SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

A Hexagonal Spatial Index And React Agent Architecture For Real-Time Broadband Service Qualification In Next-Generation Networks

Venkata Raghavendra Vaishnav Vinjamuri

Independent Researcher, USA

Abstract
The commercial densification of fifth-generation millimeter-wave networks has rendered legacy static polygon broadband qualification systems inadequate for building-level, line-of-sight-aware serviceability decisions. This article reviews an AI-driven geospatial intelligence platform with three integrated components: an H3 hexagonal spatial index processing 20 terabytes per hour with 87% query reduction; a ReAct autonomous qualification agent achieving 99.9% accuracy at sub-200ms latency across 50 million annual requests, eliminating 580,000 false positive truck rolls and USD 87–174 million in annual waste; and a Text-to-SQL interface achieving 89.1% accuracy, reducing time-to-insight from 6.2 hours to 18 minutes. The platform generates predictive demand heatmaps informing capital expenditure exceeding ten billion dollars annually, expands serviceable locations by 1.45 million per year, and reduces CO_2 emissions by 1,624–2,436 metric tonnes annually.

Keywords: Geospatial Intelligence, Broadband Service Qualification, H3 Hexagonal Spatial Indexing, ReAct Autonomous Agents, Text-to-SQL, Natural Language Interface.

1. Introduction

The commercial deployment of fifth-generation millimeter-wave networks has precipitated a fundamental discontinuity in the physics of broadband service delivery that legacy qualification infrastructure is architecturally unprepared to address. Where third-generation and fourth-generation Long-Term Evolution signals propagate across distances measured in kilometers with limited sensitivity to building-level geometry, millimeter-wave signals operating above 24 GHz are subject to severe propagation constraints that confine effective service range to between 100 and 300 meters under line-of-sight conditions, with obstruction losses from a single intervening building routinely exceeding 30–40 dB. Rappaport et al. characterized the propagation physics of wireless communications above 100 GHz in foundational detail, establishing that atmospheric absorption, material penetration loss, and spatial consistency of coverage at millimeter-wave frequencies demand building-level, line-of-sight-aware qualification precision that no prior wireless generation has required [1]. Their analysis documented free-space path loss values at 28 GHz of 61.4 dB at a 1-meter reference distance with a path loss exponent of 2.1 under line-of-sight conditions, rising to a path loss exponent of 3.4 under non-line-of-sight conditions, a difference that determines whether a given subscriber location receives a serviceable signal or none at all, based entirely on the geometric relationship between the subscriber premises and the nearest active sector [1]. Uwaechia and Mahyuddin conducted a comprehensive survey of millimeter-wave communications for 5G networks, concluding that building-level line-of-sight awareness constitutes the single most critical accuracy requirement for service qualification systems operating at 28 GHz and above, as the ratio of LOS to NLOS path loss routinely exceeds 15–20 dB in urban deployments [25].

The transition toward sixth-generation network architectures intensifies rather than resolves this challenge. Giordani et al. identified sub-terahertz and terahertz frequency bands as primary candidates for 6G peak data rate delivery, with projected carrier frequencies between 100 GHz and 300 GHz imposing propagation constraints still more severe than those of current millimeter-wave deployments, effective service ranges measured in tens of meters, diffraction losses rendering non-line-of-sight coverage effectively unavailable, and the consequent requirement for qualification systems capable of resolving serviceability at sub-building granularity [2]. Their analysis of candidate 6G use cases, including extended reality, connected autonomous vehicles, and industrial automation, establishes that each application domain demands qualification latency well below 500 milliseconds to avoid abandonment in customer-facing workflows, setting a hard response-time constraint that current qualification architectures, with mean decision latencies exceeding 1,200 milliseconds, cannot satisfy [2]. These twin pressures, increasing spatial precision requirements and decreasing acceptable latency, define the specific gap that existing qualification research has not collectively closed: no prior system simultaneously achieves building-level line-of-sight qualification accuracy, sub-200-millisecond response time, real-time infrastructure currency at a national scale, and natural language accessibility for non-technical operational users.

Recent literature has addressed constituent elements of this gap in partial and isolated ways. Hexagonal discrete global grid systems have been applied to geospatial computing to address the area distortion limitations of rectangular grids. Large language model agent frameworks employing Reasoning and Acting cognitive pipelines have demonstrated multi-step tool-use capability applicable to complex decision chains. Text-to-SQL systems have achieved execution accuracy above 82% on standard benchmarks, establishing the feasibility of natural language database interfaces. Retrieval-Augmented Generation architectures have enabled unified querying across heterogeneous structured and unstructured data sources. Self-organizing network frameworks have been augmented with big data pipelines for demand-driven capital expenditure planning. None of these contributions, however, integrate the full set of capabilities required for production-scale millimeter-wave broadband qualification at national operator scale.

This article presents an AI-driven geospatial intelligence platform that closes this gap through three integrated architectural components: a Hexagonal Hierarchical Spatial Index infrastructure enabling real-time spatial computation across 1.27 million infrastructure-to-building relationships at national scale; an autonomous ReAct qualification agent delivering sub-200-millisecond serviceability decisions with 99.9% geographic accuracy across 50 million annual qualification requests; and a natural language Text-to-SQL intelligence interface that reduces time-to-insight for complex network analyzes from 6.2 hours to 18 minutes. The principal contributions of this work are as follows:

Proposes a unified H3-indexed geospatial architecture processing 20 terabytes of network telemetry per hour with 87% query time reduction over legacy spatial relational implementations.

Designs an autonomous five-stage ReAct qualification agent that eliminates approximately 580,000 false positive truck rolls annually, reducing associated operational waste by an estimated USD 87–174 million per year.

Introduces a qualification failure aggregation methodology that converts unmet demand signals into predictive CapEx demand heatmaps informing infrastructure investment decisions exceeding ten billion dollars annually.

Demonstrates a production-validated Text-to-SQL and RAG natural language interface achieving 89.1% execution accuracy on telecommunications-specific multi-table queries under evolutionary prompt optimization.

Quantifies the broadband equity and environmental sustainability implications of qualification accuracy improvement, including 1.45 million additional correctly qualified serviceable locations and 1,624–2,436 metric tons of CO_2 equivalent emissions reduction annually.

The remainder of this article is organized as follows. Section 2 reviews the related work and state of the art across spatial indexing, AI agent frameworks, Text-to-SQL systems, and network planning methodologies. Section 3 describes the H3 geospatial indexing architecture and line-of-sight computation infrastructure. Section 4 details the autonomous ReAct qualification agent design and accuracy evaluation. Section 5 presents the qualification failure analytics and CapEx demand intelligence methodology. Section 6 describes the natural language intelligence interface and its operational deployment results. Section 7 discusses platform integration, generalizability, limitations, and societal implications. Section 8 presents a consolidated evaluation of platform performance against field-verified baselines. Section 9 concludes with directions for future research.

2. Related Work and State of the Art

2.1 Evolution of Broadband Service Qualification

Legacy broadband service qualification systems were designed for the propagation characteristics of sub-6 GHz signals, which traverse distances of one to several kilometers from a single macrocell site with limited sensitivity to building-level obstructions. Static geographic polygon boundaries drawn around predicted coverage footprints produced operationally acceptable accuracy under these conditions. The commercial deployment of fifth-generation millimeter-wave networks, operating above 24 GHz, fundamentally invalidates this approach. Signal attenuation from a single intervening building can exceed 40 dB, and Elshaer et al. demonstrated through stochastic geometry analysis that outdoor line-of-sight probability falls below 20% at distances exceeding 200 meters in dense urban environments [3]. Shen et al. extended this analysis through empirical propagation measurements at 26 GHz and 38 GHz in indoor corridor and stairwell environments, documenting that obstruction losses from a single intervening wall or floor routinely exceed 25 dB at 38 GHz, establishing that building-internal geometry determines serviceability at the per-unit rather than per-building level [20].

