

Metadata Stewardship and Quality in Telecom Sector

Bhola Gan Chaudhuri and Shalli Rani

Department of CSE, Chitkara University Institute of Engineering and Technology, Chitkara University, India

Article history

Received: 19-02-2026

Revised: 11-06-2026

Accepted: 22-09-2026

Corresponding Author:

Shalli Rani

Department of CSE, Chitkara

University Institute of

Engineering and Technology,

Chitkara University, India

Email: shalli79@gmail.com

Abstract: The spread of the telecom network infrastructure, constellations of satellites and the use of the Internet of Things (IoT) have brought about unprecedented volumes of heterogeneous information which facilitate the contemporary monitoring systems of the environment and the earth. Metadata control and information quality have also become significant in the telecom industry in controlling data reliably, enhancing data services, and providing quality analytics in the changing environment. Telecom environments are producing data environments that are large scale in terms of network, customer and service metadata; and the alternatives to convergent governance practices, common data standards and the effective stewardship mechanisms are lowering the data usability and real-time situational awareness. The difficulty is especially severe in remote sensing-driven intelligent visual analytics system in which the quality of metadata has a direct impact on discoverability, interoperability, and accuracy of decision making. This study determines these gaps by looking at strategic roles of metadata stewardship. The proposed approach in this paper is a telecom-conscious metadata stewardship through built-in automated quality measurement, semantic balancing, and governance, focusing on the policies. The framework describes a multidimensional metadata quality model that spans over completeness, consistency and accuracy, timeliness, lineage and interoperability in heterogeneous telecommunication domains. The use of machine learning-supported anomaly detectors, rule-based validation and metadata lineage screening could increase the level of transparency, auditability and regulatory compliance. Based on these comparisons, the proposed GAN-BERT-based structure presented higher accuracy (82.59%), precision (79.24%), recall (61.53%) and F1-score (79.01%) in comparison to the SSL, EM and the Co-Training models, in terms of the metadata quality prediction and the anomaly detection tasks. Telemetry open datasets have been used and simulated remote sensing metadata feeds with heterogeneous attributes related to customers, network and services were used in the experimental evaluation. The framework also enhanced the schema conformance, metadata transparency and traceability of governance, and scalability of high throughput telecom workloads. The experimental use presented in terms of telecom datasets and simulated remote sensing feeds indicates reduced metadata errors, increased schema conformance, and the maximized scalability on the high-throughput workloads, resulting in metadata stewardship as one of the enablers of intelligent visual analytics and real-time earth monitoring in telecom settings. The framework outlined offers a flexible and adaptable approach suitable for future network generations and can have significant application in terms of AI-driven governance, monitoring, and regulatory adherence for today's telecom networks.

Keywords: Metadata Stewardship, Metadata Quality, Remote Sensing, Intelligent Visual Analytics, Data Governance

Introduction

Across the telecom businesses, there are various techniques and control measures that are in place to support Metadata stewardship and quality management across different belonging compartments to meet reliability consistency and usability. Defining common metadata reference models, structures, and taxonomies to enhance the comprehensiveness, compatibility and inter-connectivity of various data sets, systems, and services (Verma and Bhatt, 2026). This encompasses standard identification of most frequently shared data attributes together with formats and naming systems for describing telecom data elements. Metadata Catalogs and Registries: Metadata repositories which contain knowledge structures such as data dictionaries, metadata indexes, data lineage information, and the forms of metadata schemas. These catalogs are utilized as central points where overall metadata assets identification, access and administration can be met (Lyashkevych, 2026).

The speed and complexity of 5G networks, Internet of Things (IoT) devices, satellite communication (satcom) systems and cloud-native telecom infrastructure have added tremendous layers and amounts of telecom metadata. For different types of customer services, billing systems, network operations, remote sensing platforms and intelligent monitoring systems, modern telecom operators create and log large volumes of metadata of different types. Unfortunately, there are many factors for poor quality metadata, inconsistent schemas, missing lineage info and lack of semantic interoperability that lead to operational inefficiencies, inlay analytics, late decision making and regulatory compliance problems. The constant shift in the telecom industry is forcing new approaches to dynamic governance maturity that are more automated, scalable, and reliant on real-time monitoring and governance of critical data and signals. Traditional metadata governance methods are primarily static, rule-based processes, and manual stewardship practices that cannot cope with telecom's dynamic industry and are no longer sufficient. Thus, in today's telecommunication infrastructures, it is essential to have intelligent AI-driven metadata stewardship mechanisms that can automatically assess quality of metadata and perform semantic validation and enable adaptive governance.

Data inconsistencies and quality issues in data formats/content can only be identified in effective metadata stewardship for telecom environments by data profiling. By using standardized metadata definitions, a consistent set of governance policies, and automated metadata extraction mechanism, the consistency, interoperability and efficient management of metadata across distributed systems can be achieved (Sharma and Pandey, 2026). Validation procedures and quality metrics are applied to metadata to assess its quality including measures such as accuracy, completeness and consistency

to ensure that metadata assets can be relied upon (Tebu and Izang, 2025). Furthermore, the metadata lineage tracking, metadata governance frameworks, access control, and metadata version management provide traceability, accountability, and good handling of metadata resources throughout their lifecycle (Paulraj, 2026).

Enabling metadata and their use in unification processes across various systems, platforms, and data sources. This includes establishing relationships between schemas of different systems, harmonization of the metadata and defining transfer protocols of data to enhance sharing of data. When telecom companies embrace proper metadata stewardship and quality management procedures, they can gain better control and quality of the telecom data elements, thus useful for decision making, compliance, operations and to improve customer experiences. Below are the most common steps used in the telecommunications sector to develop the metadata stewardship and quality process. Determine factors of the telecom operations that are associated with metadata (Mochalov et al., 2019). These are, for instance, metadata relating to the network elements, customers, calls, and services etc. The purpose and usage of each class of metadata need to be identified to determine its significance to the organization. Metadata from various sources of the organization can be collected from network management systems, billing systems, customer relationship management system, etc.

Metadata must be accurately collected in a format common to all systems, and it is useful to have associated context which originates from the metadata (source, timestamp of creation, lineage, etc.). Ensure that there is a single Metadata management center/ Metadata repository/Metadata Dictionary or a master index where all the metadata assets are held. Organize the metadata in accordance with certain agreed-upon classification schemes, such as taxonomies or ontologies, to make the material easier to find, analyze and retrieve. Capture a record of change history for metadata and make version control mechanisms to ensure changes made to metadata records are recorded.

Metadata quality assessment is the profiling and analysis of all metadata to determine how complete, accurate, consistent, and relevant the metadata is to defined standards and/or organizational needs. Data cleansing, data correction, data consistency and data enrichment of metadata with metadata, improves the usability and reliability of metadata assets. Good metadata governance also needs to be owned, stewarded, secured, monitored and improved over time, through meaningful feedback and compliance with business and regulatory needs. Furthermore, governance structures help to achieve interoperability between the diverse telecom systems, enable informed decision making and enhance transparency and accountability of the metadata management processes. The process shown in Figure 1 is called the analytical process for data governance.

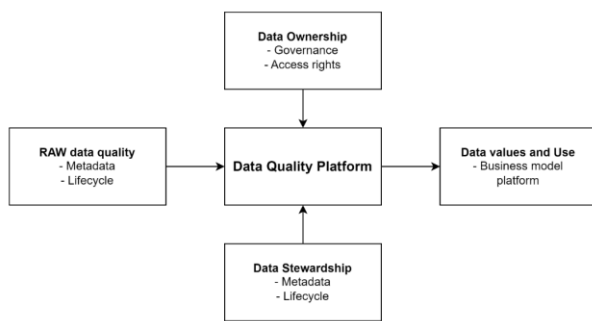


Fig. 1: Analytical process for data governance

Such processes will help telecom companies to develop the meta systematic approach to the management of Metadata and its quality that will be acceptable for a variety of business processes and demands of regulating authorities. This proof suggests that telecom companies can achieve metadata governance and quality assurance and further assure the accuracy of metadata to suit multiple business operations and compliance needs. In contrast to the current metadata governance practices that mostly rely on static rules for metadata validation, manual steering frameworks, the proposed framework marries AI-based metadata quality evaluation, semantic balancing, metadata stewardship automation, and anomaly detection using GAN-BERT for scalable and adaptive metadata stewardship for modern telecom ecosystems.

