World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(REVIEW ARTICLE)

Check for updates

GenAI-Augmented AWS Infrastructure Management for 5G Telecom Networks

Jayavelan Jayabalan *

University of Madras, India.

World Journal of Advanced Engineering Technology and Sciences, 2025, 14(03), 559-568

Publication history: Received on 26 January 2025; revised on 21 March 2025 accepted on 28 March 2025

Article DOI: https://doi.org/10.30574/wjaets.2025.14.3.0125

Abstract

The telecommunications industry is undergoing a profound transformation as the convergence of 5G networks and generative artificial intelligence (GenAI) reshapes the way communication service providers (CSPs) operate, optimize, and innovate. This research paper explores how GenAI, when augmented by scalable and intelligent AWS cloud infrastructure, enables dynamic network automation, predictive diagnostics, personalized customer engagement, and entirely new business models. Drawing from six foundational industry sources, the paper investigates the functional architecture, strategic deployment, and business value of GenAI applications within the telecom ecosystem, particularly in the context of 5G. It also examines the integration of tools like Amazon SageMaker, Bedrock, and Neptune, illustrating how they underpin GenAI workflows for CSPs. Real-world use cases, such as intelligent network prediction via Graph Neural Networks and GenAI-powered virtual assistants, are analyzed alongside challenges like legacy infrastructure, data governance, model transparency, and regulatory compliance. The paper concludes by providing strategic recommendations for future adoption and highlights the critical role of AI-as-a-service models and edge-based inferencing in the next phase of telecom innovation. This research underscores the shift of CSPs from connectivity providers to AI-driven digital service enablers, signaling a new era of intelligent, adaptive, and resilient telecommunications.

Keywords: Generative AI; 5G Networks; AWS Cloud Infrastructure; Telecommunications

1. Introduction

The rapid transformation of the telecommunications landscape has been significantly accelerated by the rise of 5G networks, which offer ultra-low latency, massive bandwidth, and ubiquitous connectivity. As global data consumption scales exponentially—projected to rise from 3.4 million petabytes in 2022 to 9.7 million by 2027—communication service providers (CSPs) face mounting challenges to maintain performance, operational efficiency, and profitability [1]. This operational complexity, paired with stagnating global revenues and the continual demand for unique service offerings, is prompting a sea-change in how telecom networks are managed and evolved. In this context, AI, and notably its newest frontier in generative AI (GenAI) have established their role as strategic enablers of telecom transformation, especially when enabled by scale in a cloud environment like AWS.

There are several facets of AI and machine learning (ML) that have supported telecoms in predictive maintenance, fraud detection and customer segmentation, moreover, GenAI goes beyond just prediction, it can create - it can generate new content, integrate dynamic interaction, and enable systems to reason in ways that humans do. This leap is particularly consequential in a 5G context, where operational agility, self-healing networks, and real-time adaptability are paramount. GenAI's ability to interpret vast, heterogeneous datasets, automate complex decision-making, and enhance user interfaces aligns precisely with the technological requirements of next-generation telecom infrastructures [1][6].

* Corresponding author: Jayavelan Jayabalan.

CSPs globally are increasingly integrating GenAI into their strategic roadmaps. As of 2023, approximately 70% of telecom executives perceive GenAI as delivering distinct and incremental value beyond traditional AI/ML models, with use cases spanning customer service chatbots, predictive network planning, and synthetic data generation for testing and security [6]. However, the integration of GenAI into telecom systems is far from trivial. The telecom sector is notorious for its legacy systems and high technical debt, factors that inhibit rapid modernization. In fact, it is projected that more than 40% of operators' IT budgets will be consumed by technical debt management by 2025 [6]. Thus, for GenAI to be truly transformative, it must be embedded within flexible, cloud-native infrastructures that allow for seamless data integration, scalable model training, and secure deployment environments.

This is where AWS enters as a pivotal enabler. AWS provides telecoms with the compute power, storage, and machine learning capabilities necessary to operationalize GenAI at scale. Through platforms like Amazon Bedrock, SageMaker, and AWS Neptune, CSPs are able to ingest data, fine-tune large language models (LLMs), and deploy GenAI-powered applications with enhanced security and governance [3][5]. In addition, edge computing using AWS global infrastructure, such as Wavelength, offers extremely low-latency processing important for real-time telecom applications such as autonomous network optimization and personalized user engagement [3].

In a sense, the merging of GenAI and AWS illustrates the evolution of 5G network management as it represents not only an increase the level of automation and intelligence but also a foundation for CSPs to transform into flexible, data-driven organizations to enable hyper-personalized and robust digital services. The sections that follow in this paper will look at the architectural, functional, and economic aspects of this convergence, with specific reference to GenAI-enabled solutions for 5G telecom networks in the AWS ecosystem.