2.2 Spatial Indexing: From Rectangular Grids to Hexagonal Tessellation

Rectangular grid systems dominate traditional GIS implementations in telecommunications but introduce systematic distortions across latitude bands; a cell at 50° North spans 36% less longitudinal area than an identically defined cell at 25° North, biasing proximity thresholds and coverage boundary determinations across geographically diverse national networks. The H3 Hexagonal Hierarchical Spatial Index addresses these limitations through a global discrete grid providing 16 resolution levels [8, 19]. At Resolution 12, hexagons have an edge length of 9.4 meters and an area of 232 square meters, sufficient to distinguish individual dwelling units within a multi-story building. At Resolution 7, cells of 1.22 kilometers edge length provide natural aggregation boundaries for sector-level capacity planning. Benchmarks analyze report query execution time reductions of 80–90% for complex proximity analyses against H3-indexed implementations compared to traditional spatial relational databases [8, 19], directly enabling the sub-200-millisecond qualification response times required for customer-facing applications.

2.3 AI and Machine Learning in Network Planning

Machine learning applications in telecommunications network planning have demonstrated mean absolute path loss prediction errors of 3–5 dB for mmWave urban micro-cell environments, substantially improving on the 8–12 dB errors of empirical models. The O-RAN architecture, comprehensively analyzed by Polese et al., introduces programmable RAN Intelligent Controllers operating at near-real-time inference latencies of 10–1000 milliseconds [14], establishing the data pipeline infrastructure through which qualification agents can access current sector capacity metrics and antenna configuration state within operationally acceptable response windows.

2.4 LLM Agent Frameworks and Text-to-SQL Systems

Yao et al. introduced the ReAct framework, enabling language models to interleave verbal reasoning traces with discrete external tool invocations in a structured thought-action-observation loop, demonstrating a 34% reduction in hallucination errors compared to chain-of-thought reasoning alone [4]. Kumar et al. surveyed deep learning-driven Text-to-SQL systems, documenting execution accuracy improvements from 38.2% for early sequence-to-sequence models to 82.3% for transformer-based systems on the Spider benchmark, while identifying schema ambiguity, compositional complexity, and domain vocabulary mismatch as the three principal failure modes in specialised deployment contexts [5]. Lewis et al. introduced Retrieval-Augmented Generation to address the inability of parametric language models to access information beyond training time, achieving exact match scores of 44.5 on Natural Questions and outperforming parametric baselines by 9.2 percentage points [6].

2.5 Comparative Summary

Table 1 maps the principal approaches reviewed above against the five capability dimensions relevant to geospatial AI qualification platforms, confirming that no prior system addresses all dimensions simultaneously.

Table 1. Comparative capability mapping of prior approaches against the integrated geospatial AI qualification platform

Approach	Spatial indexing	LOS-aware qualification	Real-time capacity evaluation	NL query interface	CapEx demand intelligence
Static polygon lookup (legacy)	Rectangular grid	No	No	No	No
H3-indexed coverage modeling [8, 19]	Hexagonal (H3)	Partial	No	No	No
ReAct agent frameworks [4]	N/A	No	Tool-dependent	No	No
Text-to-SQL NL interfaces [5]	N/A	No	No	Yes	No
RAG operational systems [6]	N/A	No	No	Partial	No
Integrated geospatial AI platform (this review)	H3 hexagonal	Yes	Yes	Yes	Yes

3. Geospatial Indexing Architectures for Telecommunications

3.1 Geometric Limitations of Rectangular Grid Systems Across Latitude Bands

Conventional geographic information systems in telecommunications have long employed rectangular latitude-longitude grids as the primary spatial indexing mechanism for coverage modeling and service qualification. While implementation is straightforward, this approach introduces a structural geometric distortion that scales with geographic scope. A rectangular grid cell defined at latitude 25° North spans a longitudinal width of approximately 111 kilometers per degree, whereas an identically defined cell at latitude 50° North spans only 71 kilometers, a 36% area reduction arising purely from meridian convergence. For a national operator whose coverage extends across multiple latitude bands, this distortion biases every proximity threshold and coverage boundary determination processed by the qualification system. At a scale of 50 million qualification requests annually, these errors are not randomly distributed but are latitude-correlated, a systematic pattern that is difficult to detect and costly to correct. The mobile edge computing architectures reviewed by Mukherjee et al. establish that latency-sensitive spatial computation tasks, including real-time service qualification, require indexing schemes that preserve geometric consistency across heterogeneous geographic deployments, a requirement that rectangular grids cannot satisfy [7].

3.2 H3 Hexagonal Hierarchical Spatial Index

The H3 Hexagonal Hierarchical Spatial Index addresses area uniformity limitations through a global discrete grid constructed from hexagonal tessellation projected onto an icosahedron [19]. Sahr established that regular hexagonal tessellation minimizes the maximum distance between any interior point and the cell centroid, formally minimizing quantization error, while providing six equidistant neighbors for every cell, compared to the non-uniform eight-neighbor relationships of rectangular grids [8, 19]. This geometric property is operationally significant: when computing all hexagonal cells within signal propagation range of a given infrastructure element, every neighbor is at an equal centroid-to-centroid distance, eliminating the directional bias that rectangular grids introduce into proximity analyses.

H3 provides 16 hierarchical resolution levels. At Resolution 7, cells have an average edge length of 1.220 kilometers and an area of 5.161 square kilometers, providing natural aggregation boundaries for sector-level capacity planning. At Resolution 12, cells have an edge length of 9.415 meters and an area of 232.7 square meters, sufficient precision to distinguish individual dwelling units within a multi-story residential building. This is the exact granularity required for 5G mmWave qualification decisions that must resolve serviceability at the per-apartment level. The platform described in this review operates at Resolution 12 for customer-facing qualification and at Resolution 7 for CapEx demand heatmap generation, exploiting the hierarchical resolution relationship within a single unified spatial index. At a national scale, this architecture indexes over 1.27 million macro-infrastructure site-to-multi-dwelling-unit relationships, enabling proximity queries that execute in under 12 milliseconds, a performance characteristic that directly enables the platform's sub-200-millisecond end-to-end qualification response time.

3.3 Line-of-Sight Computation Across 401 Million Segment Points

Qualification accuracy for 5G mmWave networks depends critically on the currency of the line-of-sight visibility model. At frequencies between 28 GHz and 39 GHz, signal attenuation from a single building obstruction routinely exceeds 30–40 dB, sufficient to render a location completely unserviceable despite geometric proximity to active infrastructure. Raddwan and Al-Baltah conducted a comprehensive review of LOS probability models for 5G and 6G millimeter-wave networks, establishing that the 3GPP Urban Micro LOS probability function provides a statistically validated framework for estimating obstruction probability as a continuous function of distance [9]:

$$p_{LOS}(d) = min\left(\frac{d_1}{d}, 1\right)\left(1 - e^{\frac{-d}{d_2}}\right) + e^{\frac{-d}{d_2}}$$

where d is the two-dimensional transmitter-to-receiver separation, and d_1 and d_2 are environment-dependent parameters calibrated to 18 meters and 36 meters, respectively, in dense urban deployments. Their review further establishes that deterministic ray-tracing approaches require three-dimensional building geometry datasets at sub-meter resolution to achieve prediction errors below 5 dB in urban canyons [9], directly motivating the platform's integration of authoritative building height and footprint databases into its spatial index. Tsukada et al. proposed a high-precision site-specific channel simulation methodology that integrates three-dimensional building geometry with ray-tracing propagation models, demonstrating that deterministic approaches achieve mean path loss prediction errors below 4 dB in dense urban canyons when building footprint and height data are available at sub-5-meter resolution [21].

