Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Human-AI Collaborative Frameworks for Intelligent Fault Detection, Proactive Quality Management, and Augmented Decision Support in Fixed Wireless Access Networks

Venkata Raghavendra Vaishnav Vinjamuri[1]

[1]*Independent Researcher, USA*

ARTICLEINFO

ABSTRACT

Introduction: The explosive growth of Fixed Wireless Access (FWA) as a primary broadband delivery mechanism has created unprecedented operational challenges for network management teams serving millions of subscriber devices across complex 5G Radio Access Network environments. Traditional reactive operational paradigms, characterized by manual fault response, fragmented monitoring systems, and engineering teams overwhelmed by non-actionable alarms, are fundamentally unable to scale to the quality management demands of modern FWA deployment.Objectives: This article aims to design, deploy, and evaluate a human-AI collaborative framework that unifies fragmented cross-domain telemetry, predicts subscriber-level quality degradation before it becomes customer-visible, closes the loop on the most common class of 5G connectivity fault, and supports engineering judgement through a constrained generative AI interface, while explicitly allocating decision authority between human and AI agents according to risk. Methods: The framework integrates four complementary components: a unified Single Source of Truth cloud-native data platform consolidating consumer profile, in-home device telemetry, and RAN performance data into a coherent real-time analytical substrate; a Network Quality Experience Score (nQES) aggregating over thirty technical features to identify subscribers at elevated churn risk; a closed-loop Missing Anchor anomaly detection and automated remediation system integrated with Self-Organizing Network infrastructure; and a Generative AI natural language interface constrained to Feature Store-grounded inference. The framework was deployed and evaluated across a production FWA network supporting 1.7 million active subscriber devices over a 180-day post-deployment period. Results: The deployed framework achieved an 80% reduction in Mean Time to Resolution (from a median of 4.2 hours to 51 minutes), a 5 percentage point reduction in voluntary churn for nQES-identified intervention cohorts, an 80% reduction in non-actionable alarm volume (from 1,200 to approximately 240 alarms per engineer per shift), and 93.67% classification accuracy on prospective service qualification assessments, a 21 percentage point improvement over the prior rules-based system. Conclusions: These results demonstrate that explicit architectural attention to the division of responsibilities between human and AI agents, operationalized through a tiered automation architecture, produces operational outcomes that neither human teams nor fully automated systems achieve independently, with the greatest gains attributable to unifying the underlying data infrastructure rather than to any single AI model.

Keywords: Human-AI collaboration, Fixed Wireless Access, Network Quality Experience Score, Anomaly Detection, Generative AI, Tiered Automation

INTRODUCTION

Fixed Wireless Access has emerged as one of the most consequential and rapidly expanding segments of the global broadband market, offering high-speed internet connectivity to residential and small business subscribers through 5G or 4G LTE radio access infrastructure rather than physical cable or fiber connections. Global FWA connections surpassed 100 million active subscribers in 2023 and are projected to exceed 300 million by 2028, with 5G-capable deployments accounting for the majority of net additions across North America, Europe, and Asia-Pacific markets. Unlike mobile broadband services, which are architecturally designed to tolerate intermittent quality variation and

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

session discontinuity, FWA is provisioned as a committed, continuously active home broadband service where subscriber expectations of reliability, latency consistency, and throughput stability are directly comparable to those of fixed-line fiber and cable alternatives. This equivalence of customer expectation, combined with the absence of a physical last-mile connection that can be inspected and replaced, creates a quality management challenge of substantially greater operational complexity than either traditional mobile broadband or legacy fixed-line service environments.

The infrastructure economics underpinning 5G FWA deployment have been thoroughly characterized in the scholarly literature. Oughton and Frias (2018) examine the cost, coverage, and rollout implications of 5G infrastructure at a national scale, demonstrating that the capital expenditure required to achieve equivalent geographic coverage through 5G radio networks is highly sensitive to frequency band selection: millimeter-wave deployments above 24 GHz require cell site densities 8-12 times greater than sub-6 GHz deployments to achieve comparable coverage radii, while sub-6 GHz mid-band spectrum, the dominant frequency tier for current FWA deployments, enables cell radii of 1-3 kilometers in suburban environments at capital costs compatible with commercially viable service economics. This infrastructure geometry directly determines the operational complexity of FWA quality management: a single 5G sector serving FWA subscribers at a 2-kilometer radius may simultaneously serve 500-2,000 active CPE devices, each presenting a distinct combination of signal conditions, in-home network configurations, and usage patterns that together determine whether the available radio resource translates to a satisfactory service experience for each individual subscriber.

