

# Strategies for AI adoption in fixed networks

Challenges, use cases and future directions

**First edition – April 2026**



## About the authors

Organisation	Contributing Authors
Fraunhofer HHI	Behnam Shariati (Editor), Angela Mitrovska, Mihail Balanici, Pooyan Safari, Hussein Zaid, Johannes Fischer, Ronald Freund
Deutsche Telekom	Matheus Sena, Rainer Schatzmayr
POST Luxembourg	Olivier Ferveur
NOKIA	Nicolas Dupuis
Telefonica	Juan Pedro Fernandez-Palacios Gimenez
China Telecom	Jian Tang
MTN Nigeria	Maria-Gorretti Dokubo, Maxwell Aniako
Infosim	Vanessa Breitenbach, Matthias Ebert, David Hock
Huawei Paris Research Center	Yvan Pointurier
Waystream	Johan Sandell
Lunet	Daniel Henriksson
RISE	Henrik Abrahamsson
Savantic	Karoly Makonyi
Dacoso	José Juan Pedreño Manresa
UC3M	Farhad Arpanaei

## Contents

<b>About the authors</b>	<b>2</b>
<b>Contents</b>	<b>3</b>
<b>Executive summary</b>	<b>5</b>
<b>Introduction</b>	<b>5</b>
<b>Evolution of AI: from statistical models to LLMs</b>	<b>7</b>
Early AI and statistical approaches	7
Rise of machine learning and deep learning	8
Emergence of generative AI and large language models (LLMs)	10
<b>Current state of AI in fixed networks</b>	<b>11</b>
Exemplary proofs of concept	13
AI-based dynamic link-capacity allocation	13
Diagnose of VDSL Line via Pattern Recognition	15
LLM-assisted and DT-driven network automation	15
Lessons learned	16
Real-world deployments	17
Gaps and challenges in widespread deployment	19
<b>Use cases and applications of AI for networks</b>	<b>20</b>
Network planning and design	20
Traffic forecasting for capacity planning	20
Data enhancement and traffic pattern recognition	20
Automated network design and optimisation	21
Network operation, maintenance and automation	21
Operators' viewpoints	21
Security, reliability and compliance	23
Stakeholder perspectives (customers, vendors, operators and regulators)	25
<b>Technical requirements and challenges in AI for networks</b>	<b>27</b>
Architecture and infrastructure requirements	27
Infrastructure monitoring	27
Data management and lifecycle	28
The importance of high-quality data for AI in fixed networks	28
AI deployment and integration	31

---

Regulatory and compliance considerations	33
<b>Networks for AI</b>	<b>35</b>
Evolving infrastructure demands for AI workloads	35
Data centre-to-data centre connectivity and cloud integration	35
High-performance computing considerations (latency, bandwidth and reliability)	37
Impact on network architectures and upgrades	39
<b>Relation to ETSI standardisation Activities</b>	<b>43</b>
<b>Conclusions and next steps</b>	<b>45</b>
<b>List of abbreviations and definitions</b>	<b>46</b>

## Executive summary

This White Paper provides an overview of adopting AI in fixed networks, moving from individual siloed research activities and proofs of concept (PoC) to real-world deployments and delivering an AI-native, intent-driven, self-operating infrastructure. It explains how advances from statistical machine learning (ML) to deep learning (DL) and large language models (LLMs) are reshaping network planning, operations, assurance, security, and customer experience. It also briefly addresses “Networks for AI,” outlining the transport and data centre interconnect upgrades required to support AI workloads. The report consolidates lessons from proofs of concept and live deployments, identifies gaps hindering scale, and discusses potential actions for ETSI and the industry to standardise interfaces, data, governance, and assurance of AI systems in multi-vendor fixed network environments. More concretely, the White Paper’s content can be outlined as follows:

- ✓ Part 1 summarises the evolution of AI with emphasis on GenAI/LLMs and their relevance to fixed networks, including operator-facing copilots, prompt engineering, fine-tuning, and retrieval-augmented grounding.
- ✓ Part 2 reviews the current state of adoption, summarising lessons from PoCs and real deployments.
- ✓ Part 3 categorises high-impact use cases across planning/design (traffic forecasting, automated design), operations/assurance (incident resolution, anomaly detection, log/config analysis), operational readiness (digital twins), closed-loop automation, sustainability, and optical access networks.
- ✓ Part 4 identifies technical requirements and gaps: architecture and telemetry, data lifecycle and governance, AI deployment/MLOps, reliability/robustness/security, and regulatory compliance (e.g., privacy by design pipelines).
- ✓ Part 5 examines “Networks for AI,” covering DC-to-DC connectivity, cloud integration, latency/bandwidth/reliability targets for AI/HPC, and the implications for optical transport and access upgrades.

## Introduction

Artificial intelligence (AI), including generative AI (GenAI) and large language models (LLMs), is rapidly reshaping how fixed networks are designed, operated, and automated. Yet, despite abundant research, PoCs, and early deployments, broad, trustworthy, multi-vendor adoption in live networks remains challenging. This White Paper clarifies the current state of AI adoption across the lifecycle of fixed networks (planning and design, control, operations, assurance, and automation), with a primary focus on capabilities that improve operational efficiency, resiliency, and user experience. Specifically, it consolidates multiple perspectives from vendors, operators, and researchers to map the evolving landscape of AI-driven network automation. While certain examples stem from ongoing industry roadmaps, the intent is to illustrate trends rather than endorse specific products.

The document:

- Reviews successful pilots and real-world deployments and presents lessons learned.
- Identifies gaps preventing scaling (e.g., data governance, interoperability, trust/assurance, skills).
- Categorises AI, GenAI, and LLM-enabled use cases for fixed networks and outlines their requirements and business incentives.
- Summarises technical requirements (architecture, data/telemetry, MLOps, security, compliance) for reliable AI integration.
- Recommends a roadmap and actions for Standards Developing Organisations (SDOs) - particularly ETSI - to catalyse interoperable, secure, and AI-native fixed networks.
- Explores “Networks for AI,” i.e., how AI workloads (training and inference) impact fixed and optical transport architectures (e.g., DC-to-DC connectivity).

The scope is limited to fixed networks (access, aggregation/metro, optical transport, and data centre interconnect) and “AI for Networks,” with a complementary view on “Networks for AI.” The intended audience includes operators, vendors, researchers, and SDOs seeking pragmatic guidance on deploying trustworthy, interoperable AI at scale in multi-vendor environments.

The document is structured as follows. The next chapter, *Evolution of AI: from statistical models to LLMs* that traces AI’s progress from symbolic and statistical methods to ML/DL, the Transformer architecture, and modern LLMs explains fine-tuning and prompt engineering and sets the technical foundation for later network-focused chapters. The chapter *Current state of AI in fixed networks* assesses how AI is becoming native to fixed networks, embedded in orchestration and closed loops and presents exemplary PoCs and real-world deployments demonstrating AI’s practicality and concludes with lessons learned, and identifies gaps and challenges. Then, the chapter on *Use cases and applications of AI for networks* categorises high-impact applications across planning/design, operations/assurance, operational readiness (digital twins), closed-loop automation, sustainability, and optical access. It also analyses motivations and concerns across stakeholders: interoperability and reliability for operators, IP/data sovereignty for vendors, privacy and service quality for customers, and risk-based oversight for regulators. The chapter *Technical requirements and challenges in AI for networks* describes architectural and infrastructure needs (including intelligent monitoring), data lifecycle platforms (collection, streaming, storage, processing, governance), AI deployment and integration (AI-native, MLOps, digital twins), and reliability/robustness/security. It concludes with regulatory and compliance considerations for operational AI. Next, the chapter *Networks for AI* examines how AI workloads reshape infrastructure: evolving demands, DC-to-DC connectivity and cloud integration, HPC performance targets (latency/bandwidth/reliability), architecture upgrades, and implications for optical transport and access. Finally, the White Paper is concluded by positioning itself with *ETSI standardisation Activities* and summarising the key takeaways.

## Evolution of AI: from statistical models to LLMs

### Early AI and statistical approaches

The field of Artificial intelligence (AI) has evolved remarkably since its inception, reflecting humanity's enduring quest to create machines that can mimic human intelligence. The journey of AI encompasses a rich set of ideas, breakthroughs, challenges, and paradigm shifts.

#### Early AI (1940 s-1970 s)

The concept of intelligent machines can be traced back to ancient myths and early philosophical discussions, but it wasn't until the mid-20th century that AI became a formal field of study. In the 1940 s, pioneers like Alan Turing began exploring the mathematical possibilities of machine intelligence. Turing's seminal 1950 paper, *Computing Machinery and Intelligence*, introduced the *Turing Test* to assess a machine's ability to exhibit human-like intelligence. The Dartmouth workshop, widely considered the birth of AI as a field, was organised in 1956 by John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon. They coined the term "AI" and set out a research agenda focused on replicating human cognition through symbolic manipulation.

Early AI research was dominated by symbolic AI or "Good Old-Fashioned AI" and logic-based approaches. Researchers believed that all intelligence could be represented symbolically and manipulated using logical operations. Developed by Allen Newell and Herbert A. Simon, the Logic Theorist was designed to prove mathematical theorems from Principia Mathematica. Also, by Newell and Simon, the General Problem Solver was an attempt to create a universal problem-solving machine that used heuristics to mimic human problem-solving strategies.

Initial successes led to widespread optimism. Researchers predicted that machines with human-level intelligence would be developed within a few decades. Despite early enthusiasm, AI research faced significant hurdles such as computational limitations, and commonsense reasoning. Early computers lacked the processing power and memory to handle complex tasks or large datasets. Also, symbolic AI struggled with ambiguity and the vastness of real-world knowledge required for common sense reasoning. Due to unmet expectations and criticisms like those from Sir James Lighthill's report, funding diminished, leading to reduced AI research activities and what is called the first AI winter between 1974 and 1980.

#### Emergence of Statistical Approaches (1980s-Present)

The limitations of symbolic AI led researchers to explore new methodologies. There was a growing interest in creating systems that could learn from data rather than relying solely on hand-crafted rules, shifting more towards empirical methods. The 1980s saw the rise of Machine Learning (ML) as a subfield of AI, focusing on algorithms that improve through experience. Inspired by biological neural networks, Artificial Neural Networks (ANNs) became popular. However, due to the limitations of the time's hardware and the backpropagation algorithm's computational demands, progress was slow. Expert systems were also rule-based systems designed to mimic the decision-making ability of human experts. While useful in specific domains, they were brittle and didn't scale well. In this atmosphere again, inflated expectations led to disappointment, and AI funding saw reductions which resulted in the second AI winter (Late 1980s-1990s).

The late 1990s and 2000s marked a significant resurgence in AI, primarily driven by statistical methods. The exponential growth of computing resources enabled more complex computations and large-scale data processing. The internet and digitalisation generated vast amounts of data, providing rich resources for training AI models. Researchers embraced probabilistic frameworks to handle uncertainty and variability in

data. Bayesian networks used probability distributions to represent the world, allowing for reasoning under uncertainty. Hidden Markov Models (HMMs) were widely used in speech and pattern recognition. HMMs modelled systems that were assumed to be a Markov process with unobserved states. Support Vector Machines, introduced in the 1990s, became a powerful tool for classification and regression tasks. Reinforcement learning focused on how agents ought to take actions to maximise cumulative reward.

Statistical methods transformed AI by enabling learning from data, handling uncertainty, and their scalability. Systems could now improve with exposure to more data, enhancing performance over time. Probabilistic models provided a way to manage ambiguity inherent in real-world data. Data-driven approaches could be scaled across various domains without extensive human intervention for rule creation.

## Rise of machine learning and deep learning

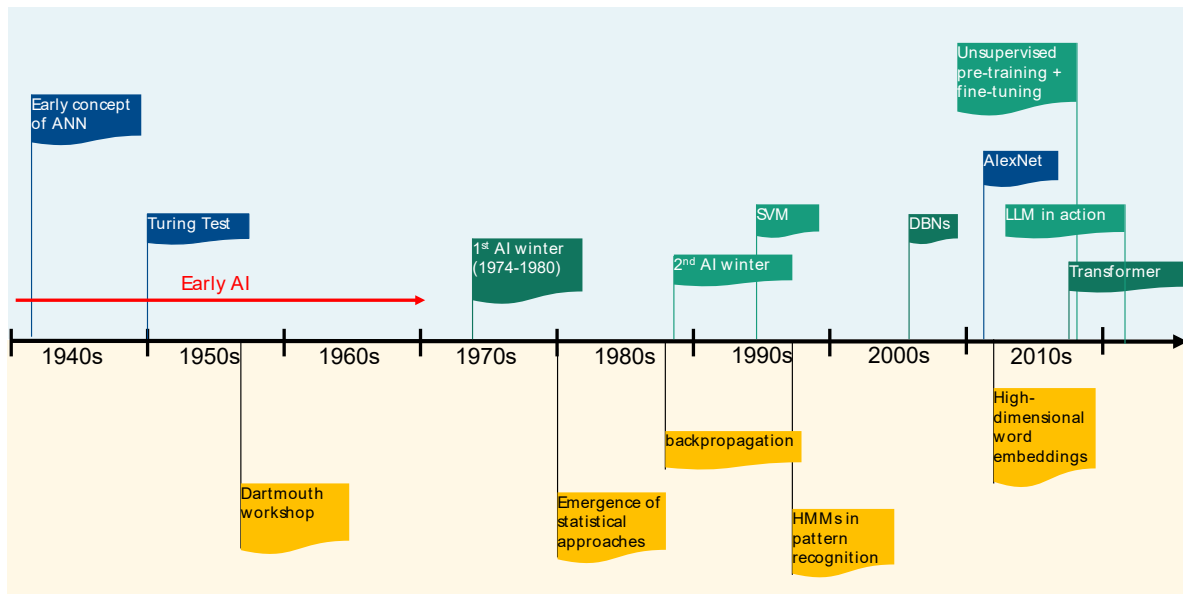
These setbacks prompted researchers to explore new methodologies. It was during this time that ML began to emerge as a promising alternative. ML shifted the focus from programming rules to enabling machines to learn from data. The idea was to develop algorithms that could identify patterns and make predictions based on input data. Concepts like neural networks, initially proposed in the 1940s with McCulloch-Pitts neurons and further developed with Rosenblatt's Perceptron in 1958, laid the groundwork for this paradigm shift. Despite early enthusiasm, neural networks faced challenges such as the inability to train multilayer networks effectively. The introduction of the backpropagation algorithm in 1986 by Rumelhart, Hinton, and Williams addressed this issue by providing a method to train deep neural networks through gradient descent. This breakthrough reignited interest in neural networks and set the stage for future developments.

One of the critical factors in the rise of ML and DL was the exponential increase in computational power and data availability. The advent of powerful CPUs and GPUs, coupled with Moore's Law, allowed for the processing of large datasets and the training of complex models that were previously infeasible. The proliferation of the internet and digital storage in the 1990s and early 2000s led to the generation of massive amounts of data. This "big data" became a valuable resource for training ML models. Algorithms could now be trained on diverse datasets, improving their accuracy and generalisation capabilities.

Deep learning, a subset of ML focused on neural networks with multiple layers (deep neural networks), began to gain significant traction in the 2000s. Researchers like Geoffrey Hinton, Yann LeCun, and Yoshua Bengio were instrumental in advancing the field. In 2006, Hinton introduced the concept of deep belief networks, which utilised unsupervised pre-training to initialise deep neural networks effectively. This approach addressed difficulties associated with training deep architectures and demonstrated they could achieve better performance than shallow models on certain tasks. The landmark moment for deep learning came in 2012 with the success of AlexNet, a deep convolutional neural network designed by Hinton's students, Alex Krizhevsky and Ilya Sutskever. AlexNet achieved a top-5 error rate of 15.3% in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), significantly outperforming previous models. This result showcased the potential of deep neural networks in handling complex tasks such as image recognition.

