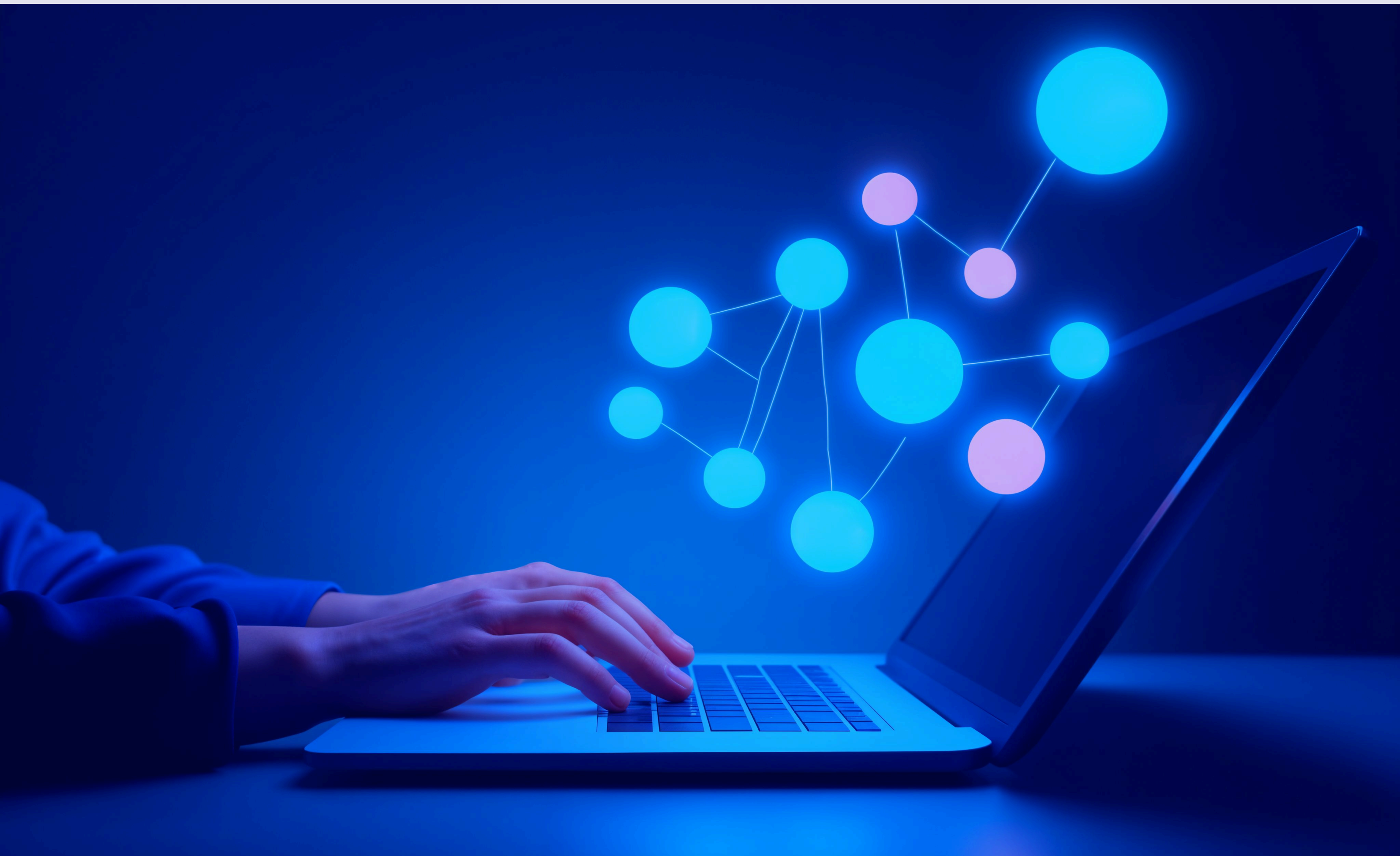
perimattic

AI in Network Optimization: From Self-Healing to Predictive Capacity Planning

JULY 2025



perimattic

Table Of Contents

Introduction

Today, connectivity underpins our daily lives and telecommunications networks play a foundational role in enabling digital interactions, commerce, mobility, and critical services. These networks must operate with precision and resilience to meet rising expectations for speed, availability, and seamless user experiences.

To deliver on this demand, communication service providers must continuously fine-tune how data moves across the network. This process, known as network optimization, **involves improving performance, efficiency, and reliability by managing and adjusting resources, traffic flow, and configurations**. It ensures uninterrupted service while preparing the network to scale and adapt to evolving technologies and future demand.

As mobile networks become more expansive and complex, traditional rule-based systems fall short in managing scale and performance. With 5G deployments and edge computing expanding, communication service providers (CSPs) must operate networks that are dynamic and predictive.

Artificial Intelligence (AI) enables this shift. By analysing live and historical network data, AI can automate optimization decisions, proactively resolve issues, and align network behaviour with business goals. The result is a high-performing, self-improving network that continuously adapts to changing conditions and future demand.

In 2025, Verizon demonstrated the real-world application of this shift. In partnership with Samsung and Qualcomm, it deployed a multi-vendor RAN Intelligent Controller (RIC) in its commercial network. This deployment integrated Samsung's AI-powered Energy Saving Manager (AI-ESM) with Qualcomm's Dragonwing™ RAN Automation Suite, enabling automated energy-efficient decisions without operator intervention. This is one of many steps toward an AI-optimized, intent-driven network architecture.