2. The Role of GenAI in Telecom Networks

For decades, the telecommunications sector has turned to traditional artificial intelligence (AI) and machine learning (ML) technologies to improve network reliability, improve the management of customer engagement, and facilitate operational workflows. Generative artificial intelligence, or GenAI, represents a new paradigm from which telecommunication service providers (CSPs) of all types and sizes will benefit from, especially as they roll out 5G network technology. Traditional AI supports classification, predictions, and decision support, whereas GenAI is specifically engineered to produce new data, generate dynamic content, and emulate human-like reasoning and interaction. As GenAI will able offer significant advancements across many areas of the telecom value chain [1][6].

The most distinguishable attributes of GenAI are the Natural Language Processing (NLP) and large language model (LLM) elements of GenAI. These capabilities help telecommunications providers deploy increasingly intelligent virtual assistants, customer-facing digital agents and chatbots. GenAI enables assistants to maintain context of the conversation, which allows for a non-linear conversational flow. Unlike rule-based bots, which are less dynamic in function and capabilities, GenAI assistants are incredibly rich in engagement and can learn from user engagement in order to better their future engagements. CSPs are already obtaining appreciable improvements in established success criteria; agent-assisted calls, resolution times, and Net Promoter Scores [1][6]. As customer expectations continue to move toward immediate, digital-first service encounters, being able to provide scalable, 24/7 capacity with GenAI is becoming a competitive differentiator.

Beyond the customer-facing use case of engagement, GenAI is also being utilized for things like internal operations, planning and change management in networks. Using unstructured data, like network logs, service tickets, and configuration files, GenAI models can identify inefficiencies, predict anomalies, and inform a superior network design. This is critically important in 5G environments; as 5G networks are and will be disaggregated and will continue to be multi-vendor by design - which adds complexity on the back end to the network. GenAI can take inputs that are extraordinarily complex, and recommend adaptive configurations to deliver the best quality of service while minimizing downtime. The shift to GenAI will also reduce the number of human specialists doing routine work (and reliance on human specialists to perform routine tasks), allowing them to focus on other skillful areas while automating design and troubleshooting [1][5].

In addition, we can say that GenAI also provides revenue-generating use cases based on hyper-personalized marketing and service offerings. This will allow telecoms to customize compositions of promotions and products to individual customers because GenAI can take behavioral analytics, spending patterns and real-time usage data to create unique personal journeys for each customer. When we improve precision targeting through generative models - specifically the ability to recommend based on predicted customer preferences and behaviors (in real time) - we provide an opportunity to lower churn, as well as surface new monetization opportunities offered through hyper-personalized content [2][6]. In the B2B space, CSPs are starting to use GenAI capabilities to bundle AI-as-a-services solutions to

smaller enterprises pursuing intelligent automation without the hassle of building their own model. The possible applications for GenAI in telecom extend even further, including in cybersecurity, where it can analyze large datasets to identify anomalies, simulate future threats and even create synthetic data for penetration testing. By providing these capabilities suite to CSP's current security operation centers (SOCs), the CSP can shift from treating threat responses as reactive, to proactive [1][4]. Although these applications are still in early development, they help to establish a running consensus that GenAI is a new layer of intelligence that amplifies the usefulness of traditional AI/ML systems.

As the technology matures, and the siloed infrastructure continues to shift toward a more frictionless approach (via scalable platforms like AWS), GenAI will become an integral value contributing factor, and even the core, of telecom innovation. It is already changing how services are delivered, how networks are being managed and how customers are being engaged. It does not take a crystal ball to predict we will only be seeing more of it in the near future.

3. AWS Cloud Infrastructure for Telecom and GenAI Enablement

As generative AI gains traction in the telecommunications industry, the true promise of generative AI will ultimately be unlocked with a cloud-native and scalable underpinning. Amazon Web Services (AWS) is central to this transformation, offering a complete portfolio of tools and services to only the set of communications service providers (CSP) enabling the deployment of GenAI solutions on top of their 5G infrastructure, with ease, secure, and flexibility. With the rapid changes to the 5G network and the amount of compute required for GenAI workloads, the benefits of cloud-native architecture for performing large storage and data ingestion, real-time inferencing, model training, and secure deploy environments become apparent. AWS provides these attributes by combining storage, compute, and AI services into one telecom optimized experience [2][3].