The operational challenge of managing FWA quality at this scale is compounded by the multi-domain structure of the data environment in which network operations teams must work. Each FWA subscriber device exists at the intersection of three technically distinct domains: the 5G radio connection between the Customer Premises Equipment unit and the serving cell sector, characterized by metrics including Reference Signal Received Power, spectral efficiency, and signal quality; the in-home wireless network created by the CPE to distribute connectivity to the subscriber's devices; and the subscriber's own usage behavior and device characteristics that determine whether the available connectivity translates to a satisfactory experience. Historically, each of these domains has resided in a separate, organizationally siloed monitoring system with incompatible data models, refresh rates measured in hours to days, and access mechanisms requiring specialist knowledge, a fragmentation that forces operations engineers to manually correlate data across multiple disconnected tools during fault diagnosis, directly inflating mean time to resolution and reducing the quality of diagnostic conclusions under time pressure.

Ayoubi et al. (2018) provide a rigorous treatment of machine learning applications for cognitive network management, establishing that the integration of real-time telemetry, ML-driven inference, and closed-loop automated remediation represents the defining operational architecture for 5G network management at population scale. Ayoubi et al. (2018) document that AI-augmented network operations systems consistently reduce fault detection latency by 60-85% and operational expenditure by 20-35% compared to legacy rule-based monitoring approaches across production deployments, while simultaneously improving the accuracy of root-cause attribution for complex, multi-layer service degradation events that exceed the pattern-matching capacity of threshold-based alarm systems. These performance differentials, achieved in production environments rather than laboratory conditions, establish the practical case for the human-AI collaborative framework presented in this article, which was designed, deployed, and evaluated across a production FWA network supporting over 1.7 million active subscriber devices.

This article presents a framework integrating four complementary components: a unified Single Source of Truth data platform eliminating cross-domain data fragmentation; a Network Quality Experience Score aggregating over thirty technical features into a composite churn-risk indicator; a closed-loop Missing Anchor anomaly detection and automated remediation system integrated with Self-Organizing Network infrastructure; and a Generative AI natural language interface providing engineers with synthesized diagnostic insights while preserving human judgment for consequential decisions. Production deployment of this framework achieved an 80% reduction in Mean Time to Resolution for network change events, a 5 percent reduction in churn for remediated customer segments, an 80% reduction in non-actionable alarm volume, and a 93.67% accuracy on prospective service qualification assessments. The article is organized as follows: Section 2 reviews the relevant scholarly literature; Sections 3 through 5 detail the

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

platform architecture, proactive intelligence components, and human-AI collaboration design principles, respectively; Section 6 presents the evaluation results; Section 7 discusses implications; Section 8 addresses broader societal dimensions; and Section 9 concludes with future directions.

OBJECTIVES

The specific objectives of this work are fourfold: (1) to eliminate the cross-domain data fragmentation that historically forced operations engineers to manually correlate consumer profile, in-home telemetry, and RAN performance data across incompatible monitoring systems; (2) to develop a composite, multi-domain predictive indicator capable of identifying subscribers at elevated risk of service-quality-driven churn before degradation becomes customer-visible; (3) to design a closed-loop anomaly detection and remediation pipeline for the most prevalent class of silent 5G-to-LTE fallback degradation, integrated with Self-Organizing Network infrastructure; and (4) to provide engineers with a generative AI decision-support interface that is constrained to structured, Feature Store-grounded inference so as to preserve operational trust while reducing diagnostic overhead.

A further objective is evaluative rather than architectural: to validate, through production deployment across a live FWA network supporting 1.7 million active subscriber devices, whether a tiered automation architecture that allocates decision authority between human engineers and AI agents according to action risk and reversibility can achieve operational outcomes, measured in Mean Time to Resolution, churn reduction, alarm volume, and service qualification accuracy, that exceed what either fully manual or fully automated approaches achieve independently, and to examine how the boundary between autonomous and human-mediated action can be governed as an evidence-driven parameter rather than a fixed design constant.

METHODS

3.1 Cloud-Native Data Platform Design and Real-Time Pipeline Engineering

The human-AI collaborative approach is enabled by a cloud-native unified data platform that integrates customarily siloed consumer profile information, in-home device telemetry, and RAN performance data into an analytics substrate layer that services over 1.7 million active FWA subscriber devices. Prior to the existence of this cloud-based data platform, no fewer than seven point solutions were required for network performance or service monitoring, each operating independently with incompatible formats, vendor-specific access methods, and refresh intervals ranging from 15-minute polling for RAN counters to 24-hour batch rollups of subscriber account attributes. One consequence of this fragmentation was increased Mean Time to Resolution, with engineers spending an estimated 60-70% of their time assembling and correlating data rather than reasoning and making decisions, precisely where human judgment is most necessary.