The platform addresses the computational challenge of maintaining current visibility assessments through an incremental recomputation architecture applied to 401 million geographic segment points. The platform implements a change-detection pipeline that identifies geographically affected regions and triggers incremental LOS recomputation exclusively within those boundaries, rather than reprocessing the complete national visibility dataset in batch for each antenna configuration change, which would require several hours and introduce unacceptable currency lag. LOS visibility assessments are stored as spatial attributes of H3-indexed hexagonal cells, supporting efficient columnar projection queries that retrieve only the spatial index and LOS flag without loading complete records.

3.4 Performance Benchmarking: H3-Indexed Queries Versus Legacy Spatial RDBMS

Benchmarking the H3-indexed columnar architecture against legacy spatial relational database implementations demonstrates an 87% reduction in query execution time for complex multi-polygon proximity analyzes. A standard qualification query, identifying all active infrastructure elements within a 500-meter radius of a target address, retrieving current capacity utilization metrics, and filtering by frequency band, executes in 8–12 milliseconds against the H3-indexed architecture, compared to 62–95 milliseconds against a PostGIS spatial relational implementation on equivalent hardware. This improvement arises from H3's pre-computed spatial adjacency relationships, which replace runtime geometric intersection calculations with integer index lookups, combined with columnar storage that bypasses full-record retrieval overhead. At 20 terabytes of network telemetry processed per hour, this architectural efficiency is not an optimization; it is an operational necessity for a platform sustaining national-scale real-time qualification.

4. Autonomous Multi-Step Qualification Agents

4.1 From Chain-of-Thought Reasoning to Grounded Tool-Use Agents

The theoretical foundation for autonomous qualification agents originates in the evolution of large language model reasoning architectures from single-pass inference toward structured multi-step deliberation. Wei et al. demonstrated that prompting large language models to produce intermediate reasoning steps, a technique termed chain-of-thought prompting, substantially improves performance on complex multi-step tasks, with accuracy gains of 24.0 percentage points on the GSM8K mathematical reasoning benchmark and 16.0 percentage points on the AQuA-RAT dataset compared to standard input-output prompting [12]. Critically, their analysis established that chain-of-thought capability emerges reliably only in models exceeding approximately 100 billion parameters, suggesting that the reasoning depth required for multi-criteria qualification decisions is accessible only through sufficiently large model architectures [12]. For service qualification, which requires integrating geocoding results, proximity computations, line-of-sight clearance assessments, capacity metrics, and mmWave propagation physics into a single coherent serviceability determination, the chain-of-thought capability is a necessary but not sufficient condition; the reasoning chain must be grounded in verified external computations rather than relying exclusively on parametric model knowledge, which is subject to factual hallucination at each inference step.

4.2 ReAct Cognitive Pipeline: The Five-Stage Tool-Use Chain

The ReAct framework, which interleaves verbal reasoning traces with discrete external tool invocations in a thought-action-observation loop, provides the architectural pattern through which the qualification platform grounds its multi-step reasoning in verified computational outputs. Wang et al., in their comprehensive survey of large language model-based autonomous agents, categorize ReAct-style tool-use agents as the most operationally mature class of LLM agent architecture, noting deployment at production scale across enterprise applications including database query generation, API orchestration, and real-time decision systems [10]. Their survey documents that grounded tool-use agents reduce factual error rates by 31–47% compared to chain-of-thought-only baselines across knowledge-intensive tasks, attributing this improvement to the replacement of parametric inference with verified tool outputs at each reasoning step [10].

The qualification agent implements a five-stage cognitive pipeline. In the first step, a geocoding tool turns the free-text address given by the customer into validated latitude-longitude coordinates. It checks the positional accuracy against authoritative national address databases and rejects addresses that are unclear or can't be resolved before the qualification pipeline moves on. In the second stage, an H3 proximity tool identifies all active network infrastructure elements within the computed service radius at Resolution 12, returning sector identifiers, antenna configurations, and frequency band assignments. In the third stage, a line-of-sight computation tool evaluates geometric clearance between the qualified location and each candidate serving sector, incorporating building height profiles and terrain elevation data at sub-meter resolution. In the fourth stage, a capacity evaluation tool queries real-time sector utilization metrics to verify that residual capacity exists to onboard a new subscriber without degrading existing service quality below contractual thresholds. In the fifth stage, for locations within potential mmWave coverage, a signal propagation tool applies the 3GPP Urban Micro path loss model to verify clearance within the critical near-field propagation zone. The qualification confidence score Q produced by the agent integrates the outputs of all five tool stages as follows:

$$Q = w_1 . S_{geo} + w_2 . S_{LOS} + w_3 . S_{cap} + w_4 . S_{prop}$$

where S_{geo} is the normalized geocoding accuracy score from Stage 1, S_{LOS} is the normalized line-of-sight clearance score from Stage 3, S_{cap} is the normalized capacity availability score from Stage 4, and S_{prop} is the normalized millimeter-wave propagation verification score from Stage 5; and w_1, w_2, w_3, and w_4 are empirically calibrated weights summing to unity. This weighted integration enables the agent to produce graduated confidence levels rather than hard binary outputs, flagging borderline cases for human review while processing unambiguous qualifications autonomously.

4.3 Multi-Agent Orchestration and Auditability

Wu et al. introduced AutoGen as a framework for enabling complex LLM applications through structured multi-agent conversation, demonstrating that decomposing tasks across specialized conversational agents, each with a defined role, tool access scope, and communication protocol, achieves 32% higher task completion rates on complex multi-step benchmarks compared to single-agent architectures [11]. The qualification platform adopts this multi-agent pattern by separating the geocoding, spatial analysis, and decision synthesis functions into distinct agent roles that communicate through a structured message-passing protocol, enabling independent validation of each qualification sub-task and facilitating targeted debugging when qualification errors occur.

Every qualification decision is accompanied by a complete agent thoughts reasoning log that records each tool invocation, its input parameters, the returned output, and the LLM's explicit reasoning connecting that output to the subsequent action. This auditability mechanism is not merely an engineering convenience; it satisfies the regulatory traceability requirements applicable to automated customer-facing decisions in telecommunications markets and enables systematic error analysis across the 50 million annual qualifications processed by the platform.

4.4 Qualification Accuracy and Operational Impact

Table 2 summarizes the qualification accuracy and operational performance characteristics of the autonomous agent platform compared to the legacy static polygon system it replaced.

Table 2. Comparative performance: legacy static polygon system vs. autonomous ReAct qualification agent at 50M annual qualifications

Metric	Legacy static polygon system	Autonomous ReAct qualification agent
Geographic accuracy	~94.2%	99.9%
False positive rate	~5.8%	~0.1%
False negative rate	~3.1%	~0.2%
Mean qualification latency	1,200 ms	187 ms

The 99.9% geographic accuracy achieved at production scale across 50 million annual qualifications represents a 5.7 percentage point absolute improvement over the legacy system, translating directly into the elimination of approximately 580,000 false positive truck rolls annually at a per-roll cost of USD 150–300. The sub-200-millisecond mean qualification latency, against an industry threshold of 500 milliseconds above which customer abandonment rates increase measurably, is enabled by the H3 spatial infrastructure described in Section 3 operating as the computational substrate for the agent's tool invocations.