There are many enablers to this ecosystem, but perhaps one of the most significant aspects is Amazon SageMaker. Amazon SageMaker allows CSPs to train and deploy machine learning and GenAI models at scale. Specifically, CSPs can utilize SageMaker to create custom large language model (LLMs) and fine-tune existing LLMs with their proprietary telecom datasets. This can be useful, for example, for applications in chatbot creation, anomaly detection, or automated network diagnostics. When paired with AWS high performance compute instance offerings, e.g., Inferentia and Trainium processors, by leveraging SageMaker, when building solutions for GenAI workloads, CSPs can improve time-to-train and cost-performance ratios [2][5]. Further, telecoms can tap foundational models via APIs using Amazon Bedrock, while leaving underlying infrastructure management to AWS, to hasten time-to-market on AI-enabled services.

Just as significant as these capabilities, is the ability to take advantage of AWS Neptune, a graph database service, which is fundamental to any sophisticated applications, including but not limited to Graph Neural Networks (GNNs). GNNs during GenAI help CSPs to examine and optimize networks as they must continuously assess thousands of relationships at any point in time between network nodes. Additionally, with Neptune CSPs can model the complex topologies that exist in network, and use GNNs to predict behavior, anticipate cell congestion, and surplus capacity constraints, and identify points of failure or risk missing service level agreements levels [5]. When combined with AWS data lakes, analytics services, and AWS OpenSearch for semantic search capabilities as well in conjunction with Neptune, enable real-time and historic data to be reconciled, resulting in better high-fidelity GenAI outputs.

To cure for ultra-low latency expectations, especially for mission-critical 5G services such as autonomous mobility and remote surgery services AWS has dedicated two objectives, Wavelength and Local Zones. AWS edge computing provides the capability to bring the AWS services known to end users closer to themselves, thus allowing for greater proximity which helps to minimize latency and bandwidth used, as well as enabling GenAI models to be deployed at the network edge. This feature is highly relevant for telecom use cases which require decision making close to the device level, such as cell tower load balancing or personalization for users with limited data connections [3].

The overall architecture of this GenAI AWS infrastructure can be viewed with the following diagram incorporating how different AWS services combine to manage and optimize telecom networks, which illustrates the integration of GenAI and Graph Neural Networks below, to achieve network predictive analytics:

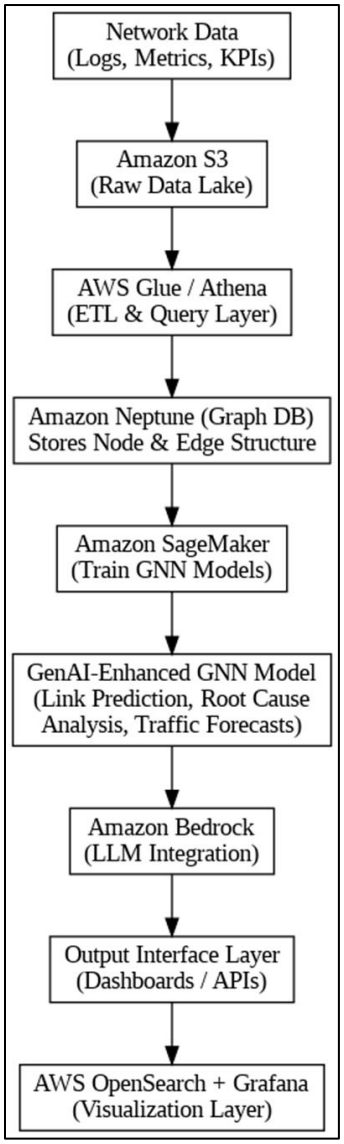

Figure 1 Integrated GenAI-GNN Pipeline for Network Prediction Tasks Using AWS Cloud Infrastructure (Adapted from [5])

The above example represents a modular yet holistic approach for deploying GenAI in a telecom environment. The pipeline begins with data ingestion using Amazon S3, the transits to SageMaker for training and inferencing, uses Neptune for graph relationships, and Bedrock to access foundational models. The meanign produced can then provide insights into operational systems or visualized via OpenSearch Dashboards, enabling CSP's to make data-drivencg decisoians at scale and with unprecedented speed.

Leveraging this sophisticated infrastructure, AWS provided not just the technological layer for GenAI applications in telecom, but also new strategic capabilities in network automation, customization of service, and predictive maintenance. All of this packaged CSPs to effortlessly transition to agile, AI-driven organizations who can more competently grasp the sensibility of the next round of digital transformation thrust by the emergence of 5G and the advent of GenAI.