Convolutional Neural Networks (CNNs), introduced by Yann LeCun in the late 1980s, are specialised neural networks designed to process data with a grid-like topology, such as images. CNNs leverage convolutional layers to automatically and adaptively learn spatial hierarchies of features, making them highly effective for image and video analysis. The success of CNNs in image recognition tasks led to rapid adoption in various applications, including facial recognition, autonomous vehicles, and medical image analysis.

Innovations such as the development of architectures like VGGNet, GoogLeNet, and ResNet continued to push the boundaries of performance in visual tasks.



**Figure 1. Timeline of some of the important breakthroughs in AI since its emergence in 1940s.**

In addition to CNNs, Recurrent Neural Networks (RNNs) and their variants, such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs), made significant strides in processing sequential data. RNNs are designed to recognise patterns in sequences of data, making them suitable for tasks like speech recognition, language modelling, and machine translation. The introduction of the Transformer model in 2017 by Vaswani et al. further revolutionised Natural Language Processing (NLP). Transformers rely on attention mechanisms rather than recurrence, enabling models to handle long-range dependencies more effectively. This architecture paved the way for large-scale language models like BERT and GPT, which have achieved unprecedented performance on a variety of NLP tasks.

In the realm of AI ethics and policy, the rise of ML and DL has sparked discussions around algorithmic fairness, transparency, and accountability. Ensuring that these powerful tools are used responsibly remains an ongoing challenge that the field continues to address. Today, the integration of ML and DL into various technologies is more prevalent than ever. The focus has shifted towards making models more efficient, interpretable, and accessible. Research into techniques like transfer learning, few-shot learning, and reinforcement learning aims to reduce the amount of data and computational resources required for training. Moreover, there is increased interest in combining symbolic AI with DL to leverage the strengths of both approaches, leading to the development of neuro-symbolic AI. Advances in hardware, such as neuromorphic computing and quantum computing, hold the potential to further accelerate ML and DL capabilities. Figure 1 highlights some of the important breakthroughs in AI since its emergence.

## Emergence of generative AI and large language models (LLMs)

The emergence of GenAI and large language models (LLMs) marks a significant milestone in the field of AI and ML. These technologies have revolutionised the way machines understand and generate human-like language, opening up a plethora of applications across various domains. Several key factors contributed to the rise of GenAI and LLMs:

- **Data Availability:** The digital age has led to an explosion of textual data online, providing rich training datasets.
- **Computational Power:** Advances in hardware, including GPUs and TPUs, have made it feasible to train deep networks with billions of parameters.
- **Algorithmic Innovations:** Breakthroughs like the Transformer architecture and improved training techniques have enhanced model efficiency.
- **Investment and Collaboration:** Increased funding from both private and public sectors, along with open-source initiatives, has accelerated research and development.

Natural Language Processing aims to enable machines to understand, interpret, and generate human language. Traditional NLP techniques relied heavily on hand-crafted features and statistical models like n-grams, which had limitations in capturing context and semantics over long text sequences. The introduction of word embeddings revolutionised NLP. Models like Word2Vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014) allowed words to be represented as continuous vectors in a high-dimensional space, capturing semantic relationships based on the context in which words appear.

A pivotal moment in the development of LLMs was the introduction of the Transformer architecture in the paper "Attention is All You Need." Unlike recurrent architectures, the Transformer used self-attention to capture long-range dependencies by weighing word relationships. Transformers enabled parallel data processing, reducing training time compared to RNNs and LSTM networks. This became the foundation for subsequent large-scale language models.

Building on the Transformer framework, researchers began scaling models to unprecedented sizes, training them on massive datasets from the internet. OpenAI®'s Generative Pre-trained Transformer (GPT) series exemplifies this trend.

- **GPT (2018):** The first model showcased the effectiveness of unsupervised pre-training on large text corpora, followed by fine-tuning on specific tasks. **GPT-2 (2019):** With 1.5 billion parameters, GPT-2 generated coherent and contextually relevant text across topics. Its release was partially withheld due to concerns over potential misuse.
- **GPT-3 (2020):** Boasting 175 billion parameters, GPT-3 exhibited remarkable proficiency in a range of tasks without task-specific fine-tuning, leveraging few-shot learning through prompts.

Table 1 summarises some of the publicly available models extensively used in the community. These models underscored the power of scale in neural language models, revealing that performance continued to improve with larger models and more data.

The emergence of GenAI and large language models (LLMs) represents a significant leap forward in AI. Stemming from decades of research and advancements, these models have transformed our interaction with technology, enabling machines to generate and comprehend language with unprecedented proficiency.

There are different ways one can utilise LLMs. Fine-tuning and prompt engineering are two key techniques for optimising the use of LLMs. Fine-tuning involves training an LLM on a specific dataset to customise its behaviour for a particular task, domain, or style. This is especially useful when the base model's general knowledge doesn't align perfectly with your requirements.

Fine-tuning adjusts the weights of the pre-trained LLM by exposing it to a domain-specific dataset. This narrows the model's focus and improves its performance for specialised tasks. However, it is usually a resource-intensive task which involves data collection, data preparation, training, and finally evaluation steps before actual deployment.

On the other hand, prompt engineering is the process of crafting inputs (prompts) to guide an LLM's output without altering the model itself. It's a lightweight, flexible method to achieve desired results. LLMs generate responses based on the prompt they receive. By carefully designing the prompt, you control the format, tone, and content of the output. There are various prompt engineering techniques such as zero-shot learning, few-shot learning, Chain-of-Thought Prompting, Role Instruction, Instructional Prompts. In zero-shot learning, one may ask the model to perform a task directly without examples, relying on its general knowledge, while in few-shot scenarios a few examples are provided in the prompt to demonstrate the task or desired output format. Chain-of-Thought Prompting encourages the model to reason step-by-step for complex problems. In role instruction the user assigns a role to the model to influence its response style and with instructional prompts you provide clear instructions to perform a specific task.

All in all, fine-tuning requires access to computational resources, data, and expertise in model training, but it offers high customisation and improved performance for specialised use cases. In contrast, prompt engineering leverages the pre-trained model as-is by crafting effective input prompts to guide the model's outputs. This approach is faster, more cost-effective, and does not require access to the model's internal architecture or retraining. However, it can be less precise for specific or niche applications. In summary, fine-tuning is ideal for deep customisation, while prompt engineering is suited for quick, flexible, and resource-efficient solutions.

While LLMs offer immense potential across various sectors, it is imperative to address the ethical and societal implications they present. Through collaborative efforts among researchers, policymakers, and stakeholders, we can harness the benefits of GenAI while mitigating its risks, ensuring that this technology advances in a manner that is responsible and beneficial for all.

## Current state of AI in fixed networks

The increasing adoption of AI in fixed networks is shifting the industry from isolated AI use cases to a more integrated, system-wide transformation. Rather than simply enhancing existing workflows, AI is becoming a native element of network architecture embedded in orchestration layers, service logic, and operational feedback loops. This evolution is not just about automation but about enabling networks to self-configure, self-optimize, and self-heal in response to real-time conditions. This transition is supported by architectural shifts toward modular, AI-native platforms that incorporate microservices, real-time telemetry, and distributed intelligence.

GenAI, and LLMs in particular, have garnered tremendous attention since 2023 in the optical network community. As an example, a complete workshop titled “How Can GenAI be Used for Network Operations?” was organised at the March 2024 edition of OFC®, the leading conference in the field<sup>1</sup>. They are emerging as a strategic interface between human operators and increasingly complex systems. Instead of requiring command-line expertise, operators can issue natural language prompts that AI agents translate into valid network actions. These interfaces are not standalone tools but are embedded in controller platforms and Digital Twins (DTs).

As optical networks still rely on manual operation, LLMs are seen as an opportunity to ease network operation, by acting as the human interface to the Network Management System (NMS) and abstracting complex concepts and tasks so that less skilled labour can operate a network, or operations can be shortened and/or automated. When combined with a DT, the output (i.e., suggestions of desired actions) of an LLM agent can first be tested within the twin, which acts as a sandbox, then be pushed to the field network. This combination AI agent (based on LLM)/DT has become popular in the research community, with several lab experimental demonstrations<sup>2</sup> and even field trials<sup>3</sup>.

As such, an LLM can be used as a copilot for optical network management. As generalised LLMs have little optical communication/networking background, they need to be given additional information on one or more of the following: physics of optical networks; optical network management rules; and product information. This can be done through providing context/prompt engineering, Retrieval-Augmented Generation (RAG), and/or fine-tuning, as in<sup>2, 3, 4</sup>. This allows an LLM-based management system to advise on certain problems faced during operation, and in certain cases, to interact directly with the equipment itself, when the network state and product information are given through one of the techniques mentioned above.

Typical tasks envisioned to be delegated to LLMs mainly pertain to network management, including but not restricted to, network design (selection of equipment matching operator and physical constraints), resource allocation (e.g., routing and spectrum allocation), physical layer optimisation (e.g., channel power equalization, setting of various equipment such as the optical amplifiers) and fault management (root cause analysis, suggestion for remediation)<sup>5</sup>.

LLMs primarily serve as an operator interface today; in controlled workflows (with DT validation/human in the loop) they can also trigger automated actions. Correct interpretation of the human request, a language-oriented capability, and calling the right function, can be quantified in terms of API calling accuracy as in<sup>2,6</sup>.

---

<sup>1</sup> A. Gumaste, R. Vilalta and A. Sharma, “Workshop: How Can generative AI be Used for Network Operations?,” workshop, OFC, San Diego, March 2024.

<sup>2</sup> C. Sun, X. Yang, N. di Cicco, R. Ayassi, V. V. Garbhapu, P. A. Stavrou, M. Tornatore, G. Charlet, and Y. Pointurier, “First experimental demonstration of full lifecycle automation of optical network through fine-tuned LLM and digital twin,” in Proceedings of the European Conference on Optical Communication (ECOC), Frankfurt, Germany, October 2024, p. paper Th3B.6 (postdeadline paper).

<sup>3</sup> X. Liu, Q. Qiu, Y. Zhang, Y. Cheng, L. Yi, W. Hu, and Q. Zhuge, “First field trial of LLM-powered AI agent for lifecycle management of autonomous driving optical networks,” 2024. [Online]. Available: <https://arxiv.org/abs/2409.14605>

<sup>4</sup> Y. Pang, M. Zhang, Y. Liu, X. Li, Y. Wang, Y. Huan, Z. Liu, J. Li, and D. Wang, “Large language model-based optical network log analysis using Llama2 with instruction tuning,” *Journal of Optical Communications and Networking*, vol. 16, no. 11, pp. 1116–1132, 2024. 226.

<sup>5</sup> White Paper, “Large-Scale AI in Telecom: Charting the Roadmap for Innovation, Scalability, and Enhance Digital Experiences,” arxiv 2503.04184, 2024.

<sup>6</sup> X. Jiang, M. Zhang, Y. Song, Y. Zhang, Y. Wang, C. Ju, and D. Wang, “Opticomm-GPT: a GPT-based versatile research assistant for optical fibre communication systems,” *Opt. Express*, vol. 32, no. 12, pp. 20 776– 20 796, June 2024. [Online]. Available: <https://opg.optica.org/oe/abstract.cfm?URI=oe-32-12-20776>

In some cases, the cognitive ability of LLMs is also leveraged to perform more advanced tasks. For instance, an LLM was able to output an algorithm and the associated code for a resource allocation problem (routing and spectrum allocation) in<sup>7</sup>. Reasoning ability is used in<sup>2, 8</sup> to analyse network logs, identify the root cause of a failure, and propose a solution.

Supporting these capabilities requires distributing intelligence closer to the edge of the network. AI inference at the access or aggregation layer allows faster responses to congestion, anomalies, or service degradation. These edge-deployed models rely on scalable infrastructure, GPU resources, and data pipelines and even the restructuring of the network itself, which is under discussion in the community under the umbrella of Networks for AI.

Finally, as AI systems become operational, challenges around data governance, model transparency, and regulatory compliance are becoming more prominent. Effective deployment requires not only technical maturity but also trust ensuring data privacy, model explainability, and system robustness across diverse vendors and technologies. These concerns, along with emerging solutions such as federated learning and privacy-preserving inference, are attracting increasing attention.

Taken together, these trends point toward a fundamental change in how fixed networks are designed, operated, and evolved. The following chapters explore in depth how these capabilities are already being implemented and what technical and organisational foundations are needed to support their broader adoption.

## Exemplary proofs of concept

In recent years, AI has transitioned from theoretical frameworks to concrete, testbed-validated applications across fixed networks. A series of AI-powered PoCs have demonstrated practical utility across various fixed network domains, including traffic engineering, performance monitoring, service provisioning, and intent-based automation. These implementations highlight a growing industry shift toward autonomous, adaptive, and data-driven optical networks.

### AI-based dynamic link-capacity allocation

An important implementation for traffic-aware resource optimisation is a dynamic link-capacity adjustment scheme using ML-based forecasting. One such concept development involves the usage of LSTM neural networks to predict both hourly peak traffic and its short-term variability in real optical enterprise links<sup>9</sup>. The system decomposed traffic into maximum values and variability indicators, trained two parallel LSTMs, and dynamically computed capacity requirements with adaptive safety margins. This approach, known as Dynamic Capacity Margin Allocation (DCMA), outperformed traditional static provisioning mechanisms by reducing average hourly over-provisioning by over 75 %. It achieved this while maintaining very low under-provisioning risk (approximately 0.45 %), even with highly bursty and fine-granular real-world traffic. The adaptive design ensured that bandwidth resources matched demand more closely, offering operational and energy efficiency gains for fixed optical networks.

---

<sup>7</sup> R. Shiraki, "LLM-assisted decision making for optical path provisioning," in Proceedings of the European Conference on Optical Communication (ECOC), Frankfurt, Germany, October 2024, p. paper W2A.172.

<sup>8</sup> Y. Wang, C. Zhang, J. Li, Y. Pang, L. Zhang, M. Zhang, and D. Wang, "AlarmGPT: an intelligent alarm analyser for optical networks using a generative pre-trained transformer," *Journal of Optical Communications and Networking*, vol. 16, no. 6, pp. 681–694, 2024.

<sup>9</sup> M. Balanici, B. Shariati, P. Safari, G. Bergk, and J. K. Fischer, "Autonomous Capacity Adjustment with Dynamic Margin Allocation for Optical Enterprise Links," in *Optical Fibre Communication Conference (OFC) 2024, M1H.2*, San Diego, CA, USA, March 2024.

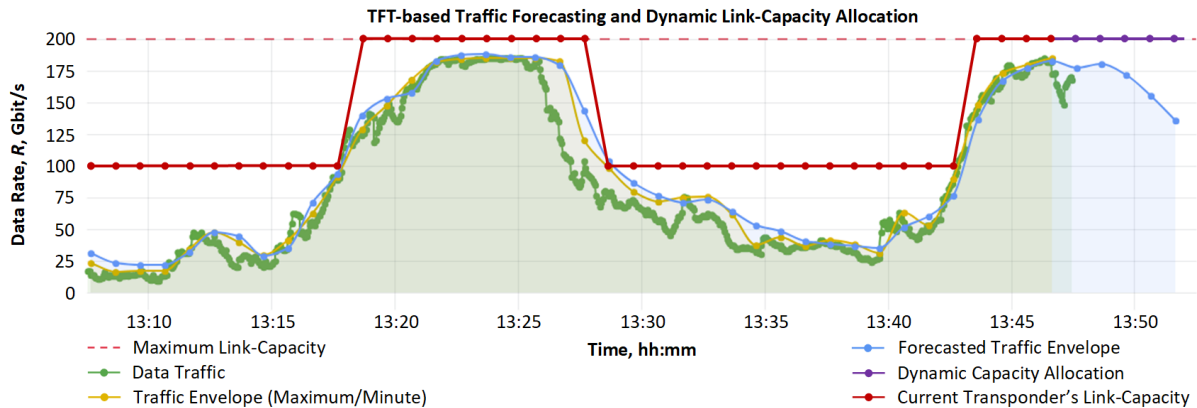


Figure 2. TFT-based traffic forecasting and dynamic allocation of link-capacity according to the predicted traffic envelope.

