ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Research on Large Model Empowered Smart City Multi-dimensional Data Governance Chain Reconstruction Technology

Yang Zhang[1,*]

[1] School of Information Engineering, Jilin Vocational College of Industry and Technology, Jilin, 132013, Jilin,China

SUMMARY: *Development of smart cities is now in a stage of deep digitalisation and all-encompassing intelligence. The multiple types of data produced during urban operation have large scales, are multi-modal, are changing rapidly, and come from various sources. The traditional data governance framework has been facing several problems, such as dispersed data collection, isolated storage systems, inefficient processing capabilities and shallow application implementation; it is thus unable to meet the high standards of timely, smart and intelligent governance in a smart city. Large-scale models have shown good performance in multimodal understanding, autonomous learning, deep reasoning and generative optimisation, and can provide strong technical support for addressing data governance bottlenecks and reconstructing end-to-end governance processes in smart cities. Assess the present state of multi-dimensional data governance in smart cities, identify the principal shortcomings of the traditional governance model, and present numerous merits that large-scale models offer for supporting data governance. Six representative links in this study are selected: data collection and aggregation; cleaning and preprocessing; integration and sharing; analysis and mining; security management; and application implementation. It seeks to build a closed-loop, intelligent and integrated data governance system, address the implementation problems and optimisation paths of this system, and thus provide theoretical support and technical reference for high-quality smart city construction, efficient data element circulation, and modernisation of urban governance.*

KEYWORDS: *Large models; Smart cities; Multidimensional data; Data governance; Link reconstruction; Intelligent governance*

1 Introduction

1.1 Research Background

With the continuous progress of the construction of digital China, smart cities have emerged as the main direction for urban modernisation, and thus are expected to address urban governance issues, enhance the quality of life for the public, and optimise the use of resources [1]. At present, the construction of smart cities includes many areas such as government services, transportation, ecological and environmental protection, public security, livelihood security and urban management. By utilising Internet of Things (IoT), cloud computing, big data and mobile internet, a large number of sensing devices, business systems and monitoring terminals have been deployed [2]. The data generated during the entire process of urban operation has

*zhangyang1986711@126.com
https://doi.org/10.65102/is2026951
Published: 30 April, 2026

many kinds and complex features, such as structured government data, semi-structured log data, and unstructured multimodal data (videos, sounds, texts and images); it comes from various sources, including government departments, enterprises, the public, IoT devices and internet platforms; its dimensions cover space and time, business activities, attributes and scenarios, etc., thus forming a huge, complicated and unevenly distributed pool of urban data resources [3].

Data are the basic production factors of smart cities, and data governance should be employed to unlock the value of data resources under conditions of security and compliance. The traditional model of smart city data governance is a single-link flow of collection-storage-processing-analysis-application and has numerous long-term problems. First of all, there is no unified standard for data collection; therefore, different departments and systems are collecting data independently, resulting in repeated collection, incomplete data acquisition and inaccurate extraction, and thus inconsistent source data quality [4]. Secondly, the distributed data storage will cause a 'data silo' problem and reduce the sharing and cooperation of knowledge among the teams. Thirdly, the method of data processing is manual; it is based on rule-based scripts, and therefore cannot handle a large amount of unstructured data; the cleaning, labelling and integration of this data are both slow and labour-intensive. Fourthly, the data analysis is still at a basic level of statistics; it lacks high-end functions for deep data exploration and intelligent inference, and thus fails to reveal underlying patterns, risks and trends that limit the full value extraction of the data [5, 6]. Finally, passive data security measures and the absence of real-time monitoring at all stages of data life may be the causes of data breaches, unauthorised use and alteration, thus posing a serious compliance risk.

In recent years, large model technologies for artificial intelligence, such as large language models and multimodal large models, have developed at a high speed and achieved a breakthrough from perceptual intelligence to cognitive intelligence. These technologies have the following basic features: large-scale data flow, multi-modal data recognition, independent feature extraction, deep logical reasoning, intelligent generation optimisation, and adaptive learning iteration. They can address the technical deficiencies of traditional data governance, rebuild the entire data governance workflow, and achieve automation, intelligence, integration and closed-loop management in data governance [7-10]. Leverage large model technology to achieve source-level quality enhancement, all-encompassing integration, efficient processing, in-depth mining, secure control and value multiplication for multidimensional data governance in smart cities. Therefore, urban governance will move away from a passive-response mode to an active-proactive mode; at the same time, it will adopt data-driven strategies based on experience and transition to refined governance. Therefore, research can be conducted on the technology for reconstructing the large-model-enabled multidimensional data governance chain.

1.2 Research Significance

Systematically identify pain points and problems in multidimensional data governance for smart cities in this paper. Based on the technical features of large-scale models, we have built a framework for reconstructing data governance processes powered by AI. The above innovations can offer new support for the theoretical foundation of big data governance and artificial intelligence, thereby promoting the development of intelligent urban governance. We put forward the following targeted technical optimisation solutions for full-lifecycle management of multidimensional data, addressing research deficiencies in the application of large-scale models to complex municipal data governance scenarios. The above results can provide a reference for the future research of this subject [11].