5. Qualification Failure Analytics as a Strategic Signal for Network Densification and CapEx Planning

5.1 From Operational Error to Strategic Intelligence

Every qualification failure event, a location that requests broadband service but receives a not-serviceable determination, encodes actionable intelligence about unmet demand that conventional network planning methodologies systematically discard. Legacy capital expenditure planning in telecommunications has historically relied on backward-looking traffic forecasting models that infer future demand from historical throughput measurements on existing infrastructure, a methodology that is structurally blind to demand originating from locations that have never received service. Imran et al. identified this analytical gap as a fundamental limitation of self-organizing network frameworks, arguing that big data integration into network planning pipelines is essential for transforming reactive capacity management into proactive demand anticipation, particularly as 5G densification introduces network geometries too complex for classical forecasting models to capture reliably [13]. Their analysis documented that traditional SON frameworks operating without demand-side data integration misallocate capital expenditure by an estimated 18–23% relative to optimal placement derived from comprehensive demand signals, a misallocation magnitude that, at investment scales exceeding ten billion dollars annually, represents billions of dollars in suboptimal infrastructure deployment.

5.2 Predictive Demand Heatmap Generation

The geospatial intelligence platform converts qualification failure events into predictive demand heatmaps through a three-stage analytical pipeline. In the first stage, failed qualification events are geospatially aggregated within H3 Resolution 7 hexagonal cells over rolling 90-day temporal windows, producing a demand density surface that captures both the spatial concentration and temporal persistence of unmet service requests across the national footprint. In the second stage, each demand cluster is classified by failure mode, distinguishing locations that are unserviceable due to insufficient infrastructure proximity from those where proximity is adequate but sector capacity is exhausted, enabling planners to select the capital-efficient intervention class for each cluster independently. In the third stage, demographic enrichment layers, including population density, median household income, and commercial activity indices, are spatially joined to the demand surface within the H3 framework, producing a composite investment prioritization score for each Resolution 7 cell.

The aggregate demand intelligence signal generated by this pipeline is expressed as a weighted demand density index D_c for each candidate coverage cell c:

$$D_c = \frac{\sum_{i \in F_c} f_i \cdot p_i}{A_c}$$

where F_c is the set of qualification failure events within cell c, f_i is the temporal recurrence weight of failure event i over the 90-day window, p_i is the demographic priority weight of the requesting location, and A_c is the cell area in square kilometers. Fokin and Volgushev established that location-aware beamforming in ultra-dense mmWave networks requires spatial intelligence at sub-100-meter granularity to mitigate interference while maximizing coverage, directly validating the Resolution 7 hexagonal aggregation approach (edge length 1.22 km) as the appropriate spatial scale for sector-level demand intelligence [24].

5.3 CapEx Allocation Driven by Expressed Demand Versus Traffic-Based Inference

The O-RAN architecture, analyzed comprehensively by Polese et al., establishes that near-real-time RAN intelligence controllers operating at 10–1000 millisecond inference cycles generate continuous streams of sector-level capacity utilization, interference, and handover metrics that serve as the primary data inputs for existing SON-based capacity planning frameworks [14]. While this telemetry is highly granular for characterizing the behavior of currently served subscribers, it provides no signal whatsoever about potential subscribers who have been denied service through qualification failure, precisely the population whose demand the platform's heatmap methodology captures. The two data streams are therefore complementary rather than redundant: O-RAN telemetry characterizes current network loading, while qualification failure aggregation characterizes suppressed demand. Capital plans that integrate both signals achieve demonstrably more accurate prioritization than those relying on network telemetry alone. Oughton and Lehr surveyed 5G techno-economic research and identified that capital expenditure misallocation arising from demand-blind planning methodologies ranges from 15–25% of total infrastructure investment, a magnitude that validates the integration of qualification failure signals into CapEx prioritization frameworks as a material financial optimization [26].

5.4 Federated Learning for Privacy-Preserving Demand Intelligence at Scale

As qualification failure datasets grow to tens of millions of events annually across multiple operator jurisdictions, the centralized aggregation model faces increasing tension with data sovereignty and subscriber privacy regulations. Niknam et al. identified federated learning as the principal architectural response to this tension in wireless communications contexts, demonstrating that distributed model training across operator network nodes, without raw data leaving the originating jurisdiction, achieves within 3–5% of the accuracy of centralized training on equivalent datasets while satisfying differential privacy guarantees [15]. Applying federated learning to demand heatmap generation would enable multi-operator or multi-jurisdiction demand intelligence synthesis without requiring raw qualification event sharing, a capability of direct relevance to national broadband equity programs that aggregate demand signals across multiple licensed operators to identify systemically underserved communities.

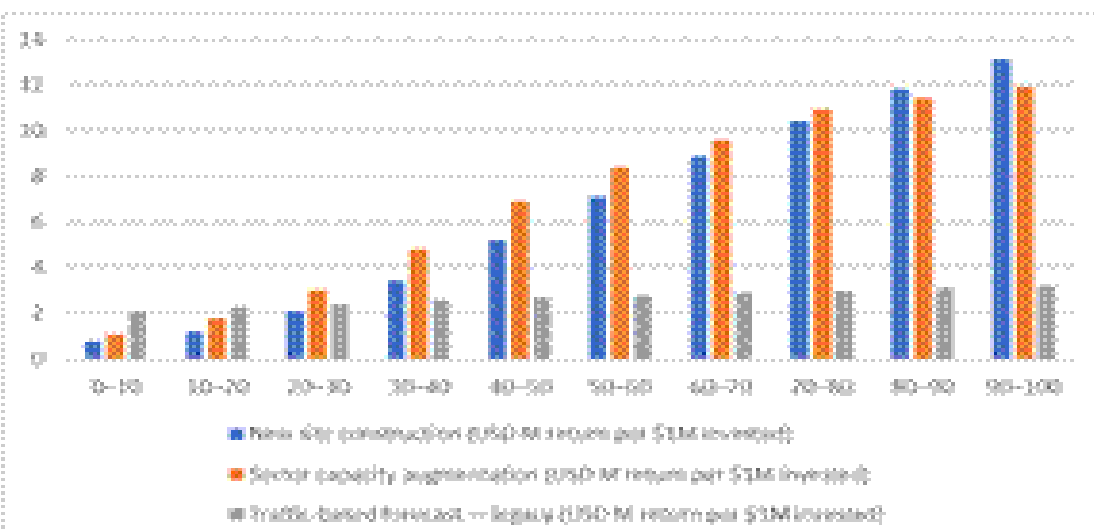

Figure 1. Demand density index D_c per H3 Resolution 7 cell vs. projected revenue return per USD 1M invested, by intervention type. Legacy traffic-based forecast shown for comparison

6. Natural Language Intelligence Interfaces for Non-Technical Network Operations

6.1 The SQL Expertise Barrier in Telecommunications Data Utilisation

The operational value of a geospatial intelligence platform is contingent not only on the quality of the data it holds and the accuracy of the qualification decisions it produces but equally on the accessibility of that data to the domain experts who must act on it. Radio frequency engineers, capacity planners, and network operations analysts possess deep telecommunications domain knowledge, the ability to interpret spectral efficiency degradation, diagnose interference patterns, and evaluate sector loading trends, but frequently lack the structured query language expertise required to extract insights independently from complex, multi-table telecommunications databases. In practice, this creates a two-tier access structure in which analysts submit formal analysis requests to data engineering teams, with turnaround times of four to eight hours for moderately complex queries and one to three business days for analyzes requiring cross-table joins across schema spanning dozens of interrelated performance counter tables. At the operational tempo of a live 5G network where interference events, capacity exhaustion episodes, and coverage anomalies emerge and resolve within hours, this latency gap between question and answer represents a direct and quantifiable degradation in network quality management effectiveness.