Today, the global AI in telecommunication market is projected to grow from USD 2.66 billion in 2025 to around USD 50.21 billion by 2034, registering a CAGR of 38.81%. While it is clear that network optimization demands are driving the growth, there is a need to look at this prospect from multiple angles.

In this whitepaper, we will discuss:

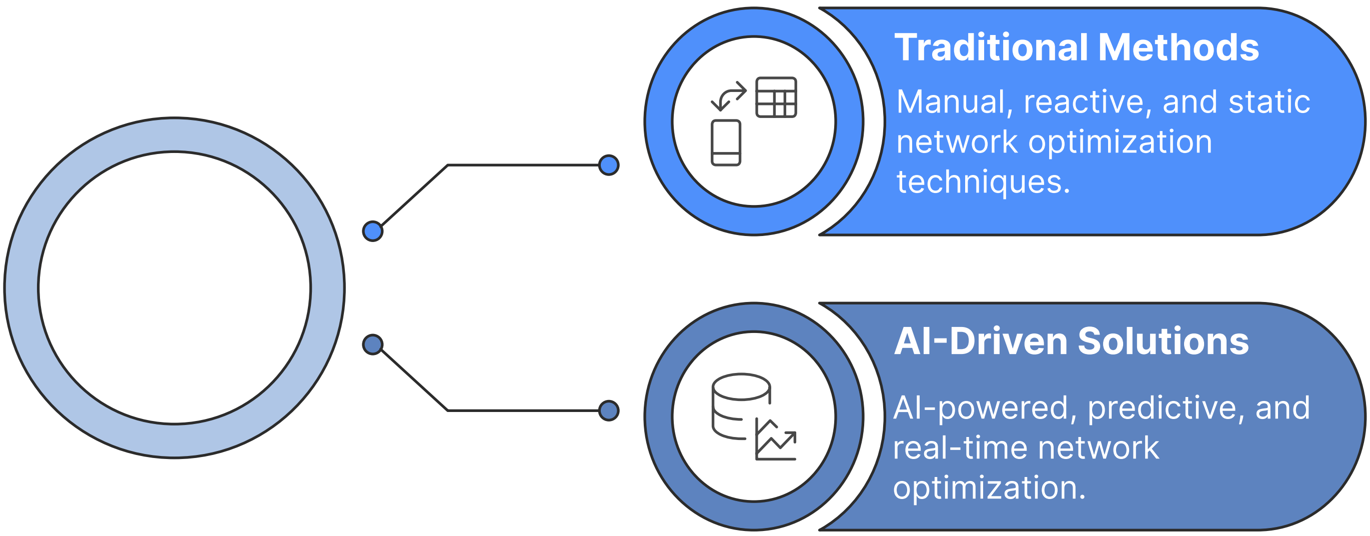
1. The shift from traditional to AI-optimized, intent-based network operations
2. How self-healing networks work and what value they offer
3. Predictive capacity planning driven by AI
4. The data and infrastructure challenges associated with AI adoption in network optimization

From Traditional to Intent-Based Network Optimization

It is not an exaggeration to say that the history of network communications is one of the most fascinating one in the tech domain. From early non-electric signalling systems to today's high-speed, software-defined infrastructures, we have come a long way.

The leap from copper wiring to fibre optics in the 1960s marked a turning point, enabling vast improvements in bandwidth and signal clarity. The mobile revolution of the 1980s and 1990s untethered communication from physical locations, giving rise to global cellular networks. Each phase of evolution has expanded the scale, complexity, and expectations placed on network infrastructure.

Today's networks are defined by virtualization, cloud-native architectures, and real-time demand from billions of connected devices. The proliferation of IoT, the rollout of 5G, and the increasing use of edge computing have created highly dynamic environments where traffic patterns shift constantly and performance requirements vary by application. Traditional, rule-based approaches to network management can no longer keep up. Modern optimization requires systems that can self-learn, adapt, and respond autonomously. This is where AI becomes central in keeping networks operational and enabling them to continuously improve to align with strategic objectives.



perimattic

Rule-Based vs. Intent-Based Networks

Feature	Rule-Based Network	Intent-Based Network
Decision Logic	Predefined static rules	AI/ML interprets intent and adjusts behaviour
Fault Handling	Manual detection and remediation	Automated issue detection and self-healing
Scalability	Limited to human capacity	Scales dynamically with network growth
Configuration Management	Manual provisioning and scripts	Automated, intent-driven configurations
Optimization Approach	Reactive	Proactive and predictive
Alignment with Business Goals	Limited	Aligned with SLAs and strategic objectives

Why CSPs Are Moving Toward Intent-Based Models

Intent-based networking allows operators to define outcomes (e.g., reduce latency for video traffic in a metro area), and the AI system automatically determines the best way to fulfil that intent in real-time.

Drivers of adoption:

- Multi-vendor environments that require dynamic integration
- 5G slicing that demands granular traffic prioritization
- Edge workloads requiring ultra-low latency
- Operational efficiency at scale without increasing headcount

How is this done:

- Identify patterns that precede service degradation
- Predict congestion before it occurs and redistribute traffic
- Automatically reroute traffic in response to early fault signatures
- Adjust capacity provisioning in real time based on demand shifts

There is a tangible change in how networks are operated: from reacting to faults after they occur to proactively ensuring optimal performance before issues arise. Instead of waiting for a base station to fail and dispatching a technician, AI-driven systems can detect early signs of degradation and trigger automated rebalancing or traffic rerouting before any user experiences service disruption.