Pajo et al. (2025) propose a hybrid architecture that alleviates data fragmentation and orchestrates cross-domain services over heterogeneous infrastructures, composed of a Cross-domain Service Orchestrator (CDSO) built on ETSI's Open Source MANO (OSM) and a Requests Handler module serving as the northbound integration point for domain-specific Service Orchestrators. The authors show that this star topology reduces connection complexity from the n(n-1)/2 growth of a fully meshed architecture to linear growth in n. This principle drives the design of the Single Source of Truth (SSOT) platform described here, which is based on a governed semantic layer across all data domains, preventing a quadratic increase in the human effort required to correlate data across domains.

The key architectural enabler of predictive and autonomous operation was the shift from 24-hour batch processing to near-real-time hourly data through Real-Time Transformation and Event-Driven Loading pipeline components. A model operating with a 24-hour data lag misses rapidly developing quality degradation events entirely: a device experiencing progressive 5G anchor degradation over a four-hour period generates a resolvable support call well before a daily batch refresh would flag it as at-risk. Reducing the data refresh lag to one hour, a 24-fold improvement, shifts the platform's predictive horizon from retrospective incident reporting to prospective intervention, enabling the nQES scoring system described below to identify at-risk devices before degradation becomes customer-visible. Gill (2022) identifies near-real-time data pipeline architecture as a foundational requirement for AI systems operating in dynamic environments, demonstrating that inference models consuming stale features experience

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

accuracy degradation proportional to the mismatch between data temporal resolution and the rate of change of the underlying phenomenon being modeled.

3.2 Semantic Layer Design and Cross-Domain Ontology

The Semantic Layer implements a formal ontology of telecommunications concepts, network elements, subscriber sessions, quality events, device states, and their interrelationships, expressed as a shared data model that all analytical applications and machine learning models consume through a consistent query interface. This abstraction eliminates the metric inconsistency that arises when different engineering teams query the same underlying phenomenon through slightly different formulations of raw database tables: prior to the Semantic Layer's introduction, three separate teams were computing “sector utilization” using subtly different denominator definitions, producing values that differed by up to 12 percentage points for the same sector at the same timestamp and generating conflicting conclusions that required manual reconciliation. The Semantic Layer resolves this by defining each metric precisely once, with a governed specification that propagates automatically to all downstream consumers.

For cross-domain analytical queries, for example correlating a subscriber's in-home Wi-Fi interference level with the concurrent load utilization of their serving RAN sector, the Semantic Layer provides traversal paths between entity types that span source system boundaries without requiring the analyst to understand the underlying physical join conditions. This reduces the time required to construct a novel cross-domain diagnostic query from several hours of schema investigation to minutes of Semantic Layer navigation, directly expanding the range of hypotheses operations engineers can investigate within the time constraints of an active incident.

3.3 Feature Store Architecture and Low-Latency Model Serving

The Feature Store component precomputes the derived features required by the platform's machine learning models, composite quality indicators, rolling statistical aggregations over 1-hour, 6-hour, and 24-hour windows, and cross-domain correlation features linking RAN sector conditions to individual device experience metrics on each hourly refresh cycle, serving them at sub-10-millisecond latency to model inference endpoints. This architecture is essential for the nQES model, which requires 30+ derived features per device per inference cycle: computing these features at query time for 1.7 million devices would impose a computational overhead incompatible with the sub-minute scoring latency required for real-time anomaly detection.

Kaur et al. (2018) establish the architectural principles for software-defined edge-cloud interplay that directly inform the Feature Store's deployment topology: computationally intensive feature aggregations, particularly rolling-window statistics over raw RAN counter streams, are computed at edge nodes co-located with regional data collection infrastructure, while the centralized cloud layer hosts the Feature Store's serving API and model inference endpoints. This edge-cloud partitioning reduces the volume of raw telemetry transferred to the central platform by approximately 78%, while ensuring the central inference layer receives fully prepared model inputs without incurring the network latency cost of transferring and aggregating raw data centrally.

4.1 Network Quality Experience Score

The Network Quality Experience Score (nQES) is a composite predictive indicator that aggregates over thirty granular technical features drawn from all three unified data domains into a single normalized risk score reflecting each subscriber device's probability of experiencing service quality degradation within the subsequent scoring window. A device may present strong Reference Signal Received Power, indicating healthy radio connectivity, while simultaneously experiencing degraded in-home Wi-Fi distribution or elevated CPE operating temperature, conditions that collectively produce a poor service experience despite a superficially healthy radio link. Ahmad et al. (2019) demonstrate in a large-scale telecom churn prediction study that composite feature models incorporating cross-domain behavioral, network, and account attributes achieve Area Under the Curve scores of 0.91-0.94, compared to 0.72-0.78 for single-domain models, a performance differential that directly motivates the multi-domain feature design of the nQES.