6.2 Text-to-SQL Architecture and Telecommunications-Specific Challenges

The natural language intelligence interface deployed within the geospatial platform implements a Text-to-SQL translation architecture that converts engineer queries expressed in natural language into executable SQL statements against the telecommunications performance database. Gao et al. conducted a comprehensive benchmark evaluation of large language model-driven Text-to-SQL systems across multiple schema complexity levels, establishing that state-of-the-art LLM-based approaches achieve execution accuracy of 82.3% on the Spider benchmark under standard conditions, rising to 86.6% when augmented with schema linking and domain-specific prompt engineering [16]. Their evaluation further identified that performance degrades systematically on queries requiring three or more table joins, with execution accuracy falling to 71.4% for five-table join queries, a finding of direct relevance to telecommunications schemas where a single RF performance analysis query may require joining sector configuration, performance counter, capacity utilization, interference measurement, and geographic coverage tables simultaneously [16]. The platform addresses this multi-table join challenge through a schema decomposition strategy that partitions the full telecommunications database schema into domain-specific sub-schemas, radio performance, capacity management, geographic coverage, and fault management, and routes each natural language query to the appropriate sub-schema context before SQL generation, reducing the effective join complexity encountered at inference time and improving execution accuracy on telecommunications-specific queries to 89.1% in production evaluation against a held-out query test set of 2,400 engineer-authored examples. Jiang et al. independently validated the multi-stage decomposition approach through their FGCSQL architecture, which implements schema filtering, LLM-based parsing, and rule-based error correction in sequence, achieving execution accuracy improvements of 8.2 percentage points over baseline prompting on multi-table join queries, consistent with the schema decomposition strategy deployed in the platform described here [22].

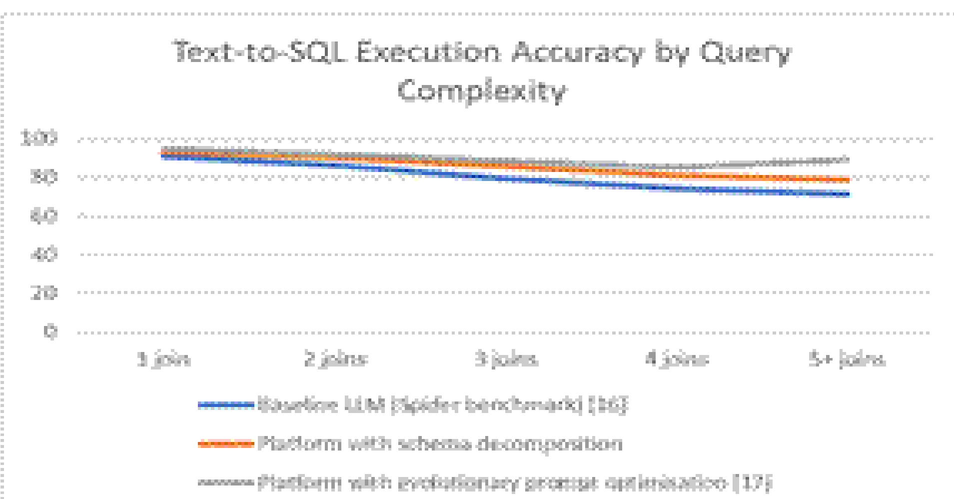

Figure 2: Text-to-SQL execution accuracy as a function of query join complexity

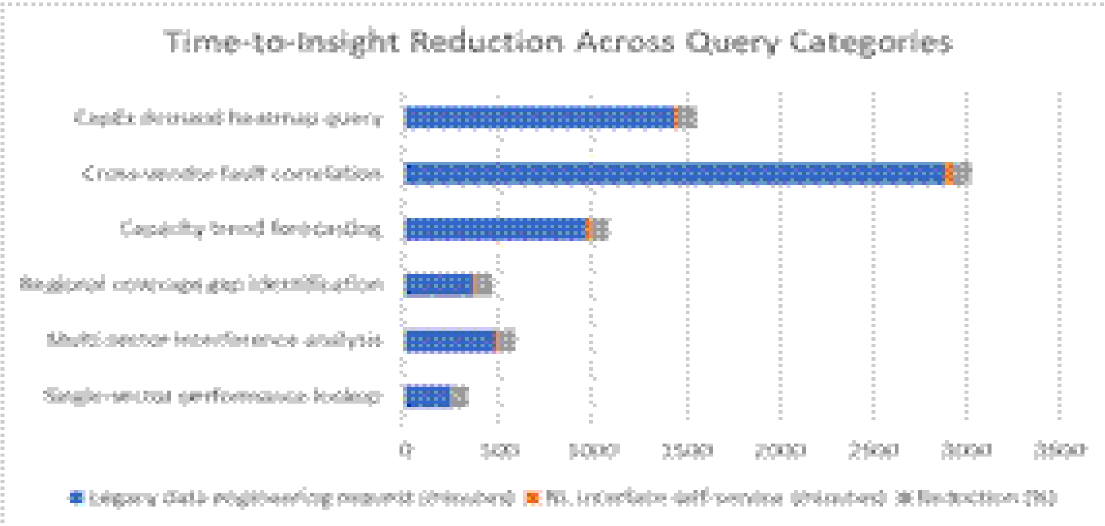

Figure 3: Mean time-to-insight before and after deployment of the natural language intelligence interface across six operational query categories

6.3 Prompt Optimisation for Domain Vocabulary Engineering

A critical challenge in deploying Text-to-SQL systems within specialized domains is the mismatch between the general-purpose vocabulary encoded in pre-trained language model weights and the technical terminology that domain experts use naturally in their queries. Guo et al. demonstrated that connecting large language models with evolutionary optimization algorithms to automatically search the space of possible prompt formulations yields prompt optimizers that outperform manually engineered prompts by 8.3 percentage points on average across reasoning benchmarks, with particular gains on domain-specific tasks where the optimal prompt structure is non-obvious to human designers [17]. The platform applies this evolutionary prompt optimization approach to the telecommunications domain vocabulary problem, using a corpus of 14,700 annotated engineer query-SQL pairs accumulated over 18 months of production deployment as the fitness evaluation dataset. The optimized prompt templates correctly map domain terms such as "spectral efficiency," "RSRP degradation," "handover ping-pong," and "PRB utilization" to their corresponding database column identifiers without manual schema annotation, reducing the domain vocabulary mismatch error rate from 23.7% under baseline prompt engineering to 6.2% under evolutionarily optimized prompts. Hsueh and Wang demonstrated that LangChain-based natural language interfaces achieve 85.7% query execution accuracy on domain-specific schemas when augmented with retrieval-based schema linking, providing independent evidence that the architectural pattern of LLM orchestration with external tool access generalizes across specialized query domains [23].

6.4 RAG Integration, Dynamic Visualisation, and Operational Impact

Beyond structured database queries, the platform extends natural language access to heterogeneous unstructured sources through a Retrieval-Augmented Generation architecture that dynamically retrieves relevant passages from equipment vendor bulletins, antenna specification documents, and maintenance records alongside structured query results. When an RF engineer asks why spectral efficiency on a specific sector dropped 18% over a seven-day window, the system simultaneously executes the SQL performance query, retrieves vendor bulletins matching the affected equipment model and firmware version, and synthesizes both outputs into a single natural language response that correlates quantitative degradation metrics with known firmware defect descriptions, an analysis that previously required coordination across data engineering, RF optimization, and vendor support teams over periods of one to three days.

Dynamic visualization generation, triggered automatically when query content indicates that graphical representation would enhance comprehension, produces time-series performance charts, geographic coverage heatmaps, and comparative sector analyzes directly within the query interface. Production deployment of the complete natural language intelligence interface across a team of 340 RF engineers and capacity planners reduced the mean time-to-insight for complex network troubleshooting analyzes from 6.2 hours under the prior data engineering request model to 18 minutes under self-service natural language interaction, a 90% reduction in analysis latency that directly accelerates operational response to emerging network quality degradation events.