Research Objectives

The key goal of this research work is to introduce Intelligent metadata stewardship and Quality governance approach for a Telecom environment with AI and ML techniques. The work is centered around enhancing the quality of metadata, improving metadata governance transparency, improving interoperability and the operational scalability of heterogeneous telecom ecosystems. This is why the following objectives are set for this study:

- Identify and create metadata quality evaluation framework in a multi-dimensional way for Telecom metadata environment
- Integrate Artificial Intelligence and Machine Learning technologies for automate metadata validation, anomaly detection and semantic balancing
- Compare the efficacy of the GAN-BERT, SSL, EM and Co-Training methods in data metadata quality prediction and optimization for data governance
- To enhance metadata interoperability, lineage and governance compliance with regard to systems already in use in telecommunications implementation
- To investigate the scalability and applicability of the metadata stewardship system of air intelligence in high throughput telecom workload

Major Contributions

This research activity culminated into a set of key achievements, summarised below:

- Introduced a new approach to metadata stewardship and governance tailored to the telecom sector that leverages artificial intelligence
- Built a multi-dimensional model of metadata quality which includes representation of completeness, consistency, timeliness, interoperability, metadata lineage and metadata accuracy
- Wrote, developed and integrated a metadata based Quality Assessment mechanism with GAN-BERT to facilitate Intelligent Anomaly Detection and Semantic Quality in the metadata
- Compared to other approaches such as SSL, EM, Co-Training and GAN-BERT implemented with telecom metadata datasets
- Enhanced intelligent analytics systems for telecom solutions in terms of increased transparency, schema conformity and scalability of metadata management
- Delivered a flexible approach that enabled Next Wave Telecom Infrastructure for the IoT, satellite communication, remote sensing worlds and more

Related Work

Modern telecom ecosystems generate a tremendous amount of metadata, which is growing in volume, variety, and complexity due to the evolution of the telecommunication infrastructure, including 5G networks, devices based on the Internet of Things (IoT), satellite communications and the use of cloud-native architectures, intelligent edge computing, and remote sensing technologies. Previous studies deal with the development of frameworks of Metadata governance, data quality management, standards for interoperability and mechanisms for controlling of data lifecycles necessary for effective integration of metadata from customers, services and resources. Most industry standard information models have been made more consistent, and are more operationally coordinated, but they continue to use traditional validation methods and manual stewardship procedures, which are less effective in large environments, and highly dynamic (Lubis et al., 2022).

The modern breakthroughs in artificial intelligence, machine learning, deep learning, and transformer models have paved the way for new challenges and opportunities in automated quality assessment, anomaly detection, predictive analytics, and semantic knowledge, as well as intelligent decision-making. These technologies are achieving excellent results in optimizing networks and automating operations but are less advanced when it comes to metadata stewardship. Despite the progress,

issues like semantic validation, interoperability management, lineage tracking, governance traceability, and scalable automation remain a constant challenge, emphasizing the need for an AI capable of aiding in adaptive metadata governance for next-generation telecom solutions (Zhao and Krishnan, 2025).

Materials and Methods

There are different models that can be applied to the telecom sector for the purpose of metadata stewardship and quality management to guarantee that all the metadata resources are adequate and precise. There is a framework called Frameworkx from the TM Forum which contains sets of guidelines dedicated to telecom operations including the use of metadata. In Frameworkx the SID stands for Information Framework, and it is dealing mainly with metadata and quality issues.

Dataset Description

The tests of the proposed metadata stewardship framework were carried out on real operational metadata from the telecom industry and mock-up remote sensing metadata streams. The telecom metadata dataset comprised diverse attributes that were related to customer management systems, billing systems, network traffic records, service provisioning logs and telecom infrastructure monitoring systems. There were structured and semi-structured metadata elements, including, for instance: Service identifiers, timing elements, network event logs, device information, call records, session parameters, and metadata lineage attributes.

Realistic metadata streams were created by means of controlled metadata augmentation and anomaly injection techniques. The datasets created have been incomplete with missing metadata field values, spurious schema incompleteness, duplicate records, delayed timestamps, and semantically questionable data to test the viability of the proposed framework in dynamic telecom settings. For machine learning evaluation, the experimental dataset was split at 80:20 as training set and testing set respectively.

Six dimensions of metadata quality (completeness, consistency, interoperability, timeliness, accuracy, and lineage traceability) were emphasized during the metadata quality evaluation process. The dimensions were chosen as they directly affect the operational reliability, service orchestration, intelligent analytics and regulatory compliance capabilities of telecom.

Metadata Quality Measurement Parameters

The proposed framework proposes to assess the quality of metadata using multiple quality assessment metrics in the multidimensional paradigm. Completeness is another indicator used to assess whether or not a

mandated metadata requirement field has been provided in a telecom data set. Consistency checks conformity of schemas and uniformity of semantics in distributed telecom systems. The accuracy is used to measure the degree of correctness of metadata in comparison with the validated public record in telecom operation. Timeliness is the time lag from metadata creation to making metadata available in governance systems. Interoperability measures the matchability of metadata across disparate telecom platforms and lineage traceability measures the ability to trace metadata from an origin to its transformation and use throughout the metadata lifecycle.

These parameters are combined, with weighted aggregation, to identify metadata anomalies and violations of good governance by calculating the 'metadata quality score'. An integrated mechanism for validating metadata with the assistance of machine learning techniques was incorporated to automatically identify the situations of inconsistencies in the metadata in real-time and categorize the metadata quality conditions.

Proposed AI-Driven Metadata Stewardship Framework

The proposed framework provides a repository of all the telecom metadata metadata governance, semantic balancing, anomaly detection and machine learning assisted quality evaluation as a single telecom metadata stewardship architecture. At first, metadata streams are gathered from information and control systems in the telecom environment, such as network infrastructure, customer relationship management systems, billing systems, remote sensing feeds and IoT monitoring environments. The gathered metadata is then preprocessed through several stages, such as normalizing the data, addressing missing values, eliminating duplicate records, and aligning the schemas.

Then feature extraction techniques are used to extract and collect appropriate metadata quality indicators like semantic similarity, metadata completeness ratio, timestamp consistency, lineage relationship etc. Extracted metadata features provided the machine learning models: Semi-Supervised Learning (SSL); Expectation Maximization (EM); Co-Training (CT); GAN-BERT architectures used for anomaly detection and the metadata quality classification.

GAN-BERT was chosen because it combines the advantages of combining transformer-based semantic understanding with adversarial learning mechanisms. The transformer module facilitates a contextual understanding of the semantics of the metadata, and the adversarial module enhances the robustness of the system to treat challenging environments of metadata such as sparse, scarce or perhaps labeled metadata that are typical in the telecom industry.

The governance layer in the framework is responsible for automatic metadata validation, policy enforcement,

schema conformance checks as well as lineage monitoring. Semantic balancing mechanisms are introduced in order to minimise metadata inconsistencies and achieve interoperability between heterogeneous telecommunication platforms. The framework also enables continuous monitoring of data whenever it is used in high throughput telecom workloads, plus adaptive governance.

Experimental Setup

The implementation of the experimental case was carried out with machine learning libraries and transformer architectures in Python. The metadata modules for pre-processing were included in the evaluation environment along with the metadata modules for anomaly detection and metadata governance validation modules, the GAN-BERT implementation. For the evaluation, Accuracy, Precision, Recall and F1-score were chosen and the efficiency of SSL, EM, Co-Training and GAN-BERT have been compared in this regard.

The experiments were conducted to measure capability of metadata quality prediction, efficiency of anomaly detection, governance traceability, and scalability in the presence of heterogeneity in the telecom metadata. The hyperparameter tuning was done with iterative methods to optimize classification performance, and the ability of metadata anomaly detection.

Algorithm Workflow

The proposed metadata stewardship workflow consists of the following major stages:

1. Collection of metadata from diverse operational sources via telecommunication
2. Metadata pre-processing and normalisation
3. Quality of metadata features extraction
4. Semantic balancing / metadata enriching
5. Anomaly detection through AI-based methods including the GAN-BERT and comparative ML algorithms
6. Good metadata quality classification and validation/control
7. Interoperability and line of tracking
8. Meta data monitoring and adaptation

The workflow allows automated metadata quality management and intelligent governance, suitable for large-scale telecom infrastructures and for the next generation of communication environments.

