policies. Potentially disruptive actions will require validation, while rollback and human escalation procedures will remain available. These safeguards are important in programmable telecommunications environments, where incorrect configuration changes can affect service reliability and security (Polese et al., 2023).

Where Saudi operator data are used, applicable telecommunications regulations and national data protection requirements will be assessed before collection. Experimental configurations, model versions, data processing procedures, and unsuccessful trials will be documented to support transparency and reproducibility.

Performance Evaluation and Statistical Analysis

Autonomy Evaluation

The performance evaluation will examine whether the proposed benchmarking framework can distinguish the operational autonomy achieved by different network management configurations. The assessment will use the Autonomy Assessment Score (AAS) and its four constituent indicators: verified closed-loop completion, human independence, disturbance robustness, and audit completeness, as defined in Section 4.

Each indicator will be calculated separately for the 5G and FTTH environments across the six operational scenarios. The composite AAS will then be computed using the predefined

weighting coefficients, subject to the framework's data quality and safety requirements. Reporting the individual indicators alongside the composite score will allow differences in operational capabilities to be examined without relying exclusively on a single numerical value.

Verified closed-loop completion will serve as the primary autonomy outcome because it establishes whether a management system can independently complete an authorized action and achieve the intended operational result. Human independence will measure the proportion of eligible events managed without human decisions, while robustness will assess autonomous completion under changing or degraded network conditions.

Audit completeness and policy compliance will provide additional evidence of accountability. Unauthorized actions, unsuccessful changes, and failed recovery attempts will be reported separately, regardless of the composite score. This approach is consistent with the emphasis on verifiable and controlled network operations in zero-touch management research (Liyanage et al., 2022).

Network Performance Evaluation

Network performance will be evaluated alongside autonomy to determine whether independent management decisions produce measurable improvements in service reliability and operational efficiency. The analysis will distinguish common service indicators from measurements specific to each network technology.

For 5G, the evaluation will consider network throughput, latency, radio resource utilization, handover or radio link failures, and service restoration time. These indicators reflect the operational requirements of mobile networks, where traffic demand, radio conditions, and user mobility can change considerably over time (Klaine et al., 2017; Polese et al., 2023).

In the present paper, these calculations describe the planned evaluation of a real controller; the numerical values in Section 7 are derived only from scripted assumptions. Across independent future runs, completion and safety outcomes should be analyzed as binary or count variables, not inferred from a fixed probability advantage assigned to one configuration.

Service outcomes will be compared against predefined operating objectives and the relevant management baseline. Although common indicators such as restoration time and availability can be assessed across both technologies, their results must be interpreted according to the network architecture, service conditions, and technical constraints involved.

The analysis will also distinguish successfully restored incidents from unresolved failures. This is necessary because calculating recovery time only for completed restorations could produce misleading estimates of management effectiveness.

Quantifying the Incremental Contribution of AI

The incremental contribution of AI will be evaluated primarily by comparing B2, the AI-enabled management configuration, with B1, the rule-based automation configuration. This comparison provides a more appropriate basis for assessing AI-specific benefits than comparisons with manual management alone.

For a selected performance indicator Y and operational scenario s, the difference between configurations is defined as:

$$\Delta Y_s = \overline{Y}_{B2,s} - \overline{Y}_{B1,s} \quad (9)$$

For scenario s, the contrast ΔY_s denotes the mean B2 outcome minus the mean B1 outcome. Y may represent completion, restoration time, service quality, or management energy, but its desirable direction is metric dependent. B2 refers to an actually trained controller only in the future protocol; the current B2_AI_proxy difference is entirely illustrative.

The interpretation of the difference depends on the indicator. A positive value may indicate improved autonomous completion or service availability, whereas a negative value may indicate reduced restoration time, energy consumption, or operational overhead.

The primary comparison will concern verified autonomous task completion. Secondary analyses will examine service restoration, human intervention, resource utilization, and service quality. Where possible, identical fault conditions and workload profiles will be evaluated under both configurations to reduce the influence of environmental variation.

Previous research has demonstrated the value of comparing learning-based resource management with established provisioning approaches (Bega et al., 2020). Accordingly, the present study will not assume that AI-enabled management necessarily performs better. Any claimed improvement must be supported by observed differences under controlled experimental conditions.

Statistical Testing and Uncertainty

Statistical analysis will establish the magnitude, precision, and consistency of differences between management configurations. Descriptive statistics will summarize autonomy

indicators, service quality, restoration time, resource utilization, and management overhead. Continuous variables will be reported using means and standard deviations or medians and interquartile ranges, depending on their distributions. Proportions will be presented with appropriate confidence intervals.

The primary comparison between B1 and B2 will use statistical procedures suited to the experimental design. For paired observations, matched comparisons will be considered. Where multiple events occur within the same experimental run or network configuration, mixed-effects models or cluster-aware bootstrap methods will account for dependence between observations.

Restoration time will be examined using both descriptive measures and time-to-event analysis where appropriate. Unresolved incidents will be treated as censored observations rather than excluded without justification. This will allow recovery performance to be evaluated without assuming that every failure is successfully restored.