For example, in a dense urban environment, AI can continuously analyse network usage patterns to anticipate peak congestion zones. It can be near a stadium before an event or a rally near city townhall. The systems can allocate additional capacity in advance to maintain speed and reliability. In rural deployments, predictive models can detect patterns associated with equipment fatigue, such as gradual drops in backhaul efficiency, and alert the operators days before a complete outage.

AI enables these capabilities by learning from vast datasets across the network: telemetry, historical incidents, configuration changes, and real-time traffic flow. Over time, it builds a probabilistic understanding of how and where problems are likely to occur. Instead of issuing alerts after SLA thresholds are breached, the system identifies leading indicators and adjusts resources, configurations, or topologies accordingly.

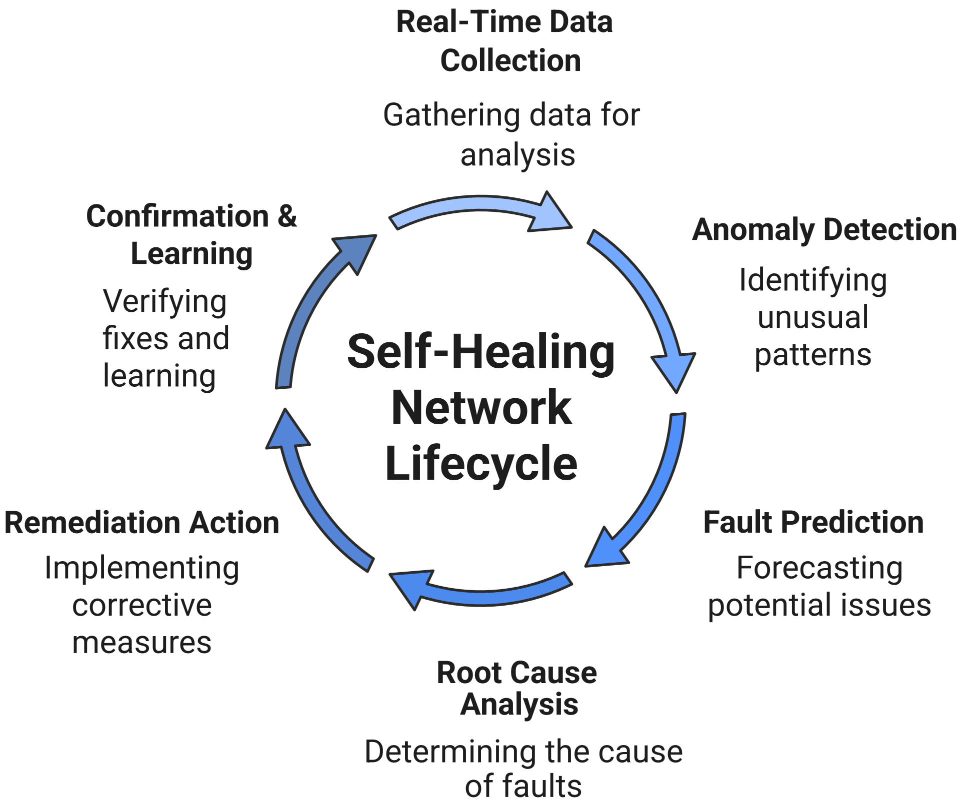
How Self-Healing Networks Work and the Value They Offer

Self-healing networks are built to **autonomously detect, diagnose, and recover from faults** with minimal or no human intervention. These networks continuously monitor their operational health, analyse anomalies in real time, and initiate corrective actions to maintain service continuity. As networks scale in complexity, particularly with 5G, edge computing, and IoT, the ability to self-recover becomes essential to achieving operational reliability and low downtime.

At the core of self-healing capabilities is **AI-driven automation**. Advanced machine learning models analyse telemetry data for latency, jitter, error rates, and throughput. This is done to identify fault precursors before failures fully manifest. Once a fault or performance degradation is detected, the system determines the optimal mitigation strategy: rerouting traffic, spinning up backup resources, adjusting bandwidth, or isolating faulty nodes.

These systems reactive and they learn from each incident to improve future responses. Over time, this closed-loop learning enables the network to adapt, reducing the frequency and severity of recurring issues.

Core Architecture of a Self-Healing Network



1. **Telemetry and Monitoring Layer**
    - Continuously collects real-time metrics from devices, links, and services
    - Supports streaming protocols (e.g., gNMI, SNMP, NetFlow)
    - Normalizes data for consumption by AI/ML systems

2. **Fault Prediction Engine**
    - Applies ML algorithms to identify behavioural patterns that indicate impending failure
    - Uses classification models (e.g., decision trees, neural networks) to categorize incidents
    - Integrates feedback loops to retrain on post-failure data and improve accuracy

3. Root Cause Analysis (RCA) Engine

- Correlates alerts across multiple layers (transport, core, access) to pinpoint the origin
- Uses graph traversal and dependency mapping to isolate fault domains
- Interfaces with historical incident databases to accelerate RCA