Here is the best way through which SID can be used for metadata stewardship and quality in telecom industries. In distributed management system (DMS) TM Forum's define as in distributed management system (DMS) TM Forum's define as:

$$\Phi = \min \sum_{y=1}^L P_y \times t_y(\theta, S, Q)$$

Where:

- Φ is optimal distributed management system
- t_y is execution time
- θ as a distribution method
- S as service schedule
- Q as an integration method
- L as the number of BPs served
- P_y is execution priority

Assume that each PCs are in two set f_i and f_j and exchanging joint request with frequency $P(i, j)$. The set of PCs $\Phi = \{F_1, F_2, \dots, F_N\}$:

$$\bigcup_{i=1}^n f_i = \bigcup_{i=1}^d F_i \text{ and } F_k \cap F_l = \emptyset, k \neq l$$

So, the frequency of conflicts on PM:

$$C_k = \sum_{i,j} P(i, j)$$

Then the total frequency can be calculated as:

$$C = \sum_{k=1}^d C_k = \sum_{k=1}^d \sum_{i,j} P(i, j)$$

So, optimality criterion defined as:

$$\Delta C = \sum_{i \in PM_i} (P(i, j) + P(j, i)) - \sum_{i \in PM_s} (P(i, j) + P(j, i))$$

The Shared Information/Data (SID) model establishes a standard definition of a set of structures, attributes and relationships for information in the various domains of telecom such as customers, services, products and network resources (Yebenes and Zorrilla, 2019).

It enables consistent metadata definitions, quality assessment and governance throughout the metadata lifecycle, and the acquisition, storage, retrieval, governance, and quality management of metadata. Data governance framework is explained in Figure 2.

The Common Information Model (CIM) builds upon these to provide reusable models and information management metadata which help interoperability and information exchange among heterogeneous telecom systems and platforms (O'Hara, 2019). Furthermore, features of the TM Forum Framework framework are available that facilitate the implementation of metadata stewardship and Quality Improvement and standardisation of governance. The benefits of adoption of SID, CIM and TM Forum standards include enhancing metadata consistency, reliability, traceability and compliance with regulation, as well as allowing telecom organizations to support effective enterprise-wide metadata management (Srivastava and Shekhar, 2020).

Data Governance tasks and activities for Telecom sector: Data governance in the telecom sector involves the control and accountability of a massive amount of heterogeneous data from multiple sources including customers, networks, and financial transactions (Krupitzer et al., 2020). Proper data management is the key to quality, security, compliance, and usefulness if any data is going to be collected and stored. The following are the major tasks and activities that are associated with data stewarding in the telecom industry.

Data Governance Framework Development: Design detailed data governance policies which are responsive to issues of data quality; data security; data privacy and data compliance relating to the telecom industry. There should also be ways through which these policies have to be complied with by the different units in the organization. Make a clear distinction between the roles & responsibilities of data ownership and the responsibilities of data management groups and individuals (Raibulet et al., 2020). Figure 2 shows data governance framework. There should be set structures to help assign responsibilities that will make data stewards hold responsibility for certain data sets

Data Quality Management: Pave way for generation of industry specific guidelines for data accuracy, data completeness, data consistency, data reliability in telecom data. It is crucial, therefore, to set standard rates and reference points commonly known as KPIs for data quality. Use automated tools that would later monitor the quality of the data on regular basis. Oversee the commencement of data quality enhancement activities to take care of the arisen problems, including errors' elimination and duplication (Ren et al., 2021).

Data Integration and Interoperability: These include identification of all the customer data, network data and

all the data with Record to Materiality value. Coordination of data from various sources in a concerted data platform should be resolved through the following approaches. It means to create specifications by the adoption of which data from the various systems and platforms incorporated in the telecom infrastructure can be interchanged. Sustain the APIs that are responsible for moving information between applications.

Metadata Management: Create standards on metadata for specific telecommunication data while at the same time documenting the data attributes. Store and manage metadata in central 'hubs' that can be accessed by anyone with a legitimate interest in the data as required. Perform metadata evaluation on a frequent basis to make certain that is up to date with the existing data. Metadata should be enriched with supplementary information that would make the data more easily searchable and more functional (Breazzano et al., 2021).

Security and Privacy Management: Therefore, it is high time to devise and enforce stringent data protection mechanisms appropriate to the telecom industry. Set up and consistently practice policies in physical computer and network controls to prevent data breaches. Comply with the requirements which are specific to telecommunication, including GDPR, CCPA and other regional regulations regarding data privacy. Apply the principles of data anonymization and data masking to minimize the risks related to customers' personal data.

Lifecycle Data Management: Promote guidelines for data within an organization from the time it is generated, applied, stored and when it is no longer needed. Implement data retention policies so that data collected will only be stored for a certain amount of time to meet legal, regulatory, and business requirements (Chen et al., 2021).

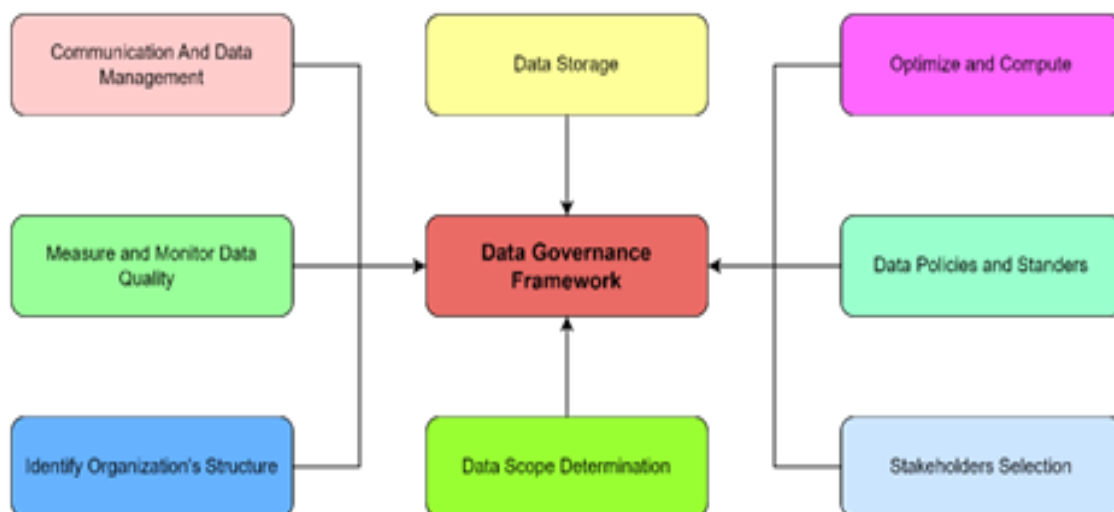


Fig. 2: Data governance framework

The other is to put in place measures that allow for proper storage of backup copies that are critical for the future of organizations. Implement policies concerning the proper disposal of information that is not relevant any longer for the organization as well as possible legal and regulatory guidelines.

Stakeholder Engagement and Training: Communicate with the stakeholders conveying the relevant policies, changes, and new initiatives within data governance. Implement feedback mechanisms to capture stakeholders' feedback on data governance practices for modifications to be made. Develop and implement awareness sessions for employee training on data stewardship and the corporation's governance policies and procedures. Key activities that need to be done are to ensure that there are awareness campaigns which would ensure that all the members of the organization embrace data driven culture.

Data Stewardship Tools and Technologies: Assess and consider those tools of data governance which can be of use in telecommunication sectors and can be selected based on requirements. Supervise the deployment of these tools, they should be properly incorporated into the other tools that are used (Purbasari et al., 2021). Ensure that data governance tools used are continually supported and updated to address emerging needs to allow comfort to the organization. Tools should be updated for new versions for constant update to fit in the new features in dealing with new data governance issues.

Data stewardship refers to the process of taking full responsibility of organizational data and its management from the time they were obtained to the time when it becomes useless. From these tasks and activities, the data stewards in the telecom sector can deploy proper practices in data governance hence improving on the quality of data, compliance to set regulations and effectively using

data for business intelligence and decisions (Tanvir et al., 2023). Thus, a Data Quality Management System or DQMS in telecom sector will help to make sure that data is accurate, complete, valid, and consistent that is helpful in operations, compliance, and decision making.

Metadata Management and Data Classification: In the telecom industry, metadata management and data classification play crucial roles in ensuring accurate documentation, discovery, governance and control of the data assets depending on how sensitive their data can be. The creation of standard metadata elements (data source, creation date, ownership, usage restriction), as well as following industry standards and organizational policies is a part of the metadata management (Myszewski et al., 2022). Periodic checks, quality assessments, and enrichment activities of metadata can ensure its accuracy, completeness, and relevance, while the use of centralized metadata storage/repositories and Data Catalog helps maximize efficiency in metadata storage and retrieval, and their management (Chen et al., 2022). The metadata management and data classification framework is shown in Figure 3.