At the practical level, the restructuring technologies and implementation solutions put forward in this paper can help to address some problems in smart cities, such as data silos, inefficient governance and a failure to utilise the value of data. The above improvements will increase data governance efficiency and improve the quality of data. Reduce the barrier to department data sharing to obtain more abundant data in collaboration among departments, regions and at all levels of collaborative governance. The solutions will support urban refinement management, emergency response systems, precision public service delivery and intelligent resource allocation to improve the operating efficiency and quality of life for the public in the city. They ensure the safe and compliant transmission of data and promote the development of a high-quality smart city [12, 13].

1.3 Research Content and Approach

First, this paper explains the basic ideas of multidimensional data governance and large model technology in smart cities, and then points out problems with the current system of data governance. Then, it will introduce the key strengths and compatibility of large models for strengthening smart city data governance. Then, based on this, research explores the reconstruction of data governance workflows driven by large models in six key stages: data collection and aggregation; cleaning and preprocessing; integration and sharing; analytical mining; security management; and application implementation, finally constructing a complete new governance system [14-16]. Problems in the technology implementation are shown, and then targeted optimisation measures and development proposals are proposed. The five divisions of the research method are "analysis of the current situation - evaluation of strengths and weaknesses - technological restructuring - challenge assessment - solution optimisation", and this paper is based on five steps to obtain reliable and feasible results. Research framework of large model empowered smart city data governance is shown in Figure 1.

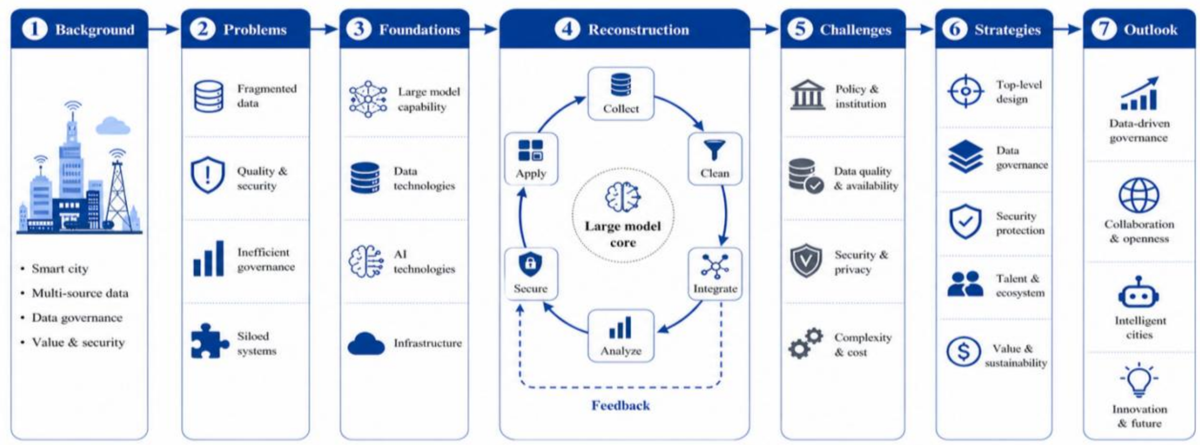

Figure 1: Research Framework of Large Model Empowered Smart City Data Governance

2 Related Concepts and Technical Foundations

2.1 Core Connotation of Multi-dimensional Data Governance in Smart Cities

Smart City Multi-Dimensional Data Governance refers to an all-encompassing system of activities and technical measures for managing various kinds of multi-source, multi-modal and multi-level data that arise in all operating circumstances of smart cities. Lifecycle management, quality control, resource integration, security assurance and value extraction are all included. The main goals are to have good data management, control and use, sharing, to break down data silos, improve data quality, release the value of data, and support scientific, rational and intelligent governance of smart cities [17].

The five features of multi-dimensional data in smart cities are listed below. Firstly, it has a multi-source nature, drawing from many parts of the country, including government departments, transportation systems, security organs, environmental protection and water-supply organizations, power enterprises and telecommunications companies, social media, public opinion outlets, IoT sensors, etc.; these are dispersed in geography and origin. Second, multi-modality: The dataset includes various forms such as tables and ledgers (structured), web pages and logs (semi-structured), as well as unstructured data, including surveillance videos, audio recordings, images, technical tickets, and remote sensing imagery; the proportion of unstructured data is over 80% of the total. Third, a fast-growing system: The urban system is a 24/7 operation that generates massive amounts of real-time data daily, reaching several terabytes (TB) or even petabytes (PB) per day; therefore, extremely high-speed processing is needed. Fourthly, high-value low-density characteristics: Although a large number of data sets contain necessary information about urban operation patterns, public needs and risk factors, an excess of invalid and redundant data hinders the value-extraction process. Fifth, strong interconnection: All kinds of data have spatiotemporal correlations, operational linkages and logical relationships, and single-dimensional analysis cannot provide a full picture of the urban status; therefore, multi-dimensional integrated analysis is required [18].