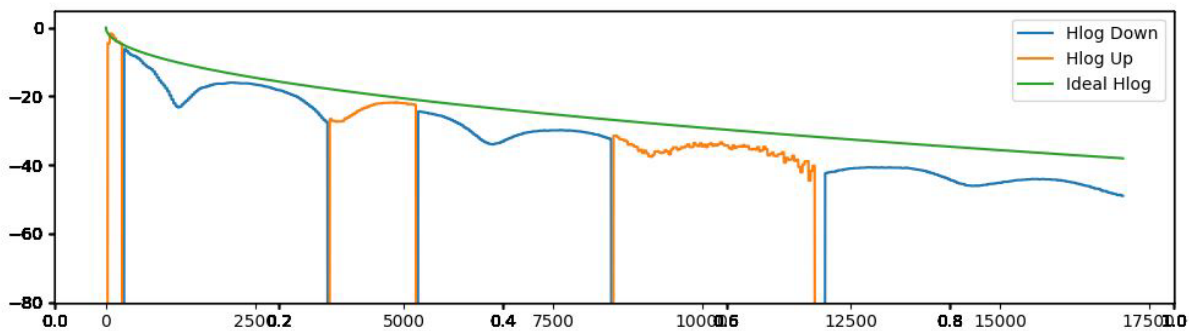


Figure 3. Frequency response of a DSL-Line with bridge tap

Further PoC implementation and demonstration of this concept<sup>10</sup> were carried out on a commercial optical metro-aggregation transport network, consisting of Reconfigurable Optical Add-Drop Multiplexers (ROADMs) and a set of optical transponders (Optical Terminals (OTs), connected into a ring architecture. The transponders are NETCONF-native and reconfigurable, allowing for their optical line-rate to be switched between 100 G and 200 Gbit/s link capacity by changing the modulation format of their optical interface. For that purpose, the forecasting of maximum ingress traffic rate (also referred to as traffic envelope) was employed using the latest state-of-the-art Temporal Fusion Transformer (TFT) - an attention-based Deep Neural Network (DNN) optimised for multi-horizon (multi-step ahead) prediction. Based on the forecasted traffic envelope values (representing the traffic maxima 5 min ahead into the future), the new scheduled/planned link-capacities are computed, such as to accommodate/allocate enough throughput for the expected traffic intensity (see Figure 2). This scheme has proven to work extremely well with both, dedicated, custom-written NETCONF-reconfiguration modules, as well as open-source SDN controllers, such as TeraFlowSDN™ (TFS), carrying out the transponders' reconfiguration<sup>11</sup>.

<sup>10</sup> M. Balanici, P. Safari, B. Shariati, A. Jafari, J. K. Fischer, and R. Freund, "Live Demonstration of Autonomous Link-Capacity Adjustment in Optical Metro-Aggregation Networks," in Optical Fibre Communication Conference (OFC) 2024, M3Z.3 Demo Zone, San Diego, CA, USA, March 2024.

<sup>11</sup> M. Balanici, B. Shariati, M. R. Raza, P. Safari, A. Jafari, V. Karunakaran, A. Autenrieth, J. K. Fischer, and R. Freund, "Autonomous Link-Capacity Adjustment using TeraFlowSDN Controller in a Disaggregated Optical Network Testbed," in European Conference on Optical Communication (ECOC) 2024, D4 Demo Zone, Frankfurt am Main, Germany, September 2024.

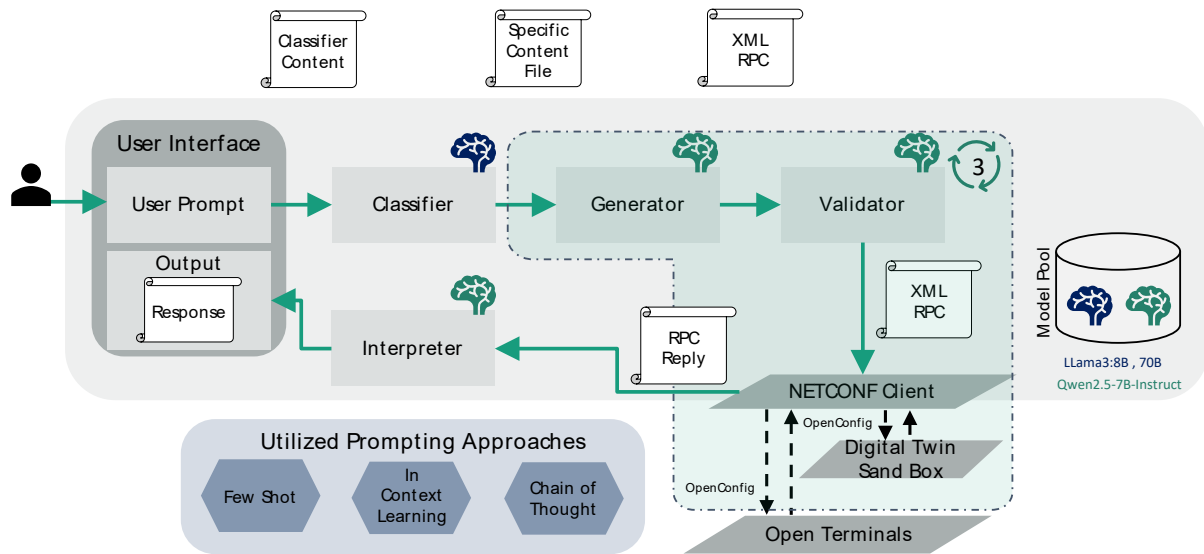


Figure 4. Architecture of the LLM-assisted network controller [Za2025].

### Diagnose of VDSL Line via Pattern Recognition

Due to the high costs associated with corrective maintenance operations, network operators have increasingly turned to AI technologies to optimise these interventions. The primary objectives of using AI are to reduce diagnostic delays and, in some cases, to prevent unnecessary field operations altogether. In the context of widespread deployment of Very High-speed Digital Subscriber (VDSL) technology, research efforts were directed toward analysing the frequency response of customer-premises modems (Figure 3). When combined with ML models, this approach enabled the identification of fault root causes, thereby enhancing the efficiency and precision of maintenance strategies.

The analysis of the frequency response was also leveraged to adapt the modulation schemes on a per-line basis. Indeed, VDSL technology is particularly susceptible to crosstalk and interference between adjacent lines, meaning that overly aggressive modulation on one line can have detrimental effects on neighbouring lines. To address this, network-wide optimisation strategies were developed to balance performance and stability. These approaches aimed to dynamically adjust modulation parameters in a coordinated manner, minimising inter-line interference while maximising overall throughput and service quality across the access network.

### LLM-assisted and DT-driven network automation

Recent PoCs have demonstrated the potential of combining LLMs with DTs to automate and enhance fixed network operations. One implementation introduced a secure, locally hosted network controller built on general-purpose LLMs using a modular multi-agent architecture<sup>12</sup>. As illustrated in Figure 4, this controller enables operators to issue high-level natural language prompts (e.g., "Provision a 100 Gbit/s service") and automatically translates them into standards-compliant, schema-validated NETCONF XML messages. Key components include task classification, prompt-based generation, output validation, and DT simulation. By avoiding cloud-based inference and fine-tuning, the system ensures data privacy and interoperability across multi-vendor environments. In a complementary PoC, researchers demonstrated full lifecycle automation

<sup>12</sup> H. Zaid, P. Safari, B. Shariati, A. Jafari, M. Balanici, and J. K. Fischer, "Multi-Agent Design for LLM-Assisted Network Management," presented at the Optical Fibre Communication Conference (OFC), San Diego, CA, USA, April 2025.

of a multi-OMS optical network using fine-tuned LLMs alongside a DT environment <sup>13</sup>. The workflow spanned four operational stages: network design, deployment, maintenance, and upgrade. Together, these PoCs highlight the emerging role of LLMs as intelligent interfaces for network management and the value of DTs in ensuring safe, testable, and automated control loops throughout the network lifecycle.

Recent research highlights the convergence of DTs, cloud-native SDN controllers, and GenAI as a foundational approach for modern optical network design, automation, and management. These DTs are dynamic virtual replicas of the physical network, capable of simulating behaviour, validating configuration changes, and supporting real-time decision-making. A key enabler is the integration of AI—especially LLMs—to drive intent-based networking (IBN) workflows. By leveraging GenAI, operators can interact with the network via natural language, while the DT simulates and validates these intents before applying them to the physical infrastructure. This minimises disruption and ensures service continuity.

Collectively, these PoCs demonstrate the applicability of AI-driven approaches in fixed networks. Across use cases, from dynamic capacity adjustment and QoT estimation to LLM-based service provisioning and DT-enabled optimisation, AI has delivered measurable benefits in accuracy, responsiveness, and operational efficiency. These early implementations provide a strong foundation for broader adoption and standardisation of AI-enabled capabilities in next-generation fixed networks.

### Lessons learned

The implemented PoCs focusing on AI for fixed networks have yielded several compelling success stories and learning opportunities, demonstrating the viability and impact of AI in transforming network operation and efficiency. One of the key lessons learned from these PoCs was that network programmability and open APIs are paramount in enabling advanced AI-driven functionalities. Without programmable interfaces and open, interoperable APIs, it would have been impossible to integrate AI engines with heterogeneous network elements. These interfaces allow for dynamic configuration, data extraction, and real-time control, laying the backbone of all AI integrations.

Equally essential was the development of telemetry agents capable of interfacing with diverse APIs, protocols, and data formats. These agents served as the critical telemetry data-gathering layer, ensuring that the AI models received accurate, timely, and comprehensive information from across the network. Their adaptability to different vendors and standards was another key aspect in achieving broad deployment feasibility.

Another critical enabler was the ability to train and deploy AI models on high-performance infrastructures, such as Edge Cloud environments, employing Virtual Machines (VMs) and Graphical Processing Units (GPUs). These platforms provided the computational resources needed to process large volumes of network data with low latency, while also supporting distributed AI processing closer to where data is generated. This was particularly important for real-time applications, where rapid model inference was necessary to respond to network events as they unfolded.

Together, these PoCs illustrated the tremendous potential of AI in improving efficiency, automation, and user experience in fixed networks. At the same time, they revealed the importance of building flexible and interoperable network architecture. Investment in open standards, cloud-native infrastructure, and robust

---

<sup>13</sup> C. Sun, X. Yang, N. Di Cicco, R. Ayassi, V. V. Garbhapu, P. A. Stavrou, M. Tornatore, G. Charlet, and Y. Pointurier, "First Experimental Demonstration of Full Lifecycle Automation of Optical Network through Fine-Tuned LLM and Digital Twin," presented at the 50th European Conference on Optical Communications (ECOC), Frankfurt, Germany, September 2024.

telemetry capabilities using a large plethora of APIs and protocols, will be essential to move beyond PoCs and scale these innovations into production environments. Going forward, collaboration among operators, vendors, and research communities will continue to be critical to drive the maturity and adoption of AI-driven network solutions.

## Real-world deployments

Over the past two years, leading telecom operators and vendors have made significant progress in integrating AI into fixed network operations and automation, moving from experimental trials to live deployments. These developments reflect a broader industry shift toward intelligent, self-operating networks that can scale with growing service demands while minimising manual overhead.

A large-scale field trial in Luleå, Sweden, conducted by Lunet, Waystream®, Savantic and RISE, explored how high-speed streaming telemetry can be used for AI-driven anomaly detection and proactive maintenance in a live fibre access network. The trial was conducted on Lunet’s city network, comprising over 1 500 Waystream access switches serving more than 22 000 households in a ring topology with 10 Gbit/s inter-node links and 1 Gbit/s downstream connections.

An open-source data pipeline (see Figure 5) was built by Savantic using Kafka® for distributed event streaming, Telegraf™ for metric collection, PostgreSQL® for storage, and Apache Superset® for visualisation. This system monitored 1 500 parameters per switch every 10 s to 60 s, scaling to more than 1 000 switches. The pipeline handled roughly 1 GB of data per switch per day (about 1 TB in total) though anomaly detection processed a filtered 32 GB/day subset. Data was aggregated with Kafka Streams and pre-processed in Python for analysis.

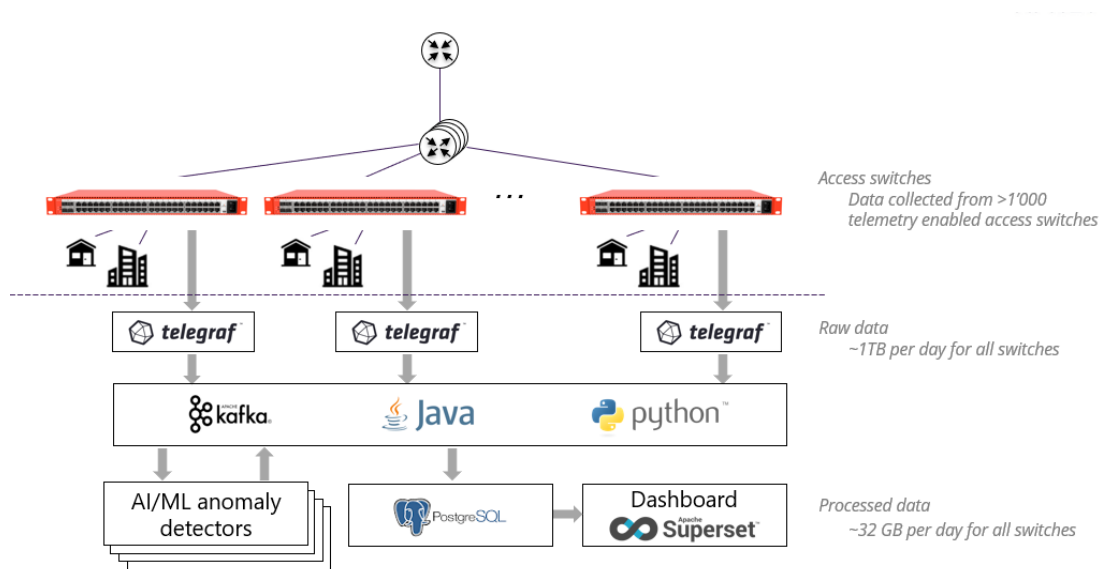


Figure 5. Data pipeline for monitoring 22 000 households

Two types of unsupervised anomaly detection were applied by RISE and Savantic. The first used univariate models to forecast downstream bitrate and the proportion of active households per switch or ring. The second employed multivariate techniques, combining an autoencoder with an ensemble of methods such as Copula-based detection, Local Outlier Factor, PCA-based analysis, and Isolation Forest, implemented via the PyOD library. The results confirmed that continuous, high-resolution telemetry collection and AI-based analysis is feasible at scale, enabling detailed, real-time visibility into the health of more than a thousand switches in a production network.

In a different activity, AT&T® has embedded GenAI into several fixed network workflows, particularly in its copper retirement and fibre maintenance operations<sup>14,15</sup>. Engineers are now supported by LLM-based autonomous assistants that assist with alarm resolution, documentation, and infrastructure assessment. For example, rather than dispatching technicians to inspect legacy infrastructure across thousands of sites, AT&T's AI systems analyse telemetry, geospatial records, and visual data to remotely assess the condition of copper lines and poles. This approach streamlines maintenance, reduces site visits, and accelerates the decommissioning of outdated assets. In parallel, AT&T uses these assistants to help identify network anomalies, correlate faults, and recommend corrective actions in near real-time, enabling faster incident resolution and reduced service downtime. This integration of GenAI into day-to-day operations supports the company's broader ambition of developing a zero touch, self-healing wireline infrastructure.