For real testbed data, cluster-aware or mixed-effects models should account for events nested within runs or network sites, and unresolved incidents should be handled with appropriate censoring. The current generator creates five scenario batches of 24 events, not independent site-level replications. Consequently, bootstrap intervals calculated from its synthetic rows describe Monte Carlo variability under assumptions and are not inferential confidence intervals for real-world controller effectiveness.

The AAS audit should report the original and alternative weight sets, equal-weight and stress-weighted scenario mixes, the completion-robustness association, and the independent components. Repeatability and inter-rater validity cannot be established by the synthetic generator; they require implementation-level observations and independent human labeling.

Together, these analyses will determine whether the proposed framework produces consistent measurements and whether observed differences are sufficiently reliable to support technical conclusions.

Performance Trade-offs and Robustness

The final evaluation will examine whether greater operational autonomy introduces additional costs, limitations, or risks. Particular attention will be given to decision latency, computing overhead, management-controller energy consumption, service reliability, and operational safety.

Robustness testing will introduce controlled disturbances involving incomplete telemetry, unexpected traffic changes, unfamiliar failure conditions, and conflicting management decisions. These tests will assess whether the controller maintains acceptable performance, recognizes uncertainty, and initiates appropriate fallback procedures.

Energy consumption will be evaluated separately for network infrastructure and management controllers. This distinction is necessary because additional computation associated with AI may increase controller energy requirements even when network resources are managed more efficiently.

The analysis will also investigate whether reductions in human intervention correspond to successful service outcomes. A management system that acts independently but produces frequent configuration errors or unresolved incidents cannot be considered operationally effective solely because it requires less supervision.

Differences between 5G and FTTH will be interpreted in relation to their physical architectures, available management interfaces, and recovery constraints. The evaluation will therefore consider the conditions under which autonomous control is beneficial, where conventional automation remains sufficient, and where human intervention is necessary.

These prespecified comparisons determine how Section 7 illustrates the reporting workflow, without treating the generated advantages as empirical test outcomes or national infrastructure effects.

Reproducible Synthetic Benchmarking Demonstration

Evidence status: All Section 7 results are synthetic illustrations from a documented Python stochastic generator. The 4,320 records are not ns-3 or PON outputs, operator measurements, or trained-model predictions. The deliberately favorable B2_AI_proxy probabilities make its apparent advantage an input to the generator, not a discovered AI effect. No hypothesis test or inferential statement about deployed networks is warranted.

Reproducibility and Scoring-Procedure Checks

The generator uses NumPy default_rng with fixed seed 20302026. It creates two domains, six scenarios per domain, five trial labels per scenario, and 24 base events per trial, yielding 1,440 base events. The same base profiles are repeated for B0_manual, B1_rules, and B2_AI_proxy to produce 4,320 configuration-specific rows. Severity is sampled as levels 1, 2, and 3 with

probabilities 0.51, 0.34, and 0.15. The five nominal load ratios are 0.35, 0.55, 0.70, 0.85, and 0.95, with Gaussian noise of standard deviation 0.075, clipped to 0.20–0.99. Disturbance probability is 0.20 + 0.08 × severity. Physical-repair exclusions occur with probability 0.10 in fault diagnosis, predictive maintenance, and restoration scenarios; telemetry availability is assigned probability 0.95.

Eligibility is determined before simulating management responses. This yields 671 eligible events per 5G configuration and 682 per FTTH configuration, out of 720 domain-specific events per configuration. Human intervention base probabilities are fixed at 1.00 for B0, 0.48 for B1, and 0.15 for B2, with additive adjustments for severity, disturbance, and missing telemetry. Base resolution probabilities are 0.91, 0.95, and 0.975, respectively, before severity- and load-related penalties. The B2 value is intentionally larger; this is an assumption, not an observed result. Full generator functions, including conditional lognormal handling durations, assumed latency/throughput multipliers, audit-completeness probabilities, controller-energy formulas, and policy blocks, are provided in the accompanying code and model_assumptions.csv.

The raw AAS is recomputed from eligible-event counts, not averaged from unweighted scenario percentages. Under base/completion-priority/governance-priority weights, the illustrative 5G scores are B0 10.06/6.71/26.83, B1 47.03/44.88/58.47, and B2 74.32/73.56/79.95. FTTH scores are B0 10.32/6.88/27.51, B1 46.27/44.40/57.42, and B2 76.48/75.35/81.47. Thus, rankings remain the same in these three weight configurations, but the raw magnitude is weight sensitive, particularly for nonautonomous B0. Scenario mix matters as well: in 5G B1, the equal-scenario score is 47.46 compared with 48.32 when fault diagnosis and restoration receive double weight; the corresponding FTTH B1 scores are 46.36 and 46.02. Across all 36 scenario-configuration cells, completion and robustness correlate at approximately $r = 0.978$, confirming substantial indicator overlap within this constructed sample.