The original data governance process is not connected to all parts and cannot be flexibly linked or intelligently managed. These systems are not highly effective in the governance of complex multi-dimensional data, have high operating costs, and thus fail to meet the deep governance requirements of smart cities.

2.2 Core Features of Large Model Technology

A large model generally refers to a deep-learning-based artificial intelligence system with a large amount of data for training and many parameters. The three types of these are language large models, vision large models and multi-modal large models. Among them, multimodal large models can handle all kinds of data, such as text, images, audio, video and numbers, at the same time, and are therefore very suitable for the requirements of multi-dimensional data governance in smart cities. These models have key attributes and are therefore necessary technical supports for restructuring the workflow of data governance [19].

First, high-volume data processing and generalisation ability. Leverage powerful computation and deep-learning-based architectures to efficiently process multi-terabyte (TB) and multi-petabyte (PB) datasets from various sources and in different formats, and avoid setting up situation-specific rules. Good Generalisation Adaptability can address all Data Governance Requirements in a Smart City Setting.

Second, Multimodal Data Understanding and Fusion Capabilities. Multimodal large models can perform unified encoding, understanding and analysis of text, images, videos, audio, structured data, etc., breaking down barriers between different modalities to achieve cross-modal data association, alignment and fusion, and thus address the problem of multimodal data integration that traditional technologies cannot handle.

Thirdly, independent learning and feature extraction ability. Large models can learn the basic features and hidden structures of a substantial amount of data independently through unsupervised and semi-supervised learning to reduce the demand for manual annotation and lower data labelling expenses. At the same time, they also find abnormal, redundant and incorrect data in the data.

Fourth, deep reasoning and decision-optimisation abilities. Large models are good at logical reasoning, causal analysis and trend prediction, so they can be used to explore a large amount of data and predict possible risks or changes in the urban area. Provide scientific support for decision-making in urban governance, promote the transformation from data to information, from information to knowledge, and from knowledge to decisions.

Fifth, Adaptive Iteration and Optimisation Capability. Large models can continuously adjust and optimise based on new data, application feedback and changes in the operating environment to enhance the accuracy and effectiveness of data governance, thereby forming a self-improving closed-loop system.

Sixth, Natural Language Interaction Ability. Large models support natural language dialogue and thus have a lower operational threshold for data governance. Non-technical staff can use conversational commands to query, analyse and process data for convenience in the administration of this office.

2.3 Compatibility Between Large Models and Smart City Data Governance

The technical attributes of large models are highly suitable for the various pain points and demands in multidimensional data governance for smart cities, thus achieving good compatibility. On the other hand, smart cities are dealing with a large number of different types of data, and these are not suitable for processing by traditional rule-based or manual methods. Large models have the capability to handle large volumes of data and perform multimodal understanding efficiently for the demands of data governance. Smart city governance also needs to have the ability to monitor and analyse data in real time. Large models are very good at fast inference and deep mining, so they are suitable for real-time monitoring, risk alerts and other applications. The Data Governance of smart cities should also break down data silos and promote cross-domain cooperation. Large models have improved the system for unifying and correlating cross-domain data. In addition, the automated governance functions will reduce manual work and lower operating costs; thus, they will be both cost-saving and efficient for the smart city.

3 Core Bottlenecks in Traditional Multidimensional Data Governance Chains for Smart Cities

3.1 Data Collection and Aggregation Phase: Lack of Standards and Inaccurate Source Data

The traditional smart city data collection systems do not have unified standards and specifications. Different departments and systems have established independent collection frameworks according to their own operating requirements; as a result, there are multiple standards, structures, scopes and frequencies for data collection, and severe problems in data fragmentation have arisen. Some areas have duplicated data collection; therefore, they are both inefficient in terms of computing resources and manpower, and they generate a large amount of redundant data. Other areas are deficient in data collection, inaccurate or delayed, and thus lack complete and accurate data on the operation of the city [20]. Many IoT sensing devices have started to have non-uniform data transmission protocols; as a result, the collection of data is scattered and irregular, reducing the quality of this data source. In addition, the lack of intelligent filtering during data aggregation leads to an accumulation of a large amount of invalid and junk data in the system; as a result, this further increases processing pressure and creates a vicious circle of "junk in, junk out". Traditional governance chain and key bottlenecks are shown in Figure 2.