Vodafone®, in partnership with hyperscale AI platforms, has adopted GenAI to enhance its fixed network lifecycle, from service design to fault management<sup>16</sup>. In operations centres, GenAI-powered assistants help engineers troubleshoot issues by querying and summarising vast stores of documentation, diagrams, and incident logs. This capability enables quicker root cause identification and more consistent issue handling across teams. Additionally, Vodafone is using GenAI to automate the generation of network design documentation. By combining network inventory data with learned templates, generative systems can produce diagrams and configuration specs that engineers can refine, accelerating the roll-out of new fibre deployments. The company has also developed AI-driven tools for predictive optimisation, which proactively flag degradation patterns based on live telemetry and recommend reconfigurations before service KPIs are impacted.

Infosim®, in collaboration with MTN Nigeria® utilised Meta's open-source Prophet to analyse interface data, focusing on anomaly detection and future traffic prediction. Prophet's strength lies in its additive model, which decomposes time-series data into a non-linear trend, various seasonal patterns (yearly, weekly, daily), and the effects of special events or holidays. For anomaly detection, Infosim and MTN repurposed this forecasting tool in an innovative way. They generated a "forecast" of historical data to model its expected behaviour. By comparing the actual data against this model, the boundaries of Prophet's confidence interval served as a dynamic threshold, flagging any point outside this range as an anomaly. Unlike conventional statistical methods such as standard deviation, it successfully identifies not only upward and downward spikes but also structural outliers, like sudden and sustained shifts in the data pattern.

---

<sup>14</sup> J. King, "AT&T's GenAI strategy? Plug it in everywhere," *Fierce Network*, 27 August 2024. [Online]. Available: <https://www.fierce-network.com/cloud/atts-genai-strategy-plug-it-everywhere>

<sup>15</sup> A. Markus, "Autonomous Assistants: The Next Step of the GenAI Revolution to Empower Employees and Serve Customers," *AT&T Blog*, 17 July 2024. [Online]. Available: <https://about.att.com/blogs/2024/autonomous-assistants.html>

<sup>16</sup> S. Said and M. Guido, "How Vodafone is using gen AI to enhance network lifecycle," *Google Cloud Blog*, 22 November 2024. [Online]. Available: <https://cloud.google.com/blog/topics/telecommunications/vodafone-gen-ai-enhances-network-lifecycle>

Furthermore, when compared to more complex algorithms like Transformers, the relatively simple Prophet model delivered similarly accurate and replicable results with significantly less configuration and training effort.

For its primary role in forecasting, Prophet's ability to easily integrate unique events, such as major software updates or sporting events, allows for highly accurate traffic predictions. For worst-case scenario planning, the solution was enhanced with a custom feature. It identifies the largest gap between a past traffic peak and the model's expected value for that moment. This difference is then added to all future predictions to simulate the impact of a similar record spike at any time in the prediction. As MTN is Nigeria's largest telecommunications provider, it manages a gigantic network, including tens of thousands of microwave links alone that continuously generate analysable data. To apply the anomaly detection method at scale, the Prophet-based approach was combined with a time-series k-means clustering algorithm. This step efficiently grouped individual time-series measurements into distinct clusters based on their similarity in shape and pattern. For each cluster, a representative time-series, known as the centroid, was calculated. Instead of modelling every individual series, the team applied the Prophet model to this single centroid. The resulting confidence interval served as a universal threshold for all original measurements within that cluster, flagging any data point that fell outside this shared boundary. This powerful two-step process allowed the team to efficiently group and analyse a large volume of time series data for anomalies, requiring comparatively little computational effort. The solution was presented as part of the TM Forum® 's 2025 Moonshot Challenge.

Together, these efforts illustrate a meaningful shift in fixed network management, from static, rule-based systems to intelligent, adaptive platforms powered by GenAI. Operators are no longer using AI solely for monitoring or ticket triage; they are entrusting AI agents with critical responsibilities such as provisioning, documentation, diagnostics, and even service assurance. The result is a new operational paradigm: fixed networks that not only react faster to issues but also anticipate and resolve them before service degradation occurs. As GenAI continues to mature and integrate with SDN, telemetry, and orchestration frameworks, its role will only deepen in shaping the next generation of intelligent, self-operating fixed broadband networks.

## Gaps and challenges in widespread deployment

Despite the promising results achieved through PoCs, several technological and operational challenges remain before these solutions can be deployed in production networks and at large scales. One of the primary gaps is the lack of standardisation across APIs and data models. Although open APIs are key enablers, inconsistencies across vendors and domains continue to create significant integration overhead and limited interoperability. This issue is also compounded by the presence of legacy network systems, that were not originally designed for programmability or telemetry purposes, making it difficult to extract operational and configuration data in real-time or apply AI-based control uniformly.

Another major challenge remains the scalability and adaptability of telemetry agents. While these agents are successfully developed and deployed in PoCs, ensuring they perform reliably across a broad range of network environments, protocols, and data formats remains a complex task. Furthermore, the quality and granularity of data used to train AI models often varies, with datasets being in some cases incomplete, or lacking proper labeling, which affects model accuracy and reliability.

Operational integration of AI also brings its own set of challenges. Open-source tooling to manage AI model lifecycles, including monitoring, debugging, and updating models, remains immature and often disconnected from traditional network management systems. At the same time, the adoption of AI

introduces new security risks and privacy concerns, especially when telemetry data is sent to third-party platforms or cloud services. Ensuring data protection and secure model operations is an ongoing task.

Lastly, human factors also play a critical role. Many network operators remain hesitant to fully trust autonomous AI-driven decisions, underlining the importance of transparency, explainability, and human in the loop mechanisms. Additionally, a significant organisational challenge is the shortage of interdisciplinary skills required to deploy and maintain AI systems in networks. Bridging the gap between traditional network engineering, who often lack software development skills, and modern data science and AI engineering is essential for successful implementation and operation.

Addressing these challenges will require continued innovation, stronger standardisation efforts, cross-domain collaboration, and a focus on developing secure, scalable, and trustworthy AI solutions tailored for real-world network environments.

## Use cases and applications of AI for networks

This chapter looks at categories of use cases and applications that can be brought about by AI in fixed networks.

### Network planning and design

AI technologies are being applied extensively in fixed broadband and optical transport networks to **enhance network planning and design**. These use cases focus on anticipating future needs and optimising network buildouts, resulting in more efficient, reliable infrastructure. Key applications include **traffic demand forecasting, capacity planning, and automated network design**.

#### Traffic forecasting for capacity planning

ML models analyse historical traffic patterns and external data (e.g. demographics, upcoming events) to predict future demand. For example, operators can foresee localised traffic spikes and plan to adapt in advance<sup>17</sup>. This proactive capacity planning ensures the network will meet demand growth with fewer bottlenecks. AI-driven forecasts also feed into “what-if” simulations (DTs) to evaluate upgrade scenarios. Telcos are now able to **simulate how network improvements (new fibre routes, additional bandwidth, new sites) will impact customer experience and ROI before investing**<sup>18</sup>. By comparing simulated outcomes, planners can prioritise the projects that matter most for users and business goals. This data-driven approach has allowed some operators to **optimise capital expenditure plans by 10–15 %** while still meeting capacity needs.

#### Data enhancement and traffic pattern recognition

An alternative to the creation and/or externalisation of high volumes of very granular network traffic data at the edge, leveraging AI and domain-specific know-how, shows smart up sampling capabilities and improved performance by artificially increasing the resolution of aggregated volumes of exchanged bytes. AI produces generated realistic-like network traffic data traces that are consistent with the aggregated

<sup>17</sup> T. Krásová, "Ciena pitches generative AI for network planning," *Light Reading*, 1 August 2023. [Online]. Available: <https://www.lightreading.com/ai-machine-learning/ciena-pitches-generative-ai-for-network-planning>

<sup>18</sup> B. Gaffey, D. Patel, S. Cubela, and T. Lajous, "Pushing Telcos' AI Envelope on Capital Decisions," *McKinsey & Company*, 28 February 2025. [Online]. Available: <https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/pushing-telcos-ai-envelope-on-capital-decisions>

volume of bytes measured at the initial pacing and present patterns close to the expected ground truth<sup>19</sup>. Applications of such an AI model to independent use cases, such as AI models that recognise and quantify the service from those traces, exhibit improved quantification performance compared to when executed on the initial non-sampled traffic data. This could be an enabler for the detection of transient events without requiring changes in the types and amounts of currently available aggregated traffic data. Ultimately, this also addresses the amount of data to store by permitting some aggregation or, alternatively, monetizing historical coarse-grained network traffic data via offline up sampling.

### Automated network design and optimisation

AI is accelerating the design of fibre and transport networks by automating complex planning tasks that were traditionally manual. Telefónica®'s Brazilian unit (Vivo®) provides a real-world example with its "Fractal" project, which injects AI into the network capacity creation process. Fractal's AI algorithms analyse multiple information sources (e.g., subscriber locations, existing fibre routes, topology data) and **autonomously identify optimal expansion paths and equipment placements**. This provides designs that maximise resource utilisation (e.g., fibre and port usage) and minimise unnecessary construction, while reducing human error. Notably, the system can dynamically adapt its planning strategy to different geographies – urban or rural – finding the best solution for each region. The AI-driven tool recommends fibre routes that are fully diverse (disjoint), ensuring new links improve resilience against outages. Reports show that this intelligent automation has significantly sped up the planning phase and produced network topologies with higher reliability, better coverage, and lower cost than manual methods<sup>20</sup>. In essence, AI is helping engineers "design it right the first time" by **suggesting the most resilient, cost-effective topology** for fibre metro networks.

Overall, AI in fixed network planning/design is moving from theory to practice. **Traffic forecasting and capacity planning** solutions are helping operators stay ahead of demand curves (avoiding over-provisioning or last-minute panic upgrades). **Automated design tools** are cutting down planning cycles from months to days and producing more robust network designs. The net result is that fibre and optical transport providers can roll out infrastructure smarter and faster – optimising capital deployment while delivering a high quality of service. Many of these AI-driven planning solutions are repeatable and have been **proven in live networks or commercial products** indicating that AI-assisted network planning is becoming a practical reality.

## Network operation, maintenance and automation

### Operators' viewpoints

#### AI in the Fulfilment Domain

In the Fulfilment area, responsible for the provisioning and activation of services, AI can streamline and improve key operational tasks:

**Mapping customer-facing services to resource-facing services:** Generative models, such as code transformers, can be used to automatically translate high-level service intents into specific activation commands, bridging the gap between business logic and technical implementation.

**Network guardian functions:** AI can detect inconsistencies or potential errors in configuration and activation orders by comparing automation-generated commands against predefined expectations or policy templates. This prevents misconfigurations and accelerates error resolution.

<sup>19</sup> N. Dupuis, A. Van Damme, P. Dierickx, O. Delaby, "Upsampling Aggregated Network Traffic Data with Denoising Diffusion Probabilistic Models", IEEE Network Operations and Management Symposium NOMS, 2024.

<sup>20</sup> F. Barreros and G. K. Sinohara, "Automating Network Capacity Creation with AI: Vivo's Fractal Project," *Telefónica Blog*, 31 March 2025. [Online]. Available: <https://www.telefonica.com/en/communication-room/blog/automating-network-capacity-creation-ai-vivos-fractal-project/>

### **AI in Service Assurance**

The Service Assurance domain benefits particularly from AI's ability to process and correlate vast amounts of operational data at speeds far beyond human capability. AI enables several high-impact use cases:

- **Support for incident resolution:** AI acts as a co-pilot for support teams, assisting human or automated agents in identifying root causes and suggesting remedial actions during service disruptions.
- **Pattern recognition and failure prediction:** By learning from historical fault data, AI models can identify recurring patterns and provide predictive insights to prevent service degradation.
- **Knowledge aggregation and retrieval:** LLMs can serve as powerful aggregators, synthesizing information from diverse sources (e.g., vendor documentation, support tickets, online knowledge bases) to aid troubleshooting in real time.
- **Anomaly detection and trend analysis:** AI can detect outliers or unexpected behaviours in network data streams, flagging potential problems before they escalate. It can also analyse short-, medium- and long-term trends to inform capacity planning and preventive actions.
- **Log File Analysis:** AI models can aid with the severity classification and prediction of log messages within large-scale networks. By automating log analysis, AI enables faster root cause analysis and proactive network maintenance.
- **Config Analysis:** AI models compare configs to their trained knowledge on best practices and historical baselines to automatically optimise performance and detect misconfigurations before they affect service. In advanced implementations, AI can also automatically resolve configuration issues, enabling self-healing capabilities and reducing the need for human intervention.

### **AI in Operational Readiness**

AI can also support readiness activities, which are essential before a network is brought into full service:

- **Pre-emptive maintenance:** Predictive models trained on historical performance and failure data can anticipate where and when issues are likely to occur, allowing operators to act proactively.
- **Pre-testing via DTs:** AI can generate and interact with DTs of network segments to simulate configurations and activations, identifying possible issues before deployment in the live environment.

### **AI in Closed-Loop Automation**

In closed-loop automation, AI acts as the critical intelligence that links service assurance with service fulfilment, enabling autonomous network operations. Key examples include:

**Feedback-driven service restoration:** AI models analyse anomalies or trends identified in the assurance layer and determine the services affected, recommending or executing the required reconfiguration actions.

**Evaluating configuration success:** By analysing feedback from service assurance reports, AI can assess the effectiveness of fulfilment actions, detect misconfigurations, and trigger automated corrections.

### **AI for Sustainability**

As communication networks expand, a new challenge emerges ensuring energy efficiency while maintaining performance. AI plays a crucial role in optimising network operations to reduce energy consumption. Modern telecom infrastructures are no longer isolated, they are increasingly integrated with

complex energy and cooling systems. AI enables dynamic coordination between these systems, predicting demand, adjusting cooling, and managing power flows in real time. This intelligent orchestration fosters a more sustainable digital ecosystem, where network growth aligns with environmental responsibility.

### **Drivers for Introducing AI into Optical Access Networks**

The Optical Access Network (OAN) serves as the closest access point to end users within an operator's infrastructure. By deeply integrating AI technologies, operators can significantly improve the efficiency of fault diagnosis in network operations and maintenance, enhance the level of automation and intelligence in services, and ultimately boost user satisfaction. Meanwhile, the OAN's pervasive access capabilities and reliable data transmission provide a solid foundation for delivering a better experience for various AI-driven applications. Its goal is to enable adaptive, self-optimising, and self-healing capabilities across the optical access domain, including indoor optical networks, thereby enhancing access efficiency and improving the user experience.

### **Security, reliability and compliance**

Advanced ML models are also transforming security and reliability in fixed networks by enabling real-time, high-resolution telemetry data analysis across the entire network lifecycle. Unlike traditional rule-based systems, ML models can analyse high volumes of data in real-time, detect complex patterns and optimise decision-making across the entire network lifecycle.

In the domain of security, ML-based solutions are increasingly being deployed to protect optical transport networks against cyber and physical threats. ML models, such as CNNs and recurrent neural networks, can be employed to detect anomalies in encrypted traffic and recognise distributed denial-of-service (DDoS) patterns that would otherwise be difficult to capture using static systems. Vendors like Nokia® and Ciena®, have integrated such ML-based capabilities into their commercial platforms to monitor fibre integrity to detect and localise issues and fibre tapping attempts in real-time<sup>21</sup> or to identify abnormal traffic patterns at the switch level<sup>22</sup>. Such capabilities improve network security and integrity, contributing to a more resilient and secure infrastructure.

ML-based solutions can also improve fault management in fixed networks by shifting the operational mode from reactive troubleshooting to predictive analytics. Using supervised and unsupervised ML algorithms, operators can anticipate failures before they happen, reducing the time to repair and minimising service disruptions. Furthermore, ML-based QoT estimation nowadays can be performed in near real-time, enabling dynamic service placement and restoration even in highly congested network topologies.