Data categorization systems create classifications like public, internal, confidential or highly confidential data to label information to handle appropriately in accordance to business and regulatory needs (Danielsson et al., 2022). Classification policies outline methodologies for the classification, reclassification and tracking of data throughout its existence and machine learning-based techniques could complement the efficient classification of large-scale telecom data. The storage, transmission, and usage of classified information are regulated by security policies and Role Based Access Control (RBAC), which only allow identified users to access information of that sort (Meritxell et al., 2022).

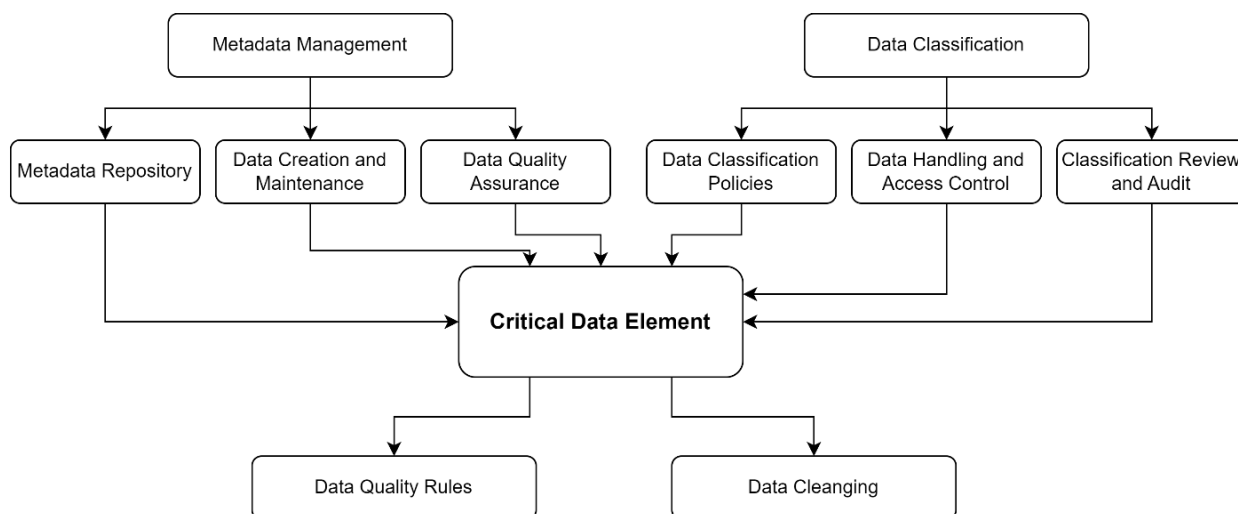


Fig. 3: Metadata management and data classification

The capacity of an integration to provide better data discoverability, effective data governance and security access management, as well as meeting regulatory requirements, can help achieve the overall goals of telecom data stewardship (Ouyang et al., 2022).

For RBAC0 user role assignment relation $UA \subseteq U \times R$, permission role assignment relation $PA \subseteq P \times R$. So, $Roles(u)$ defined as, $Roles(u) = \{r \in R: (u, r) \in UA\}$. Where, U is user, R is role, and P is permission. So, set of permission as:

$$Prms(r) = \{p \in P: (p, r) \in PA\} \quad \text{and} \quad Prms(R') = \bigcup_{r \in R'} Prms(r)$$

RBAC1 defined as:

$$\downarrow Roles(u) = \{r' \in R: \exists r \in Roles(u), r' \leq r\}$$

By optimizing metadata and adopting effective data classification techniques, telecom organizations can increase the quality and compliance of its data and thus can work optimally and with efficiency.

Data Quality Standards and Policies: Implementing data quality standards and policies within the telecom industry is paramount to ensuring that data has credibility, accuracy, consistency, completeness, is timely, relevant and reliable; and is important for operational efficiency, regulatory compliance and informed decision making. Data quality frameworks must specify acceptable level of quality, validation, reconciliation and monitoring

processes to guarantee that data denotes real world entities and do not additionally suffer from variations in the systems. (Arif et al., 2023).

Fundamentally, it depends on the clear assignments of data owner(s), data steward(s), and data quality managers; policies for data profiling, data validation, data storage, data integration, and data lifecycle management (Sudigyo et al., 2023). Frequent audits, data protection measures, encryption protocols, and access controls make sure that sensitive telecom data is kept secure and complies with regulatory standards like GDPR and CCPA. However, organizations should also put in place protocols for ethical data use, regular audits, communication with stakeholders and data policy reviews to ensure they continue to be good managers of quality and governance. The architecture of data quality as a service is shown in Figure 4.

Thus, by determining and enforcing these rules and regulation of model of data quality, telecom organizations can attain the desired qualities to the organizational data such as accuracy, consistency, completeness, timeliness, relevancy, and reliability, that is, achieving good data governance and improved organizational performance.

Data Quality Assessment System: As for the GAN-BERT tasks to be used in the telecom industry may be used for the generation of synthetic call data records, or customer interactions, or other network performance, judging the data quality by the accuracy, recall, F1 score can be also applied to compare the characteristics of the generated data with the real telecom data (Emmadi et al., 2023). Here's how you can apply these metrics.

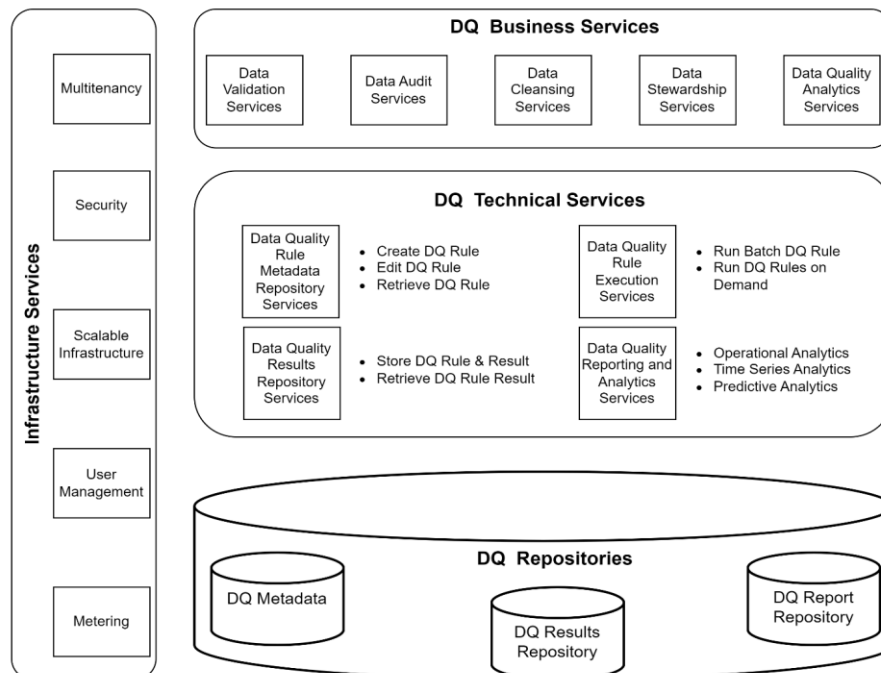


Fig. 4: Architecture of Data Quality as a Service

Accuracy: In this case, accuracy means the level at which telecom data is reconstructed keeping the telecom attributes including call duration, types of calls, customers, network characteristics among other attributes into consideration Accuracy can be determined from the ratio of the number of accurately generated data points to the number of data points in a validation set of telecom data (Belkhale et al., 2023):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

Recall: As a preview, in the given dataset context for telecom data generation, it estimates how well the proposed GAN-BERT model can embrace existing characteristics that exists in real telecom data. It measures the extents of accurately generated instances (True Positive) to the total sum of accurately generated instances (True Positive) and the instances that are not accurately generated (False Negative):

$$Recall = \frac{TP}{TP + FN}$$

F1 Score: F1 score is the average of precision and recall where both are given equal importance and does not favor one over the other. It is especially helpful when the proportion of accurately generated instances and the proportion of the instances inaccurately generated, in the generated telecom data, is not equivalent:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Where precision is the ratio of true positives to the sum of true positives and false positives:

$$Precision = \frac{TP}{TP + FP}$$

It is possible to calculate these metrics with the help of proper validation methods by comparing the generated telecom data with the real telecom data. Furthermore, while performing the data generation and evaluation it might be also relevant to apply domain-specific criteria that would reflect the complexity and peculiarities of telecom operations in the best way. Figure 5 shows data quality assessment system.