The nQES is formally expressed as a weighted aggregation of normalized feature inputs passed through a non-linear scoring function: $nQES_i = f(\Sigma_{j=1}^{N} w_j \cdot \varphi_j(x_{ij}))$, with $nQES_i \in [0, 100]$, where x_{ij} denotes the jth

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

feature value for device i, φ_j(·) is a normalization transform mapping each feature to a common scale, w_j is the learned feature weight, and f(·) is a sigmoid-family activation function bounding the composite output to a 0-100 interpretable range. The model was trained against ground-truth customer experience outcomes, support call generation, voluntary service cancellation, and customer satisfaction survey responses, across a labeled dataset of 4.2 million device-month observations. Deployed as a continuous monitoring signal refreshed on each hourly cycle, the nQES enables proactive outreach for devices scoring below a validated risk threshold before degradation becomes severe enough to generate a support contact.

Table 1. nQES Feature Set by Data Domain

Domain	Feature Category	Key Features	Predictive Role
RAN Performance	Radio frequency	RSRP, RSRQ, SINR, spectral efficiency	Core signal quality indicators: low RSRP + high sector load is strongest single-domain churn predictor
RAN Performance	Capacity	Sector utilization, capacity headroom, congestion index	Identifies capacity-driven degradation invisible to per-device metrics
RAN Performance	Handover stability	Handover frequency, ping-pong rate, anchor stability score	Elevated handover rates precede Missing Anchor fallback events
In-Home Telemetry	CPE hardware	CPU, memory, temperature (°C), uptime	Temperature >75°C + high sector load raises churn probability 2.3×
In-Home Telemetry	Connectivity	Interruption frequency, mean duration, protocol error rate	Captures customer-visible breaks absent from RAN metrics
In-Home Telemetry	Wi-Fi distribution	Client count, interference index, per-client bandwidth	Distinguishes in-home failures from RAN-originated degradation
Consumer Profile	Account attributes	Tenure, service tier, prior support contacts	Tenure <6 months + ≥2 contacts identifies highest-risk cohort
Consumer Profile	Usage behaviour	Peak throughput, baseline usage, consistency index	Deviation from subscriber baseline serves as early degradation signal
Consumer Profile	Geographic context	Sector assignment, distance-to-tower, infrastructure change events	Correlates device risk with sector-level operational events

Note: 30+ features trained on 4.2 million device-month observations. RSRP = Reference Signal Received Power; RSRQ = Reference Signal Received Quality; SINR = Signal-to-Interference-plus-Noise Ratio; CPE = Customer Premises Equipment.

4.2 Missing Anchor Detection and Closed-Loop Automated Remediation

The Missing Anchor anomaly detection system targets a common, highly disruptive failure mode in 5G FWA networks in which devices silently revert from optimum 5G Non-Standalone (NSA) operation to sub-optimal 4G LTE-only operation due to 5G anchor degradation or misconfiguration. Devices remain connected and monitored, log no fault alarm, but are served at 40-70% of their baseline 5G throughput, a class of degraded device that is structurally invisible to standard threshold-based monitoring tools.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Dimou and Noubir (2025) present ARGOS, an O-RAN-compliant intrusion detection system deployed in the Near Real-Time RAN Intelligent Controller to detect rogue base station downgrade attacks using unsupervised models on cross-layer features from Modem Layer 1 logs and Measurement Reports from commercial off-the-shelf user equipment. Evaluating autoencoders, denoising autoencoders, variational autoencoders, and isolation forests trained only on benign observations, they find that variational autoencoders achieve the best accuracy-efficiency trade-off, 99.5% accuracy with a 0.6% false positive rate. This methodological approach directly informs the Missing Anchor detection pipeline, which mirrors ARGOS's three-stage architecture.

The automated pipeline runs an ensemble anomaly classifier on the candidate feature stream in the Feature Store to detect the statistical signature of Missing Anchor fallback, degraded throughput, absence of 5G-specific protocol indicators, and geographic proximity to sectors with known anchor instability. At the diagnosis stage, the suspected device location is checked against sector-level network management context to confirm the malfunction matches the network state and to prevent false-positive escalations. At the confirmed diagnosis stage, the Self-Organizing Network management system autonomously reconfigures the 5G anchor to restore service end-to-end without human intervention, with a median automated response time of four minutes. Klaine et al. (2017) show that closed-loop ML-based reconfiguration systems of this kind converge to fault recovery 55-80% faster than comparable manual workflows; Kaur et al. (2018) describe the edge-cloud deployment pattern in which anomaly detection runs on regional edge nodes while SON reconfiguration commands are issued from the centralized cloud control layer, yielding a sub-minute end-to-end response latency across the geographically distributed 1.7-million-device subscriber base.