These checks demonstrate internal calculation reproducibility only. The full event-level CSVs, model assumptions, data dictionary, fixed seed, regeneration script, figure script, and sensitivity table are supplied separately. Because the generator directly assigns favorable B2 behavior, neither ranking stability nor apparent effect size validates the scoring model, the controller, or real-world autonomy. The manual configuration is explicitly labeled not autonomous under the minimum qualification rule, despite retaining a nonzero diagnostic AAS.

CogNexus
Multilingual, International, Peer-Reviewed, Open Access Journal
Vol. 2(1) 2026, 192-244

Illustrative Autonomy Comparisons

The synthetic domain-level raw AAS values are 10.06, 47.03, and 74.32 for 5G B0, B1, and B2, respectively; for FTTH they are 10.32, 46.27, and 76.48. These numbers are outputs of the stated probability model, not estimated network performance. B0 fails the minimum autonomy criterion because its independent completion and independence measures are both zero.

Among the eligible events, synthetic verified-completion rates are 40.39% (B1) and 71.98% (B2) in 5G, and 40.76% (B1) and 72.58% (B2) in FTTH. The 31.59- and 31.82-percentage-point differences mainly reflect assigned intervention and resolution probabilities. They must not be described as demonstrated gains from AI.

The corresponding synthetic independence ratios are 45.90% and 78.99% for 5G B1 and B2, and 44.43% and 80.79% for FTTH. Because the human-escalation assumptions favor B2, these figures illustrate the calculation and plotting conventions rather than independently support a controller-level inference.

Figure 3 displays these results as an explicitly synthetic reporting example. No significance tests, testbed confidence intervals, or calibrated AI performance claims are attached to the bars.

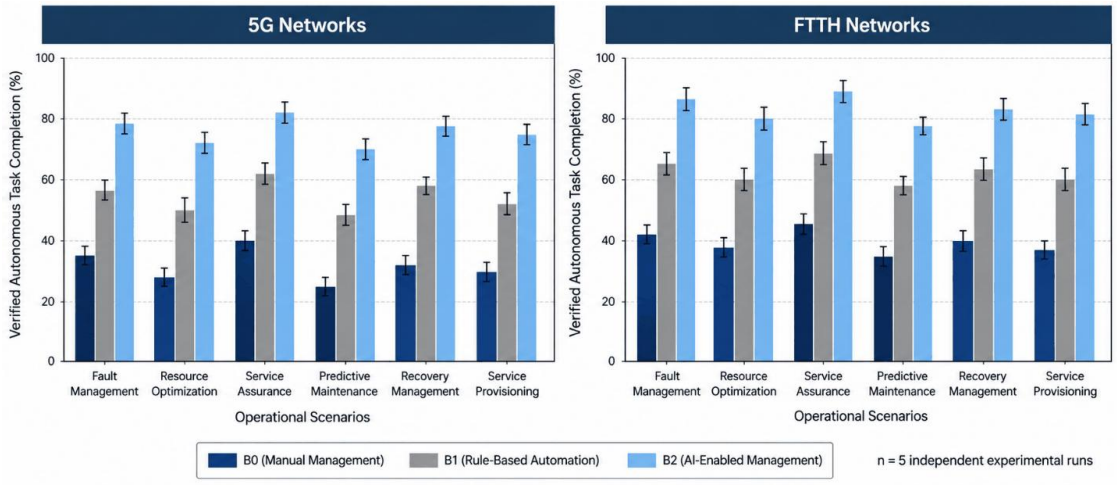


Figure 3. Synthetic demonstration only: verified autonomous completion by management configuration and scenario.

Illustrative Fault Restoration Outcomes

The assumed restoration-time generator produces 5G mean durations of 56.59 minutes (B0), 31.21 minutes (B1), and 18.71 minutes (B2) among successfully resolved fault diagnosis and restoration cases. The corresponding FTTH means are 81.87, 45.39, and 25.80 minutes. These

values arise from mode-specific duration multipliers and do not establish that an AI system is faster.

For transparency, unresolved incidents accompany the means. Among 240 fault-related rows per domain and configuration, 5G has 20, 22, and 15 unresolved B0/B1/B2 incidents; FTTH has 29, 21, and 14. Incidents without restoration receive right-censored status at the illustration's 180-minute analysis horizon. Differences in physical access topology cannot be inferred from these assigned durations.

Figure 4 visualizes the synthetic time-to-restoration profiles. Since both recovery distributions and unresolved-event probabilities were specified in advance, the figure is useful only for illustrating how a future measured dataset could be reported.

Synthetic restoration-time summaries exclude unresolved cases by construction. Any field study must report those cases, use actual event timestamps, and provide appropriately censored comparisons before claiming recovery improvements.