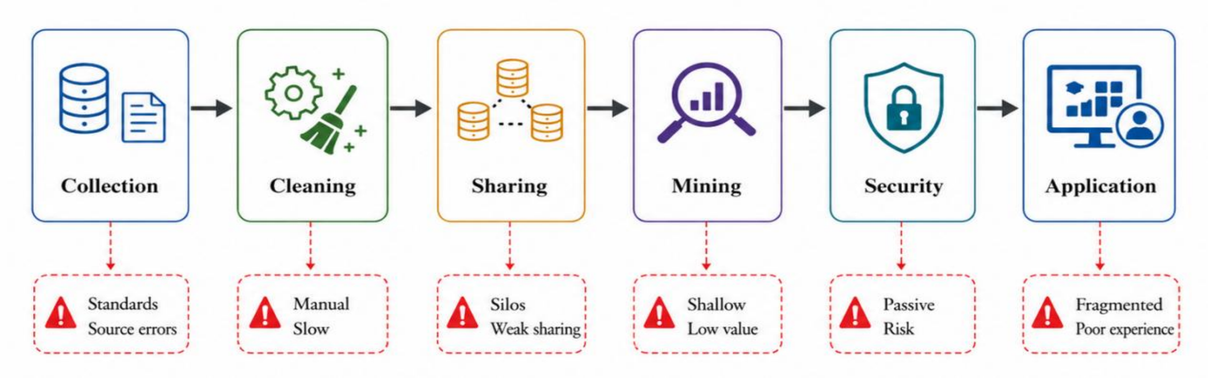

Figure 2: Traditional Governance Chain and Key Bottlenecks

3.2 Data cleaning and preprocessing stage: Manual dependency, low efficiency

Data cleaning and preprocessing need to be done to improve the quality of the data. The former is a manual operation with a rigid rule-based script, thus having a very low processing efficiency and high error rates. Unstructured data needs to be manually cleaned, annotated and categorised; thus, it is relatively slow and labour-intensive. At present, dealing with a large volume of video, audio and text data manually will not meet the real-time requirements. Fixed rule scripts only support structured data processing, cannot address the complex cleaning requirements of multimodal data, and are thus unable to handle fuzzy, abnormal or missing data effectively. There are also no automated mechanisms for deduplication, error correction and data completion; thus, the results are subject to errors and lack the required accuracy for building a high-quality, standardised data repository. Some datasets have been further distorted by repeated manual processing and are no longer suitable for subsequent analysis and applications.

3.3 Data Fusion and Sharing Phase: Severe Barriers and Insufficient Collaboration

A major problem of data silos in smart city data governance is that the traditional data-integration mode has severe departmental barriers, system deficiencies and regional division. Government agencies, public-private partnerships and local governments all have separate data storage systems and do not share data collaboratively. Strict Data Access Control and an inefficient data circulation model are also problem causes. There are no unified platforms for government data, social data and IoT data; thus, cross-departmental and cross-sectoral coordination at a cross-scenario level is difficult to achieve. Furthermore, because there is no unified fusion framework for different kinds of data, structured and unstructured data cannot be combined, multi-dimensional data synergy is limited, and the value of the data is greatly reduced. Compounding the above problems, insecure data-sharing protocols and a lack of information sharing among departments due to security risks have led to data silos [21].

3.4 Data Analysis and Mining Phase: Superficial Analysis with Limited Value

Traditional methods of data analysis and mining are based on surface-level statistics and visualizations, and the algorithms for simple summation, counting and comparison are fixed. These ways do not have strong functions such as deep reasoning, association mining and trend prediction. Many big data are unorganized and thus cannot provide specific support or other problems. Urban operating risk, fluctuating public demand and limited resources will all need to be addressed through post-event analysis, rather than proactive prediction. The analysis process for this data requires specialised technical staff, so it is cumbersome to run and cannot meet the demand for timely and dynamic urban management. Fragmented and incomplete analytical results also hinder the development of a full-featured urban operating consciousness and thus cannot support rational decision-making. Therefore, the entire value of the data has not been realised, and many data resources remain unused.

3.5 Data Security Control Phase: Passive Protection with Prominent Vulnerabilities

The old model of data security management has mainly used passive defence, such as firewalls, permissions and encrypted storage; it does not cover all parts of the data life cycle dynamically and intelligently. There is no end-to-end traceability and regulatory system for the data-collection and application processes, so risks such as data breaches, misuse, tampering and illegal transactions cannot be identified in a timely manner. Identification and protection of sensitive data and personal information are still being carried out manually; therefore, the efficiency is relatively low and there is a high risk of omission, which leads to serious compliance problems. In addition, the security system is not adaptable and cannot respond quickly to new types of cyberattacks or data-theft methods; thus, many security risks have arisen recently. In addition, data security management is still disconnected from data governance and application processes; therefore, a dual-track model has emerged that either prioritises security over the application, or applications over security, and thus fails to achieve a balanced arrangement of the two [22, 23].

3.6 Data Application Implementation Phase: Fragmented Scenarios and Poor User Experience

The traditional applications of data have low diversity and are fragmented; thus, these systems operate in isolation and lack good data interoperability to promote integrated governance. Public services, traffic management, emergency response and other sections have not been linked through a common data platform, so their operations are not integrated, the response speed is slow, and data accuracy is low. For example, citizens repeatedly submit the same documents for administrative services; traffic congestion is not dealt with in advance; the emergency response system lacks an all-weather data infrastructure, etc., and thus handling efficiency declines. Furthermore, the data applications are not personalised and intelligent enough to meet the different needs of different users in a personalised manner. As a result, there will be poor experiences for the people and regulatory organs, and the goals of data-driven governance cannot be fully achieved.