5.1 Generative AI for Augmented Network Operations

The generative AI components occupy the direct human-AI collaboration interface layer, the point at which the platform's analytical capabilities are translated into operationally actionable guidance. A conversational diagnostic assistant enables engineers to query the platform's full cross-domain analytical capabilities through natural language, retrieving affected device nQES scores, serving sector utilization trends, recent RAN configuration change events, and comparative throughput distributions as a synthesized, contextually coherent analysis. Internal deployment measurements indicate that cross-domain diagnostic queries previously requiring 45-90 minutes of manual data correlation are completed through the conversational interface in under eight minutes, a reduction in diagnostic overhead exceeding 80% for the most complex investigative tasks.

A second component applies generative AI to the ticket management workflow, automatically clustering incoming support contacts by underlying network condition category, correlating clusters with nQES patterns and sector performance metrics, and generating natural language root cause summaries with recommended resolution approaches. The Broken Trend Optimization module, which applies anomaly detection to network performance trend sequences to suppress non-actionable alarm progressions, reduced non-actionable alarm volume by approximately 80% in production, from a pre-deployment mean of 1,200 alarm notifications per engineer per shift to approximately 240, directly reducing the cognitive overload that Lee and See (2004) identify as a primary driver of automation disuse in high-alarm-rate environments. The generative AI interface is deliberately constrained to inferences grounded in the structured performance data held within the Feature Store rather than parametric knowledge, consistent with Shu et al.'s (2021) finding that fact-grounded generation systems outperform unconstrained generation by 23 percentage points in factual consistency on structured knowledge synthesis tasks.

5.2 Tiered Automation Architecture and Human Oversight Preservation

Effective human-AI collaboration requires explicit, principled decisions about which AI-generated recommendations execute autonomously versus which require human review. Hemmer et al. (2021) establish that human-AI complementarity is most reliably achieved when task allocation matches respective competence profiles: AI systems should handle high-volume, pattern-recognition-intensive subtasks, while humans retain authority over ambiguous, high-stakes, and contextually novel decisions. This principle is operationalized through a three-tier automation architecture classifying all potential automated actions by risk profile, summarized in Table 2.

Tier One actions, those with well-characterized, reversible, single-device impact profiles, execute autonomously without engineer approval; the Missing Anchor remediation pipeline is the canonical example. Tier Two actions,

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

affecting shared sector-level infrastructure or carrying meaningful uncertainty about impact scope, are automatically prepared with AI-generated supporting evidence and projected impact assessments but require explicit engineer approval. Tier Three actions, software rollbacks, cross-sector parameter changes, or any action whose impact model indicates it could affect more than 10,000 subscriber devices at once, undergo full engineering team review with mandatory stakeholder notification and documented rollback procedures. Amershi et al. (2019) establish that systems should make their uncertainty explicit and support efficient human correction near the boundary of reliable competence, a principle instantiated in the Tier Two workflow, where the Service Qualification model operates at 93.67% classification accuracy, a 21-percentage-point improvement over the prior rules-based system, while explicitly surfacing its confidence score so representatives retain full authority to override its recommendation where local context renders the inference incorrect.

Table 2. Tiered Automation Architecture: Decision Criteria and Outcomes

Dimension	Tier 1 - Autonomous	Tier 2 - Human-Approved	Tier 3 - Team Review
Risk profile	Single-device; fully reversible	Shared infrastructure; moderate uncertainty	Potential widespread disruption (>10,000 devices)
Human role	None; AI executes automatically	Engineer reviews and approves before execution	Full team review; stakeholder notification; rollback plan required
AI confidence	>95%	80-95%	<80% or novel failure mode
Decision latency	Median 4 minutes	15-45 minutes	2-8 hours
Example action	Missing Anchor 5G reconfiguration via SON	Sector parameter adjustment; service qualification	Cross-sector software rollback; regional infrastructure change
Monthly volume	~42,000 remediations	~1,800 recommendations	~35 major events
Key outcome	80% MTTR reduction; 5% churn reduction	93.67% qualification accuracy (+21 pp vs prior system)	4.1/5.0 engineer satisfaction with AI scenario analysis

Note: Latency and volume figures from 180-day post-deployment evaluation. MTTR = Mean Time to Resolution; SON = Self-Organizing Network; pp = percentage points.

RESULTS

The production deployment of the human-AI collaborative framework across a live FWA network supporting 1.7 million active subscriber devices generated four primary measurable outcomes that collectively validate the architectural design principles articulated above. Each outcome is reported against a pre-deployment baseline established over a 90-day observation window immediately preceding full platform activation, with post-deployment measurements drawn from a subsequent 180-day production period under comparable network load and subscriber volume conditions.

The most operationally significant outcome was an 80% reduction in Mean Time to Resolution for network change events, from a pre-deployment baseline median MTTR of 4.2 hours to a post-deployment median of 51 minutes. This improvement is attributable to two compounding mechanisms: the elimination of manual cross-system data

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

correlation through the unified SSOT platform, which reduced the diagnostic data assembly phase of incident investigation from a mean of 68 minutes to under 9 minutes, and the automated Missing Anchor remediation pipeline, which resolves the most prevalent class of 5G connectivity degradation events without entering the human-mediated incident queue at all. Soldani and Manzalini (2015) establish MTTR reduction as the primary efficiency metric for next-generation network operations systems, arguing that the transition from reactive to proactive operational paradigms requires end-to-end incident lifecycle automation across detection, diagnosis, and remediation stages, precisely the three-stage pipeline validated by these results.