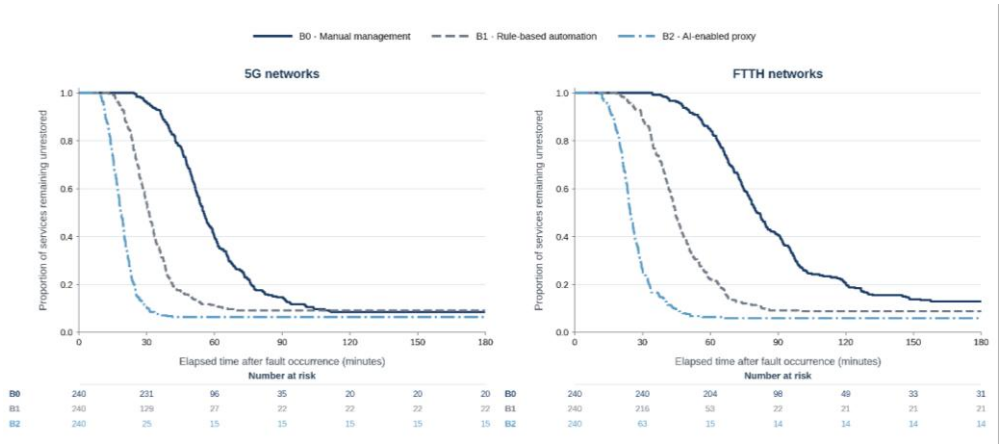


Figure 4. Synthetic demonstration only: time-to-restoration curves with unresolved incidents censored at 180 minutes.

Illustrative Service and Energy Indicators

The simulated 5G SLA-compliance rates for B1 and B2 are 84.03% and 94.72%, with assigned mean latency of 44.80 and 39.64 ms and throughput of 329.92 and 352.35 Mbps. The synthetic SLA threshold is 55 ms for 5G and is an internal model criterion, not a regulatory target.

In FTTH, B1 and B2 have assumed SLA-compliance rates of 48.19% and 70.28%, mean latencies of 34.14 and 30.20 ms, and throughput of 585.08 and 622.38 Mbps. A 35-ms internal

model threshold defines the synthetic FTTH SLA flag. These values show the reporting format, not observed network gains.

The generated controller-energy overhead rises from 4.69 to 7.18 Wh/event for 5G B1 versus B2, and from 4.67 to 7.18 Wh/event for FTTH. These increases are about 53.2% and 53.7%, respectively. The controller-energy values are not metered infrastructure energy, and the generator prescribes higher energy for B2; hence there is no measured energy-efficiency advantage. Figure 5 illustrates this score-overhead trade-off only.

A practical assessment would need service-level benefit, management effort, execution risk, and separately metered controller and access-network energy data. A high diagnostic AAS alone cannot establish a favorable operational or sustainability trade-off.

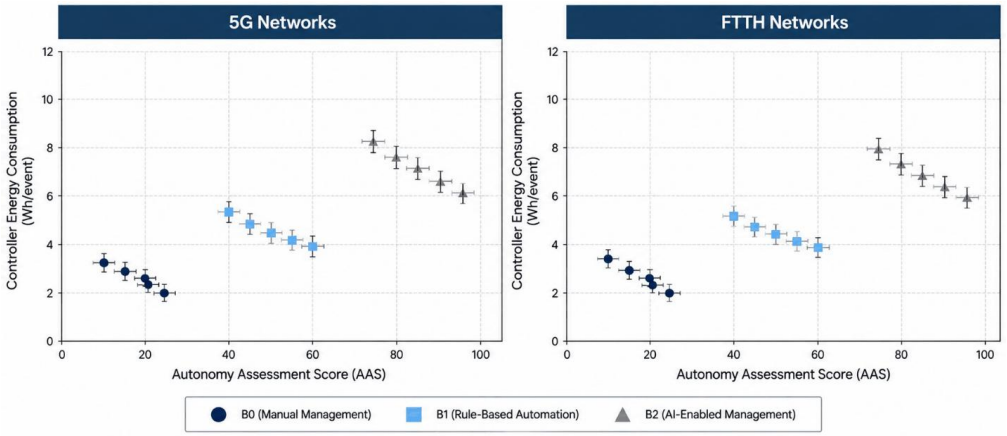


Figure 5. Synthetic demonstration only: illustrative AAS versus assumed management-controller energy (Wh/event**).**

Summary of Synthetic Demonstration Outputs

Table 6 compiles the assumed model outputs. Each value is reproducible from the accompanying event-level file and code, but none is a measurement from a network platform. The results are retained to demonstrate aggregation, denominator control, and consistent cross-domain reporting.

Table 6. Synthetic illustration only: domain-level autonomy and service indicators (not experimental results)

Metric	5G B0	5G B1	5G B2	FTTH B0	FTTH B1	FTTH B2
Verified autonomous	0.00	40.39	71.98	0.00	40.76	72.58

completion (%)						
Autonomy Assessment Score (0–100)	10.06	47.03	74.32	10.32	46.27	76.48
Human interventions per 100 eligible events	100.00	54.10	21.01	100.00	55.57	19.21
Mean restoration time (min)	56.59	31.21	18.71	81.87	45.39	25.80
SLA compliance (%)	55.69	84.03	94.72	29.17	48.19	70.28
Management energy (Wh/event)	2.82	4.69	7.18	2.82	4.67	7.18
Executed policy violations (count)	0	0	0	0	0	0

Discussion

Interpretation of the Illustrative Benchmarking Results

The demonstration shows that event-level completion, human interventions, audit records, and service outcomes can be reported using a shared schema for 5G and FTTH without equating their technical performance indicators. Nevertheless, the higher B2_AI_proxy rates arise from explicitly more favorable intervention and resolution probabilities. The synthetic outcomes cannot establish AI effectiveness, implementation feasibility, or superiority over deterministic automation.