4 Large Model Empowered Technical Framework for Multi-dimensional Data Governance Chain Restructuring in Smart Cities

Large-scale models are enabling smart cities to rebuild multi-dimensional data governance workflows and are aiming to break the traditional linear and segmented model of governance.

The first is a closed-loop intelligent governance system for "source data collection - intelligent preprocessing - cross-domain integration - deep mining - security control - scenario-based application - feedback-driven iteration" [24, 25]. Multimodal large models are used as the core to remove barriers at all levels of governance and achieve fully automated, intelligent and collaborative operations throughout the process. In this way, data governance will improve the efficiency, quality and security standards of the data, and the full benefits of data assets can be realised. Technical Architecture for Smart City Data Governance is shown in Figure 3.

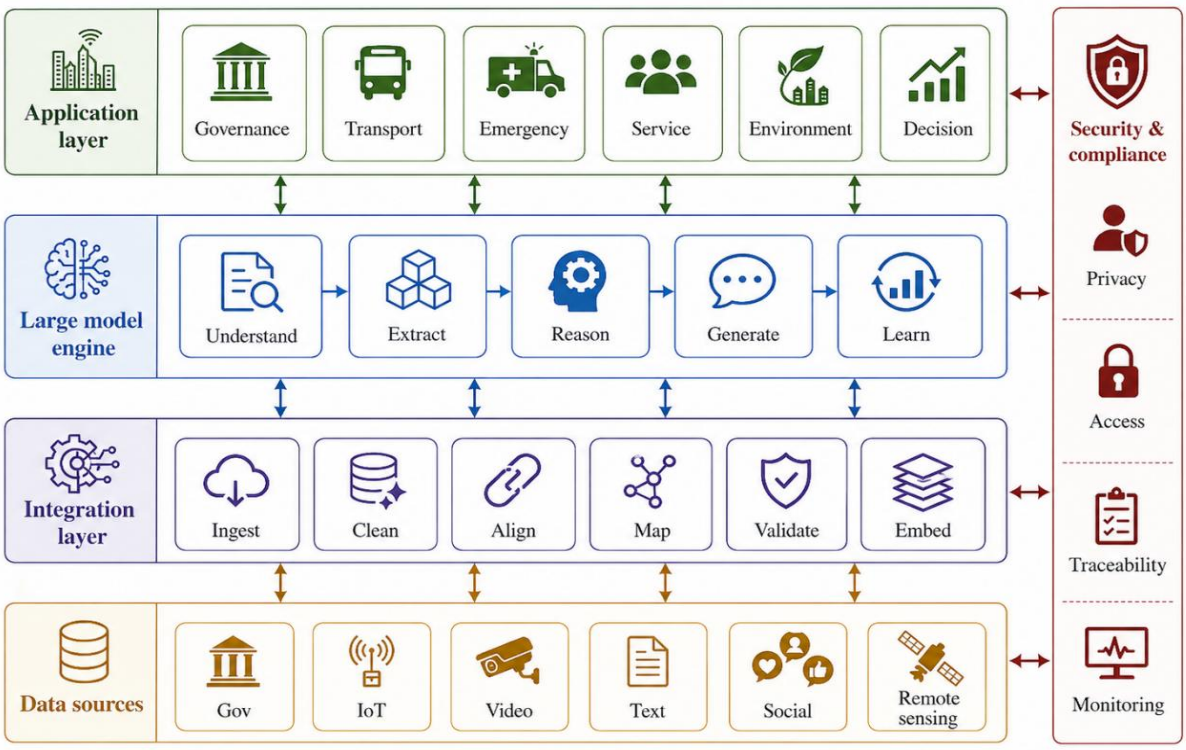

Figure 3: Research framework of large model-enhanced smart city data governance

4.1 Restructuring of Data Collection and Aggregation Phase: Standardized Collection and Intelligent Integration

To solve the problems of absent standards for traditional data collection and unreliable sources, a unified intelligent data collection and aggregation system based on large models has been constructed to achieve standardised, accurate and efficient data collection at the source.

We will build an all-encompassing data collection and scheduling platform powered by large models. Based on all of the above requirements for urban data, we will build standardized data collection protocols that cover the scope of collection, formats, dimensions, frequency and transmission methods, and achieve integrated standards for the collection of government data, IoT data, social data and internet data. Fully utilise the multimodal understanding capabilities of large models to achieve intelligent adaptation of various collection devices and business systems according to the requirements of different protocols and formats in the data collection terminals, thereby establishing a single window for domain-wide data access.