The nQES-driven proactive intervention program achieved a five percentage point reduction in voluntary churn for customer segments identified as at-risk and contacted before service degradation generated a support call. This is an absolute reduction in the 90-day voluntary cancellation rate, the difference between the intervention cohort's churn rate and that of a matched control group, rather than a relative reduction proportional to baseline churn. Across the deployment's subscriber base, a one percentage point reduction in monthly churn corresponds to approximately 17,000 retained subscribers per month at the 1.7 million device scale, making the five percentage point churn reduction a commercially substantial outcome independent of its operational significance. The churn reduction is measured as the difference in 90-day voluntary cancellation rates between the nQES-identified intervention cohort and a matched control group of comparable-risk devices that did not receive proactive outreach during the evaluation period, providing a controlled estimate of the model's causal impact rather than a simple pre-post comparison susceptible to confounding seasonal trends.

The Broken Trend Optimization module's 80% reduction in non-actionable alarms, from a pre-deployment mean of 1,200 alarm notifications per engineer per shift to approximately 240, produced measurable downstream effects on diagnostic accuracy as well as workload. Post-deployment incident post-mortems indicate that engineer miss rates for genuine high-priority alerts decreased from 14% to 3% following alarm volume reduction, consistent with the signal detection theory prediction that reducing irrelevant stimulus volume improves detection sensitivity for target signals at constant response threshold. The Service Qualification model's 93.67% classification accuracy on prospective subscriber location assessments, validated against a held-out test set of 84,000 location evaluations, represents a 21 percentage point improvement over the 72.4% accuracy of the prior rules-based qualification system it replaced, directly reducing both failed installations and declined qualified locations.

The evaluation methodology acknowledges three limitations that constrain causal inference. First, the deployment was not randomized at the subscriber level; practical operational constraints precluded withholding platform capabilities from a randomly assigned control group at scale, meaning that the pre-post comparison for MTTR and alarm volume outcomes may partially reflect concurrent network infrastructure improvements rather than platform effects exclusively. Second, the churn reduction estimate relies on propensity-score matched controls rather than random assignment, introducing residual confounding risk from unobserved subscriber attributes correlated with both nQES risk scores and churn propensity. Third, the 180-day evaluation window, while substantial, does not capture seasonal variation in FWA network load patterns that may affect metric stability across annual cycles. These limitations are consistent with the inherent constraints of evaluating operational AI systems in live production environments and are noted to support appropriate interpretation of the reported outcomes rather than to diminish their practical significance.

DISCUSSION

The operational results are interpretable within the theoretical frameworks reviewed above, and their pattern illuminates which design principles most directly drive performance improvement in AI-augmented network operations. The 80% MTTR reduction is most directly attributable to the data unification investment, the SSOT platform, rather than to the AI inference components layered above it, suggesting that the primary barrier to effective human-AI collaboration in network operations is not algorithmic capability but data infrastructure readiness. This finding aligns with Rafique and Velasco's (2018) identification of data fragmentation as the primary deployment barrier for ML-based network automation: the AI systems in this framework perform at their documented levels because they operate on a coherent, real-time, cross-domain feature substrate that prior operational architectures did not provide.

Journal of Information Systems Engineering and Management
2026, 11(3s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The five percent churn reduction for nQES-triggered interventions validates the composite multi-domain scoring approach over single-metric threshold monitoring, consistent with Ahmad et al.'s (2019) demonstration that cross-domain feature models outperform single-domain models by 13-22 AUC points in telecommunications churn prediction. More theoretically significant is the interaction between the nQES and the tiered automation architecture: the nQES functions as a Tier Two decision support tool, generating high-confidence risk stratifications that human operations representatives act upon, rather than as a Tier One autonomous action trigger, reflecting Hemmer et al.'s (2021) principle that human-AI complementarity is maximized when human judgment is retained for decisions carrying customer relationship implications, while AI autonomy is reserved for technically bounded, reversible reconfigurations.

The tiered automation architecture's most practically important finding is that the boundary between autonomous and human-mediated action is not static: as operational confidence in the Missing Anchor detection pipeline accumulated, measured by a false positive rate that declined from 8.3% in the first deployment month to 2.1% by month six, the appropriate Tier One action boundary expanded, allowing a broader class of 5G anchor reconfiguration variants to be handled autonomously without increasing operational risk. This adaptive autonomy property is identified by Seeber et al. (2020) as a defining characteristic of mature human-AI teaming systems, distinguishing them from brittle automation that requires manual re-engineering to adjust its operational scope as deployment experience matures.