4. GenAI for 5G Network Infrastructure Management

The advent of Generative AI into 5G network infrastructure management represented an enormous shift of how communication service providers (CSP) could plan, operate, optimize, and maintain their complex infrastructure. Once comprised of somewhat repeating steps void of greater complexity, the 5G environment represents an immense escalation of complexity due to the need for dense cells, use of virtualization and the possible combination of simple

routers and switches to completely different vendor routing, switching and potentially ill-defined endpoints, which is why legacy, rule-based systems are not enough. GenAI is a new layer of intelligence for planning, managing and deleting operational methods, offering comprehensive learning and automated results from disparate networks. We can now imbue cognitive augmentation into real-time decision making as users and applications are scaling and data traffic is growing at unprecedented proportions. Real-time adjuncts to credibility are set into service offerings that require a frantic responsiveness from, for example, industrial automation to augmented reality [1][4].

One of the most emergent applications for GenAI in network infrastructure could be found in the context of predictive modelling and real-time optimization. Graph Neural Networks (GNNs), are a subclass of deep learning algorithms that take graph-structured input data, which allows CSPs to properly account for elaborate relationships between nodes in the network and to monitor their behavior over time. The GNNs described in this project are particularly valuable when deployed onto AWS stacks, which provide a complete pipeline for each model building workflow: information ingestion into AWS via Amazon S3, graph storage using AWS Neptune, and then model development and deployment in the AWS environment through Amazon SageMaker [5]. With this pipeline, CSPs can identify bottlenecks, predict failures, and dynamically optimize signal routing, and potentially offer uninterrupted service to customers, even under adverse service loads.

One of the most important distinctions in deploying GNNs as part of Generative AI (GenAI) tools for CSPs is transductive vs. inductive learning, where each learning type has certain advantages based on the network topology and whether the prediction task involves uncertain information about predicted events. Transductive learning is essentially attempting to predict for the known nodes and for known relationships already coded into the learning structure into the GNN, which is more desirable within static environments. Inductive learning, on the other hand, attempts to generalize patterns in the known nodes and their relationships to predictions for nodes that are previously unseen, and potentially for new, unseen node relationships as well, which is more desirable since learning environments tend to be more dynamic with respect to generative architectures for applications like 5G. The following table distinguishes transductive and inductive learning in terms of the two learning approaches and how they fit into telecom applications.

Table 1 Transductive and Inductive Learning Techniques for Making Predictions in Graph Neural Networks (Adapted from [5])

Learning Type	**Characteristics**	**Application in 5G Networks**
Transductive	Learns embeddings for known nodes and edges only	Useful for static environments or monitoring existing node behavior
Inductive	Generalizes learning to unseen nodes and new relationships	Ideal for dynamic networks with evolving topologies and new nodes

GenAI can also be utilized to automate network lifecycle management processes such as planning, configuration, deployment, and optimization. Telecoms can use cloud-based services (like AWS) and orchestration tools to help automate cell tower provisioning, spectrum allocation, and energy optimization using data and data-driven GenAI models. For example, why shouldn't the GenAI model interact with all unstructured and structured data the operator collects (including telemetry data, optimization logs, performance statistics, and general environmental data)? All such data could help suggest or implement the best changes for benefiting operational efficiency (e.g., lowering power consumption and optimizing user loads) while still adhering to service-level agreements that are part of common telco services [2][5].

GenAI can also help simulate and generate synthetic network traffic to stress-test and validate security for deployments. It can situate various network attack scenarios or congestion patterns, so telecoms can build their infrastructure to withstand potential threats and possible operational breakdowns. These aspects of simulation and research are especially compelling in an environment where reliance on cloud and virtualized network functions are growing and vulnerabilities (or risk) can often apply to both software-defined networks and physical nodes [1][4].

In general, GenAI empowers, not only automation, but also creates opportunities for a shift towards providing self-optimal networks (SON) that are designed to be responsive to real world changes. When combined with the elasticity and scalability of cloud AWS services, GenAI provides CSPs with unprecedented levels of operational efficiency, service quality, and network resiliency. The combination of these tools and technology enables CSPs to evolve from connectivity

supplier to intelligent network orchestrator. They can adapt to the emerging needs of a rapidly digitalizing and connected world.

5. Business and Operational Impacts of GenAI-Augmented AWS for 5G

The application of Generative AI to current day 5G network operations, especially in scaling and intelligent cloud infrastructures, such as AWS, presents a wide range of implications on the business and operational models of communication service providers (CSPs). This transition is not simply a technology transition; it is a strategic transition. The investment in Generative AI is causing a tectonic shift in CSPs' perceptions and paradigms of customer engagement, operational efficiency and revenue generation. CSPs are distinguishing themselves from traditional industrial organizations and, instead, establishing themselves as entities that are more agile, software-centric "TechCos." In order to mitigate challenges marking the technical landscape and rapidly expanding customer expectations with stagnating revenue and rising costs [1][2], CSPs are looking to adopt Generative AI to begin and delivery revenue revenue stream gains, as they transition into TechCos, requiring increasingly innovative and creative approaches to leverage their experience, knowledge, people and technology at scale.