A large-scale model can intelligently analyse the data-collection requirements and dynamically schedule. Since urban governance has many types of cases and the amount of data and the structure of past data collection are constantly changing, the frequency and scope of system data collection need to be flexibly adjusted to avoid duplication or omissions/errors in the data. It will increase the collection rate of traffic condition data in the peak hour and improve the environmental monitoring of bad weather. The system will also automatically exclude the unnecessary and invalid collection requests to save resources.

We have established an intelligent data collection quality assurance system based on large-scale models for real-time initial verification of collected data and have identified missing, distorted or abnormal data points. These abnormal cases are immediately sent back to the collection terminals for follow-up collection or re-collection to ensure the quality of the data at the source. At the same time, the models will be used to classify and label the collected data based on the type of data, source, application scenario and etc. Together, the two will conduct data collection and some processing at the same time to reduce the work in the next step.

4.2 Reconstruction of data cleaning and preprocessing stage: Fully automated processing with high-quality quality enhancement

No longer relying on the previous method of manual cleaning and standardised processing, large models will be used to achieve full automation and intelligence in data cleaning preprocessing to enhance processing efficiency and improve data quality.

A multimodal large model is used to unify the data, and intelligent cleaning processes are implemented for structured, unstructured and semi-structured data. The large model automatically handles the deduplication, error correction, data completion and format conversion of structured data, and also identifies abnormal values and logical inconsistencies. Intelligently adjust based on history and related data, then use algorithms to fill in missing data precisely.

Large-scale models are used to automatically annotate, classify, reduce noise in and extract key information from unstructured data such as videos, audio recordings, images and text. For example, automatically identify violations and anomalies in surveillance videos to extract key frames and details; transcribe, correct errors in, and perform sentiment analysis on voice-based work orders; classify official documents and public feedback, extract keywords, and generate summaries. In addition, the above models can effectively eliminate spam data, redundant information and noise, and retain only high-value, high-quality data to reduce storage pressure.

We have built an intelligent data quality assessment system that uses large models to evaluate all kinds of indicators of the preprocessed data, such as completeness, accuracy, consistency, timeliness and usability. Data in the system is classified by quality and stored or used at different levels; only high-quality, compliant data will be integrated into the system. It does not require manual operation and has a high-speed processing ability for a large amount of real-time data.

4.3 Reconstruction of Data Fusion and Sharing Process: Comprehensive Integration and Deep Collaboration

Using a large model at its core, we will build a cross-domain, cross-modal, and cross-tier data fusion and sharing system to eliminate data silos completely, enable full-featured multi-dimensional data integration, and achieve seamless circulation in all domains.

We will build an integrated platform for urban data using large-scale models to establish a standardised data encoding system and correlation mapping, and enable centralised access and management of dispersed data in various departments, systems and devices. Using the cross-modal fusion functions of large models, we will achieve alignment, correlation and integration of various data modalities, such as text, images, videos and numbers, to build an all-encompassing data resource that includes both spatial-temporal characteristics and business operation data.

An intelligent permission management and sharing exchange mechanism has been constructed in the system to distribute data access, invocation and sharing rights for large models dynamically according to departmental responsibilities, business needs and security levels; thus, "on-demand authorisation, minimal sufficiency and end-to-end controllability" have been realised. Departmental barriers have been removed, and one-click cross-departmental, cross-domain and cross-region data invocation and real-time sharing have been achieved without manual approval, significantly boosting collaboration efficiency. In addition, it can build a data knowledge graph to show the relationships among all the datasets visually and find new connections for deeper analysis.

Large models can achieve both data de-sensitisation and integration of government-enterprise data, public data, and personal information simultaneously, and automatically process desensitisation for sensitive information and personal privacy. Under the guarantee of security, it will be possible to carry out lawful data integration and sharing to solve the problems of "not daring to share" and "unable to share".

4.4 Reconstruction of Data Analysis and Mining Phase: Deep Inference, Global Perception

Leveraging the strong reasoning and autonomous learning capabilities of large models, deep analysis and prediction can be carried out on the data to move from mere statistical description to in-depth exploration, and from reactive incident response to proactive risk control; thus, the full value of this data is realised.

Large-scale models are conducting all-round situational awareness analysis of integrated multi-dimensional data to monitor the operating environment of urban areas in real time from all aspects, such as transportation, environmental protection, public administration, security systems and social welfare. Using all kinds of correlation and causality, these systems can identify operating modes, pain points, and risks in urban management, such as the reasons for traffic congestion, sources of environmental pollution, trends in public concerns, and possible safety hazards.

A large-scale model is used to analyse the previous years' data, real-time data and external environmental changes to construct a predictive framework and early warning system. The above systems can foresee urban disasters and changes in traffic and environment, predict public demand, etc., and issue early warnings. For example, they can predict the risk of flooding and fire, traffic accidents, as well as peak-hour traffic and congested roads, proactively address shortcomings in public service resources, and shift from a reactive response mode to proactive risk management.