Two limitations warrant explicit scholarly attention. The nQES model's training on historical churn and support call outcomes introduces the risk of encoding systematic biases present in historical service patterns; if prior reactive operational practices resulted in faster resolution for certain subscriber segments, the model may learn spurious correlations between segment membership and quality risk that do not reflect genuine network conditions. Ongoing monitoring of nQES score distributions across demographic and geographic subscriber segments is a necessary governance practice that the current evaluation period does not yet provide sufficient data to fully characterize. Additionally, the generative AI diagnostic interface's reliance on Feature Store-grounded inference, while reducing hallucination risk, does not eliminate it; engineers must maintain calibrated skepticism toward AI-generated root cause narratives rather than treating them as authoritative conclusions.

Fixed Wireless Access occupies a structurally distinct position in the broadband access landscape: for rural and economically underserved communities where the capital economics of physical fibre or coaxial cable deployment are prohibitive, 5G FWA frequently represents the only realistic path to high-quality broadband connectivity within a viable commercial timeframe. The FCC's 2023 Broadband Data Collection indicates that approximately 14.5 million U.S. households remain without access to fixed broadband capable of delivering 25 Mbps download speeds, with rural counties accounting for 79% of these underserved locations. In these markets, an FWA subscriber who experiences persistent service degradation unaddressed by reactive complaint-response operations has no alternative provider to switch to, meaning that service quality failures translate directly to digital exclusion rather than competitive churn. The nQES-triggered proactive outreach model identifies at-risk devices by technical condition rather than by complaint behavior, removing the systematic bias toward better-served, more complaint-active subscriber segments that characterized the prior reactive model, a distinction Sufyan et al. (2023) note is particularly important given that rural 5G FWA deployments operate under systematically more challenging propagation conditions than urban deployments.The human-AI collaborative framework does not eliminate the network operations engineering workforce; it qualitatively restructures the work engineers perform. The 80% reduction in non-actionable alarm volume and the automated resolution of Missing Anchor conditions, which collectively accounted for an estimated 55% of Tier One incident volume in the pre-deployment period, shift engineer attention from high-frequency, low-complexity data triage toward lower-frequency, higher-complexity diagnostic reasoning, model oversight, and cross-functional coordination. Engineers operating within the tiered automation framework develop a distinctive competency profile, interpreting AI-generated diagnostic narratives, identifying model failure modes, and providing calibrated override judgments, competencies that Seeber et al. (2020) argue require deliberate organizational investment in relational and interpretive training rather than an assumption that technical familiarity with the interface is sufficient.This article has presented and evaluated a human-AI collaborative framework for FWA network operations that integrates unified data infrastructure, composite quality scoring, closed-loop anomaly remediation, and generative AI decision support into a coherent operational architecture validated across 1.7 million

active subscriber devices. The production results, an 80% reduction in Mean Time to Resolution, a five percent churn reduction for nQES-identified intervention cohorts, elimination of 80% of non-actionable alarms, and 93.67% service qualification accuracy, collectively demonstrate that the integration of these four components produces operational outcomes that substantially exceed what either human engineering teams or fully automated systems achieve independently. The central theoretical contribution is the operationalization of human-AI complementarity through a tiered automation architecture that treats the boundary between autonomous AI action and human-mediated decision-making as a governed, evidence-driven variable that expands as operational confidence accumulates, rather than a fixed design parameter. Future work should extend the SSOT platform's data substrate to network slicing management and edge computing workload optimization, strengthen the causal interpretation of churn and MTTR outcomes through longitudinal evaluation across full annual load cycles, and address the open governance question of how tiered automation boundaries should be set, under what accountability structures, and with what audit trail, as AI autonomy in critical communications infrastructure continues to expand.

REFRENCES

[1] Ahmad, A. K., Jafar, A., & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. Journal of Big Data, 6(1). https://doi.org/10.1186/s40537-019-0191-6

[2] Amershi, S., Inkpen, K., Teevan, J., Kikin-Gil, R., Horvitz, E., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., & Bennett, P. N. (2019). Guidelines for human-AI interaction. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems—CHI'19. https://doi.org/10.1145/3290605.3300233

[3] Ayoubi, S., Limam, N., Salahuddin, M. A., Shahriar, N., Boutaba, R., Estrada-Solano, F., & Caicedo, O. M. (2018). Machine learning for cognitive network management. IEEE Communications Magazine, 56(1), 158–165. https://doi.org/10.1109/mcom.2018.1700560

[4] Chen, X., Liu, C.-Y., Proietti, R., Li, Z., & Yoo, S. J. B. (2022). Automating optical network fault management with machine learning. IEEE Communications Magazine, 60(12), 88–94. https://doi.org/10.1109/mcom.003.2200110