In practice, ML-based solutions have been able to analyse trends in bit error rate (BER), optical power drift, amplifier noise, etc., to spot fibre aging or equipment degradation. For instance, Huawei® has deployed an ML-based optical service fault prediction solution that is able to predict optical channel degradation with approximately 85 % accuracy<sup>23</sup>, allowing traffic to be rerouted before customers were affected. In addition to predictive maintenance, ML-based approaches can also enhance network self-healing and resilience. ML-based solutions can be used to correlate and filter network system alarms to pinpoint root causes, shortening repair and outage time. Huawei's ML-based root cause analysis solution compresses alarms into

---

<sup>21</sup> Nokia, "How AI/ML drives the evolution toward autonomous broadband networks". [Online]. Available: <https://onestore.nokia.com/asset/214180>

<sup>22</sup> Ciena, "Ciena Powers High-Performance Computing at SC23," 10 November 2023. [Online]. Available: <https://www.ciena.com/about/newsroom/press-releases/ciena-powers-high-performance-computing-at-sc23>

<sup>23</sup> Huawei, "Huawei Releases its Optical Service Fault Prediction Solution to Build Highly Reliable Optical Networks with AI," 26 June 2018. [Online]. Available: <https://www.huawei.com/en/news/2018/6/optical-service-fault-prediction-solution-ai>

a single actionable alert<sup>24</sup>, dramatically streamlining troubleshooting. Beyond just diagnosis, ML-based solutions can also enable autonomous recovery through autonomous light path restoration in meshed optical networks after fibre cuts or multiple failures.

However, while the development of ML models in fixed networks enhances security monitoring and network reliability, it also increases the attack surface and introduces novel security vulnerabilities. Attackers may easily exploit AI-based systems by feeding adversarial inputs to provoke unsafe decisions and behaviour<sup>25</sup>. Furthermore, ML models rely on vast volumes of operational data from the network, which may inadvertently include personally identifiable information<sup>26</sup>. Network data are considered confidential from the perspective of operators and regulatory bodies and are subject to privacy-preserving laws such as the General Data Protection Regulation (GDPR) in the EU, the California Consumer Privacy Act (CCPA) in the United States of America (USA) and the Telecommunications Act in the United Kingdom (UK). These regulatory frameworks mandate rigorous security measures, including requirements for data anonymization, encryption, and the implementation of strict governance mechanisms.

As ML mechanisms for fixed networks are still evolving, there is a lack of consensus among stakeholders on a range of issues related to standardisation of data generation processes, dataset specifications, data structures, ML models evaluation and performance metrics. Regulators have historically focused on network quality of service and security and are just now starting to consider how data-driven decisions made by AI-based systems fit into these obligations. Due to the lack of standardisation and the need for compliance with regulations, there is an additional lack of real-world data that can be used for training ML models, which slows down the broad development of AI-based systems for fixed networks.

The integration of AI into fixed networks holds immense potential for increased security and robustness. Real-world deployments by industry leaders demonstrate the viability of ML-based security monitoring, predictive maintenance, and autonomous fault management. Yet, realising the full benefits of AI requires a careful balance of innovation, compliance, and accountability, through coordinated efforts across the ecosystem to develop transparent standards, data privacy aware pipelines and robust model governance. Figure 6 provides a conceptual illustration of a closed-loop AI-based automation ecosystem for fixed networks that incorporate the topics explained before.

---

<sup>24</sup> Huawei, "Intelligent OTN O&M". [Online]. Available: <https://carrier.huawei.com/en/products/fixed-network/nce/NCE-T/intelligent-otn-om>

<sup>25</sup> ETSI TR 104 066 V1.1.1 Securing artificial intelligence; Security Testing of AI

<sup>26</sup> A. Mitrovska, B. Shariati, A. Jafari, P. Safari, J. Karl Fischer and R. Freund, "Network data sharing: a governance framework for ensuring data sovereignty and privacy compliance," in *Journal of Optical Communications and Networking*, vol. 17, no. 11, pp. 1019-1031.

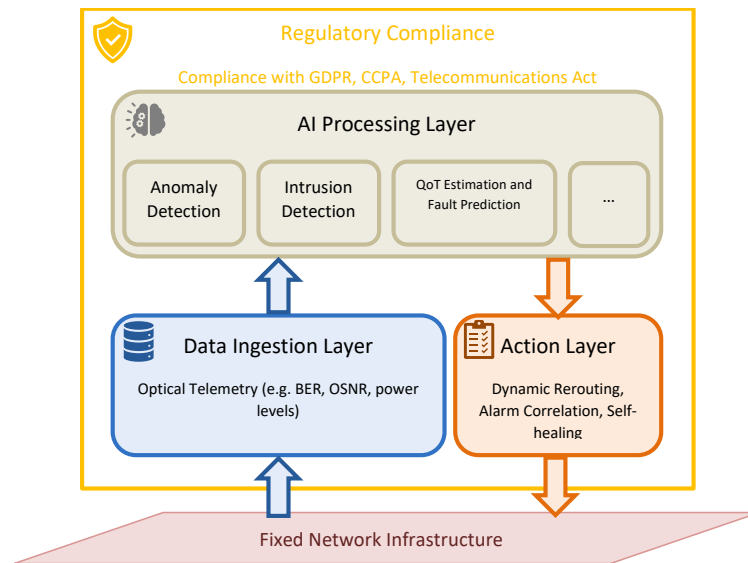


Figure 6. Closed-loop AI-based automation enabling security and reliability in fixed networks, with a regulatory compliance

## Stakeholder perspectives (customers, vendors, operators and regulators)

While the integration of AI in fixed networks offers many benefits for network management, security and reliability, it also redefines the existing stakeholder roles of operators, customers, vendors and regulators, necessitating collaboration and transparency. ML-based management and optimisation solutions introduce non-deterministic behaviour and data confidentiality concerns, which can impact the operational effectiveness, security, and regulatory compliance of fixed networks. Thus, each stakeholder group (operators, vendors, customers, and regulators) has its own motivations, concerns, and expectations regarding AI integration<sup>27</sup>.

For operators running the network, AI holds promise for automation of operations, cost reduction and performance optimisation. They see potential in AI-based systems to be used for predictive maintenance, traffic forecasting, and dynamic network optimisation, which can improve service delivery and customer satisfaction. However, for them the development of AI-based systems introduces risk in terms of data confidentiality, reliability and sovereignty issues, as outlined in the previous section. Additionally, according to a report by TM Forum, operators lack the workforce skilled in both AI and networking, which further complicates the commercial adoption of AI-based systems<sup>28</sup>. Overall, operators need AI-based systems that are interoperable, low-risk, and can be easily integrated with legacy systems, combined with governance mechanisms that can enable sovereign and privacy-preserving sharing of network data for ML model training and validation purposes.

Software and hardware vendors, on the other hand, view AI as both an opportunity and a competitive edge. Many vendors are investing in the introduction of AI-based capabilities in their products to offer intelligent

<sup>27</sup> Z. Wang, G. Wei, Y. Zhan and Y. Sun, "Big Data in Telecommunication Operators: Data, Platform and Practices," in *Journal of Communications and Information Networks*, vol. 2, no. 3, pp. 78-91, September 2017, doi: 10.1007/s41650-017-0010-1

<sup>28</sup> <https://inform.tmforum.org/research-and-analysis/reports/autonomous-networks-business-and-operational-drivers>

decision-making solutions to operators. In general, vendors advocate for adoption of AI and recognise that AI can reduce costs over time and increase the return on investment. However, vendors also face challenges in AI-based systems development. In multi-vendor environments vendors avoid sharing device-specific data, as they consider it critical to their business sovereignty and competitiveness<sup>29</sup>. This can potentially hinder the development of interoperable AI solutions that can span different vendors' equipment.

Customers of fixed networks benefit indirectly from AI through better network quality, faster issue resolution, and personalised services. AI can aid in outage prediction and dynamic traffic routing, leading to more reliable services for users. However, customers may have concerns about privacy and fairness in AI-based systems. As network data can contain personally identifiable information, for customers, it is important that ML model training does not violate their privacy. Additionally, AI-based systems might make classification errors, such as classifying normal behaviour as a security threat. Thus, for customers, the adoption of AI-based systems should not compromise the quality or integrity of their service. Customers would have a positive experience only when AI-based systems enhance the user experience while at the same time ensuring compliance with data protection regulations.

Regulatory bodies (government regulators, data protection authorities, and standardisation organisations) mandate fair competition, data privacy, and user protection. Regulators, although cautious, see the potential of AI to improve network efficiency, but also focus on preventing harm. The biggest challenge of AI-based systems for regulators is compliance with data privacy laws. Any data handling for training ML models should be in accordance with regulatory frameworks such as GDPR, and by utilising solutions that enforce privacy by design. As telecommunication networks are critical national infrastructure, regulators also need to issue requirements for AI-based systems to be robust against cyber threats.

Many regulators are actively studying AI's implications, and international bodies are developing standards to guide trustworthy AI in fixed networks<sup>30</sup>. Soon, regulators may require certification of AI-based systems in networks, similarly to any other critical network equipment. In general, the regulatory perspective on AI-based systems mostly centres on risk mitigation. Regulators want to enable innovation with AI, but under rules that safeguard consumers and the network's integrity.

In summary, all stakeholders see the transformative benefits that AI can offer for fixed networks, but each of them emphasises different challenges and requirements for successful adoption, as shown in Figure 7.

---

<sup>29</sup> N. Hashemi, P. Safari, B. Shariati and J. K. Fischer, "Vertical Federated Learning for Privacy-Preserving ML Model Development in Partially Disaggregated Networks," 2021 European Conference on Optical Communication (ECOC), Bordeaux, France, 2021, pp. 1-4.

<sup>30</sup> Gordon, J. et al. Summary: Workshop on machine learning for Optical Communication Systems. NIST Special Publication 2100-04. <https://doi.org/10.6028/NIST.SP.2100-04> (2020)



Figure 7. Different stakeholders' promises and challenges of adoption of AI in fixed networks.

## Technical requirements and challenges in AI for networks

### Architecture and infrastructure requirements

#### Infrastructure monitoring

The integration of intelligent monitoring methods into fixed telecom networks presents a paradigm shift in how network infrastructure is designed, deployed, and operated. These methods not only detect abnormal operational conditions such as malfunctioning devices but can also sense environmental disturbances, effectively transforming the network into a distributed sensing platform.

To understand the implications for network architecture, consider the example of fixed fibre-optic networks. Due to their deployment flexibility and high-throughput capabilities, fibre networks are among the most valuable assets for network operators. However, the vast scale, often involving thousands of kilometers of deployed fibre, makes efficient monitoring a significant challenge. Infrastructure monitoring becomes critical for detecting issues like fibre cuts, amplifier malfunctions, or signal degradation, which directly affect service continuity and performance.

Traditionally, network monitoring has relied on tools such as Optical Time Domain Reflectometers (OTDRs) to locate faults, or on remote access to amplifier telemetry for assessing operational status. While effective, these solutions come with notable limitations. OTDRs, for instance, are cost-intensive to deploy widely and

often require either dedicated dark fibres (which may not be affected by a fault) or optical bypass components to avoid interference from active devices like amplifiers. Additionally, managing telemetry from numerous distributed nodes adds computational and architectural complexity, making large-scale monitoring harder to scale.

To address these issues, intelligent monitoring techniques based on network tomography are emerging as a powerful alternative. These methods rely on analysing received optical signals at the end terminals, such as transponders, by digitising and post-processing the data to infer the state of the entire link. This receiver-side monitoring reduces some per-node measurements in optical transport contexts, but comprehensive AI operations still require multi-source telemetry (Operations Support Systems (OSS)/ Business Support Systems (BSS), logs, access/metro, etc.).

Moreover, tomographic monitoring is inherently vendor-agnostic and well-aligned with the concept of network disaggregation, as it does not depend on accessing internal telemetry from intermediate nodes, which may vary across vendors. This architecture simplifies interoperability and supports scalable monitoring in modern multi-vendor environments.

Importantly, these techniques contribute directly to DT development for optical networks. By continuously analysing telemetry data, operators can create real-time digital models of the physical network. These models enable operators to simulate network changes and evaluate their potential impact before implementing them in the live environment.

Practical use cases where network tomography has proven effective include:

1. Identifying excessive fibre losses after field maintenance,
2. Visualising gain disturbances in optical amplifiers through longitudinal power profiling,
3. Mapping different fibre types deployed along a route,
4. Detecting fibre intrusion attempts, such as the insertion of taps by observing power anomalies.

Enabling intelligent monitoring through tomographic techniques has significant implications for network architecture. It reduces physical infrastructure requirements, minimises operational overhead, and supports a more agile and secure monitoring layer—laying the foundation for smarter, more resilient optical networks.

## Data management and lifecycle

### The importance of high-quality data for AI in fixed networks

AI systems, especially those powering fixed network operations, are only as effective as the data they are built upon. The performance of AI models, whether used for fault prediction, traffic forecasting, or service provisioning, depends directly on the quality, granularity, consistency, and completeness of the underlying data. Unlike traditional static rule-based systems, data-driven approaches rely on continuous telemetry, historical event logs, alarm streams, and configuration records to train, validate, and update learning algorithms.

In fixed networks, data originates from a wide array of sources: routers, switches, optical transponders, and specialised platforms such as OSS, which manage network resources and services, and BSS, which handle customer-facing processes like billing and orders. Additional sources include support ticket databases that log incidents and service requests, and increasingly, Customer Premises Equipment (CPE) such as home routers and optical terminals, which provide valuable telemetry from the end-user environment. Effective data management practices are necessary not only to curate these disparate inputs into usable form but

also to ensure data integrity and timeliness, both of which are essential for maintaining AI system reliability, fairness, and auditability.

### Key building blocks of a fixed network data management system

To ensure that this data can be collected, processed, and analysed in a reliable and scalable manner, operators must build robust data management platforms that can handle volume, velocity, and variety. At the heart of such platforms are several foundational components that work together to support the full AI lifecycle, from raw data ingestion to real-time analytics and model feedback (see Figure 8). These components are not necessarily sequential, but they are logically layered to reflect the flow and transformation of data through the system. It is worth mentioning that it also follows the principles defined in this ETSI specification<sup>31</sup>.

#### Data collection layer

The foundation of the data lifecycle begins with collecting raw operational data from across the fixed network. This layer is responsible for capturing telemetry, logs, metrics, and traffic flow records from diverse and often multi-vendor environments. Common components and technologies include:

- **Telemetry agents** are embedded in network devices, using protocols such as gNMI, SNMP, NETCONF or vendor-specific interfaces;
- **Syslog listeners** gather event logs in real time;
- **Packet capture (PCAP) or NetFlow®/IPFIX collectors** to analyse traffic characteristics.

This layer must ensure low-latency, lossless data capture while also tagging metadata (e.g., timestamps, device ID, location) for downstream use.

#### Streaming and messaging layer

Once data is collected, it needs to be ingested, buffered, and routed through a system that can handle high throughput with minimal lag. The streaming layer decouples producers (network devices) from consumers (i.e. AI pipelines) and enables real-time processing at scale. Popular tools used in this layer include:

- **Apache Kafka** used for high-throughput, distributed message streaming;
- **RabbitMQ® or MQTT** used for lightweight or IoT-style telemetry flows;
- **Packet capture (PCAP) or NetFlow/IPFIX collectors** to analyse traffic characteristics;
- **Apache Pulsar™** offers native support for both queuing and publish-subscribe.

This layer plays a critical role in supporting reactive AI architectures and is often integrated with observability stacks and anomaly detection engines.

---

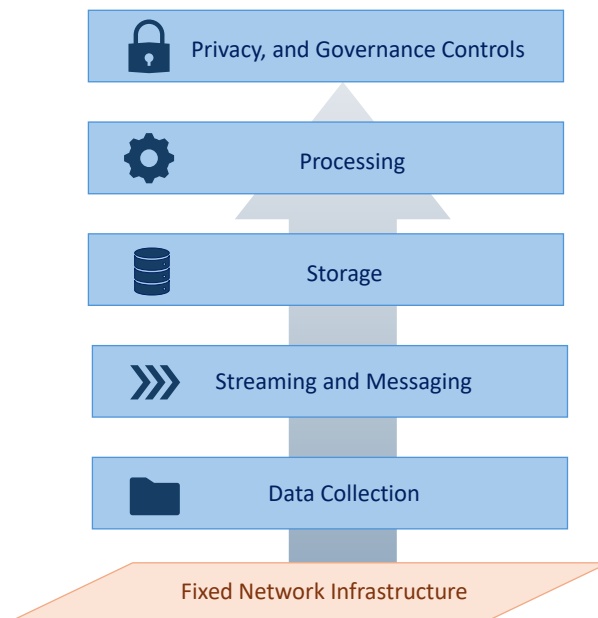
<sup>31</sup> [ETSI GS F5G 011](#): "Fifth Generation Fixed Network (F5G); Telemetry Framework and Requirements for Access Networks".