Natural language interaction analysis is available in the system; therefore, various data analyses can be performed by the management department via conversational commands to obtain analytical results, charts and reports quickly without writing code or using special software. At the same time, large-scale models can automatically generate in-depth analysis reports and decision-making recommendations to provide scientific, precise and practical support for urban governance, forming a closed-loop system of data-to-decision.

4.5 Restructuring of Data Security Control Processes: Full-Chain Protection with Dynamic Controllability

Build a large model-driven, end-to-end dynamic and intelligent data security management system to achieve in-depth integration of data governance and security protection, and balance the utilization of data with the safety of data.

The large-scale model has achieved end-to-end security traceability for the data lifecycle and provides complete traceability and record-keeping of all links in data collection, transmission, storage, processing, sharing, application, etc. All data calls, changes and shares are recorded, and in case of a security incident, it will be simple to identify the origin and determine who is responsible.

Intelligent recognition of sensitive data and security risks: A large model automatically scans all data for sensitive information, such as personal privacy, trade secrets, and classified data, and supports multi-level protection. Monitor for abnormal behaviour in real time during the data-flow process, detect risks such as unauthorised access, data loss and malicious tampering, issue alerts and apply automatic interception measures promptly to achieve early detection and response to risks.

The system has been extended to add a dynamic security adaptation mechanism, and in the face of changes in the network environment or new types of attacks, protective measures are automatically reinforced to address new security risks and ensure the security of data transmission. At the same time, it ensures the compliant circulation of data by strictly adhering to data security and personal information protection regulations, automatically reviewing data sharing and application behaviour to prevent unauthorised operations, and thereby achieving coordinated progress of data security and application development.

4.6 Restructuring of Data Application Implementation Phase: Scenario Integration and Service Precision

Leverage restructured high-quality data resources and intelligent analysis capabilities to enable large models to realise all-encompassing implementation of smart city applications in all scenarios, achieving integrated governance and precision-oriented services.

Large-scale models of government services integrate administrative data from various departments to achieve one-stop online processing, a single window system, and automatic entitlement issuance without the need for application submission. The system can retrieve the necessary data automatically, so citizens do not need to submit documents repeatedly; thus, it saves them the trouble of repetition. The system will also have a 24/7 smart Q&A and business guidance centre.

All-weather grid-based fine-grained management in urban management can be used to build a large-scale model for real-time monitoring of urban conditions, municipal facilities and environmental sanitation. The system will automatically detect the violation or facility malfunction, issue a work order immediately, and monitor the resolution status of the issue in the city proactively.

Many kinds of multidimensional data are collected in large-scale models of transportation, such as road conditions, vehicle information, pedestrian movement, weather data, etc., to optimise the timing of traffic lights in real time, predict areas that will be congested, issue travel warnings, plan optimal routes, reduce traffic pressure, and improve the efficiency of passage.

Aggregated data on incident locations, nearby facilities, and weather or geographical features in the area of emergency response can be employed to quickly develop a large-scale model for response planning, coordinate rescue resources, and promote multi-departmental cooperation to improve the speed and effectiveness of emergency response and disposal.

Based on the analysis of public needs, resource distribution and service deficiencies in the area of livelihood security, targeted delivery of public services such as education, healthcare, elderly care and employment is being realised to provide personalised and precise livelihood services for the public, thereby promoting the happiness and satisfaction of the public.

4.7 Closed-loop feedback iteration phase: Continuous optimization and self-improvement

The reconstructed governance framework introduces a closed-loop feedback iteration mechanism to build an all-encompassing closed-loop system. Continuously collect real-time data on the results of large-scale model application, governance effectiveness evaluation and user feedback. Add new data sources and scenario modifications to change the collection standard, processing algorithm, integration model, analysis framework and security policy independently through repeated adjustments. The above methods will improve the accuracy of data governance and operate more efficiently; thus, the governance system will be more advanced and intelligent over time. It will continue to meet the changing needs of the smart city and governance at present and in the future, and have a long-term sustainable data governance system. A closed-loop governance chain driven by large models is shown in Figure 4.