[5] Dimou, S., & Noubir, G. (2025). ARGOS: Anomaly recognition and guarding through O-RAN sensing. ArXiv.org. https://arxiv.org/abs/2506.06916

[6] Gill, S. S. (2022). AI for next generation computing: Emerging trends and future directions. Internet of Things, 19(1), 100514. https://doi.org/10.1016/j.iot.2022.100514

[7] Habib, M. A., Enrique, P., Ozcan, Y., Elsayed, M., Bavand, M., Gaigalas, R., & Erol-Kantarci, M. (2025). Harnessing the power of LLMs, informers and decision transformers for intent-driven RAN management in 6G. ArXiv.org. https://arxiv.org/abs/2505.01841

[8] Hemmer, P., Schemmer, M., Vössing, M., & Kühl, N. (2021). Human-AI complementarity in hybrid intelligence systems: A structured literature review. PACIS 2021 Proceedings. https://aisel.aisnet.org/pacis2021/78/

[9] Iturria-Rivera, P. E., Zhang, H., Zhou, H., Mollahasani, S., & Erol-Kantarci, M. (2022). Multi-agent team learning in virtualized open radio access networks (O-RAN). Sensors, 22(14), 5375. https://doi.org/10.3390/s22145375

[10] Kaur, K., Garg, S., Aujla, G. S., Kumar, N., Rodrigues, J. J. P. C., & Guizani, M. (2018). Edge computing in the industrial internet of things environment: Software-defined-networks-based edge-cloud interplay. IEEE Communications Magazine, 56(2), 44–51. https://doi.org/10.1109/mcom.2018.1700622

[11] 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

[12] Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors: The Journal of the Human Factors and Ergonomics Society, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

[13] Mohammed, A. M., & Syed. (2025). Intelligent network management through deep reinforcement and federated learning. Journal of Cognitive Computing and Cybernetic Innovations, 1(3), 36–43. https://doi.org/10.21276/jccci.2025.v1.i3.6

[14] Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654), 187–192. https://doi.org/10.1126/science.adh2586
[15] Omheni, N., Koubaa, H., & Zarai, F. (2025). Artificial intelligence for 5G and 6G networks: A taxonomy-based survey of applications, trends, and challenges. Technologies, 13(12), 559. https://doi.org/10.3390/technologies13120559
[16] Oughton, E. J., & Frias, Z. (2018). The cost, coverage and rollout implications of 5G infrastructure in Britain. Telecommunications Policy, 42(8), 636–652. https://doi.org/10.1016/j.telpol.2017.07.009
[17] Pajo, J. F., Egeland, G., Kahvazadeh, S., Khalili, H., Tolan, M., McCloskey, R., Xie, M., & Erdal, O. B. (2025). Hybrid approach to enabling cross-domain service orchestration over heterogeneous infrastructures. Sensors, 25(18), 5804. https://doi.org/10.3390/s25185804
[18] Papidas, A. G., & Polyzos, G. C. (2022). Self-organizing networks for 5G and beyond: A view from the top. Future Internet, 14(3), 95. https://doi.org/10.3390/fi14030095
[19] 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, 1–1. https://doi.org/10.1109/COMST.2023.3239220
[20] Rafique, D., & Velasco, L. (2018). Machine learning for network automation: Overview, architecture, and applications [Invited tutorial]. Journal of Optical Communications and Networking, 10(10), D126–D143.
[21] Seeber, I., Bittner, E., Briggs, R. O., de Vreede, T., de Vreede, G.-J., Elkins, A., Maier, R., Merz, A. B., Oeste-Reiß, S., Randrup, N., Schwabe, G., & Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. Information & Management, 57(2), 103174. https://doi.org/10.1016/j.im.2019.103174
[22] Shu, K., Li, Y., Ding, K., & Liu, H. (2021). Fact-enhanced synthetic news generation. Proceedings of the AAAI Conference on Artificial Intelligence, 35(15), 13825–13833. https://doi.org/10.1609/aaai.v35i15.17629
[23] Soldani, D., & Manzalini, A. (2015). Horizon 2020 and beyond: On the 5G operating system for a true digital society. IEEE Vehicular Technology Magazine, 10(1), 32–42. https://doi.org/10.1109/mvt.2014.2380581
[24] Sufyan, A., Khan, K. B., Khashan, O. A., Mir, T., & Mir, U. (2023). From 5G to beyond 5G: A comprehensive survey of wireless network evolution, challenges, and promising technologies. Electronics, 12(10), 2200. https://doi.org/10.3390/electronics12102200
[25] Tomala, M., & Staniec, K. (2023). Modelling of ML-enablers in 5G radio access network—Conceptual proposal of computational framework. Electronics, 12(3), 481. https://doi.org/10.3390/electronics12030481