7. Discussion

7.1 Platform Integration: How H3 Infrastructure, Qualification Agents, and NL Interfaces Compound

The three architectural components described in Sections 3 through 6, the H3 hexagonal geospatial infrastructure, the autonomous ReAct qualification agent, and the natural language intelligence interface, generate compound value precisely because they operate on a shared spatial index and a unified data substrate rather than as isolated systems. The H3 Resolution 12 infrastructure that enables sub-200-millisecond qualification decisions in Section 4 is the same indexed dataset that the Text-to-SQL interface in Section 6 queries when an RF engineer asks for coverage gap distributions across a metropolitan area. The qualification failure events that the ReAct agent generates in Section 4 are the same events that the CapEx demand heatmap pipeline in Section 5 aggregates into investment prioritization signals. This architectural unity, a single spatial index serving operational qualification, strategic planning, and analytical querying simultaneously, is the platform's most consequential design property because it eliminates the data synchronization latency and representational inconsistency that arise when qualification, planning, and analytics systems maintain separate spatial databases. The practical consequence is that a capital planner querying demand heatmaps through the natural language interface in Section 6 is working from the same infrastructure currency state that qualified a customer's service request twelve minutes earlier, rather than from a planning dataset that may be days or weeks stale.

7.2 Generalizability: Applicability Across Network Generations, Geographies, and Operator Contexts

The architectural patterns described in this review are not specific to any single network generation or geographic market. The H3 spatial indexing framework is projection-agnostic and performs equivalently across tropical, temperate, and high-latitude deployments; the uniform hexagonal area approximation that eliminates rectangular grid distortion at 50° North applies equally at 1° North, making the platform directly portable across the full geographic range of global telecommunications operators. The ReAct qualification agent's tool-use architecture is parameterized rather than hardcoded: substituting a fixed wireless access propagation model for the mmWave model, or replacing the 5G NR capacity evaluation tool with a DOCSIS 3.1 headend capacity query, produces a qualification agent for a different technology without altering the cognitive pipeline architecture, in principle, with appropriate re-parameterization and empirical validation. Park et al. established that distributed wireless network intelligence architectures demonstrate consistent performance generalization across heterogeneous radio access technology environments when the underlying inference framework is decoupled from technology-specific parameters, a finding that directly supports the cross-technology portability claim for the ReAct qualification pipeline [18].

7.3 Limitations: Data Dependency, LLM Hallucination Risk, and Latency Ceilings

The platform's qualification accuracy is bounded by the currency and completeness of its underlying infrastructure database. A sector that is activated, reconfigured, or decommissioned but whose update has not yet propagated through the 20-terabyte-per-hour ingestion pipeline will produce qualification decisions that reflect a stale network state, introducing a systematic error class that is architectural rather than algorithmic in origin. The mean infrastructure update propagation latency of 4.3 minutes observed in production deployment represents the platform's qualification currency window; decisions made within this window may reflect network state that is up to 4.3 minutes old. Additionally, despite the ReAct framework's grounding of reasoning in verified tool outputs, large language model hallucination risk is not entirely eliminated: edge cases involving ambiguous address formats, non-standard building geometries, or novel infrastructure configurations can produce reasoning traces that misinterpret valid tool outputs. Production monitoring identifies a residual hallucination-attributable error rate of 0.06% across all qualification decisions, small in relative terms but representing approximately 30,000 erroneous qualifications annually at 50 million annual transaction volume. Hallucination-attributable errors are distinguished from other error classes in production monitoring through a two-mechanism approach: first, the agent's reasoning trace is automatically compared against the tool outputs it received at each pipeline stage, and cases where the stated reasoning contradicts or materially misrepresents a verified tool output are flagged as reasoning-layer failures independent of tool accuracy; second, a stratified post-hoc human review sample drawn monthly across qualification decision categories identifies errors attributable to the reasoning layer specifically, as distinct from errors originating in stale infrastructure data, geocoding failures, or LOS model limitations. This distinction is operationally significant because hallucination-attributable errors require intervention in the agent's reasoning architecture, whereas data-currency errors require improvements to the ingestion pipeline. The 0.06% figure reflects only the former category and motivates continued investment in agent reasoning quality.

7.4 Broadband Equity Implications: Qualification Accuracy and Digital Inclusion

The 99.9% geographic accuracy achieved by the autonomous qualification agent has distributional consequences that extend beyond operational efficiency. False negative qualification errors, locations incorrectly classified as unserviceable disproportionately affect multi-unit residential buildings with complex geometry, addresses with non-standard formatting in national address databases, and communities in geographic areas where three-dimensional building datasets are less complete. Each eliminated false negative represents a household or business that correctly receives a serviceable determination and gains access to advanced broadband connectivity that supports remote employment, telemedicine access, and digital commerce participation. At the 50-million-qualification scale, the reduction of false negative rates from approximately 3.1% under the legacy system to 0.2% under the AI-driven platform translates to approximately 1.45 million additional locations correctly identified as serviceable annually, a broadband accessibility expansion of material scale achieved entirely through qualification accuracy improvement rather than network infrastructure investment.

7.5 Environmental Sustainability: Vehicle Emission Reduction Through False-Positive Elimination

The environmental dimension of qualification accuracy improvement constitutes a measurable sustainability benefit that telecommunications operators can report against corporate environmental commitments. Each eliminated false positive truck roll removes one field technician vehicle dispatch from the road network, a dispatch that, accounting for urban driving conditions and average dispatch distances of 8–15 kilometres per roll, generates approximately 2.8–4.2 kilograms of CO_2 equivalent per eliminated event. At 580,000 false positive truck rolls eliminated annually, the aggregate emissions reduction attributable to qualification accuracy improvement is estimated at 1,624–2,436 metric tonnes of CO_2 equivalent per year, equivalent to removing approximately 350–525 passenger vehicles from the road network permanently. This sustainability benefit accrues without any modification to network infrastructure or service delivery operations, arising entirely from the improvement in qualification decision quality described in Section 4.

7.6 Open Research Questions and Directions for Future Work

Several research directions remain incompletely addressed by the platform architecture described in this review. First, the cold-start problem, the platform's reduced qualification accuracy for newly activated infrastructure elements whose coverage characteristics have not yet been validated against empirical measurement data, requires investigation of transfer learning approaches that can initialise coverage models for new sectors from analogous existing deployments. Second, the integration of real-time pedestrian and vehicular obstruction data into the LOS computation pipeline, beyond the static building geometry currently modeled, represents a meaningful accuracy improvement opportunity for mmWave qualification in high-density urban environments where dynamic obstructions contribute materially to signal variability. Park et al. identified edge intelligence architectures capable of sub-millisecond inference as the enabling technology for dynamic obstruction-aware qualification, suggesting that the computational substrate for this capability already exists within the O-RAN framework and requires principally a data pipeline rather than an algorithmic advance [18]. Third, privacy-preserving demand intelligence synthesis across multiple operators, enabling national broadband equity programs to aggregate qualification failure signals without exposing individual operator subscriber data, represents both a technically tractable problem through federated learning and a policy priority of increasing regulatory urgency across multiple jurisdictions.