4. Healing Execution Layer

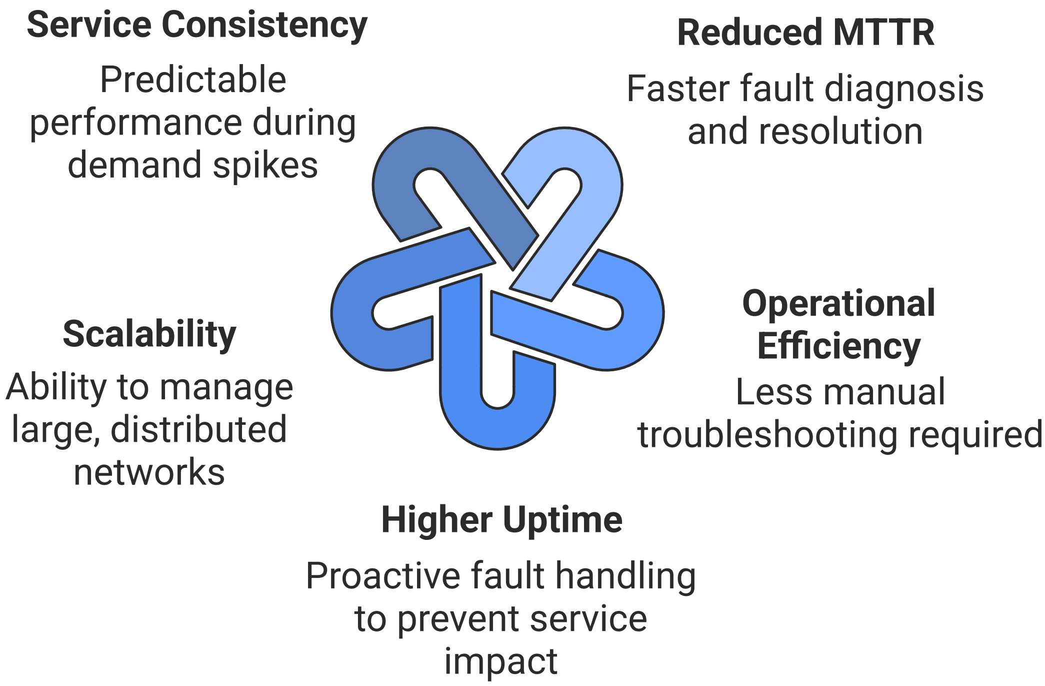
- Deploys remediation scripts or triggers predefined playbooks
- Actions may include route reconvergence, service migration, or interface resets
- Interfaces with SDN controllers for automated path adjustments

5. Policy and Governance Layer

- Ensures healing actions comply with operational policies and regulatory constraints
- Logs all actions for auditability and post-incident reviews

perimattic

Value Delivered by Self-Healing Networks



A study titled Self-Healing Networks with AI-Based Fault Prediction in IoT Ecosystems demonstrates that the integration of advanced machine learning techniques with autonomous network management protocols significantly improves fault tolerance, reduces downtime, and enhances system reliability....

In 2020, Spark, a New Zealand-based telecommunications provider, deployed the first phase of its next-generation Optical Transport Network (OTN 2). This infrastructure includes self-healing capabilities that automatically reroute traffic in the event of link failures, environmental damage, or physical disruptions, such as those caused by natural disasters.

- Operates at 800 Gbps, an 8x increase over prior infrastructure
- Connects cities and towns through fibre-based backbones, transporting customer traffic across all domains (mobile, broadband, enterprise)
- Ensures resilience and performance for Spark’s expanding 5G services
- First deployment of this type in New Zealand, showcasing self-healing as a backbone feature

perimattic

Predictive Capacity Planning with AI

Predictive capacity planning is the process of forecasting future network demand and provisioning resources in advance to maintain optimal performance. Unlike reactive scaling, where capacity is increased after congestion occurs, predictive planning uses AI models to anticipate future bottlenecks and adjust infrastructure or routing before service degradation happens.

It involves:

- Forecasting traffic trends across time, location, and user profiles
- Predicting resource needs across layers (spectrum, backhaul, compute)
- Prioritizing expansion projects based on demand projections

How AI Drives Predictive Planning

AI enables dynamic, data-driven decisions by analysing:

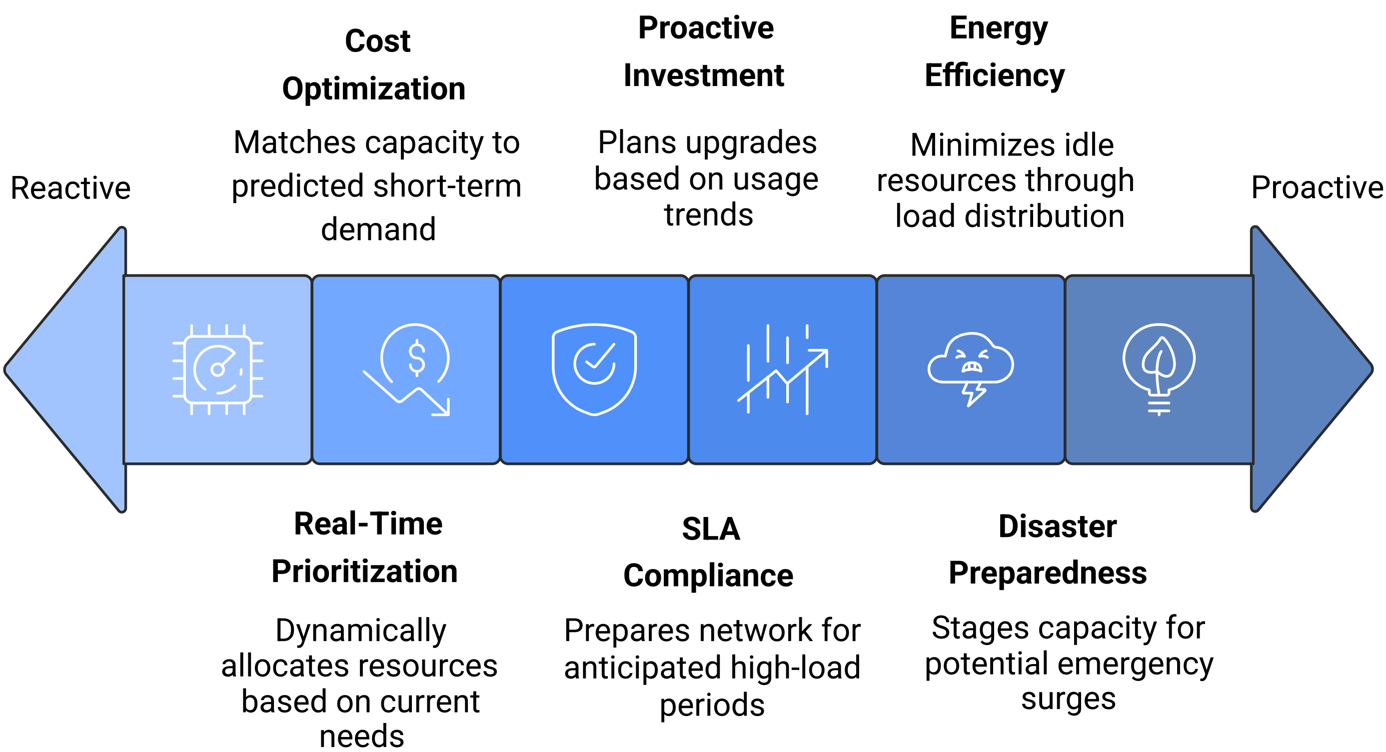
- Historical usage patterns
- Real-time traffic telemetry
- External variables like events, weather, or regional movement
- Seasonality in traffic and user behaviour

Machine learning models such as LSTM (Long Short-Term Memory) networks, ARIMA models, and reinforcement learning agents can continuously update capacity forecasts based on live input. These models evolve over time, making capacity planning not just automated, but context-aware and adaptive.

perimattic

Benefits of AI-Driven Predictive Capacity Planning

Network capacity planning ranges from reactive to proactive



There are several real-world examples where AI-first network optimization, self-healing and/or predictive capacity planning is successfully in action.

India's Jio has built a powerful platform called JioBrain, which integrates high-speed 5G connectivity with machine learning capabilities such as anomaly detection, predictive forecasting, and automation. JioBrain can deploy ML models at both the network edge and the cloud, enabling:

- Capacity planning that adapts in real time
- Seamless scaling for healthcare, education, and entertainment applications
- Edge inference for low-latency AI use cases like video optimization or AR experiences

JioBrain uses **generative AI** to parse vast telemetry datasets and simulate network behaviour under various conditions, helping engineers pre-emptively mitigate congestion or service degradation. This platform marks a milestone in how AI can be embedded in the infrastructure as well as the services running on top of it. Jio has shown how AI can be a game changer for network optimization in a country like India. It is a massive country where scale and geographic diversity are often obstacles and AI enables faster, more adaptive network management.

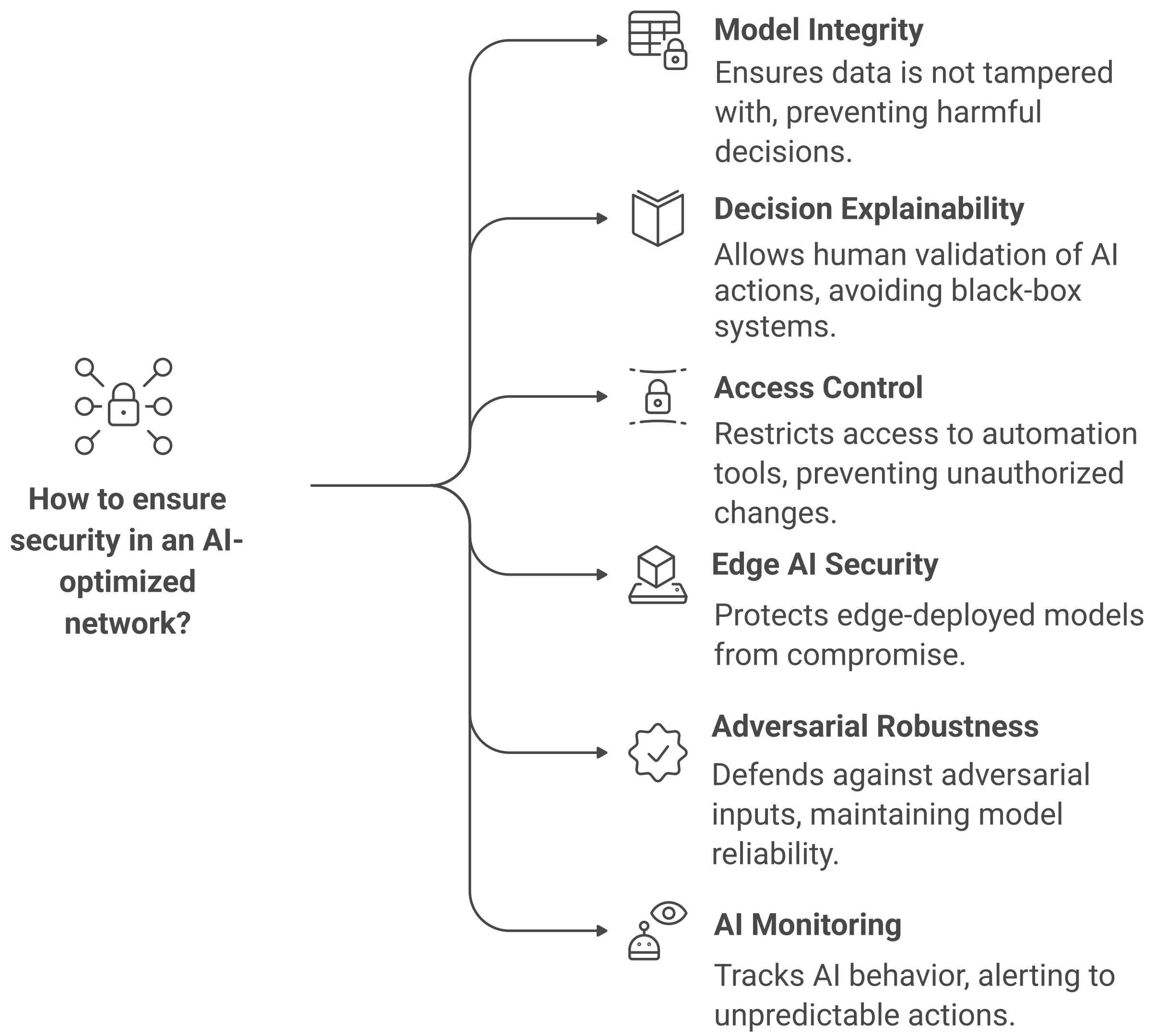
Another example is how AT&T integrates AI/ML into every phase of its network lifecycle. This includes:

- Selecting future **cell tower locations** based on forecasted demand.
- Optimizing spectrum allocation to avoid underuse or congestion.
- Automating configuration of new **network slices** to balance enterprise and consumer loads.
- Using **AI simulation environments** to test and validate network changes before roll-out.

This systemic use of AI enables AT&T to scale with precision and deliver consistent user experiences despite a constantly shifting traffic situation.

Security Considerations in AI-Optimized Networks

As AI becomes embedded in core network functions, from fault detection to predictive scaling, it also introduces new vectors of vulnerability. A compromised AI system can misconfigure network paths, suppress valid alerts, or redirect traffic in harmful ways. Therefore, designing secure AI systems is a foundational requirement for self-healing and predictive architectures.



There is a clear need for new levels of transparency, security, and human oversight. Afterall, **AI is not a replacement for engineering expertise**, it’s an augmentation layer that, when implemented thoughtfully, enables more resilient, scalable, and intelligent network operations.

Data and Infrastructure Challenges in AI-Driven Network Optimization

The integration of AI into network optimization brings with it significant technical, operational, skill-based, and financial hurdles. As telecom networks become increasingly software-defined and distributed, the successful deployment of AI systems requires investments in technology, people, process redesign, and long-term change management.

One of the most immediate challenges is cost. Implementing AI capabilities into a network stack involves high upfront investments in infrastructure, software platforms, skilled personnel, and integration efforts. Telecom operators often need to upgrade their data pipelines, deploy additional compute resources for training and inference, and overhaul legacy monitoring tools to support real-time telemetry. These are not one-time expenses; they require **ongoing tuning, model retraining, and operational support**.

Upfront costs include not just the deployment of AI tools but also the modernization of data infrastructure. AI models demand vast volumes of high-quality, structured, real-time data, which often necessitates re-architecting legacy systems that were never built for dynamic, automated decision-making. As real-time data collection increases across all network layers: core, transport, RAN, and edge, the storage, transfer, and processing of this data becomes a growing operational concern.

Infrastructure cost extends to cloud or hybrid cloud deployments. Many AI models, particularly those supporting self-healing or predictive capacity planning, are compute-intensive. Organizations must decide whether to invest in on-premises hardware with high-performance GPUs or use elastic cloud services, which may present new challenges in latency, data residency, and integration complexity.

In parallel, staffing costs and upskilling demand focused investment. **AI in telecom is not a plug-and-play solution. It requires a workforce that understands both network engineering and data science**. Training network operations teams to interpret AI-generated insights, manage model drift, and oversee system behaviour requires structured upskilling programs or new hiring altogether.

Data management is another persistent challenge in deploying self-healing capabilities. Self-healing network functionality relies on vast datasets to detect anomalies, forecast failures, and initiate corrective actions autonomously. But managing this data at scale poses several hurdles:

- Ensuring **data integrity** and **accuracy** is fundamental, as AI decisions are only as good as the data they're based on.
- **Data security and privacy** must be preserved throughout the data lifecycle, especially when data is transmitted between edge devices and cloud systems.
- There is a risk of **false positives and false negatives** in AI-driven fault detection, which can lead to unnecessary remediation actions or missed incidents.
- **Model robustness** becomes an issue in highly dynamic environments, where network conditions change rapidly and require models that are not just accurate but also adaptable.

A self-healing architecture must be integrated seamlessly with existing network management systems, which is rarely straightforward. Many telecom infrastructures include legacy systems that operate on proprietary protocols or outdated architectures. AI modules must be made interoperable without compromising system stability or security.