Generative AI's most significant impact on any organization is probably related to its ability to improve productivity through automation and augmenting many manual and labor intensive processes. For instance, in a customer support orientation, Generative AI reduces the need for a large human support resource base through providing customer support agents with highly efficient AI virtual agents that can independently resolve inquires-both internally and externally. These agents are no longer limited to predetermine static responses however it will take advantage of real-time learning, and context, which aids in the factual correctness and delivery response. This in turn helps the agent reduce the work requirement thus the operational expenditure, but keep customer satisfaction the same or improve metrics [1][6].

In addition, GenAI when synthesized across AWS also exhibits a varying level of elasticity & scalability that performs as an optimal function around the success of 5G. AWS services like Bedrock or SageMaker, allow the CSP to explore, train, and deploy their models throughout this model, which could allow the CSP to conduct pilots of use cases and move them into production without required capital extensivity up front. Therefore if a CSP does not like the model, they can stop the model and start over without great risk or loss of capital. There is increasingly little CapEx based models available through GenAI, and more OpEx bundling which is overwhelmingly great for the opportunity to market new services or features quickly, which will be becoming increasingly important to survive in the ultra-competitive telecom environment [2][5].

Beyond internal efficiencies, GenAI opens up new avenues for revenue generation. CSPs are increasingly bundling their AI capabilities as part of “AI-as-a-Service” (AIaaS) offerings for enterprise clients, particularly in sectors like retail, healthcare, and logistics that are seeking low-barrier entry into AI adoption. These offerings are facilitated through the same AWS platforms used internally, turning CSP infrastructure investments into monetizable services. This strategic pivot allows CSPs to move up the value chain and diversify income beyond traditional connectivity models [4][6].

The regional uptake of GenAI in telecoms also emphasizes its developing business significance. From the data, CSPs in North American markets exhibit the highest current average adoption levels of GenAI use cases at 21% followed closely by European operators at 19%, with the Asia-Pacific region trailing slightly at 16%. The graphs below show the current and anticipated average adoption levels that will presumably increase significantly from today over the next two years.

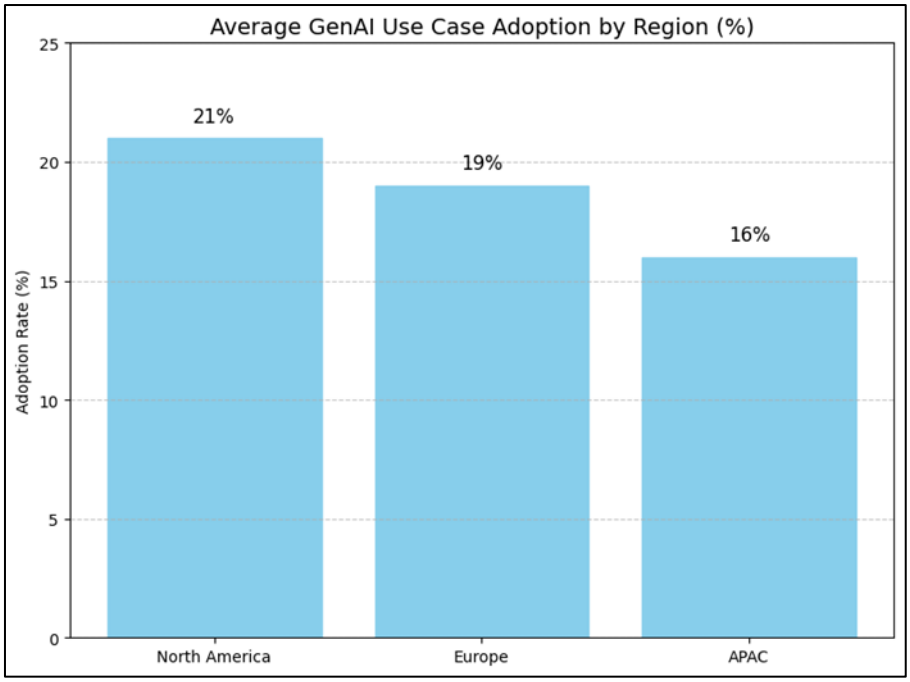

Figure 2 Average Adoption of Generative AI Use Cases by Region - Communication Service Providers (Adapted from [6])

These numbers indicate that CSPs worldwide must act quickly to bridge the adoption gap of GenAI and leverage it as an enabler of business transformation. Initially, when entering the GenAI landscape, early adopters of GenAI focused on applications that were primarily low risk with high-return, such as chatbots and documentation automation. More recently, there is a renewed focus on GenAI applications that are complex and/ or revenue generating including product personalization and intelligence network services.