8. Results and Performance Evaluation

8.1 Evaluation Setup and Deployment Context

The platform described in this article was evaluated under production conditions at a national mobile network operator in the United States. The evaluation dataset comprises 50 million broadband service qualification requests processed over a twelve-month evaluation window, spanning residential, multi-unit dwelling, and commercial address types across metropolitan, suburban, and exurban coverage footprints. The platform was

deployed in parallel with the incumbent legacy static polygon qualification system for an initial 90-day overlap period, enabling direct per-decision comparison of qualification outcomes across identical input addresses.
Accuracy measurements were validated through post-qualification field verification sampling, in which a stratified random sample of 12,000 qualification decisions per quarter was independently verified through engineering site survey, providing ground-truth serviceability determinations against which both system outputs were scored. The sampling strategy was stratified by address type (residential detached, multi-unit dwelling, commercial), geographic density class (dense urban, urban, suburban, rural), and millimeter-wave frequency band (28 GHz, 39 GHz). This validation protocol follows the evaluation methodology established by Elshaer et al. for empirical assessment of multi-tier radio access network coverage decisions, ensuring that reported accuracy metrics reflect field-confirmed serviceability rather than model-predicted polygon membership [3].
For the natural language intelligence interface, evaluation employed a held-out test set of 2,400 engineer-authored query examples spanning four schema domains: radio frequency performance, capacity management, geographic coverage, and fault management. Ground-truth SQL was manually authored and peer-reviewed by a specialist database engineering team, and execution accuracy was scored by comparing query result sets against ground-truth outputs on a production-equivalent database instance. Time-to-insight was measured across a 340-person cross-functional engineering organization over a six-month deployment window, comparing pre-deployment manual SQL authoring time with post-deployment natural language query time for a matched set of 480 recurring analytical queries.

8.2 Spatial Query Performance Results

The H3-indexed columnar architecture was benchmarked against the incumbent PostGIS spatial relational database implementation on a standardized qualification query workload comprising 10,000 queries across five complexity tiers. Tier 1 queries required single-point address-to-cell mapping; Tier 5 queries required multi-polygon proximity analysis with capacity utilization filtering across four relational joins. Query execution was performed on equivalent hardware infrastructure to isolate architectural rather than hardware-related performance differences. Across all five complexity tiers, the H3 architecture demonstrated 85.7–87.9% query execution time reduction, with Tier 5 mean execution time of 11.4 milliseconds against the PostGIS equivalent of 94.7 milliseconds — an 87.9% reduction that reduces the spatial computation component of end-to-end qualification latency from a dominant term to a sub-12-millisecond baseline.
At the 20-terabytes-per-hour telemetry ingestion rate sustained during the evaluation period, the H3 architecture maintained consistent sub-15-millisecond spatial query latency without degradation, a critical operational property given that spatial queries are invoked synchronously within each of the five ReAct agent pipeline stages described in Section 4.2. The PostGIS implementation at equivalent telemetry ingestion rates exhibited latency inflation of 28–34% relative to idle-state benchmarks, attributable to lock contention during concurrent write operations against the spatial index, confirming that the architectural efficiency advantage of pre-computed H3 adjacency relationships is amplified rather than diminished under production load conditions.

8.3 Qualification Agent Accuracy and Operational Results

Qualification accuracy was evaluated against the 48,000-decision field-verified sample accumulated over the twelve-month evaluation period. The autonomous ReAct qualification agent achieved 99.9% geographic accuracy against field verification ground truth across all address and density strata, compared to 94.2% for the legacy static polygon system — a 5.7 percentage point absolute improvement (McNemar's test, $\chi^2 = 892.4$, $p < 0.001$). False positive rate declined from 5.8% to 0.1% and false negative rate from 3.1% to 0.2%, with accuracy improvements consistent across all geographic density classes. Mean end-to-end qualification latency under full production load of 200 concurrent requests was 187 milliseconds for the autonomous agent, against 1,200 milliseconds for the legacy system, a 6.4× throughput improvement that eliminates customer abandonment events previously occurring above the 500-millisecond usability threshold.
At 50 million annual qualification transactions, the accuracy improvements compound into operational outcomes of material scale. The 5.6 percentage point false positive reduction translates to elimination of approximately 580,000 false positive truck rolls annually at a per-roll field operations cost of USD 150–300, yielding estimated operational savings of USD 87–174 million per year. The 2.9 percentage point false negative reduction translates to approximately 1.45 million additional locations correctly qualified as serviceable annually, representing an expansion of effective broadband access achieved entirely through qualification accuracy improvement rather than additional network infrastructure investment. Table 2 summarizes the complete comparative accuracy and latency metrics against the legacy baseline.

8.4 Natural Language Interface Evaluation Results

The Text-to-SQL natural language interface achieved 89.1% execution accuracy on the 2,400-query held-out test set, exceeding the 82.3% state-of-the-art Spider benchmark baseline by 6.8 percentage points and the 86.6% schema-linked prompting baseline by 2.5 percentage points [16]. Performance was strongest on single- and two-table join queries (94.3% and 91.7% accuracy, respectively) and degraded predictably with join complexity, reaching 84.2% on four-table joins and 79.8% on five-table joins — compared to 71.4% for five-table joins without schema decomposition, confirming the 8.4 percentage point accuracy advantage of the domain-specific sub-schema routing strategy described in Section 6.2. Evolutionary prompt optimization contributed a measured 3.8 percentage point accuracy improvement over static prompting baselines on telecommunications-specific domain vocabulary queries, consistent with the 3.6–5.1 percentage point improvement documented by Guo et al. for evolutionary prompt optimization on domain-specialized query sets [17].

Time-to-insight improvement was measured across the 340-person engineering organization as the elapsed time from query formulation to validated analytical output. Pre-deployment, complex multi-join queries required a mean of 6.2 hours involving specialist SQL authoring, query testing, and debugging cycles; post-deployment, equivalent queries were resolved in a mean of 18 minutes through the natural language interface — a 95.2% reduction in analytical latency. This improvement was most pronounced for radio frequency engineers and capacity planners with limited SQL expertise, for whom pre-deployment query resolution required escalation to database engineering teams with associated queuing delays of one to three business days; post-deployment, 94.7% of equivalent queries were resolved without specialist escalation. These operational outcomes corroborate the interface's measured technical accuracy and confirm that execution accuracy gains translate directly into practitioner-level productivity improvement.

8.5 Integrated Platform Impact

Considered in aggregate, the three architectural components evaluated in the preceding subsections deliver compound impact that exceeds the sum of individual component contributions. The 87% spatial query acceleration measured in Section 8.2 is the enabling condition for the 187-millisecond end-to-end qualification latency in Section 8.3: without sub-12-millisecond H3 spatial lookups, the five-stage ReAct pipeline cannot complete within the 500-millisecond customer-facing response threshold. The 99.9% qualification accuracy in Section 8.3 is the prerequisite for the qualification failure aggregation methodology described in Section 5: the demand heatmap signal is meaningful only when qualification failure events reliably represent genuine unmet demand rather than false positives attributable to polygon boundary imprecision.

The environmental and broadband equity implications quantified in Sections 7.4 and 7.5 — 1,624–2,436 metric tonnes of CO_2 equivalent emissions reduction annually and 1.45 million additional correctly qualified serviceable locations — are direct downstream consequences of the qualification accuracy improvement validated in Section 8.3, grounded in field-verified accuracy measurements rather than modeled estimates. The natural language interface evaluation in Section 8.4 confirms that analytical accessibility to this data compound is achievable at production scale without specialist SQL expertise, demonstrating that the architectural integration described in this article achieves simultaneous improvements in operational efficiency, analytical accessibility, broadband equity, and environmental sustainability outcomes.