In this way, thanks to moving from ad hoc, non-standard, and reactive metadata management to systematic, prescribed, and predictive one, the telecom sector has boosted metadata stewardship and stewardship quality (Khoh et al., 2023). This evolution enhances good data management, service delivery as well as compliance

with the various standards set by the regulatory authority. Use of Date Picker, AutoComplete and other form controls for automatic metadata creation, validation, and update, to minimize on the likelihood of errors. Compliance to standards which include ISO 19115, ISO 19139 for compatibility with the other systems. Periodical check and constant assessment to ensure that the quality of the metadata is commendable. Coordination of metadata with respect to the various platforms and services within a given environment to make the exchange of information, efficient and effective. Controlling of training sessions and policies that will be acted out by the staff to minimize on metadata malpractices.

More on the same we have Automation and Standardization leading to improved Accuracy and decreased errors. Process Efficiency because the several repetitious tasks will be minimized through automation. It is a term used to define better adherence to regulatory requirement. Data utility through enhanced data availability and compatibility of data. Such practices enable telecom companies manage their data more efficiently and address issues that include enhanced decision-making, efficiency in service delivery and compliance to regulatory requirements.

Data Augmentation Techniques used in AI & ML for Data Quality: Data Augmentation is a critical technique in ensuring the quality of telecom data and enhancing the performance of Artificial Intelligence (AI) and Machine Learning (ML) models. Data augmentation by creating more training samples from existing datasets boosts diversity in datasets, introduces robustness and generalization. This is useful in telecom settings where data marked with specified labels might be scarce, poor or unevenly distributed.

There are multiple ways using AI & ML to augment your data and enhance the quality of your metadata. The advantages of the Semi-Supervised Learning (SSL) strategy are obvious, as it includes both labeled and unlabeled data to enhance the learning process and generate more training information, which make the model more adaptable and more accurate for classification.

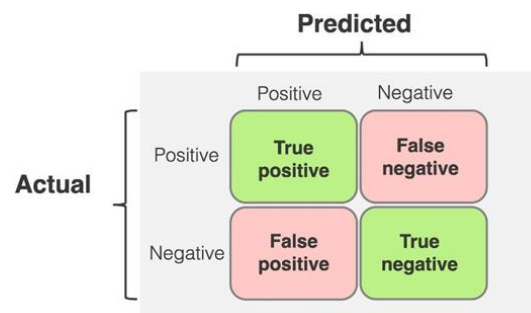


Fig. 5: Data quality assessment system

Co-Training takes multiple images of the same data and enables multiple models to learn a complement of information and provide enhanced samples to facilitate training. Expectation-Maximization (EM) is a method of probabilistic modelling that can be used when parts of the label or the latent variables are missing for a given sample, can generate further samples with the help of the missing information and is used to estimate their mean.

One of these methods is Generative Adversarial Networks with Bidirectional Encoder Representations from Transformers (GAN-BERT), which is particularly promising for data augmentation and metadata quality assessment. By fusing the generative power of GANs and the semantic representation ability of BERT, GAN-BERT generates more realistic and semantically relevant samples which are much closer to the original data distribution and fulfill the original distribution. The augmented data then can be employed to enhance the accuracy of classification, anomaly detection capability, and the quality of metadata prediction.

Let y denote the true class for X . Let v generic input for D . Real X processed by BERT from the generator G . Let PG be the distribution input v . Let $(p_D(\hat{y} = y|v, y = 0))$ be the provability of v classified by determinator D as fake. Let $PD(\hat{y} = y|v, y \in \{1,2,3\})$ provability of input v .

So, the discriminator loss function ℓ_D is given as:

$$\ell_D = \ell_{D_{Score}} + \ell_{D_{RF}}$$

Where, $\ell_{D_{Score}} = -E_{v,y \sim PB}[\log p_D(\hat{y} = y|v, y \in \{1,2,3\})]$ and:

$$\ell_{D_{RF}} = -E_{v,y \sim PB}[\log(1 - p_D(\hat{y} = y|v, y = 0))] - E_{v,y \sim PB}[\log p_D(\hat{y} = y|v, y = 0)]$$

Let $f(v)$ be the activation of the hidden layer D given input v . Then the generator G define as *feature matching* generator loss be:

$$\ell_{G_{feature\ matching}} = \|E_{v \sim PB}[f(v)] - E_{v \sim PG}[f(v)]\|^2$$

So, Complete generator loss function ℓ_D is given as:

$$\ell_D = \ell_{G_{feature\ matching}} + \ell_{G_{caught}}$$

Where:

$$\ell_{G_{caught}} = -E_{v,y \sim PG}[\log(1 - p_D(\hat{y} = y|v, y = 0))]$$

The proposed GAN-BERT framework is well suited for telecom applications as it can cope with heterogeneous and partial labeled metadata and enhance transparency of the semantic consistency and governance. Moreover, its potential in tasks like Network Traffic Analysis,

Customer Churn Prediction, Predictive Maintenance, Network Optimization, and Service Quality Assessment showcases its relevance in today's telecom landscape.

Proposed Methodology

The following outlines the fundamental steps that need to be taken to create a more sound and practical approach of applying AI and machine learning in the management of metadata stewardship and quality. First, it is essential to gain a broad idea of what kind of metadata exists in the given field and in what forms it can exist. Metadata will consist of data descriptions, data where-used lists, data quality values, and other related content. Acquire suitable datasets which possess information regarding their metadata. Make sure you did not mess up with the format of the data, or failure to annotate it correctly (Myers et al., 2024). That step is utmost important for training accurate models with Machine Learning algorithms. Find out if there are useful attributes from the metadata, which can be used to describe quality of the metadata as well as its management. This can include such operations such as completeness, accuracy, consistency, and timeliness of attributes extracted. Add 'ground truth' tags to metadata regarding the quality or regarding its stewardship level. Figure 6 shows data quality proposed architecture. This step may need the help of domain knowledge to decide correct quality state parameters and values to be set.

Select proper Machine Learning algorithms for the given problem or work. This could be in the form of judgmental approaches like decision trees, random forecast, or elaborate learning types like deep learning for errands like luminance pattern or language analysis. This means that the selected models need to be trained with the annotated metadata (Alboqmi et al., 2024). hyperparameters tuning and validation techniques must be given emphasis to guarantee the best performance out of the model. Check on the performance of the trained models using the right indicators. This could be simple metrics such as accuracy, precision, recall or F1-Score, or even more complicated metrics that are selected for metadata quality assessment. Revised the models from time to time according to the feedback received and new data available. Over time, structural changes may occur in the metadata resulting into modification in the models based on new demands and qualities. Implement the trained models into automation and optimization of the metadata stewardship process. This could also entail aspects such as automated identification of quality issues, suggestions on improvements that could be made or parts that require human intervention. The following is a list of best practices that should be followed when building, training, and deploying AI models as follows: If the models are no longer effective then retrain or update the models to make them effective again.