The AAS audit also exposes a construct problem: manual management obtains a nonzero raw score from audit completeness even though no event is completed autonomously. Requiring nonzero completion and independence, plus mandatory safety checks, prevents this diagnostic

index from being mistaken for an autonomy classification. The close relationship between total completion and perturbed-event completion requires future work on indicator redundancy, alternative weighting, and construct validity.

Comparison With Existing Research and Standards

Previous studies demonstrate learning-based traffic prediction and resource optimization in mobile networks and fault management applications in optical systems (Bega et al., 2020; Musumeci et al., 2019). Those algorithm-level contributions do not establish independent end-to-end management. The present framework proposes separate evidence for authorization, execution, verification, human intervention, and service consequences, while making no claims about performance of an implemented AI model.

TM Forum IG1252 provides a maturity assessment of operational processes, and ETSI ZSM provides architectural principles for zero-touch management (TM Forum, 2023; ETSI, 2019). The proposed AAS is not an official TM Forum level and cannot be mapped onto Level 0 through Level 5 without empirical calibration. Its possible added value lies in reporting repeatable event counts and service outcomes alongside, rather than instead of, established process-level assessments.

The proposed AAS does not replace these approaches. Rather, it introduces a continuous, event-based measurement intended to accompany standards-based maturity assessment. Its originality must ultimately be evaluated through comparison with existing methods, particularly regarding measurement validity, reproducibility, and the additional operational information provided.

Implications for 5G and FTTH Network Management

The framework has potential applications in fault management, resource optimization, service assurance, and network recovery. In 5G environments, autonomous management may support faster responses to congestion, radio resource constraints, and changing service demands. Polese et al. (2023) identified programmable interfaces and intelligent control as important capabilities for advanced radio access management. However, the benefits of these capabilities depend on the reliability of the underlying analytics and the availability of appropriate execution mechanisms.

For FTTH networks, the principal opportunities concern optical fault diagnosis, subscriber service monitoring, bandwidth allocation, and automated recovery from configuration-related

failures. Recent research also highlights the continuing evolution of high-capacity passive optical networks, which increases the importance of effective management and monitoring (Rizzelli & Gaudino, 2026).

A common benchmarking methodology could help operators evaluate these management capabilities without overlooking differences between mobile and optical access infrastructure. Nevertheless, autonomy should remain subject to operational limitations. Physical fiber damage, hardware replacement, and certain security-sensitive operations may require direct human intervention.

The practical value of the framework therefore lies in identifying which operational tasks can be managed autonomously, whether those tasks are completed successfully, and how autonomous decisions affect service reliability and resource efficiency.

Application Roadmap for Saudi Vision 2030

The framework is relevant to Saudi Vision 2030 as a prospective operational measurement protocol, not as evidence of achieved national autonomy. National connectivity summaries from the Communications, Space and Technology Commission (CST, 2025) may establish service-access context, but only authorized operator telemetry and independently verified incident records can support autonomy scores. Saudi Vision 2030 (2025) provides policy context, not a network-event dataset.

For an operator pilot, candidate evidence sources include RAN performance-management counters, core-network analytics where available, OLT/ONU alarms and optical telemetry, trouble-ticket systems, change-management and approval records, independent end-to-end probes, and separately metered controller energy. Access would require operator agreements, data minimization, compatible clocks, vendor mappings, and compliance checks. Public CST reports cannot substitute for protected controller, subscriber, or maintenance records.

Implementation would proceed through an operator-approved inventory and interface audit; a shared event and eligibility dictionary; baseline collection for manual and rule-based operations; shadow-mode scoring without automatic actuation; controlled deployment of an independently tested AI controller; and periodic verification of safety, restoration, resource cost, and regional coverage. Table 7 maps infrastructure priorities to feasible evidence sources and limitations, without attributing any synthetic differences to Saudi networks.

Table 7. Mapping of proposed autonomy benchmarking indicators to Saudi Vision 2030 digital infrastructure priorities

Saudi infrastructure priority	Candidate authorized evidence source	Operator implementation and limitation
Reliable digital connectivity	Operator incident logs; independent availability probes	Baseline and verify service recovery by site; cross-check maintenance tickets; private operator data are not publicly accessible.
Operational efficiency	NOC ticket timestamps; human approvals; change records	Map intervention decisions to event IDs; audit staffing and reporting coverage; work practices differ by carrier.
Advanced digital services	RAN/5GC PM counters; OLT/ONU counters; independent service probes	Compare latency, throughput and SLA attainment within service classes; avoid direct raw 5G-versus-FTTH KPI comparisons.
Sustainable infrastructure	Metered RAN/OLT/ONU energy; metered controller energy; traffic counters	Measure energy per traffic and Wh/controller event separately; public energy estimates cannot substitute for metering.
Geographic service reliability	Region-coded incident records and authorized site inventory	Stratify sample by urban and non-urban access sites; nonrepresentative access precludes nationwide inference.
Digital infrastructure governance	Policy-engine logs; change authorizations; rollback records	Verify attempted blocks and executed violations against signed logs; proprietary and sensitive records require approval.