On the other hand, GenAI also enables continuous innovation to thrive in an ecosystem that fuses decision-making and collaboration across functions. In this case, GenAI democratizes the access to relevant data and insights across platforms hosted by AWS creates new opportunities for independent sales, engineering, and customer support teams to form, operate, and implement a single layer of intelligence. This cohesion results in a more agile and responsive organization that is poised to adapt to market changes and evolving regulatory standards.

In summary, how Generative AI (GenAI) can impact telecom business operations is multi-faceted—optimizing workflows, improving customer experiences, developing revenue streams, and shifting the perspective of CSPs from traditional suppliers to more agile, intelligent service providers. The combined capability of AWS' cloud infrastructure and GenAI's generative capabilities provide the anchor needed for the transformation that ensures telecom operators can remain resilient and adaptable to constantly changing market conditions and digital disruption.

6. Challenges in Implementation

As positive as it sounds, even with so much potential offered by Generative AI (GenAI) within 5G telecom infrastructures there are barriers to deployment due to the challenges that are inherent within telecom environments and GenAI's computational and ethical considerations. Fundamentally, this creates a number of technical, organizational and regulatory-compliance issues that communication service providers (CSPs) need to be aware of. These issues also hinder the scalability, efficiency, and attractiveness of GenAI enabled systems built on top of cloud platforms such as AWS.

Data governance is perhaps the biggest barrier. GenAI models need massive datasets, but data in telecom is quite complex and often contains sensitive user information, call records, location data, and content of communications. The sector is bound by stringent data privacy regulations such as the General Data Protection Regulation (GDPR) in Europe and evolving national laws across Asia-Pacific and North America. According to recent industry research, CSPs universally cite data governance and ownership as the most pressing gaps in their GenAI strategies. The challenge is not

merely about having access to data but ensuring its lawful use, storage, and integration across highly siloed and often legacy infrastructure environments [6].

Closely tied to data governance is the issue of legacy systems and technical debt. Many telecom operators still rely on outdated infrastructure that is poorly equipped to support the demands of GenAI, particularly the high-volume, high-velocity data pipelines necessary for model training and inference. It is estimated that over 40% of telecom IT budgets are currently consumed by efforts to manage technical debt, which includes maintaining incompatible systems, fragmented databases, and monolithic network architectures [6]. These issues not only slow down modernization efforts but also restrict the ability to fully leverage cloud-native solutions offered by AWS, such as real-time inferencing on edge devices or automated network optimization through SageMaker and Neptune [5].

The other time is model interpretability and trust. GenAI models like large language models are often referred to as "black boxes" due to their opaque decision-making capabilities. This poses risks to mission-critical telecom scenarios, where GenAI is leveraged in areas such as network security, configuration, management, and customer service. Even if the chance of model hallucinations are very low, the potential for generating reasonably plausible but incorrect outputs may create incorrect configurations or misleading information passed to customers or technicians. As a result, CSPs are opting for a "human-on-the-loop" process and are taking away not only GenAI's essential power but also the intended speed and efficiency gains [1].

A skills shortage represents another structural challenge. GenAI technologies need a specific skill set that includes deep learning, cloud architecture, natural language processing, and model fine-tuning. Most telecom operators are just beginning to develop those capabilities internally. Though AWS provides managed services and pre-trained models, internal expertise is still critical for tasks such as fine-tuning, customizing, and governing [2][4]. In addition, a skills shortage means that rapid innovation, trying new use cases, and nimbleness with rapidly evolving customer expectations or regulatory desires are constrained.

Second, there is considerable uncertainty around cost containment. Deploying and scaling GenAI systems, especially those with custom models or proprietary training datasets, can chew through resources quickly. Many CSPs are still figuring out if the investment is going to be in building their own foundation models versus paying for off-the-shelf solutions. Some industry evidence indicates that while few CSPs plan to build a model from scratch, most have significant expectations for increased GenAI with plan costs rising from below 1% to as much as 6% of total technology budgets within the next 2 years [6].

Finally, regulatory compliance continues to be a moving target. Governments across the world are rapidly deploying legal frameworks for the use of AI and defining requirements for algorithmic transparency, bias reduction, and content labeling, etc. This can add complexity, especially when companies are doing business in jurisdictions that are still building out their AI policy. The need for CSPs to pivot in their approach to comply with varying regulations requires constant changes in operational, legal and construction processes underpinning the use of GenAI applications [6].

In short, while GenAI offers unprecedented opportunities to modernize telecom's technology and operations, its implementation depends upon balancing the significant challenges posed by data governance, legacy infrastructure, model viability, talent shortages, costs, and regulation. Each barrier requires a planned and coordinated approach, and together with sound cloud infrastructure support similar to that provided by AWS, practices in responsible AI governance, cross-disciplinary expertise, and an agile regulatory business models environment.