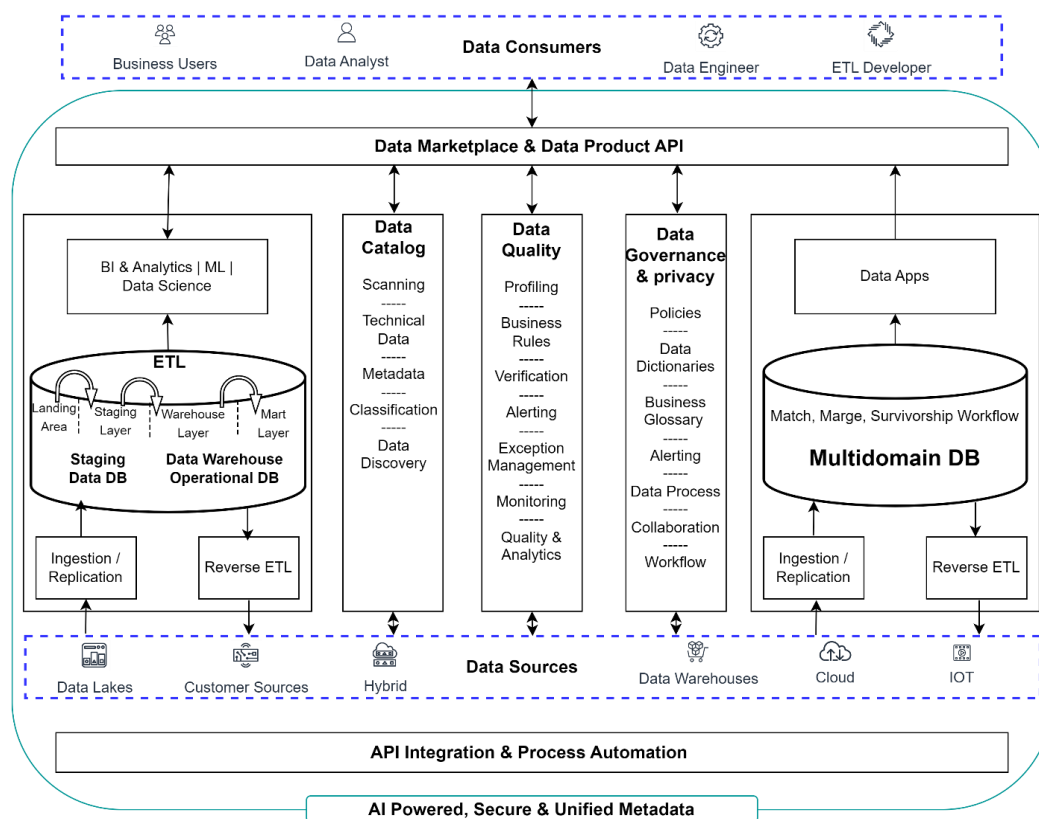


Fig. 6: Data quality proposed architecture

In all the stages, it is pertinent to engage use domain specialists, data stewards and other stakeholders to ensure that identified AI solutions support the established organizational objectives and standards. Furthermore, issues relating to partiality and data privacy, or protection should be well dealt with in all stages of the methodology section. The metadata problems can also be predicted by AI models so that necessary corrections can be made as and when required. Machine learning searches for metadata trends and forecast when a shift will occur and can then act accordingly. Translates unstructured information or data into metadata which includes appending automatically generated more meaningful data to the data received (Yandrapalli, 2024). Data mining with the help of AI helps in the identification of metadata inconsistencies across various datasets, and correction of such. Several instances of metadata problems are identified using real-time machine learning algorithms and thus resolved.

Ironing out means that the manual effort is cut out and efficiency gains are made. AI and machine learning solutions are adjustable to the amount of information that is received by the system. Automated metadata refining due to the learning and memory functions which can improve the accuracy of metadata description throughout subsequent repetitions (Noman et al., 2024). The

application of AI and machine learning in telecoms has the potential of improving the quality of collected metadata, thus improving efficient data management, and more importantly improving the quality of decision making in their companies.

Results and Discussion

Regarding telecom sector, utilization of AI and machine learning for metadata stewardship and quality has following advantages. Metadata, for instance, can be process by AI algorithms to look at hundreds of pages and recognize if there are contradictions, wrong data input or data which is missing. Such kind of approach will help the telecom companies to maintain their databases more reliable and trustful when data quality problems will be indicated by the system. There are many routine processes in metadata stewardship, which may be automated to avoid many tasks that require human intervention. Due to this, the telecom companies' resources are utilized better and only an adequate amount is devoted to the specifics of the data keeping requirements rather than daily efforts (Pradhan et al., 2024). As will be further explained, correct metadata can help to speed up and direct the decision-making processes.

Experimental Performance Analysis

The proposed GAN-BERT-based metadata stewardship framework showed greater performance in comparisons of quality prediction and anomaly detection capabilities than other frameworks, such as SSL, EM, and Co-Training. The classification accuracy was highest for the GAN-BERT model with an accuracy of 82.59% and better precision and F1-scores than the other models. The ability of GAN-BERT to represent semantics using its transformer architecture improved the understanding of metadata relationships, contextual semantics and hidden metadata inconsistencies in heterogeneous telecom datasets.

Because SSL relies on a small set of samples of labeled metadata and has less semantic representation ability, the SSL model had less favorable performance. The EM-based approach has been able to estimate the parameters and improve the performance of metadata classification, but it achieved less robustness in dealing with the incomplete and inconsistent telecom metadata records. Likewise, the Co-Training method was able to have decent classification performance, however it was less adaptable in situations where the metadata is updated constantly, having multiple different schemas and different metadata structures.

The proposed GAN-BERT framework showed better anomaly detection skills for metadata, namely detecting duplicated record, schema mismatch, missing data types in metadata, delay in data types, and semantic inconsistencies, among telecom OSs. The adversarial learning mechanism enhanced the capabilities of this framework: Getting better at generalizing between partially labeled and sparse metadata conditions that are prevalent in telecom ecosystems.

Scalability and Governance Analysis

The scalability analysis showed the proposed framework could ensure the stable performance of aggregating metadata quality assessments under high throughput telecom workloads where the metadata is continually generated. The system used to ingest metadata from a variety of sources, including customer management systems, billing infrastructure, network monitoring, remote sensing and IoT, and successfully manages the data Heterogeneity.

The governance layer increased metadata traceability, schema conformance and interoperability in distributed Telecommunication infrastructures. With the automated lineage tracking feature, the process for tracking metadata changes and lifecycle activities was improved, thereby increasing the transparency and auditability of the process. The implementation of the semantic balancing mechanisms also minimized metadata conflicts and telecom interoperability problems between heterogeneous systems.

Moreover, the automated governance framework reduced the manual effort required for metadata validation by automating both metadata quality assessment and anomaly detection process with AI. This helped improve operational efficiency, minimize inconsistencies in metadata and empower analytics decisions in telecom environments.

Hyperparameter and Model Behavior Analysis

Experimental results suggested that the hyperparameters (learning rate, batch size, transformer embedding, and training epochs) play a critical role in GAN-BERT performance. Prominent hyperparameter tuning enhanced semantic learning capability and accuracy of metadata anomaly detection. Higher semantic sensitivity with smaller batch sizes and higher computational time spent, and higher scalability and training times with larger batch sizes.

Another experiment revealed that the transformer-based contextual embedding mechanism can better discover semantic relationships among different attributes of the telecom metadata compared to those of the machine learning-based methods. Additionally, the models were further strengthened and made robust through the use of the adversarial training process in the presence of noisy and incomplete metadata environment.

While the proposed framework showed good capability in metadata governance and anomaly detection, the scalability was decreased as the volume of the metadata and their dimensions of the semantic features increased, increasing computational complexity and increasing training time. Other optimization techniques like distributed training, federated learning and light weight transformer architecture will provide further to enhance the scalability of a real-life telecom deployment.

Discussion

The results of the experiments show that metadata stewardship based on artificial intelligence can lead to a significant improvement in metadata governance, interoperability and reliability within a telecom system. The seamless embedding of automated feature recognition and anomaly detection through semantic validation brings flexibility into metadata management that can be applied to the dynamic and diverse telecommunications environment.

The proposed framework also helps in enhancing intelligent visual analytics, telecom monitoring systems, regulatory compliance management and automated governance operations. Improved service orchestration capability, predictive analytics and efficiency of operational decisions making can be realized by the Telecom Operator by minimizing inconsistencies and normalizing metadata and making it transparent.

The current study has a few weaknesses, however. When considering the experimental evaluation, there was

a reliance on simulated telecom metadata environments and controlled techniques for the introduction of anomalies, based on the unavailability of shared telecom metadata data sets. Moreover, the cost of computation for transformer types can impact deployment considerations in highly constrained edge environments in telecommunications.

Future research may include testing the validator in real-world deployments of telecoms, integration with a streaming analytics platform, metadata governance architectures that rely on federated metadata, optimization of the lightweight transformer, and intelligent systems with large language models for adaptive governance of telecoms.

While the proposed framework has shown to perform well in a simulated environment of telecom metadata, future work will include testing of the framework in the operational environment of telecoms, where possible, by collaborating with industry and using telecoms publicly available data sources. This validation will help validate the metadata quality prediction capability, anomaly detection, the effectiveness of governance and scalability under real operational conditions. Future studies will also explore deployments in real telecom infrastructures to assess interoperability, compliance and governance performance in a multitude of network environments such as 5G, IoT, cloud-native and edge computing. A number of approvals, access rights, infrastructure and research funds are required for practical deployment and validation, because of the sensitive nature of telecom data and regulatory constraints. Furthermore, in production environments there would be a need to coordinate with the senior management, the Security Operations Centers (SOC) and the Network Operations Centers (NOC) for testing and deployment because it could impact the daily operations. This would give more empirical evidence on its practical impact and applicability, as well as validate the proposed framework on a large scale.