The mapping should be used only after the operator specifies the affected service class, measurement window, network sites, privacy basis, and comparability rules. Regional

disparities cannot be inferred without geographically representative data. Outcomes from one carrier, vendor, or metropolitan testbed would not establish nationwide performance.

The current analysis provides no direct Saudi operator evidence. Data-sharing restrictions, commercial confidentiality, proprietary telemetry fields, and uneven access to rural or remote-site logs may constrain a future validation study and must be disclosed in any Vision 2030 interpretation.

Practical Deployment Considerations

Implementing the proposed framework in operational telecommunications environments would require reliable telemetry, accessible management interfaces, consistent equipment inventories, and clearly defined control permissions. Differences among equipment vendors and legacy management platforms may complicate the collection of standardized event records, particularly in heterogeneous 5G and FTTH deployments.

Data quality represents another important consideration. Incomplete alarms, inaccurate timestamps, and inconsistent service records can affect fault diagnosis, action verification, and autonomy scoring. The framework therefore requires data validation procedures and explicit treatment of events with insufficient evidence. These concerns are consistent with research emphasizing dependable monitoring and data processing in intelligent optical networks (Rafique & Velasco, 2018; Musumeci et al., 2019).

Operational safety must also remain central to deployment. AI-enabled controllers should operate within predefined authorization boundaries, maintain auditable decision histories, and provide rollback or escalation mechanisms when service objectives cannot be achieved. Before deployment, operators would need to establish acceptable risk thresholds, assess model performance under unfamiliar conditions, and verify compatibility with existing network management systems.

A staged deployment would first validate event identifiers, clocks, and policy logs, then collect manual and rules baselines, evaluate AI decisions in shadow mode, and finally authorize limited closed-loop actions after predeclared safety checks. Fault eligibility and unresolved-event reporting should be audited by an independent operations team. If controller energy rises, any operational benefit must be assessed net of compute overhead and service risk.

These steps position AAS as a potentially useful governance and research instrument rather than a confirmed index of Saudi infrastructure improvement. Its adoption depends on

interoperability, independently verified measurements, weight validation, and an appropriate security and regulatory review.

Limitations and Threats to Validity

The contribution of this paper is a scoring methodology and an illustrative synthetic calculation, not an implemented AI comparison. Its limitations include construct overlap, subjective weighting, simulated event eligibility, unknown empirical calibration, and restricted policy transferability.

Internal validity is limited because the generator assigns B2 a lower base human-intervention probability (0.15 versus 0.48 for B1) and a higher base resolution probability (0.975 versus 0.95). Those assumptions produce part of the apparent performance advantage and prevent causal or efficacy claims. Five trial labels are generated batch groupings, not independent simulator experiments.

Construct validity requires particular caution because overall completion and completion under disturbances reuse related successes. In the synthetic scenario cells their correlation is approximately 0.978. Alternative weights change numerical results markedly, especially where audit completeness contributes without independent action. The raw AAS therefore requires separate component reporting and safety/minimum-autonomy qualification; its relationship to TM Forum maturity levels has not been validated.

External validity remains untested. No ns-3 experiment, calibrated FTTH model, deployed AI controller, operator telemetry, or Saudi field study was performed. Scenario probabilities, latency thresholds, energy values, and service durations are illustrative assumptions and may not resemble vendor-specific 5G or PON behavior.

Reproducibility of the generator does not resolve measurement validity. Actual applications will require aligned clocks, vendor-specific telemetry translation, independent service checks, traceable policy changes, and predefined rules for missing data and physical repairs. Commercial and national data-access restrictions may also prevent representative geographic sampling.

Model limitations cannot be estimated because B2 is a probabilistic proxy rather than a trained model. Future research must test calibration, training-test separation, drift, confidence-based

abstention, and resilience to rare disturbances. Similarly, the generated zero policy-violation count is a property of the scripted safeguard, not evidence of deployed safety.

Controller energy was assumed to increase under B2 and infrastructure energy was not metered. Consequently, sustainability implications are unresolved. The proposed framework is therefore limited to a reproducible methodological demonstration until its metrics, controller mechanisms, and operational consequences are independently validated.

Conclusion and Future Research

This article specifies a technology-aware framework for evaluating operational autonomy in 5G and FTTH networks through verified closed-loop completion, human independence, disturbance robustness, and audit completeness. Its proposed AAS complements established management standards while keeping service performance, energy, and safety requirements visible as separate outcomes.

The accompanying 4,320 synthetic records demonstrate the mechanics of paired event generation, metric calculation, graphical reporting, and weight sensitivity. They do not test actual AI control. The apparent advantage of B2_AI_proxy is partly programmed into intervention and resolution probabilities, and the overlap between completion and robustness requires further construct testing.

The main contribution is thus a transparent evaluation specification, not empirical confirmation of faster restoration or higher autonomy. A manual configuration can receive audit-based index points, so autonomous qualification must depend on nonzero completion and independence plus mandatory safety checks rather than a raw score threshold alone.