### Storage layer (data lakes and time-series databases)

To support both real-time and historical analytics, collected data must be stored in a manner that balances performance, cost, and accessibility. Depending on the nature of the data, operators may deploy a mix of storage solutions:

- **Time-series databases** like InfluxDB® and Prometheus® for metric data (e.g., signal strength, latency);
- **Log indexing/search engines** like Elasticsearch® for fast querying of textual logs;
- **OLAP/columnar databases** like PostgreSQL or ClickHouse® for structured analytics.

Efficient storage design should also include lifecycle policies for archiving, compaction, and tiered access, especially when working with regulatory constraints on data retention.



*Figure 8. Key building blocks of a fixed network data management system*

### Processing and transformation layer

Raw data must be processed before it becomes useful for AI models. This layer performs ETL (Extract, Transform, Load) operations, feature engineering, enrichment, and aggregation:

- **Apache Spark™** or **Apache Flink®** for distributed, scalable data processing;
- **Apache Airflow®** or **Dagster™** to orchestrate batch and streaming pipelines;
- **Python-based scripts** or **ML feature stores** for computing features required by specific AI tasks (e.g., moving averages, anomaly scores).

This layer may also include integration with DT environments, where processed data is simulated and validated before model feedback is applied to the live network.

## Security, privacy and governance controls

Modern data governance in AI-powered fixed networks goes beyond access control, addressing how data is protected, shared, and used across systems. Several mechanisms are essential:

- **Data Anonymisation and Tokeniation:** Sensitive information such as subscriber IDs or locations is masked to comply with privacy laws
- **Federated Learning:** Enables AI models to be trained across distributed data sources without centralising raw data, preserving privacy and supporting data sovereignty.
- **Data Marketplaces:** Controlled environments where curated datasets can be shared or monetised under strict usage policies, with auditability and licensing enforcement.
- **Privacy-Preserving Techniques:** Tools like differential privacy and secure multi-party computation allow insights to be extracted without exposing individual data points.

These controls help telecom operators deploy AI responsibly ensuring compliance, safeguarding trust, and enabling collaboration without compromising data integrity.

## AI deployment and integration

The transition toward intelligent, autonomous broadband infrastructure is reshaping how fixed networks are managed, optimised, and scaled. Rather than using AI merely to automate isolated processes, operators and vendors are rethinking the architecture of the network itself to be AI-native. AI deployment in fixed networks is evolving from static predictive analytics to continuous learning systems embedded directly within operational workflows, underpinned by mature MLOps practices, intelligent data pipelines, and closed-loop control systems.

### From tool to core capability: AI-native fixed networks

AI-native networks, as defined by Ericsson<sup>®</sup>, transcend the paradigm of AI as an optimiser and reframe it as a foundational design principle. In such networks, AI is a central capability integrated across service orchestration, fault management, customer experience, and assurance. This shift is driven by increasing complexity, real-time demands, and the need for scalable intelligence in response to growing service diversity and customer expectations. Ericsson's "AI-Native" architecture introduces the concept of microservices-based AI agents embedded in network functions. These agents support distributed learning and decision-making across multiple layers—from access nodes to cloud-native cores. Unlike traditional rule-based operations, these agents can perceive intent, evaluate conditions, and adapt behaviours dynamically<sup>32</sup>.

This vision is complemented by the "sense-think-act" model, articulated by Nokia, which divides AI-driven operations into telemetry collection (sense), decision logic (think), and automated execution (act). Central to this framework is the cognitive controller and DT network (DTN). DTNs simulate real-world network behaviour using high-frequency data and ML-based forecasting to validate actions before deployment. This capability enables intent verification, congestion mitigation, and proactive maintenance without human intervention<sup>21</sup>.

### MLOps as the Backbone of Scalable AI Deployment

For fixed networks to operationalise AI at scale, robust MLOps (ML Operations) pipelines are essential. The NGMN Alliance<sup>®</sup> outlines a comprehensive view of MLOps for autonomous networks, emphasizing lifecycle automation, version control, reproducibility, and closed-loop feedback. Fixed network operators are beginning to adopt these principles through platforms that support:

---

<sup>32</sup> M. Iovene, L. Jonsson, D. Roeland, M. D'Angelo, G. Hall, M. Erol-Kantarci, and J. Manocha, *Defining AI Native: A Key Enabler for Advanced Intelligent Telecom Networks*, Ericsson White Paper, BCSS-23:000056 Uen, February 2023.

- Model training and retraining on high-frequency telemetry from access nodes;
- Model validation using domain-specific simulation environments like DTNs;
- Deployment orchestration via CI/CD pipelines with roll-out strategies;
- Monitoring and drift detection to ensure model relevance over time.

For example, BT®'s AI Accelerator initiative uses platform-based AI deployment to streamline model development and operational impact across network assurance tasks. The accelerator supports cross-domain integration and reduces AI deployment timelines from months to days<sup>33</sup>. Meanwhile, Nokia's AIOps suite empowers fibre network operators to embed AI models into real-time telemetry pipelines, leveraging anomaly detection, capacity forecasting, and bandwidth optimisation through integrated data and model services<sup>34</sup>.

### Key use cases: real-world impact of AI integration

AI-driven closed-loop automation is among the most transformative innovations in fixed network operations. Nokia demonstrates how network congestion caused by uneven traffic patterns can be mitigated in real time by dynamically adjusting scheduling priorities in the OLT. AI Models detect congestion risks through continuous monitoring and autonomously trigger reconfigurations to preserve the quality and correct faulty states.

DTs, on the other hand, facilitate both strategic and operational decisions. They validate new user onboarding, preempt service degradations, and test service tier upgrades virtually before field execution. This not only minimises risk but enhances service agility and resource utilisation. Vodafone Germany® has also leveraged DTs to automate intent verification and optimise real-time decision-making across fixed access networks. According to Vodafone, the deployment of DTs enhanced their ability to simulate network conditions and automate closed-loop control mechanisms without manual interventions<sup>35</sup>.

### Towards operational AI maturity

AI is becoming a foundational part of fixed networks, moving beyond trials to live, evolving systems. Operators now focus on reliable deployment, ongoing monitoring, and alignment with service goals. Technologies like MLOps, DTs, and closed-loop automation are essential to building adaptive and efficient networks. Their maturity marks a shift toward autonomous operations, where networks can react, learn, and improve with minimal human input. This evolution sets the stage for fixed infrastructure that is not just AI-enabled, but AI-driven, flexible, resilient, and ready for future demands.

### Reliability, robustness and security

AI-based systems require interpretability, transparency and accountability to be considered reliable alternatives to non-ML solutions. However, many ML models, particularly those relying on DL, typically employ a black box methodology, making them an unreliable approach for critical commercial applications.

---

<sup>33</sup> Datatonic, "BT Group Unveils AI Accelerator to Shorten Rollout of New artificial intelligence from Six Months to Six Days," *Datatonic*, 3 October 2022. [Online]. Available: <https://datatonic.com/case-studies/bt-ai-accelerator-artificial-intelligence/>

<sup>34</sup> F. de Greve, "AIOps is a vital tool for fibre network operators," *Nokia Blog*, 2 September 2024. [Online]. Available: <https://www.nokia.com/blog/aiops-is-a-vital-tool-for-fibre-network-operators/>

<sup>35</sup> R. Chambers, "Vodafone Germany unlocks next-level network automation using digital twins," *Total Telecom*, 17 June 2025. [Online]. Available: <https://totaltele.com/vodafone-germany-unlocks-next-level-network-automation-using-digital-twins/>

Additionally, supervised learning often relies on historical ticket logs, technician notes, or outage records, which are sources that may be inconsistent, incomplete or biased. To address this issue, AI solution providers and operators could consider utilising interpretability techniques such as layer-wise relevance propagation (LRP) to trace how input features of the data contribute to the final decision<sup>36</sup>.

Another aspect of reliability is continuous learning and adaptation: networks evolve so AI models need periodic retraining or updating to stay effective. ML operations (MLOps) practices treating AI models as continuously maintained software are important to keep models performing reliably. In essence, ensuring reliability requires treating the AI solution with the same rigor as any network element: testing it under stress, preparing backup plans, monitoring its performance and performing comprehensive lifecycle management. With that in mind, standardisation bodies are beginning to define metrics for AI reliability and even categorise levels of autonomy in network management to help quantify and guarantee the reliability of AI-based systems deployment and operation.

In addition to being reliable, ML models must also be robust to the dynamic and unpredictable conditions of live networks. Thus, they have to be trained and tested on data from varying network conditions to minimise the chances of failure when faced with unusual inputs. Additionally, these models need to be resilient to adversarial attacks, where an attacker might intentionally feed the model adversarial inputs. To achieve the needed robustness and resilience, operators and ML solution providers can employ methods like federated learning, differential privacy and secure multi-party computation to protect against external adversarial attacks. An important step towards ensuring that ML models are resilient to adversarial attacks is to perform robustness testing by simulating adversarial attacks and adversarial training. By including these considerations in the development cycle of AI-based systems, operators and solution providers can preserve the privacy of the data, the model and the AI-based system.

## Regulatory and compliance considerations

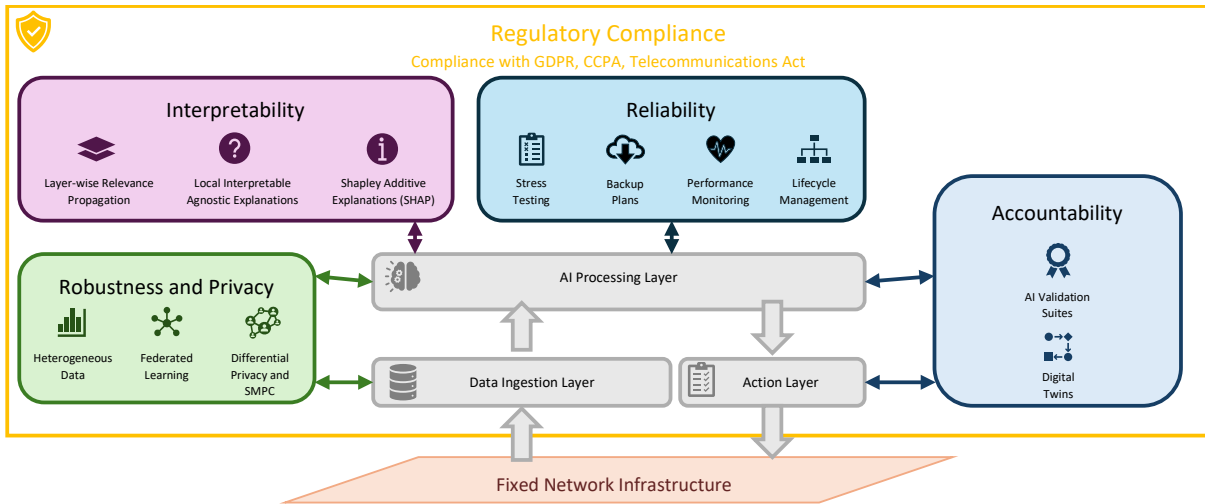
In addition to reliability, robustness and security considerations, AI-based systems must meet regulatory and compliance considerations as part of their design. From a technical perspective these requirements need to be translated into specific features and constraints of AI-based systems for fixed networks.

Given the sensitive nature of network data, data privacy is one of the main concerns addressed by regulations such as the GDPR. To manage telemetry data responsibly, the GDPR mandates rigorous security measures, including requirements for data anonymization and encryption. Such measures are essential not only for enabling legal compliance, but also for maintaining consumer trust, safeguarding against data breaches and preserving national security. While physical encryption of data offers robust protection, its implementation can impose high financial and processing costs, due to additional hardware requirements. Thus, stakeholders must utilise alternative approaches to adhere with regulatory frameworks. One way to enable legal compliance is by introducing governance in the stakeholder's systems by establishing robust policies and frameworks aligned with GDPR and CCPA that ensure data integrity, quality, privacy, security and compliance throughout the entire data lifecycle. A critical component of these measures is data anonymisation, which plays a key role in safeguarding sensitive information while allowing for compliant data sharing.

---

<sup>36</sup> Ayoub, O., Troia, S., Andreoletti, D., Bianco, A., Tornatore, M., Giordano, S., & Rottondi, C. (2022). Towards explainable artificial intelligence in optical networks: the use case of lightpath QoT estimation. *Journal of Optical Communications and Networking*, 15(1), A26-A38.

In addition to privacy, AI-based systems also require accountability, such that if an AI-based system makes a mistake, stakeholders can demonstrate that they took appropriate steps to prevent the mistake from happening. Technologically, this requires the integration of AI validation suites AI or DTs for compliance testing. By performing post hoc inspection and observing the model’s output to a certain input, stakeholders can determine whether the model behaved correctly and made an appropriate decision. Accountability for AI-based systems can also be achieved through the inclusion of interpretability and complete transparency of the development process.



**Figure 9. Interpretability, reliability, robustness, privacy and accountability requirements and realizations for AI in fixed networks.**

While cross-industry AI standards and guidance exist and regulators are studying the topic, sector-specific, fixed-network regulations and certification schemes are still largely absent. There is a need to define the boundary regulatory conditions for AI-based systems in fixed networks, and for that industry collaboration is key. Industry and governmental bodies must jointly decide on regulatory policies for ML models. These regulatory policies may include requirements for metadata inclusion for every dataset with information regarding the dataset purpose, its generation method, its information content, potential data biases, recommended use-cases and legal considerations. The same approach could be taken for developed ML models, where the performance of the model in a variety of operating conditions must be provided. This kind of standardisation can go a long way in enabling accountability in AI-based systems.

In conclusion, the technical architecture for AI in fixed networks must have privacy by design by incorporating methods that minimise personal data use, secure data in transit and at rest, and auditing of data usage. By incorporating these features, stakeholders can leverage data for AI while protecting data privacy and abiding by regulatory obligations. Figure 9 illustrates such an ecosystem.

## Networks for AI

### Evolving infrastructure demands for AI workloads

The increasing integration of AI into network operations is driving evolving infrastructure demands that traditional network environments are not fully equipped to handle. AI workloads – especially those involving real-time analytics, LLMs, and DTs – require significant computing, storage, and networking resources. As a result, networks must evolve toward more distributed and scalable architectures, such as edge cloud and hybrid cloud models, to meet the low-latency and high-throughput requirements of AI processing.

One key demand is the deployment of high-performance computing infrastructure closer to the data sources. This enables real-time AI inference at the edge, which is essential for latency-sensitive applications involving forecasting and fault detection tasks. To accommodate continuous AI model training and updating, networks also need dynamic resource provisioning and orchestration capabilities. This includes containerization, GPU scheduling, and workload mobility across cloud and edge nodes. Moreover, robust APIs and network programmability are essential to ensure that AI models can access and control network functions as needed.

Security and data governance become even more critical as AI workloads often involve sensitive operational data. Infrastructure must be designed with end-to-end security, encryption, and compliance controls. Moreover, observability tools must evolve to monitor not only network performance but also AI model behaviour and resource usage. In this context, meeting the demands of AI in networks requires a shift toward intelligent, elastic, and highly programmable infrastructure.