Table 1 shows performance data for different algorithms and Figure 7 shows data augmentation performance graph. Thus, telecom operators can use metadata stewardship plus AI technologies to improve their understanding of network performance, customers' behavior, and market shifts that help them act faster in response to emerging business environments.

Using machine learning models, it is possible to some of the problems that could occur and prevent them from reaching critical stages. For instance, the use of predictive

analytics algorithms can help detect the likelihood of network traffic jams through analysis of the metadata and thus enable the operators to proactively deal with the probability of service interruption.

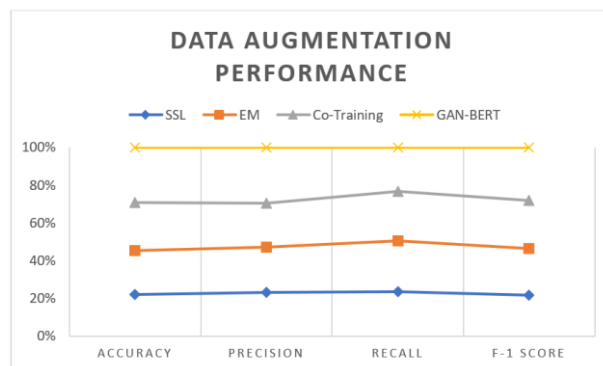


Fig. 7: Data performance graph

Telecom companies are therefore able to manage the available resources such as bandwidth, spectrum, infrastructure investments with special reference to the usage and demographics of the network. This allows the resources which are required to be used where they are most wanted, and this makes the operations smooth and cost effective than the normal operations. Analyzing metadata using natural language processing can help telecom providers get better understanding of customers' wants, trends of usage, and demands. This is a way in which they can be able to offer service that directly meets the needs of the customer and market their products and services in a more efficient manner hence enhancing the level of satisfaction that customers have towards the company (Pulicharla, 2024). Meta dynamics can assist telecom players in sustaining the regulation and industry standards by following up on metadata utilizing machine learning algorithms. Also, AI based security analytics can differentiate between antagonizing forces like fraudsters, hackers, or Data breach in real-time.

The use of AI and machine learning algorithms is good for scalability and can easily incorporated when dealing with large volumes of metadata in terms of volume, variety, and velocity within the telecom industries. Since data ecosystems are dynamic, thus metadata stewardship solutions powered by artificial intelligence are also elastic in the sense that they can adapt to the emerging changes.

Table 1: Data augmentation performance for different algorithms

Model	Accuracy	Precision	Recall	F-1 Score
SSL	0.6258	0.6243	0.6258	0.6013
EM	0.6567	0.6452	0.7128	0.6917
Co-Training	0.7218	0.6258	0.6952	0.7115
GAN-BERT	0.8259	0.7924	0.6153	0.7901

With the advent of AI and machine learning, metadata stewardship and quality management in the telecom sector for analysis can bring a change for companies by taking it to the next level by improving the business and providing enhanced services to the consumers.

Limitations

Even though the proposed framework had some good outcomes in the area of metadata stewardship and metadata management in the telecom environment, there were some areas of limitations. The experiment was mainly performed in simulated telecom metadata environments and in controlled injection mechanisms of anomalies because of limited availability of publicly available telecom metadata datasets. Therefore, framework validation of large scale real world telecoms operational infrastructures has not been accomplished as of yet.

Furthermore, transformer-based models, like GAN-BERT, demand significant amount of computation power and training time and could have an impact on deployability in telecom edge sites with computation power constraints. The framework can also have different performance when it has metadata sources with different diversity, complex schemas and labeled samples of metadata.

Additionally, in-depth analysis of the groundwork for real-time distributed deployment, federated governance architectures, while incorporating most aspects of cyber security and support for a streaming metadata optimization, is left beyond the scope of the current study.

Conclusion

The study introduced an AI based-Metadata Stewardship and Quality Governance framework for flexible and heterogeneous telecom ecosystems where telecom networks, IoT infrastructures, remote sensing platforms and cloud-native services are a source of different types of metadata. The suggested framework combined the assessments of metadata quality, semantic balancing, anomaly detection, interoperability management, and automation of governance by applying machine learning techniques such as SSL, EM, Co-Training models, as well as the GAN-BERT model.

The experimental evaluation showed that the GAN-BERT-based learning framework is particularly effective in terms of metadata quality prediction and anomaly detection compared to the conventional methods. In the proposed scenario, the proposed model has achieved accuracy of 82.59%, precision of 79.24%, recall of 61.53% and F1-score of 79.01% where it have shown its better capability for identifying various inconsistencies, semantic conflicts, incomplete record, and governance violations in telecom scenarios. Besides, it enhanced the transparency of metadata like its conformance to schema, interoperability, lineage traceability, and scalability of governance in high throughput telecom workloads.

AI-powered semantic validation and AI-powered automation of governance showed promise of minimizing the manual amount of metadata management consumed and streamlining telecom analytics system decision making. The proposed framework also advanced the intelligent monitoring and regulatory compliance management, adaptive metadata governance, and intelligent visualization of next generation telecom infrastructures.

But the present study is limited in some aspects. Because of limited availability of large datasets of public telecom metadata, simulated telecom metadata environments were primarily used for the experimental evaluation. Further, transformer-based architectures like GAN-BERT make the model more complex and slow to train especially for large applications. Thus, it is still worth exploring computational efficiency, and validation of the solution on real telecom infrastructures, are still worth pursuing in future research.

Ongoing research work could explore real-time/deployment validation of the telecom, federated metadata governance architecture, light transformer optimization, telecom streaming metadata analytics, large language model-based telecom metadata intelligence, and adaptive AI-driven multi-dynamic metrics governing the telecoms. The proposed framework offers an excellent backbone to create intelligent metadata stewardship and scalable telecom data governance in the future communication landscape.

Future Scope

Land scale, operational metadata environments and streaming telecom infrastructures can be explored for telecom deployment in a real world use case in future research. The ability to integrate with federated learning frameworks, low component training architectures, edge AI systems and large language models-driven metadata intelligence can further enhance scalability and adaptable governance capabilities. Future research directions could include real-time analytics of metadata, cybersecurity-enabled metadata governance, distributed lineage tracking, as well as AI-driven regulatory compliance mechanisms for the next-generation telecom ecosystem.

Acknowledgment

I am grateful to my organization for providing a positive environment that allowed me to explore this topic and continue learning. I also appreciate my peers who took the time to review my ideas and offer honest feedback, which helped improve the clarity and direction of this article.

Funding Information

Thank you to the publisher for their support in the publication of this research article. We are grateful for the

resources and platform provided by the publisher, which have enabled us to share our findings with a wider audience. We appreciate the efforts of the editorial team in reviewing and editing our work, and we are thankful for the opportunity to contribute to the field of research through this publication.

Author's Contributions

Bhola Gan Chaudhuri: Prepared figures and provided Data curation, validation, investigation and edited the manuscript. Both authors reviewed the manuscript. wrote the main manuscript text and provided Conceptualization, Methodology, Software.

Shalli Rani: Has done the project administration and Supervision. Wrote the main manuscript text and provided Conceptualization, Methodology, Software.

Ethics

The authors have not received any financial support or funding to report.