Future work should implement the prespecified 5G-LENA and XGS-PON management environments, deploy and calibrate a trained controller under the same action permissions as rule-based automation, and collect independently verified service and power data. Pilot experiments should determine sample size and assess uncertainty, scoring weights, correlated dimensions, unresolved incidents, and model drift. After technical validation, authorized multi-vendor Saudi operator trials could evaluate geographic and policy relevance.

Until such evidence is available, the framework is best used for designing and auditing benchmarking studies, with all numerical examples explicitly identified as synthetic and no claims of national infrastructure improvement.

References

3rd Generation Partnership Project. (n.d.). Architecture enhancements for 5G System to support network data analytics services (3GPP TS 23.288). https://www.3gpp.org/DynaReport/23288.htm

Afolabi, I., Taleb, T., Samdanis, K., Ksentini, A., & Flinck, H. (2018). Network slicing and softwarization: A survey on principles, enabling technologies, and solutions. IEEE Communications Surveys & Tutorials, 20(3), 2429–2453. https://doi.org/10.1109/COMST.2018.2815638

Amirabadi, M. A., Nezamalhosseini, S. A., Kahaei, M. H., & Chen, L. R. (2025). A comprehensive survey on machine and deep learning for optical communications. IEEE Access, 13, 88794–88846. https://doi.org/10.1109/ACCESS.2025.3569559

Arnaz, A., Lipman, J., Abolhasan, M., & Hiltunen, M. A. (2022). Toward integrating intelligence and programmability in open radio access networks: A comprehensive survey. IEEE Access, 10, 67747–67770. https://doi.org/10.1109/ACCESS.2022.3183989

Asiri, A., & Wang, B. (2025). Deep reinforcement learning for QoT-aware routing, modulation, and spectrum assignment in elastic optical networks. Journal of Lightwave Technology, 43(1), 42–60. https://doi.org/10.1109/JLT.2024.3446762

Bega, D., Gramaglia, M., Fiore, M., Banchs, A., & Costa-Perez, X. (2020). DeepCog: Optimizing resource provisioning in network slicing with AI-based capacity forecasting. IEEE Journal on Selected Areas in Communications, 38(2), 361–376. https://doi.org/10.1109/JSAC.2019.2959245

Boutaba, R., Salahuddin, M. A., Limam, N., Ayoubi, S., Shahriar, N., Estrada-Solano, F., & Caicedo, O. M. (2018). A comprehensive survey on machine learning for networking: Evolution, applications and research opportunities. Journal of Internet Services and Applications, 9, 16. https://doi.org/10.1186/s13174-018-0087-2

Broadband Forum. (n.d.). Technical library: Access network and user services management specifications (TR-383; TR-369). https://www.broadband-forum.org/technical-library/

Chen, X., Li, B., Proietti, R., Lu, H., Zhu, Z., & Yoo, S. J. B. (2019). DeepRMSA: A deep reinforcement learning framework for routing, modulation and spectrum assignment in

elastic optical networks. Journal of Lightwave Technology, 37(16), 4155–4163. https://doi.org/10.1109/JLT.2019.2923615

Chen, X., Proietti, R., & Yoo, S. J. B. (2019). Building autonomic elastic optical networks with deep reinforcement learning. IEEE Communications Magazine, 57(10), 20–26. https://doi.org/10.1109/MCOM.001.1900151

Communications, Space & Technology Commission. (2025). Saudi internet report 2024. https://www.cst.gov.sa/en/knowledge-center/reports/saudi-internet-24

Cruzes, S. (2026). Telemetry and agentic AI: Foundations for optical network automation. IEEE Access, 14, 8800–8838. https://doi.org/10.1109/ACCESS.2025.3649768

European Telecommunications Standards Institute. (2019). Zero-touch network and service management: Reference architecture (ETSI GS ZSM 002). https://www.etsi.org/technical-groups/zsm/

Foukas, X., Patounas, G., Elmokashfi, A., & Marina, M. K. (2017). Network slicing in 5G: Survey and challenges. IEEE Communications Magazine, 55(5). https://doi.org/10.1109/MCOM.2017.1600951

Giordani, M., Polese, M., Mezzavilla, M., Rangan, S., & Zorzi, M. (2020). Toward 6G networks: Use cases and technologies. IEEE Communications Magazine, 58(3), 55–61. https://doi.org/10.1109/MCOM.001.1900411

Gu, R., Yang, Z., & Ji, Y. (2020). Machine learning for intelligent optical networks: A comprehensive survey. Journal of Network and Computer Applications, 157, Article 102576. https://doi.org/10.1016/j.jnca.2020.102576

International Telecommunication Union. (n.d.). 10-gigabit-capable symmetric passive optical network (XGS-PON) systems (Recommendation ITU-T G.9807.1). https://www.itu.int/rec/T-REC-G.9807.1/en

Klaine, P. V., Imran, M. A., Onireti, O., & Souza, R. D. (2017). A survey of machine learning techniques applied to self-organizing cellular networks. IEEE Communications Surveys & Tutorials, 19(4), 2392–2431. https://doi.org/10.1109/COMST.2017.2727878