### Data centre-to-data centre connectivity and cloud integration

AI workloads, particularly those involving large-scale model training or distributed inferencing, are already imposing unprecedented requirements on data transmission, compute orchestration, and system responsiveness. Central to meeting these demands is the development of robust, high-capacity, and agile data centre-to-data centre (DC-to-DC) connectivity and cloud integration architectures. These elements form the invisible but essential fabric enabling real-time collaboration between compute clusters, efficient data replication, and seamless integration between private, public, and edge cloud environments.

By interconnecting several DCs, organisations can exchange datasets among servers to train ML models with better performance, while at the same time overcoming local power, space, or cooling limits by treating geographically separated sites as one big computing cluster<sup>37</sup>. Additionally, with the emergence of AI regulations, such as the EU AI Act, and regulations such as the GDPR, there is a need to protect data privacy by keeping data on the network edge. These trends give rise to distributed applications like distributed LLM fine-tuning and training. At the same time, cloud applications are also hosted by the same DCs, further burdening the traffic congestion between them<sup>38</sup>. Thus, this rising trend requires overcoming key infrastructure challenges, including the deployment of advanced optical networking.

---

<sup>37</sup> NVIDIA, "Turbocharge LLM Training Across Long-Haul data centre Networks with NVIDIA Nemo Framework," 8 May 2025. [Online]. Available: <https://developer.nvidia.com/blog/turbocharge-llm-training-across-long-haul-data-centre-networks-with-nvidia-nemo-framework/>

<sup>38</sup> P. A. Baziana, "Optical data centre Networking: A Comprehensive Review on Traffic, Switching, Bandwidth Allocation, and Challenges," in *IEEE Access*, vol. 12, pp. 186413-186444, 2024, doi: 10.1109/ACCESS.2024.3513214

Inter-DC traffic has been projected to increase sixfold in the next five years, with AI workloads being the main driver of new DC-to-DC connectivity. DC decision makers and operators anticipate that distributed LLM training will occur across geographically distributed facilities, underscoring that multi-DC clusters will become the norm rather than the exception<sup>44</sup>. Rather than deploying their own dark fibre, a majority of DC decision makers expect to rely on managed optical fibre networks utilising carrier-operated high-capacity networks for long-haul DC connectivity<sup>44</sup>. This means that the emergence of GenAI and foundational models does not only redefine the computational requirements of DCs and cloud environments but also necessitates rethinking the supporting network architecture. These high-performance computing workloads require seamless integration between the underlying infrastructure, and specialised hardware interconnected through ultra-reliable and low-latency networks, to allow massive volumes of training data and frequent synchronisation. As a result, the network interconnects must support ultra-high bandwidth – ranging from 100 Gbit/s to several Tbit/s – while maintaining deterministic low latency, high availability, and dynamic scalability.

These requirements extend beyond traditional enterprise workloads and call for a new class of network infrastructure optimised for AI's data-centric behaviour. Currently DC networks, such as those of Cisco<sup>®</sup> and Google Jupiter utilise 400 Gbit/s technology, however, to keep up with the demands of increasing traffic, optical interconnects will have to move towards the throughput level of Tbit/s. Operators and big-data DC companies view the evolution in these pluggables as a key to scaling network efficiently and reducing power consumption and space in DC-to-DC connectivity. A key challenge in these scenarios, especially for use-cases such as distributed LLM training is the need for frequent synchronisation of model parameters across nodes, with data exchanges that are latency-sensitive and throughput-intensive. Thus, the DC-to-DC connectivity and infrastructure must be able to minimise jitters and guarantee end-to-end latency, while providing robust redundancy.

Alongside physical interconnects, cloud integration has also emerged as a top priority. AI workflows are inherently hybrid and multi-cloud: enterprises may choose to train models on-premises for security or cost-efficiency, while offloading inferencing or real-time analytics to public cloud platforms or edge locations. As such, the ability to interconnect private DCs with multiple cloud providers, while also traversing between cloud regions efficiently is also of importance. Technologies such as AWS Direct Connect<sup>®</sup>, Azure ExpressRoute<sup>®</sup>, and Google Cloud Interconnect<sup>™</sup> provide dedicated, high-throughput links from enterprise environments to cloud providers. Yet, in the AI context, these links must evolve to support more nuanced requirements, such as real-time data offloading to the edge, federated model training, and dynamic workload migration across cloud domains. As a result, the integration of Cloud Exchange Platforms, data-aware routing, and cloud-native orchestration tools enables enterprises to move data and workloads where they are most effective, based on latency, availability, energy efficiency, and regulatory compliance. Furthermore, end-to-end programmability via APIs, orchestration frameworks, and intent-based networking, will allow these infrastructures to integrate directly with AI pipelines, dynamically adjusting to each stage of model development, training, or deployment. This enables fully automated provisioning of resources, seamless scaling of capacity, and real-time optimisation of data flows that underpin AI applications.

In the above context, DC-to-DC connectivity and cloud integration are not just supporting technologies in the AI era, but rather critical key enablers of scalability and global-scale adoption. Building elastic, intelligent, and secure interconnection fabrics will be a defining challenge for infrastructure providers in the coming decade. The convergence of optical networking, software-defined control, and cloud-native principles will ultimately determine the performance and competitiveness of future AI ecosystems. To support these developments, modern DC interconnects must be reshaped for AI-centric performance demands, as explored in the next section.

## High-performance computing considerations (latency, bandwidth and reliability)

Latency, bandwidth and reliability are critical performance metrics for high-performance computing workloads. These workloads, especially when distributed, can impose immense pressure on traffic flows within and across DCs. As a result, scalable optical interconnects with low latency, elastic bandwidth and fault-tolerant reliability are foundational.

Latency optimisation in such scenarios requires minimising both propagation delay and processing overhead. Target latency metrics for high-performance AI networks are typically in the range of <10 ms round-trip for regional traffic, and <1 ms for local training clusters and edge inferencing. Physical-layer innovations such as DWDM enable ultra-low-latency links by eliminating the need for immediate optical-electrical-optical switching. In practice, many operators limit multi-DC distributed training to metro-scale distances (tens to low hundreds of kilometers), keeping the latency to only a few milliseconds or less. China Telecom®, for example, conducted a 175-billion-parameter LLM training across two distributed DCs 120 km apart, connected by 800 Gbit/s optical link, and achieved approximately 99 % of the training efficiency of a single-site cluster<sup>39</sup>. Research done by NVIDIA® on multi-site training (with DCs approximately thousand kilometers apart), found that using adaptive resource orchestration, hierarchical communication patterns and asynchronous operation can sustain high efficiency across thousands of GPUs spanning several sites<sup>37</sup>. These approaches along with overlapping communication and computation can potentially pave the way towards efficient and low-latency DC-to-DC distributed training, even across continents, paving the way for multi-region AI workloads.

Thus, minimising latency involves optimisation across multiple layers, rather than just one single domain, such as:

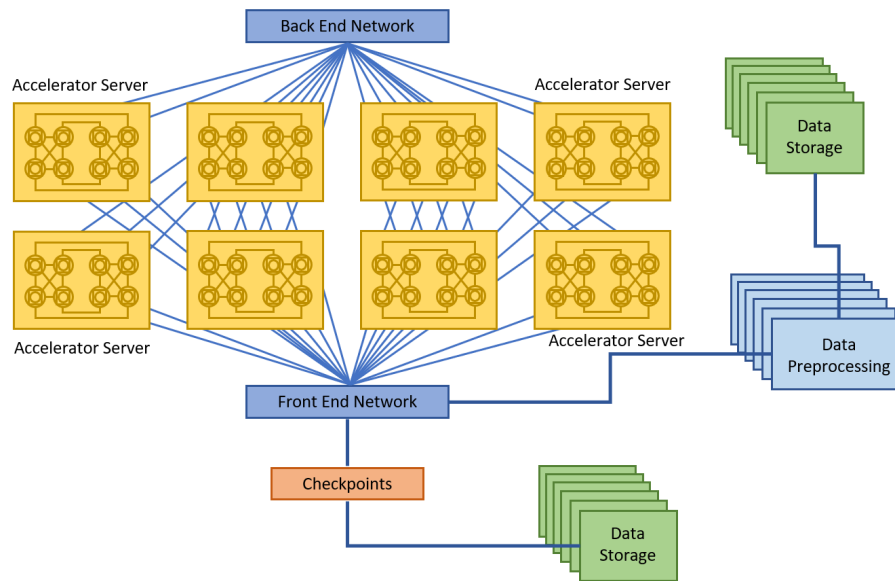
- Optical transport: deploying low-latency coherent DWDM systems, eliminating unnecessary OEO conversions, and using shortest-path routing across the optical mesh.
- Packet networks: reducing queuing and processing delays via segment routing, cut-through switching, and QoS prioritization for AI traffic classes.
- Edge-cloud integration: Deploying edge nodes closer to inference endpoints, minimising the round-trip time for real-time decisions.

In addition to latency considerations, bandwidth elasticity is also crucial for accommodating heterogeneous and large AI workloads. Each LLM training round may require exchanging gigabytes of parameter updates among training nodes, meaning that the network must deliver high-throughput communication. If the bandwidth is insufficient, there is a risk that the training stalls or degrades. SDN and optical circuit switching are increasingly employed to dynamically allocate bandwidth in real time. Closed-loop control mechanisms allow operators to monitor utilisation and instantly provision capacity where needed.

To support these needs, modern network topologies must evolve to become non-blocking and dynamically adaptable. For instance, Meta's AI backend network is isolated from the general-purpose traffic, and each rack's training switch uplinks to cluster switches with multiple 400 GbE optical links, ensuring that GPU-to-

---

<sup>39</sup> Y. Liu, A. Zhang, X. Wang, L. Feng, K. Lv, H. Liu, X. Sheng, X. Huo, and J. Li, "Field Trial of Multi-Datacentre Distributed Training for LLM Based on Bandwidth Convergence and Two Parallel Strategies over 120 km High-reliability 800 Gbit/s C+L OTN," in Optical Fibre Communication Conference (OFC) 2025, Technical Digest Series (Optica Publishing Group, 2025), paper Th1A.3.



**Figure 10. Block scheme of Meta's AI training cluster [Alduino2025].**

GPU communication spans only 2 to 3 low latency hops<sup>40</sup>. Additionally, as AI gradient exchanges can create bursty traffic patterns with link utilisation peaking to 100 % in a repeatable pattern, headroom, large buffer switching techniques and traffic scheduling are necessary. Meta has reported improved throughput and efficiency by reusing historical workload patterns to compute the most optimal path and avoid hot spots. The block scheme of the Meta's AI training cluster is illustrated in Figure 10.

In terms of further bandwidth requirements, modern HPC networks for AI must support:

- High aggregate throughput across the fabric, enabled by multi-terabit optical interconnects and 100 G+/400 G+ Ethernet links.
- Bandwidth elasticity, allowing dynamic scaling of inter-node or inter-DC links based on real-time workload intensity.
- Non-blocking architectures, avoiding oversubscription bottlenecks in top-of-rack (ToR) or spine-leaf designs.
- Technologies such as Remote Direct Memory Access (RDMA) over Converged Ethernet (RoCE), NVLink<sup>®</sup>/NVSwitch<sup>®</sup>, and smart NICs help reduce data movement overheads and improve effective bandwidth utilisation.

Reliability is another important cornerstone of high-performance computing. Distributed AI workloads cannot handle long disruptions, and there is a need to ensure zero packet loss in fixed networks. Failures in large clusters are inevitable, and network issues can cause minutes of stall time across training jobs. Thus, hyperscalers often develop fault tolerant training techniques that can isolate and work around failed nodes, without aborting the entire job. In addition to fault tolerance, distributed AI-training requires data privacy and compliance. Even though the data stays put and is not moved from its local store, adversaries might intercept the model updates and obtain proprietary information or training artifacts. Thus, optical links must be protected via robust encryption modules, fibre intrusion detection and monitoring solutions.

<sup>40</sup> Meta, "RoCE networks for distributed AI training at scale," 5 August 2024. [Online]. Available: <https://engineering.fb.com/2024/08/05/data-centre-engineering/roce-network-distributed-ai-training-at-scale/>

Additionally, techniques such as differential privacy and secure multi-party computation could be utilised to safeguard the training process.

As a result, to meet the compounded demands of latency, bandwidth, and reliability, AI infrastructures architects must adopt a co-design approach, integrating hardware, software, and network intelligence. The key strategies in designing such infrastructures include programmable optical networks with telemetry-fed control loops to dynamically optimise routes and capacities; AI-native network overlays, supporting workload tagging, resource slicing, and intent-based configuration; edge acceleration, pushing training and inference closer to data sources, reducing backbone dependency; and last but not least, multi-tier caching and in-network computing, reducing repeated data transfers across the fabric. In summary, high-performance computing networking for AI workloads is characterised by aggressive performance targets and reliability. These network capabilities allow large-scale LLM training runs with tests of thousands of GPUs to operate efficiently for weeks at a time. As models and clusters continue to grow, the network is increasingly a critical bottleneck and a crucial enabler for AI training. Continued innovation in optical technologies, intelligent capacity planning and fault tolerance is more essential than ever to meet the demands of AI workloads.

## Impact on network architectures and upgrades

The relatively recent large-scale adoption and further development of massive models such as GPT-4 and PaLM, has placed extraordinary demands on the underlying DC networks. These workloads are not only compute-intensive but also require high-throughput, low-latency, and scalable interconnects to ensure efficiency across thousands of interconnected GPUs or TPUs. As a result, network architectures for AI have undergone significant upgrades and innovations to meet the exponential growth in both data and compute.

At the heart of modern AI infrastructure is the need for high bandwidth and bisection throughput. This is crucial for operations like distributed training, which rely on collective communication protocols such as all-reduce. Network links in leading AI DCs have already reached 200 to 400 Gbit/s, with 800 Gbit/s deployments currently becoming more common. Equally critical is ultra-low latency; technologies like InfiniBand® consistently deliver sub-5  $\mu$ s latencies, making them a mainstay in high-performance computing and large-scale AI clusters. Meanwhile, Ethernet—traditionally less favoured for such latency-sensitive tasks – is closing the gap using RDMA over Converged Ethernet (RoCEv2), achieving latencies below 10  $\mu$ s.

Lossless network fabrics have also become essential in this environment, as packet drops and retransmissions severely degrade distributed training performance. Both InfiniBand and RoCE employ congestion control and buffer-flow techniques to maintain high reliability. Energy efficiency is another growing concern, with power increasingly becoming a limiting factor. Energy-proportional networking using standards like IEEE 802.3az is being adopted to reduce power draw, especially in hyperscale environments where efficiency translates directly into lower operational costs.

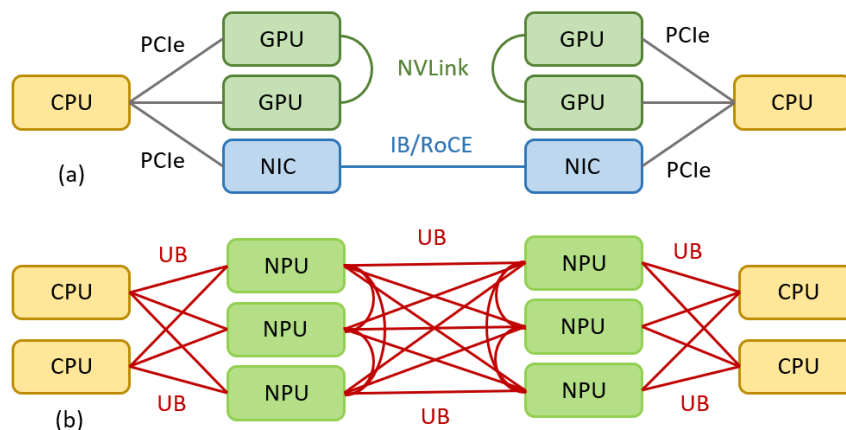
Among current technologies, InfiniBand – particularly at HDR (200 Gbit/s) and NDR (400 Gbit/s) – remains a leader. It forms the backbone of high-performance clusters like NVIDIA's Selene and Meta's RSC, often deployed in Clos topologies for predictable performance. On the other hand, high-speed Ethernet paired with RoCE is gaining traction due to its interoperability and cost benefits. Industry consortia such as the Ultra Ethernet Consortium, involving Arista®, Broadcom®, and Cisco, are pushing Ethernet toward AI-optimised standards. Arista's 400 and 800 Gbit/s Ethernet switches, marketed as "AI spine," are being deployed by hyperscalers to support backend training networks.