According to an EY study,

- 50% of organizations report difficulty in identifying the right type of generative AI vendor, and 53% struggle with choosing the right 5G deployment model.
- This lack of clarity on AI model selection and deployment strategy often slows down the transition to proactive, AI-driven operations.
- Compounding this is a broader concern: many enterprises do not believe that technology suppliers fully understand their business imperatives.
- The top area where they seek improvement is in deeper collaboration and clearer articulation of how emerging technologies translate into operational value.

Skill gaps add to the complexity. AI-powered optimization is not a traditional IT problem. It spans across network engineering, AI modelling, infrastructure management, and governance. Many telcos are facing the reality of having to build cross-functional teams from scratch, combining engineers, data scientists, and security experts. These skills are often in short supply and come at a premium in the global talent market.

Barriers to AI-Driven Self-Healing Networks

Recent research from the International Journal of Applied and Behavioural Sciences has consolidated a wide range of challenges faced by telecom operators deploying AI-powered self-healing networks in IoT-rich environments. These challenges, though cantered around IoT, are directly applicable to broader telecom infrastructure:

- **False Positives and Negatives**: AI models may generate inaccurate alerts, making algorithm tuning an ongoing challenge.
- **Limited Data Availability:** Incomplete or low-quality data restricts the ability to train robust models.
- **Security Concerns:** AI systems are vulnerable to adversarial attacks and data tampering.
- **Complexity of IoT Ecosystems:** Diverse device types, communication protocols, and environmental conditions increase integration complexity.
- **Interoperability Issues:** Ensuring that AI tools communicate effectively with heterogeneous network components remains difficult.
- **Resource Constraints:** Limited bandwidth, memory, or compute at the edge affects model performance.
- **Ethical and Legal Concerns:** Autonomous decision-making raises questions around accountability and compliance.
- **Adaptability to Evolving Threats:** AI models must evolve continuously to handle emerging attack vectors and failure patterns.
- **Dependence on Connectivity:** If connectivity fails, edge-based healing mechanisms may also be disabled.
- **Complex Debugging:** Tracing root causes through layers of automation and AI inference makes troubleshooting difficult.
- **High Implementation Costs:** Especially for smaller players or operators with fragmented networks.
- **Resistance to Change:** Organizational culture and fear of automation create resistance among existing teams.

perimattic

Plausible Solutions

As AI-adaption becomes mainstream necessity, we propose various solutions that can positively address the aforementioned challenges.

Prioritize Data Infrastructure Before Model Deployment

Before investing in advanced AI models, telecom operators should prioritize building a stable and scalable data infrastructure. This includes streamlining data ingestion across devices, establishing normalized data schemas, and integrating real-time telemetry pipelines. Without reliable, clean, and high-volume data, even the most advanced models will fail to deliver value.

Establishing centralized data lakes and standard APIs allows AI teams to iterate quickly, test models across consistent datasets, and apply federated learning across edge and cloud domains. Operators should also invest in metadata management to enable traceability and transparency in data lineage, critical for debugging and auditability.

Implement Modular AI Architectures

Deploying AI models in modular, interoperable components helps reduce the risk of failure and enables phased rollout. Start with narrow use cases like anomaly detection in a specific segment of the RAN or transport layer, then scale based on performance validation. Adopt containerized deployments and microservices to simplify integration with legacy infrastructure.

These modules should be designed to interact via secure APIs and policy-based orchestration layers. This allows network operators to gradually build toward more advanced use cases like root cause prediction or autonomous healing without overhauling the entire infrastructure.

Plausible Solutions

As AI-adaption becomes mainstream necessity, we propose various solutions that can positively address the aforementioned challenges.

Prioritize Data Infrastructure Before Model Deployment

Before investing in advanced AI models, telecom operators should prioritize building a stable and scalable data infrastructure. This includes streamlining data ingestion across devices, establishing normalized data schemas, and integrating real-time telemetry pipelines. Without reliable, clean, and high-volume data, even the most advanced models will fail to deliver value.

Establishing centralized data lakes and standard APIs allows AI teams to iterate quickly, test models across consistent datasets, and apply federated learning across edge and cloud domains. Operators should also invest in metadata management to enable traceability and transparency in data lineage, critical for debugging and auditability.

Implement Modular AI Architectures

Deploying AI models in modular, interoperable components helps reduce the risk of failure and enables phased rollout. Start with narrow use cases like anomaly detection in a specific segment of the RAN or transport layer, then scale based on performance validation. Adopt containerized deployments and microservices to simplify integration with legacy infrastructure.

These modules should be designed to interact via secure APIs and policy-based orchestration layers. This allows network operators to gradually build toward more advanced use cases like root cause prediction or autonomous healing without overhauling the entire infrastructure.

Address Skills Gaps with Hybrid Teams and Upskilling Programs

Building cross-functional AI teams is key. Hybrid teams that include telecom engineers, data scientists, cybersecurity professionals, and AI operations specialists should collaborate on model development, deployment, and oversight. Simultaneously, invest in structured internal upskilling programs focused on AI literacy for network teams and domain training for data science teams.

Collaborations with academia and research institutions can help accelerate internal capabilities. Telecom operators should also consider industry certifications and role-based training pathways to ensure AI deployment is sustainable and supported internally.