References

- Alboqmi, R., Jahan, S., & Gamble, R. F. (2024). A Risk Adaptive Access Control Model for the Service Mesh Architecture. *2024 IEEE 3rd International Conference on Computing and Machine Intelligence (ICMI)*, 1–6.
<https://doi.org/10.1109/icmi60790.2024.10585800>
- Arif, M. S., Schade, P., Lehfeldt, R., Notay, V., Carstens, G., & Kösters, F. (2023). From simulation to dissemination: automation of data and metadata management. *IOP Conference Series: Earth and Environmental Science*, 1136(1), 012006.
<https://doi.org/10.1088/1755-1315/1136/1/012006>
- Belkhale, S., Cui, Y., & Sadigh, D. (2023). Data Quality in Imitation Learning. *Advances in Neural Information Processing Systems* 36, 80375–80395.
<https://doi.org/10.52202/075280-3524>
- Breazzano, C., Coraci, D., & Busa, R. (2021). MT-GAN-BERT: Multi-task and generative adversarial learning for sustainable language processing. *CEUR Workshop Proceedings*. 5th Workshop on Natural Language for Artificial Intelligence (NL4AI 2021) co-located with the 20th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2021), Aachen, Germany.
- Chen, H., Chen, J., & Ding, J. (2021). Data Evaluation and Enhancement for Quality Improvement of Machine Learning. *IEEE Transactions on Reliability*, 70(2), 831–847.
<https://doi.org/10.1109/tr.2021.3070863>
- Chen, H., Pieptea, L. F., & Ding, J. (2022). Construction and Evaluation of a High-Quality Corpus for Legal Intelligence Using Semiautomated Approaches. *IEEE Transactions on Reliability*, 71(2), 657–673.
<https://doi.org/10.1109/tr.2022.3156126>
- Danielsson, B., Santini, M., Lundberg, P., Al-Abasse, Y., Jönsson, A., Eneling, E., & Stridsman, M. (2022). Classifying Implant-Bearing Patients via their Medical Histories: a Pre-Study on Swedish EMRs with Semi-Supervised GAN-BERT. *Proceedings of the Language Resources and Evaluation Conference*, 5428–5435. <https://doi.org/10.63317/32kbqp54ox5j>
- Emmadi, S. C., Agrawal, P., Samudrala, S., Shimpi, V., & Natu, M. (2023). Theory meets practice approach for Event Correlations. *2023 IEEE International Conference on Big Data (BigData)*, 1918–1921.
<https://doi.org/10.1109/bigdata59044.2023.10386269>
- Khoh, W. H., Pang, Y. H., Ooi, S. Y., Wang, L.-Y.-K., & Poh, Q. W. (2023). Predictive Churn Modeling for Sustainable Business in the Telecommunication Industry: Optimized Weighted Ensemble Machine Learning. *Sustainability*, 15(11), 8631.
<https://doi.org/10.3390/su15118631>
- Krupitzer, C., Temizer, T., Prantl, T., & Raibulet, C. (2020). An Overview of Design Patterns for Self-Adaptive Systems in the Context of the Internet of Things. *IEEE Access*, 8, 187384–187399.
<https://doi.org/10.1109/access.2020.3031189>
- Lubis, M., Nugroho, F. A., Lumingkewas, L. W., & Lubis, A. R. (2022). Mapping of TOGAF ADM and TMForum Framework in the Telecommunication Industry. *Proceedings of Sixth International Congress on Information and Communication Technology*, 216, 493–503.
https://doi.org/10.1007/978-981-16-1781-2_44
- Lyashkevych, V. (2026). Multi-Drift Predictive Monitoring for Evolving Information Systems. *Computer Systems and Information Technologies*, 2, 171–184. <https://doi.org/10.31891/csit-2026-2-15>
- Meritxell, G.-O., Sierra, B., & Ferreira, S. (2022). On the Evaluation, Management and Improvement of Data Quality in Streaming Time Series. *IEEE Access*, 10, 81458–81475.
<https://doi.org/10.1109/access.2022.3195338>
- Mochalov, V., Bratchenko, N., Linets, G., & Yakovlev, S. (2019). Distributed Management Systems for Infocommunication Networks: A Model Based on TM Forum Framework. *Computers*, 8(2), 45.
<https://doi.org/10.3390/computers8020045>
- Myers, D., Mohawesh, R., Chellaboina, V. I., Sathvik, A. L., Venkatesh, P., Ho, Y.-H., Henshaw, H., Alhawawreh, M., Berdik, D., & Jararweh, Y. (2024). Foundation and large language models: fundamentals, challenges, opportunities, and social impacts. *Cluster Computing*, 27(1), 1–26.
<https://doi.org/10.1007/s10586-023-04203-7>
- Myszewski, J. J., Klossowski, E., Meyer, P., Bevil, K., Klesius, L., & Schroeder, K. M. (2022). Validating GAN-BioBERT: A Methodology for Assessing Reporting Trends in Clinical Trials. *Frontiers in Digital Health*, 4, 878369.
<https://doi.org/10.3389/fdgth.2022.878369>

- Noman, A. H. M., Mustaquim, S. M., Molla, S., & Siddique, I. M. (2024). Enhancing Operations Quality Improvement through Advanced Data Analytics. *Journal of Computer Science Engineering and Software Testing*, 10(1), 1–14.
<https://doi.org/10.46610/jcses.2024.v10i01.001>
- O'Hara, K. (2019). *Data trusts: Ethics, architecture and governance for trustworthy data stewardship*.
<https://doi.org/10.5258/SOTON/WSI-WP001>
- Ouyang, Y., Wang, L., Yang, A., Gao, T., Wei, L., & Zhang, Y. (2022). Next Decade of Telecommunications Artificial Intelligence. *CAAI Artificial Intelligence Research*, 1(1), 28–53.
<https://doi.org/10.26599/air.2022.9150003>
- Paulraj, D. (2026). AI-Driven Autonomous Telecom Operations for High-Availability Carrier Networks. *International Journal of Future Innovative Science and Technology*, 9(1), 49.
- Pradhan, S. K., Heyn, H.-M., & Knauss, E. (2024). Identifying and managing data quality requirements: a design science study in the field of automated driving. *Software Quality Journal*, 32(2), 313–360.
<https://doi.org/10.1007/s11219-023-09622-8>
- Pulicharla, M. R. (2024). Data Versioning and Its Impact on Machine Learning Models. *Journal of Science & Technology*, 5(1), 22–37.
<https://doi.org/10.55662/jst.2024.5101>
- Purbasari, A., Rinawan, F. R., Zulianto, A., Susanti, A. I., & Komara, H. (2021). CRISP-DM for Data Quality Improvement to Support Machine Learning of Stunting Prediction in Infants and Toddlers. *2021 8th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*, 1–6.
<https://doi.org/10.1109/icaicta53211.2021.9640294>
- Raibulet, C., Fontana, F. A., & Carettoni, S. (2020). A preliminary analysis of self-adaptive systems according to different issues. *Software Quality Journal*, 28(3), 1213–1243.
<https://doi.org/10.1007/s11219-020-09502-5>
- Ren, Y., Hu, W., Wang, Z., Zhang, X., Wang, Y., & Wang, X. (2021). A hybrid deep generative neural model for financial report generation. *Knowledge-Based Systems*, 227, 107093.
<https://doi.org/10.1016/j.knosys.2021.107093>
- Sharma, Dr. D., & Pandey, Dr. B. (2026). Data Management Techniques in IoT Using API. *Eco-Networking for IoT*, 153–165.
https://doi.org/10.1007/979-8-8688-2317-6_6
- Srivastava, K., & Shekokar, N. (2020). Design of Machine Learning and Rule Based Access Control System with Respect to Adaptability and Genuineness of the Requester. *EAI Endorsed Transactions on Pervasive Health and Technology*, 6(24), e1.
<https://doi.org/10.4108/eai.24-9-2020.166359>
- Sudigyo, D., Hidayat, A. A., Nirwantono, R., Rahutomo, R., Trinugroho, J. P., & Pardamean, B. (2023). Literature study of stunting supplementation in Indonesian utilizing text mining approach. *Procedia Computer Science*, 216, 722–729.
<https://doi.org/10.1016/j.procs.2022.12.189>
- Tanvir, R., Shawon, M. T. R., Mehedi, M. H. K., Mahtab, M. M., & Rasel, A. A. (2023). A GAN-BERT Based Approach for Bengali Text Classification with a Few Labeled Examples. *Distributed Computing and Artificial Intelligence, 19th International Conference*, 20–30.
https://doi.org/10.1007/978-3-031-20859-1_3
- Tebu, G., & Izang, A. (2025). Customer Churn Prediction in the Telecommunication Industry Over the Last Decade: A Systematic Review. *Asian Journal of Research in Computer Science*, 18(4), 256–271.
<https://doi.org/10.9734/ajrcos/2025/v18i4618>
- Verma, R., & Bhatt, R. (2026). Hybrid DCN-transformer framework with role-based access control (RBAC) policy for threats classification in cloud. *International Journal of Information Technology*, 18(1), 507–514.
<https://doi.org/10.1007/s41870-025-02743-2>
- Yandrapalli, V. (2024). AI-Powered Data Governance: A Cutting-Edge Method for Ensuring Data Quality for Machine Learning Applications. *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)*, 1–6.
<https://doi.org/10.1109/ic-etite58242.2024.10493601>
- Yebenes, J., & Zorrilla, M. (2019). Towards a Data Governance Framework for Third Generation Platforms. *Procedia Computer Science*, 151, 614–621.
<https://doi.org/10.1016/j.procs.2019.04.082>
- Zhao, J., & Krishnan, S. (2025). *Metadata Management for AI-Augmented Data Workflows*.
<https://doi.org/10.48550/arXiv.2508.06814>