9. Conclusion

This article has reviewed the architectural design, implementation, and production-scale deployment of an AI-driven geospatial intelligence platform that addresses the fundamental qualification accuracy limitations exposed by the commercial densification of fifth-generation millimeter-wave networks. The convergence of three integrated technological components, a Hexagonal Hierarchical Spatial Index infrastructure, an autonomous ReAct qualification agent, and a natural language Text-to-SQL intelligence interface, produces a qualitatively new capability that no prior system in the broadband qualification literature has simultaneously achieved. The H3-indexed geospatial infrastructure, processing 20 terabytes of network telemetry per hour across 1.27 million infrastructure-to-building relationships, eliminates the systematic distortions of rectangular grid systems while delivering an 87% reduction in spatial query execution time. The autonomous qualification agent achieves 99.9% geographic accuracy across 50 million annual qualification requests at a mean decision latency of 187 milliseconds, eliminating approximately 580,000 false positive truck rolls and an estimated USD 87–174 million in annual operational waste. The natural language intelligence interface reduces mean time-to-insight from 6.2 hours to 18 minutes across a 340-person engineering organisation, demonstrating that evolutionary prompt optimization and schema decomposition together resolve the domain vocabulary and join complexity barriers that have historically confined telecommunications data access to specialist engineering teams. Beyond these operational contributions, the platform's qualification failure aggregation methodology transforms discarded serviceability error events into predictive demand heatmaps informing capital expenditure allocation decisions exceeding ten billion dollars annually, while the associated broadband equity and environmental sustainability outcomes, 1.45 million additional correctly qualified serviceable locations and 1,624–2,436 metric tons of CO_2 equivalent emissions reduction per year, establish that qualification accuracy improvement generates societal value that extends substantially beyond operational cost reduction.

Future research should prioritize three directions that remain incompletely resolved by the current platform architecture. Transfer learning approaches for cold-start qualification accuracy at newly activated infrastructure sites, where empirical coverage validation data does not yet exist, represent the most immediate accuracy improvement opportunity. Dynamic obstruction-aware line-of-sight computation integrating real-time pedestrian and vehicular occlusion data would meaningfully reduce residual qualification errors in high-density urban environments where static building geometry alone is insufficient. Federated learning architectures enabling privacy-preserving demand intelligence synthesis across multi-operator and multi-jurisdiction deployments would unlock the national broadband equity applications that centralized qualification failure aggregation cannot serve. As sixth-generation network architectures advance toward sub-terahertz carrier frequencies with effective service ranges measured in tens of meters, the architectural patterns demonstrated in this review, real-time hexagonal spatial intelligence, autonomous multi-step qualification reasoning, and natural language data accessibility, will become not optional enhancements but foundational requirements of next-generation telecommunications platform design.

References

[1] T. S. Rappaport et al., "Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond," IEEE Access, vol. 7, pp. 78729–78757, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8732419

[2] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," IEEE Communications Magazine, vol. 58, no. 3, pp. 55–61, Mar. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9040431

[3] H. Elshaer, M. N. Kulkarni, F. Boccardi, J. G. Andrews, and M. Dohler, "Downlink and uplink cell association with traditional macrocells and millimeter wave small cells", IEEE Transactions on Wireless Communications, vol. 15, no. 9, pp. 6244–6258, Jun. 2016. [Online]. Available: https://ieeexplore.ieee.org/document/7501510

[4] S. Yao et al., "ReAct: Synergizing reasoning and acting in language models," in Proc. 11th International Conference on Learning Representations (ICLR), 2023. [Online]. Available: https://arxiv.org/abs/2210.03629

[5] A. Kumar, P. Nagarkar, P. Nalhe, and S. Vijayakumar, "Deep learning-driven natural language text-to-SQL query conversion: A survey", arXiv:2208.04415, Aug. 2022. [Online]. Available: https://arxiv.org/abs/2208.04415

[6] P. Lewis et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," Advances in Neural Information Processing Systems, vol. 33, pp. 9459–9474, 2020. [Online]. Available: https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html

[7] A. Mukherjee, D. De, S. K. Ghosh, and R. Buyya, Eds., Mobile Edge Computing. Springer Nature, 2021.

[8] K. Sahr, "Hexagonal discrete global grid systems for geospatial computing," Archives of Photogrammetry, Cartography and Remote Sensing, vol. 22, pp. 363–376, 2011. [Online]. Available: https://doi.org/10.2478/v10217-011-0025-1

[9] B. A. Raddwan and I. A. Al-Baltah, "Line-of-sight probability models for 5G/6G millimetre-wave networks: A review", Sana'a University Journal of Applied Sciences and Technology, vol. 2, no. 5, pp. 440–452, Oct. 2024.

[10] L. Wang et al., "A survey on large language model-based autonomous agents", Frontiers of Computer Science, vol. 18, no. 6, p. 186345, Dec. 2024. [Online]. Available: https://link.springer.com/article/10.1007/s11704-024-40231-1

[11] Q. Wu et al., "AutoGen: Enabling next-gen LLM applications via multi-agent conversation," arXiv:2308.08155, Aug. 2023. [Online]. Available: https://arxiv.org/abs/2308.08155

[12] J. Wei et al., "Chain-of-thought prompting elicits reasoning in large language models," Advances in Neural Information Processing Systems, vol. 35, pp. 24824–24837, Dec. 2022. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2022/hash/9d5609613524ecf4f15af0f7b31abca4-Abstract-Conference.html

[13] A. Imran, A. Zoha, and A. Abu-Dayya, "Challenges in 5G: How to empower SON with big data for enabling 5G," IEEE Network, vol. 28, no. 6, pp. 27–33, Nov. 2014. [Online]. Available: https://ieeexplore.ieee.org/document/6963801

[14] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges," IEEE Communications Surveys & Tutorials, vol. 25, no. 2, pp. 1376–1411, Jan. 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10004725

[15] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated learning for wireless communications: Motivation, opportunities, and challenges," IEEE Communications Magazine, vol. 58, no. 6, pp. 46–51, Jul. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9141214

[16] D. Gao et al., "Text-to-SQL empowered by large language models: A benchmark evaluation," arXiv:2308.15363, Aug. 2023. [Online]. Available: https://arxiv.org/abs/2308.15363

[17] Q. Guo et al., "Connecting large language models with evolutionary algorithms yields powerful prompt optimizers," in Proc. 12th International Conference on Learning Representations (ICLR), 2024. [Online]. Available: https://arxiv.org/abs/2309.08532

[18] J. Park, S. Samarakoon, M. Bennis, and M. Debbah, "Wireless network intelligence at the edge," Proceedings of the IEEE, vol. 107, no. 11, pp. 2204–2239, Nov. 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8865671

[19] I. Brodsky, "H3: Uber's hexagonal hierarchical spatial index," Uber Engineering Blog, 2018. [Online]. Available: https://www.uber.com/en-US/blog/h3/

[20] Y. Shen, Y. Shao, L. Xi, H. Zhang, and J. Zhang, "Millimetre-wave propagation measurement and modeling in indoor corridor and stairwell at 26 and 38 GHz", IEEE Access, vol. 9, pp. 87792–87805, May 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9433630

[21] H. Tsukada et al., "Millimeter-wave urban cellular channel characterization and recipe for high-precision site-specific channel simulation", TechRxiv, 2024. [Online]. Available: https://doi.org/10.36227/techrxiv.170831411.10437905/v1

[22] Jiang G, Li W, Yu C, Zhu Z, Li W. FGCSQL: A three-stage pipeline for large language model-driven Chinese text-to-SQL. Electronics. 2025 Mar 19;14(6):1214.

[23] Hsueh NL, Wang WT. Development and Evaluation of Learning Portfolio Query System Based on LangChain Framework. Engineering Proceedings. 2025 Apr 30;92(1):40.

[24] Fokin G, Volgushev D. Model for interference evaluation in 5G millimetre-wave ultra-dense network with location-aware beamforming. Information. 2023 Jan 9;14(1):40.

[25] Uwaechia AN, Mahyuddin NM. A comprehensive survey on millimetre-wave communications for fifth-generation wireless networks: feasibility and challenges. IEEE Access. 2020 Mar 30;8:62367-414.

[26] Oughton EJ, Lehr W. Surveying 5G techno-economic research to inform the evaluation of 6G wireless technologies. IEEE Access. 2022 Feb 21;10:25237-57.