www.pipelinepub.com
Volume 21, Issue 9

How Data Engineering is Powering Trusted AI in Telecoms

By: Mari Nilsson Bjorkman

Artificial Intelligence (AI) is no longer a futuristic concept in the telecommunications sector - it is the driving force behind the industry's transformation. AI is revolutionising how telcos operate and deliver value, from predictive maintenance and automated service delivery to enabling innovations like IoT services and smart cities.

However, this revolution hinges on one critical factor: data. More specifically, the quality, governance and trustworthiness of the data that powers AI models.

In the AI era, data engineering has become the unsung hero of telco transformation. It is the foundation upon which reliable, transparent and explainable AI is built. As telecommunications organisations accelerate their AI journeys, the need for robust data engineering practices has never been more vital. Without them, AI applications risk delivering inaccurate predictions, inefficient operations, and even regulatory violations, potentially eroding profitability and customer trust.

The cost of poor data quality

The consequences of poor data quality are already being felt. Research shows that bad data costs organisations over £10 million annually on average due to operational inefficiencies, flawed analytics and poor decision-making.

In the context of AI, these costs can escalate quickly. Unlike traditional IT systems, AI models do not just process data—they learn from it. If the training data is incomplete, inconsistent, or biased, the resulting models will replicate and amplify those issues in real-time.

This is especially concerning in telecoms, where AI makes complex decisions across sprawling networks. Whether it's reallocating bandwidth, identifying service disruptions, or targeting customers with personalised offers, the risks of AI misjudgement are significant.

In such a high-stakes environment, trusted data is not a luxury - it's a necessity.

The impact of the rollout of 5G

5G has exponentially increased the volume, velocity, and variety of data that telecoms must handle. From mobile devices and base stations to connected cars and smart sensors, the infrastructure generates petabytes of data daily. This explosion of data brings opportunities for AI to enhance performance, but it also introduces new engineeringchallenges. For example, AI-powered predictive maintenance can analyse sensor data to predict when equipment might fail, allowing repairs to be scheduled before issues occur - minimising downtime and avoiding unexpected breakdowns. Likewise, AI is used in dynamic network optimisation to manage traffic flow, adjust bandwidth and reroute data in real time, ensuring stable and efficient service.

However, if the data feeding these systems is noisy, duplicated, or outdated, the predictions become flawed, leading to costly errors and degraded performance. Data engineers are on the frontlines of solving these challenges. They are tasked with designing pipelines that clean, standardise, and integrate diverse data sources while ensuring data freshness and availability for downstream AI models. The accuracy and effectiveness of AI are directly tied to the rigour and sophistication of these pipelines.

The importance of data governance

One of the most pressing data engineering concerns related to AI in the telecom industry is governance. Companies handle highly sensitive customer and network data, making compliance with privacy regulations such as GDPR and the UK's Digital.

Regulation Framework is critical. But governance is not just about legal compliance; it's about creating transparent, auditable AI systems that can be trusted by customers, regulators and internal stakeholders alike.

This starts with data lineage, which includes understanding where data comes from, how it has been transformed, and who has accessed it. Advanced metadata management platforms are increasingly being deployed to provide this visibility, enabling organisations to track data across its entire lifecycle. By maintaining data integrity and enforcing access controls, telcos can reduce the risk of unauthorised use, bias, or data leakage.

In AI workflows, this governance must extend to the model level. Data engineers and data scientists must work together to document the data used to train models and set foundations for easy monitoring to catch model degradations for timely refreshing/retraining.

This level of transparency builds confidence in AI-driven decisions and enables organisations to demonstrate accountability, which is a growing expectation from both customers and regulators.

The process of automating the data supply chain

AI is now being used to improve data engineering processes. Traditional data processing methods cannot cope with the complexity of today's telecom environments since manual cleaning, integration, and transformation are not only time-consuming but also error-prone.

Machine learning algorithms can be deployed at various stages of the data pipeline to detect anomalies, fill in missing values and standardise formats - automatically and at scale. This ensures that only high-quality data reaches AI models, reducing drift and improving long-term model performance.
Automation also enhances agility, allowing telcos to adapt quickly to changes in data sources, formats, or business requirements.

Real-time data ingestion tools are further transforming the telco landscape. Using streaming architectures, telcos can feed live data into AI models, enabling split-second decisions that optimise customer experience and network performance.

However, without proper validation and monitoring, real-time pipelines can introduce their own risks. Therefore, automation must be combined with continuous testing and observability, allowing engineers to catch and correct issues before propagating.

The role of synthetic data

Privacy remains a significant challenge in AI for telcos, especially when using customer data for model training. Organisations are under pressure to protect personally identifiable information (PII) while still gaining insights that drive innovation. One solution that is gaining momentum is synthetic data.

Synthetic data is artificially generated but statistically representative of real-world datasets. It allows organisations to train, test and optimise AI models without exposing sensitive or regulated data.

In telecoms, synthetic data can be used to simulate rare network failures, test new routing algorithms, or prototype customer journeys—all in a controlled and compliant environment. This approach enhances privacy and model robustness.

Real-world data is often imbalanced or incomplete, which can bias AI outcomes. Synthetic data allows engineers to create balanced datasets that reflect a wide range of scenarios, leading to fairer and more resilient AI systems.

Embracing an integrated approach and breaking down silos

Advanced tooling and automation alone won't guarantee AI success - people and culture matter just as much. Telcos must foster a culture of cross-functional collaboration between data engineers, data scientists and domain experts.

Each brings a unique perspective: engineers ensure data quality and infrastructure, scientists design and train models, and business stakeholders provide context and objectives.

Successful AI initiatives start with shared understanding. This means establishing common data definitions, co-developing governance frameworks, and leveraging collaborative platforms where teams can collaborate on datasets, models, and metrics.

Data engineers should be involved early in the AI development lifecycle, not just at the implementation stage, so they can help shape data strategies that support the end goals. Organisations that embrace this integrated approach are better positioned to deliver AI that is not only innovative but also reliable and sustainable.

Entering a new era of AI-ready data engineering

As telcos continue to evolve into digital service providers, the demands on their data infrastructure will only increase.

AI will become more embedded, more autonomous and more central to decision-making. Data engineering must also advance to support this evolution - becoming more intelligent, automated, and aligned with AI requirements.

This means adopting new ways of organising and managing data that make it easier to grow and adapt as needs change. It also involves applying smarter, more efficient ways of working to speed up the development and rollout of AI tools. All this will support innovation, ultimately enabling teams to experiment with data and AI to solve new challenges or improve business processes.

Just as importantly, it's about giving data teams the right tools to work more independently, monitor how data is used and maintain strong oversight to ensure everything runs smoothly and responsibly.

Ensuring responsible use of AI At its core, the telecoms industry's mission is simple: to ensure that every AI-driven decision is grounded in accurate, explainable and trusted data.

Data engineering is not just a technical discipline. It is a strategic enabler of trustworthy and responsible AI. By prioritising data quality, governance and automation, telcos can build the infrastructure required for the next generation of intelligent networks.

In doing so, they will not only optimise operations and increase customer satisfaction but also build a resilient, innovative and future-ready business.