Liyanage, M., Pham, Q.-V., Dev, K., Bhattacharya, S., Maddikunta, P. K. R., Gadekallu, T. R., & Yenduri, G. (2022). A survey on zero touch network and service management (ZSM) for

5G and beyond networks. Journal of Network and Computer Applications, 203, Article 103362. https://doi.org/10.1016/j.jnca.2022.103362

Luong, N. C., Hoang, D. T., Gong, S., Niyato, D., Wang, P., Liang, Y.-C., & Kim, D. I. (2019). Applications of deep reinforcement learning in communications and networking: A survey. IEEE Communications Surveys & Tutorials, 21(4), 3133–3174. https://doi.org/10.1109/COMST.2019.2916583

Mrabet, H. (2022). A survey of applied machine learning techniques for optical orthogonal frequency division multiplexing based networks. Transactions on Emerging Telecommunications Technologies, 33(4), e4400. https://doi.org/10.1002/ett.4400

Musumeci, F., Rottondi, C., Nag, A., Macaluso, I., Zibar, D., Ruffini, M., & Tornatore, M. (2019). An overview on application of machine learning techniques in optical networks. IEEE Communications Surveys & Tutorials, 21(2), 1383–1408. https://doi.org/10.1109/COMST.2018.2880039

Nevin, J. W., Nallaperuma, S., Shevchenko, N. A., Li, X., Faruk, M. S., & Savory, S. J. (2021). Machine learning for optical fiber communication systems: An introduction and overview. APL Photonics, 6, Article 121101. https://doi.org/10.1063/5.0070838

Ordonez-Lucena, J. A., Ameigeiras, P., Lopez, D. R., Ramos-Munoz, J. J., Lorca, J., & Folgueira, J. (2017). Network slicing for 5G with SDN/NFV: Concepts, architectures, and challenges. IEEE Communications Magazine, 55(5), 80–87. https://doi.org/10.1109/MCOM.2017.1600935

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://doi.org/10.1109/COMST.2023.3239220

Raca, D., Zahran, A. H., Sreenan, C. J., Sinha, R. K., Halepovic, E., Jana, R., & Gopalakrishnan, V. (2020). On leveraging machine and deep learning for throughput prediction in cellular networks: Design, performance, and challenges. IEEE Communications Magazine, 58(3). https://doi.org/10.1109/MCOM.001.1900394

Rafique, D., & Velasco, L. (2018). Machine learning for network automation: Overview, architecture, and applications. Journal of Optical Communications and Networking, 10(10), D126–D143. https://doi.org/10.1364/JOCN.10.00D126

Raza, S. M., Minerva, R., Crespi, N., Alvi, M., Herath, M., & Dutta, H. (2025). A comprehensive survey of network digital twin architecture, capabilities, challenges, and requirements for edge–cloud continuum. Computer Communications, 236, Article 108144. https://doi.org/10.1016/j.comcom.2025.108144

Rizzelli, G., & Gaudino, R. (2026). High-speed PON solutions: Recent evolutions and expected trends. Journal of Optical Communications and Networking, 18(2), A134–A149. https://doi.org/10.1364/JOCN.575601

Saudi Vision 2030. (2025). Vision 2030 annual report 2024. https://www.vision2030.gov.sa/media/r3ij40wu/en-annual-report-vision2030-2024.pdf

Shariati, B., Ruiz, M., Comellas, J., & Velasco, L. (2019). Learning from the optical spectrum: Failure detection and identification. Journal of Lightwave Technology, 37(2), 433–440. https://doi.org/10.1109/JLT.2018.2859199

Thangappan, T., Therese, B., Suvarnamma, A., & Sai Swapna, G. (2020). Review on dynamic bandwidth allocation of GPON and EPON. Journal of Electronic Science and Technology, 18(4), Article 100044. https://doi.org/10.1016/j.jnlest.2020.100044

TM Forum. (2023). Autonomous network levels evaluation methodology (IG1252, Version 1.2.0). https://www.tmforum.org/resources/how-to-guide/ig1252-autonomous-network-levels-evaluation-methodology-v1-2-0/

Tzanakaki, A., Anastasopoulos, M. P., & Simeonidou, D. (2019). Converged optical, wireless, and data center network infrastructures for 5G services. Journal of Optical Communications and Networking, 11, A111–A122. https://doi.org/10.1364/JOCN.11.00A111

Wang, Z., Sun, Y., & Yuan, S. (2022). Intelligent radio access networks: Architectures, key techniques, and experimental platforms. Frontiers of Information Technology & Electronic Engineering, 23, 5–18. https://doi.org/10.1631/FITEE.2100305

Xiong, Z., Zhang, Y., Niyato, D., Deng, R., Wang, P., & Wang, L.-C. (2019). Deep reinforcement learning for mobile 5G and beyond: Fundamentals, applications, and challenges. IEEE Vehicular Technology Magazine, 14(2), 44–52. https://doi.org/10.1109/MVT.2019.2903655

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://doi.org/10.1109/COMST.2019.2904897