7. Future Outlook and Recommendations

As Generative AI continues to grow and affect how businesses operate, its presence in the telecommunications sector will be made more ubiquitous and more disruptive. The integration of GenAI with AWS infrastructure for 5G telecom networks is not a short-term innovation trend; it is a durable disruption that is about to reshape the operational, strategic, and business models of the industry sector. In the post-5G future, the evolution of CSPs to AI-native organizations through intelligent automation and AI-as-a-service will lead directly to how quickly and effectively they can embed GenAI in their networks and services and customer experience.

A potential future pathway that offers the most promise is the growth of sovereign large language models (LLMs) by telecommunications operators. Some CSPs have already started to fund the development of GenAI by developing its own models that take into account local languages, customer preferences, and regulatory regime surrounding data. In places such as Europe and the Asia-Pacific, where data residency laws and language diversity have made it extremely challenging to use generic globally trained models. By building sovereign LLMs, CSPs can deliver compliance and offer

training and services to their customers that are uniquely localized and personalized [6]. Additionally, these models will also help facilitate differentiated offerings that are more in line with regionally based and cultural expectations. This properly harnessed, provides company loyalty and gives the offering a competitive advantage.

Another significant area of development is the rise of AI-as-a-service (AIaaS) platforms. With specific packaging services hosted on AWS, telecom operators are well-positioned to deliver GenAI capabilities to small- and medium-sized businesses (SMBs). Essentially, telecom operators can take advantage of their current network and delivery infrastructure to offer services like virtual assistants, customer support automation tools, and domain specific predictive analytics engines. Expanding portfolio of offerings to include value-added AI services represents a significant opportunity for revenue generation as telecom services continue to commoditize [1][4]. Also, asymmetrical flexibility with AWS services like Bedrock, SageMaker, and Wavelength means these AIaaS offerings can be effectively scaled across geographies and business use cases.

Edge processing or inferencing usage will likely shape the future of GenAI in telecom. Increasingly, applications requiring real-time processing, such as autonomous vehicles, industrial automation and immersive media will lead to edge based GenAI models. AWS Wavelength and Local Zones are already making this transition a reality, by moving low-latency compute closer to the end user. In the coming years, CSPs will seek to deploy GenAI workloads at the edge so that they can allow rapid decision-making and real-time service delivery from the cell tower or local data center [3][5]. This means that the AI would be delivered in a more decentralized architecture, thereby improving the user experience and minimizing pressure on the core network and lowering costs.

CSPs need to strategically implement GenAI workloads by thinking along phases of value delivery. For example, assessing use cases that would have high impact, prioritizing the use cases based on business value and technical feasibility, and cycling at sprint length to produce prototypes of use cases.

As GenAI matures and becomes a primary focus of CSPs, developing a strong governance model framework that assesses data ethics, transparency of models, and ensuring all regulatory assessments are followed would benefit the CSP immensely. CSPs will also need to team up. CSPs will likely need to engage with hyperscalers, academic institutions, and governmental institutions to agree on standards, establish best practices, and create deals intent on achieving interoperability that ensures responsible AI use at scale [1][6].

Finally, CSPs have to develop (internal) talent and AI literacy. In order for GenAI to reach its full potential, telecoms need to build multidisciplinary teams to foster competitive advantage - comprising of domain experts, data scientists, and cloud-native engineers. Upskilling programs, partnerships with universities, and cross-enterprise innovation labs can be used to develop capabilities. To do this, leadership must create an experimental environment that allows for the testing of new use cases, business models, and solutions without the fear of failure [2][4].

In summary, 5G telecom networks will evolve at the intersection of GenAI and cloud. With AWS providing the technology, and GenAI providing capabilities in automation, personalization, and analytics, CSPs will able enter a new era of intelligent network management and consumer engagement. By tackling their current issues now, and preparing for the future, telecom operators can become AI-first organizations that are engaged with the leading digital enterprises of the future.

8. Conclusion

The application of Generative AI and the transformational impact of 5G telco infrastructures, enabled by the scalability and flexibility of AWS cloud services, may be the most remarkable advance we will see in the modern communications sector. The convergence of the two cases taken together is far more than a pure technology upgrade—it represents a paradigm shift in the way communication service providers run, customer interactions, and create new forms of value. This paper has documented the diversity of implications of MSP GenAI-managed AWS infrastructures, addressing the systemic implications for network optimization, operational efficiencies, customer experience, and revenue generation.