SmartNICs and Data Processing Units (DPUs) represent another key development. These devices, such as NVIDIA’s BlueField®-3, offload networking tasks from CPUs and GPUs, offering hardware-accelerated RDMA, security, and even in-network compute capabilities. Inside compute nodes, interconnects like NVIDIA’s NVLink provide up to 900 GB/s GPU-to-GPU throughput, significantly outpacing traditional PCIe links.

In academic and industrial collaborations, newer topologies like Unified-Bus (UB)-Mesh are emerging. In contrast to baseline systems that employ diverse interconnected techniques, such as PCIe, NVLINK, IB and RoCE, UB-Mesh uses the novel Unified-Bus for all-components interconnection – an approach improving flexibility of IO resource allocation and enabling efficient hardware resource pooling. UB also provides a seamless cross-layer optimisation<sup>41</sup>. This hierarchical multi-dimensional full-mesh topology offers improved locality, 2× cost efficiency, and higher availability than traditional Clos networks. Optical co-packaged switches and chiplet-based architectures are also being explored to break through the bandwidth and power limitations of traditional designs, enabling multi-die scalability essential for exascale AI clusters. Figure 11 shows the traditional architecture based on hybrid interconnects in comparison to UB-Mesh architecture deploying Unified Bus interconnects.

These shifts have broad implications. InfiniBand continues to outperform in raw metrics, but Ethernet is becoming the default upgrade path for many organisations, especially as 400 and 800 Gbit/s hardware becomes more widely available. Vendors like Arista are betting on Ethernet’s flexibility, while others like Cornelis Networks® offer next-generation alternatives based on Intel®’s former Omni-Path architecture, bridging performance and open standards.

Looking ahead, enterprises must assess their current network fabric – whether InfiniBand, Ethernet, or a hybrid of the two – and align future upgrades with workload scale and performance needs. For smaller to mid-sized clusters, Ethernet with RoCE and DPUs can provide sufficient performance with better cost-efficiency. Larger HPC-style deployments may still benefit from the deterministic performance of InfiniBand. Organisations should prepare for transitions toward 800 Gbit/s Ethernet, integrate RDMA and smart offloading, and explore emerging optical and chiplet-based solutions for long-term scalability.



**Figure 11. Traditional architecture based on hybrid interconnects (a) vs. UB-Mesh architecture deploying Unified Bus interconnects (b).**

<sup>41</sup> H. Liao et. al., “UB-Mesh: a Hierarchically Localised nD-FullMesh Datacentre Network Architecture,” arXiv:2503.20377v3, May 2025.

Moreover, network topologies must also be revisited: while Clos remains dominant, alternatives like UB-Mesh may yield better results for AI-specific traffic patterns.

Next-generation switch ASICs (e.g., Broadcom Tomahawk<sup>®</sup> 5, NVIDIA Spectrum-X<sup>™</sup>) provide up to 51 Tbit/s of bandwidth and native support for AI-related features such as adaptive routing, flowlet switching, and in-band telemetry. At the server edge, SmartNICs and DPUs (like NVIDIA BlueField<sup>®</sup> and Intel Mount Evans) are taking over transport stack processing and accelerating collective operations, freeing up CPU cycles and reducing tail latency<sup>42 43</sup>.

The current and future AI demands will continue to reshape the data centre network landscape from the inside out. Interconnect technologies are evolving across multiple layers – from intra-node to inter-node – requiring a holistic approach to performance, scalability, and efficiency. Staying competitive in this rapidly advancing space will depend not only on compute power but on investing in the right network architecture and planning for continuous evolution.

### Optical Transport Networks for AI

The development of AI impacts optical networks in two fundamental aspects. New AI models enable new applications, which in turn change the requirements on the underlying (transport) network infrastructure; the new capabilities of AI can also be leveraged to design and implement optical networks. For instance, applications as diverse as 3D video creation, online education (for home users), image optimisation, intelligent calls (for general customers), smart manufacturing and R&D (for business users) will co-exist on the same network infrastructure. AI, and the associated computing, and therefore the associated network infrastructure, will become a commodity that is essential for daily life and business. In the larger context of the optical networking scene, as it becomes pervasive, AI will become at the same time a spectator (the performance of AI applications is impacted by the underlying network, “networks for AI”) and an actor (AI changes how networks are implemented and operated, “AI for networks”). More in depth, optical networks will evolve in the following five aspects.

1. Awareness: as AI services have diverse requirements, the optical network should stop being a big fat pipe and become smarter. To achieve this, the network should first be able to sense or identify the application requirements, possibly with AI algorithms. Then, the network fabric should self-adjust so that large capacity, latency-insensitive AI video flows are processed and transmitted separately from the low capacity, latency-sensitive intelligent car driving flows. More technically, optical networks should support the sensing and support of differentiated Service Level Agreements (SLAs) guarantees. As the optical (L0) layer typically only provides a large pipe, finer granularity will be provided by the electrical layer (L1) so that several slices with different SLAs are supported on the same infrastructure.
2. Always on-demand: the aforementioned services with guaranteed SLAs should be established (and released) on-demand by the end users to maximise the utilisation of the infrastructure and give the end-user fast access to dedicated resources fitted to their needs. This will require a tight cooperation across the optical (L0) and electrical (L1) layers.
3. Assurance: As optical networks, at L0 and L1, are inherently deterministic, strong guarantees can easily be achieved. Once a bit is sent on an optical network, its reception time is only driven by the distance to its destination and not by the amount of traffic that is sent as in congestion-prone packet-switched networks. Jitter can be driven arbitrarily low once a connection is established. The

<sup>42</sup> Broadcom, “Tomahawk 5 Product Brief,” 2023.

<sup>43</sup> NVIDIA, “BlueField-3: Accelerated Networking for AI Workloads,” 2023.

rare cases of soft outages such as fibre plant/equipment degradation will be mitigated through flexible, dynamic resource allocation, while the even rarer cases of hard failures such as fibre cut can be made either almost undetectable through fast restoration, or fully undetectable (zero-bit loss) through more complex and costly optical/electrical protection schemes.

4. Autonomous O&M: rather than simply reacting to the said soft or hard failures, even higher network availability will be enabled by AI-assisted network operation and management, whereby continuous monitoring feeds large AI models for the proactive detection of possible problems, and automated mitigation in case a problem is (proactively or reactively) detected.
5. AI-Native: moving even further in the use of AI, future networks will embed AI deep into their operation, so that functions currently implemented without AI see their limitations lifted by AI. As AI models are becoming more powerful, they can tackle more complex problems and generalise operation functions that are currently tailored to specific cases. Within the terminal, AI can perform traffic classification or even help with DSP; within the equipment, AI can help with resource (e.g., fibre) availability prediction, as mentioned above, or resource allocation, static or dynamic; at the management level, AI can analyse risks, perform troubleshooting and even propose remediation actions. In that respect, AI will natively be embedded in the network fabric.

#### ***Optical access networks for AI***

OAN networks inherently possess intelligent and diverse capabilities that enable them to provide AI services directly. Furthermore, many AI-powered devices are expected to emerge in homes and industrial settings. These devices will generate massive volumes of data requiring distributed training in coordination with cloud platforms. Consequently, there will be stringent requirements for efficient and stable data transmission, as well as precise control over latency and jitter.

To meet these demands, the optical access network must evolve beyond being a mere 'dumb pipe'. It must be capable of offering deterministic guarantees. This can be achieved through intelligent service traffic steering technologies, which adaptively ensure differentiated computing access services based on the specific requirements of various new computing workloads. In doing so, the service experience will be elevated from a best-effort model with uncertain quality to one with guaranteed and deterministic performance.

## Relation to ETSI standardisation Activities

This White Paper complements and informs ongoing ETSI work on AI-enabled, automated, and trustworthy network operations, and highlights areas where ETSI can accelerate interoperable adoption in fixed networks:

- [ETSI ISG F5G](#) (Fifth Generation Fixed Networks): The paper’s use cases, technical requirements, and roadmap directly support ISG F5G’s mission to evolve fixed networks. It motivates potential F5G work items on AI-native fixed network management, digital-twin information models for fixed/optical domains, data/telemetry models for AI-assisted OAM, and guidelines for LLM-assisted operator interfaces and intent translation in multi-vendor environments. These proposals would complement existing ISG F5G Group Reports and Group Specifications on use cases, architecture, and evolution of fixed networks.
- [ETSI ISG ENI](#) (Experiential Networked Intelligence): The shift to intent-driven, closed-loop, and cognitive operations aligns with ENI’s AI-enabled management architecture. This White Paper leverages and can feed into the following:
  - ETSI GR ENI 001 (Use Cases)
  - ETSI GS ENI 005 (System Architecture) It highlights the need to extend ENI concepts for fixed access/optical specifics (e.g., QoT, ROADM control, OLT/ONT domains, multi-layer coordination) and to codify model-assurance and explainability hooks for operational acceptance.
- [ETSI ISG ZSM](#) (Zero-touch network and service management): Cross-domain automation and closed loops referenced in this paper map to:
  - ETSI GS ZSM 001 (Requirements)
  - ETSI GS ZSM 002 (Reference Architecture) The White Paper proposes feeding fixed-network AI/LLM use cases, assurance-to-fulfilment loops, and intent validation via DTs into ZSM PoCs and specifications, including alignment on service models, APIs, and KPIs for AI-driven autonomy.
- [ETSI TC SAI](#) (Securing AI): The trust, safety, and governance issues raised—adversarial robustness, data supply chain, and model assurance—align with:
  - ETSI GR SAI 004 (AI Threat Assessment)
  - ETSI GR SAI 005 (Mitigation) The paper recommends applying SAI’s guidance to fixed-network AI (e.g., privacy-preserving telemetry, attack surfaces in LLM-assisted operations, and certification/readiness criteria for AI components used in OAM).
- [ETSI ISG MEC](#) (Multi-access Edge Computing): Edge-deployed inference and distributed telemetry processing discussed here align with:
  - ETSI GS MEC 003 (Framework and Reference Architecture) The paper highlights MEC as a natural anchor for low-latency AI inference in fixed/access domains and proposes coordination with F5G/ENI/ZSM on data pipelines and lifecycle management at the edge.

- Synergy with ETSI Open Source (e.g., OSG TeraFlowSDN): PoCs in this paper using SDN controllers (e.g., NETCONF/YANG, model-driven APIs) align with ETSI's open-source activities as implementation vehicles for standards-based automation and closed loops in fixed/optical networks.
- The ETSI Operational Co-ordination Group on artificial intelligence (OCG AI) acts as a coordination group for the standardisation activities related to AI that are handled in the technical bodies, committees and ISGs of ETSI. However, it does not yet explicitly cover AI for Fixed Networks, specifically the related activities discussed within ETSI ISG F5G. It would be valuable for a stream of activity focused on AI for fixed networks will be kick-off within OCG AI.

In summary, the White Paper offers: (1) validated fixed-network AI/LLM use cases to populate and prioritise ETSI work; (2) concrete gaps for interface, data, and assurance standardisation across ENI, ZSM, SAI, MEC, and F5G; and (3) proposals for new or extended work items in ISG F5G focused on AI-native fixed networks, digital-twin information models, dataset governance for training/validation, and model assurance and explainability in operations.

---

## Conclusions and next steps

This White Paper has reviewed the transformative shift underway in fixed networks as AI evolves to become fundamentally embedded in network architecture itself. The industry is moving decisively toward AI-native, intent-driven, self-operating infrastructures that can dynamically self-configure, self-optimize, and self-heal in response to real-time conditions.

Current deployments and proofs of concept demonstrate that AI technologies - from statistical machine learning to DL and LLMs - are already delivering tangible benefits across the fixed network lifecycle, with successful implementations in traffic forecasting, dynamic capacity allocation, anomaly detection, and closed-loop automation showing measurable improvements in efficiency and reliability. LLMs are emerging as strategic interfaces between human operators and increasingly complex systems, translating natural language prompts into valid network actions while DTs provide essential validation environments. However, significant barriers remain for widespread adoption, including inconsistent APIs and data models across vendors, incomplete or inconsistent datasets, immature MLOps tooling, security risks, and limited model explainability.

ETSI has a critical role to play in accelerating standardisation efforts around AI-native fixed network management frameworks, DT information models, data and telemetry specifications for AI-assisted operations, and LLM-assisted operator interfaces. The organisation should also develop assurance frameworks for model validation and explainability, foster cross-SDO collaboration with groups like ZSM and SAI, create practical implementation guidelines for operators, and address the emerging requirements for optical transport upgrades needed to support distributed AI workloads.

The convergence of AI and fixed networks represents not merely an incremental improvement but a fundamental architectural shift that will enable closed-loop automation where networks continuously optimize performance based on business intent, integrate LLMs as standard management interfaces, and evolve toward self-healing capabilities that predict and prevent failures before service degradation occurs.

The next steps will be critical for establishing the foundational standards that enable trustworthy, interoperable AI adoption. Through coordinated industry action, ETSI can play a pivotal role in accelerating the transition to intelligent, autonomous fixed networks that deliver superior performance while meeting evolving regulatory and operational requirements. The journey toward AI-native fixed networks has already begun in operator networks worldwide, and with continued collaboration among operators, vendors, researchers, and standards bodies, today's promising proofs of concept can become tomorrow's standard operational practice.

## List of abbreviations and definitions

Acronym	Definition
AI	Artificial intelligence
ANN	Artificial Neural Network
BSS	Business Support System
CNN	Convolutional Neural Networks
CPE	Customer Premises Equipment
DCMA	Dynamic Capacity Margin Allocation
DDoS	Distributed Denial of Service
DL	Deep learning
DNN	Deep Neural Network
DT	Digital Twin
GenAI	Generative AI
GPT	Generative Pre-trained Transformer
GPU	Graphical Processing Units
GRU	Gated Recurrent Units
HMM	Hidden Markov Models
IBN	Intent Based Networking
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
LLM	large language model
LSTM	Long Short-Term Memory
ML	Machine learning
MLOps	ML Operations
NLP	Natural Language Processing
OAN	Optical Access Network
OSS	Operations Support System
OT	Optical Terminal
OTDR	Optical Time Domain Reflectometry
PoC	proofs of concept
QoT	Quality of Transmission
RNN	Recurrent Neural Networks
ROADM	Reconfigurable Optical Add Drop Multiplexers
TFS	TeraFlowSDN
TFT	Temporal Fusion Transformer
VM	Virtual Machine



The Standards People

ETSI  
06921 Sophia Antipolis CEDEX, France  
Tel +33 4 92 94 42 00  
info@etsi.org  
www.etsi.org

**This White Paper is issued for information only. It does not constitute an official or agreed position of ETSI, nor of its members. The views expressed are entirely those of the author(s).**

ETSI declines all responsibility for any errors and any loss or damage resulting from use of the contents of this White Paper.

ETSI also declines responsibility for any infringement of any third party's Intellectual Property Rights (IPR), but will be pleased to acknowledge any IPR and correct any infringement of which it is advised.

**Copyright Notification**

Copying or reproduction in whole is permitted if the copy is complete and unchanged (including this copyright statement).

© ETSI 2026. All rights reserved.

DECT™, PLUGTESTS™, UMTS™, TIPHON™, IMS™, INTEROPOLIS™, FORAPOLIS™, and the TIPHON and ETSI logos are trade marks of ETSI registered for the benefit of its members.

3GPP™ and LTE™ are Trade marks of ETSI registered for the benefit of its Members and of the 3GPP Organisational Partners.

GSM™, the Global System for Mobile communication, is a registered Trade Mark of the GSM Association.