Establish AI Governance and Trust Frameworks

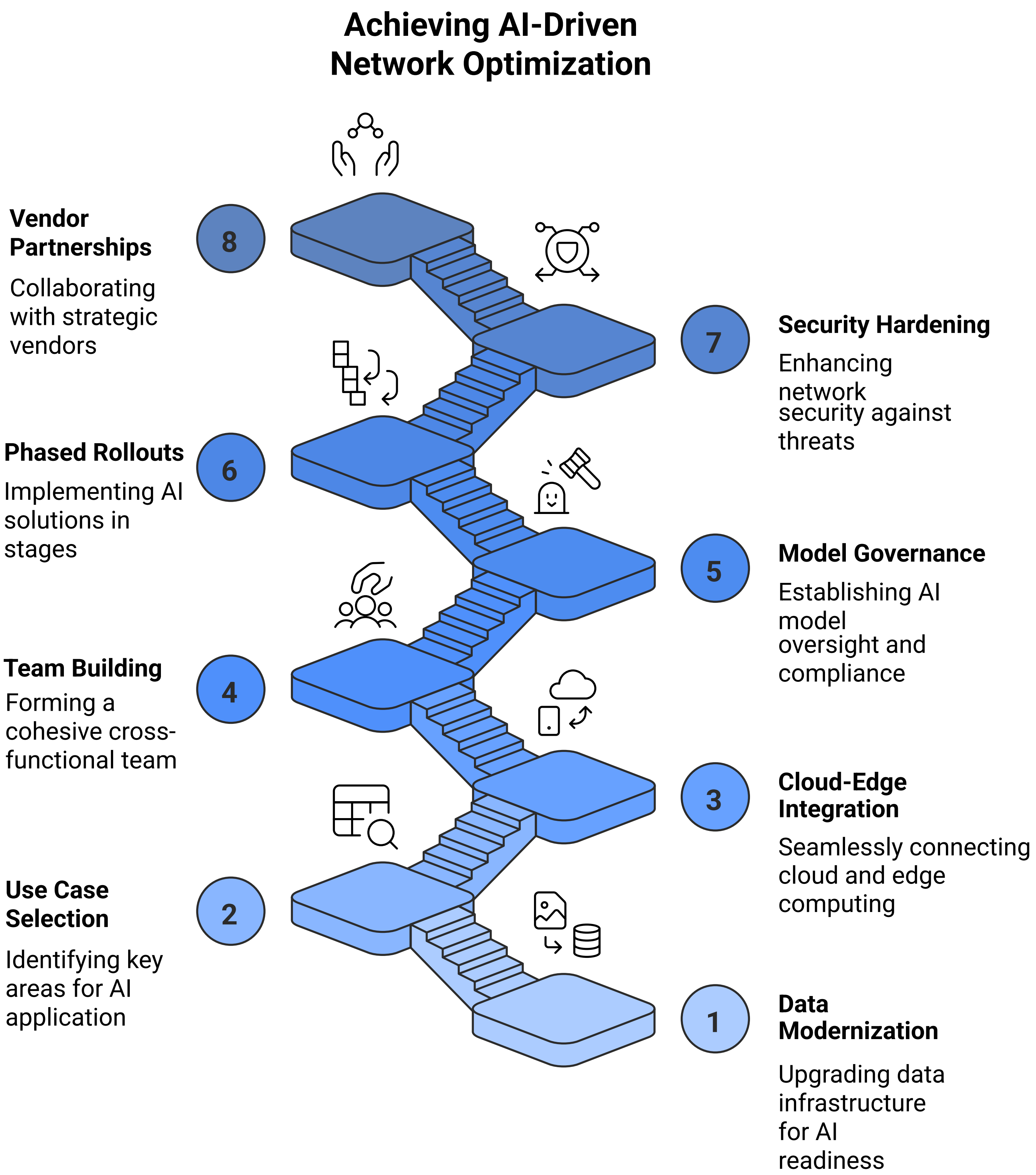
To manage risks, operators must implement robust AI governance frameworks. This includes:

- Ongoing validation of model performance across scenarios.
- Bias detection and correction protocols.
- Explainability tools for model interpretability.
- Incident response playbooks for AI malfunction or error propagation.

Clear accountability structures are essential. Human-in-the-loop mechanisms should be maintained for high-impact actions until models have been extensively validated in production environments. Model decisions must be auditable, especially in regulated markets where SLA violations carry financial consequences.

perimattic

Future Roadmap and Strategic Takeaways



Conclusion

As telecommunications networks grow more distributed, intelligent, and service-driven, the role of AI in network optimization is shifting from experimental to essential. Self-healing systems are reducing downtime and manual intervention. Predictive capacity planning is aligning network behaviour with future demand. These capabilities are enabling CSPs to deliver consistent service at scale, while unlocking operational efficiency and resilience.

However, these benefits are contingent on addressing core challenges: data reliability, infrastructure readiness, skills gaps, and security risks. AI models must be deployed with transparency, oversight, and a deep understanding of the telecom environment. The transformation is as much strategic as it is technical.

AI is becoming the orchestration layer for next-generation networks. For decision-makers, now is the time to assess how AI fits into your broader operational roadmap. Consider the potential of agentic AI systems, those capable of learning, adapting, and taking context-aware action autonomously. These systems represent the next step toward networks that anticipate and shape the demands.

perimattic

Custom, Secure And Scalable Digital Solutions For Your Business

Perimattic builds digital solutions that change how your business works from automation to cybersecurity and everything in between.

Contact Us Now

US +1 (442) 319-7124

UK +44 (20) 3769 0051

IND +91 92142 66896

sales@perimattic.com

www.perimattic.com

Follow Us