GenAI integrates new cognitive capabilities into existing network management processes. Traditional AI approaches focus on a limited scope of predetermined tasks, such as fault detection and customer segmentation, whereas GenAI provides dynamic, adaptive, and creative functions that are essential in the unpredictable, and sometimes chaotic space of 5G networks, to allow CSPs to autonomously engineer end-to-end network diagnostics, generate optimized configurations from an assortment of data, and provide pre-emptive, and predictive service management capabilities. These capabilities are crucial in the decentralization, virtualization, and data density of the 5G environment. Cloud and IT platforms such as AWS SageMaker, Bedrock, and Neptune, provide CSPs with the capabilities to train, customize, and

deploy GenAI models at scale. These models can support network and customer-facing services with improved operational speed and accuracy.

The use of GenAI also has implications for customer engagement. CSPs are leveraging GenAI to transform customer service from a mere cost center into a strategic differentiator, employing a range of intelligent agents, from virtual assistants that mimic human experience to recommendation engines deploying machine learning in real-time to provide individually tailored, personalized, multi-channel experiences. Intelligent agents work for CSPs 24 hours a day, 7 days per week, learn after each interaction, and provide a seamless service regardless of channel. CSPs have been struggling with negative effects stemming from persistent customer churn and service offerings that are fragmented across the different channels. When combined with the tools AWS makes available through its AI offerings (and its cloud delivery) intelligent agents can also be individualized by market fast they can be innovated and utilized on a substantial scale to service meeting local regulatory compliance and agility.

On the commercial side, GenAI can facilitate entirely new models of revenue. As traditional connectivity services become commoditized, CSPs are investigating the AI-as-a-Service market for its enterprise clients, including small- and mid-size businesses. These services include smart analytics platforms and customer engagement bots that will be hosted through AWS. CSPs are looking to capitalize on monetizing their investment in infrastructure, rather than relying on product-based revenue models. This paradigm shift from product-based revenue models to service-based models represents not only a new method of making money but reflects CSPs evolution from being connectivity service providers to becoming digital service suppliers offering full-stack service solutions.

Of course, the challenge is not insignificant. As noted, with regard to data governance, regulatory compliance, technical debt, and talent shortages, these conditions remain demonstrable hurdles. These constraints need to be overcome by careful effort and collaboration of that nature. The increasing complexities of ethical use of AI, data sovereignty, and cross-border compliance also add layers to risk management, particularly in jurisdictions with narrow legal frameworks. Nevertheless, CSPs can identify and mitigate many of these aspects with good governance, investment in AI skills and through partnerships with technology enablers such as AWS.

The future of telecom is intelligent automation, evidence-based decision-making, and customer-focused innovation, enabled by GenAI. As AWS continues to evolve its GenAI stack and telecoms deepen their AI capabilities, the sector will witness a profound transformation in how networks are designed, managed, and monetized. This paper has shown that GenAI, when integrated strategically with AWS cloud services, is not merely a tool for operational enhancement, but a catalyst for holistic, enterprise-wide reinvention.

In essence, the next generation of telecommunications will not just be faster or more connected—it will be smarter, more adaptive, and profoundly more human in how it understands and serves its users. GenAI is the enabler of this vision, and AWS provides the foundational infrastructure to make it a reality. Together, they represent the cornerstone of the future digital telco.

9. References

[1] World Economic Forum. (2025). Artificial Intelligence in Telecommunications. Geneva: World Economic Forum. In collaboration with Accenture. https://reports.weforum.org/docs/WEF_Artificial_Intelligence_in_Telecommunications_2025.pdf

[2] Wipro & Amazon Web Services. (2024). Applying Data, AI/ML and Generative AI to the Telecommunications Industry. AWS Partner Network Blog. Retrieved from https://aws.amazon.com/blogs/apn/applying-data-ai-ml-and-generative-ai-to-the-telecommunications-industry/

[3] Tan, A. (2023, October 19). How AWS is building a tech stack for generative AI. Computer Weekly. https://www.computerweekly.com/news/366556254/How-AWS-is-building-a-tech-stack-for-generative-AI

[4] Amazon Web Services. (2024). Generative AI for Telecom – AWS Solution Overview. https://aws.amazon.com/telecom/generative-ai/

[5] AWS Industries. (2024). Unleashing the Power of Graphs: Operating 5G Networks with GNN and Generative AI on AWS. https://aws.amazon.com/blogs/industries/transforming-network-management-graph-gnn-and-generative-ai-on-aws/

[6] Altman Solon. (2023). Telecommunications Generative AI Study. September 2023. In collaboration with AWS. https://pages.awscloud.com/rs/112-TZM-766/images/Altman%20Solon_AWS_Telecoms%20Generative%20AI%20Study.pdf