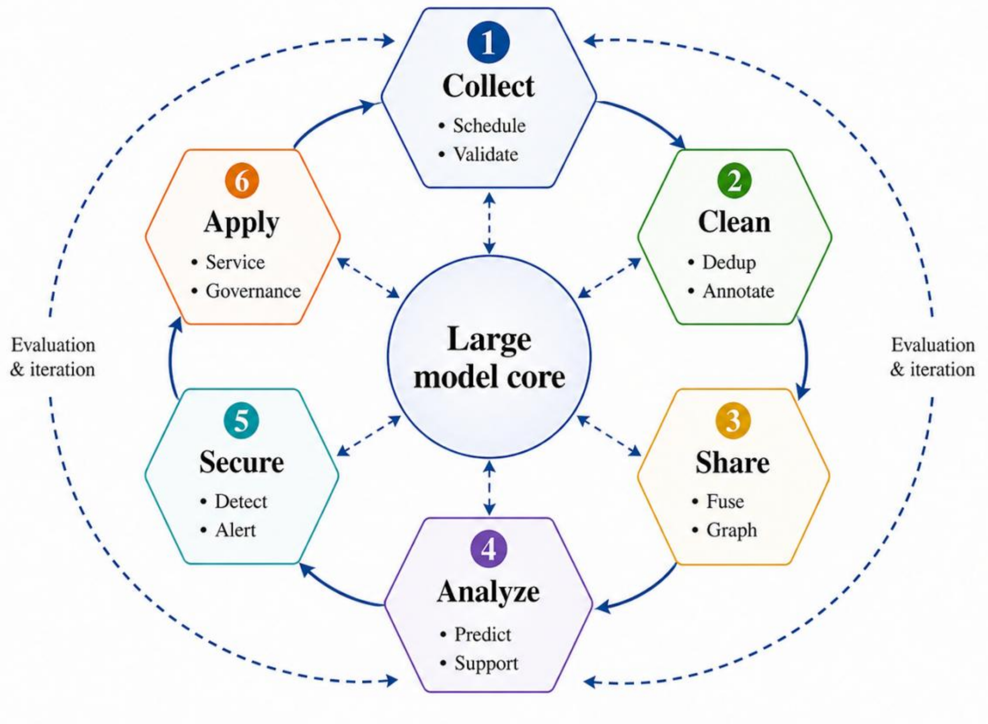

Figure 4: Closed-loop Governance Chain Driven by Large Models

5 Challenges in Reconstructing Multi-dimensional Data Governance Chains for Smart Cities Empowered by Large Models

5.1 Computing resource bottleneck and high deployment costs

Training and inference of large-scale models require vast computational resources, extremely large storage space, high-speed network infrastructure, and thus have considerable hardware investment and operating expenses. Smart cities have large amounts of data that need to be handled, and localized deployment and real-time inference of multimodal large models require strong computing power. Many medium and small cities are under serious financial difficulties and cannot afford the high cost of deployment and operation. In addition, due to the uneven distribution of computing power, there are limited functions at the grassroots level; therefore, real-time data governance at the edge is not possible, and both the efficiency and timeliness of governance are reduced.

5.2 Data quality varies significantly, posing challenges for model training

Although large models can improve the quality of data governance, they require a large volume of high-quality, accurately labelled urban data for initial training. However, current urban data sources have varied quality, contain numerous invalid or noisy data, and lack labelled datasets. As a result, fine-tuning and training performance of large models is suboptimal; therefore, inference errors, distorted outcomes and model hallucinations often occur. Further reduce the generalisation ability and inference accuracy of the model by neglecting incomplete cross-departmental and cross-domain data integration.

5.3 Model Safety and Controllability Risks

Large Models inherently have security risks. Models, cyberattacks and data theft are all prone to causing information leakage and governance failure. The models are not transparent in their operation, and it is difficult to identify the source of an error. There is a risk of model hallucination, and as a result, false information or incorrect conclusions that harm the scientific basis of urban governance may be generated. Another is that the permissions for the large model are too liberal; thus, the data may be misused or accessed by unauthorised parties.

5.4 Shortage of technical talent and inadequate implementation adaptation

To build a large-scale model integration platform for smart city data governance, talents with interdisciplinary knowledge are needed who can use AI technology and urban governance, as well as big data management. At present, there is a serious lack of such specialized talents, and grassroots management departments do not have the technical teams required to perform model deployment, debugging, maintenance and optimisation. In addition, general-purpose large models are not suitable for specific local urban areas, and custom fine-tuning is both difficult and time-consuming; therefore, the government cannot be effectively applied.

5.5 Institutional and systemic barriers impede collaborative advancement

The parties to smart city data governance are the government, enterprises and the public. Some of the organisational problems remain, such as a lack of clear division of labour, missing links in the coordination chain, and strong departmental interests. Some departments are not motivated to share data and resist granting data-access permissions, so the data integration is not complete. There is a lack of unified coordination in the efforts of cross-departmental governance, and thus, the new governance model cannot be fully implemented effectively.

5.6 Incomplete compliance supervision system

The current legal system and regulatory mechanism for large model application and urban data governance are still lacking; there are no specific provisions on data ownership confirmation, data trading, data sharing and liability allocation. In the course of strengthening data governance through large model use, at present, there are no particular rules for the use of compliant data, privacy protection and security responsibilities; as a result, serious regulatory problems and compliance risks have arisen.

6 Strategies for Reconstructing and Optimizing Multi-dimensional Data Governance Chains in Smart Cities Empowered by Large Models

6.1 Coordinate computing resources to reduce deployment costs

We will build a computing power system with municipal-level coordination, tiered deployment and edge collaboration. By integrating government cloud resources, enterprise cloud infrastructure and public computing resources, we will construct an urban intelligent computing platform for unified scheduling, on-demand allocation and efficient use of computing resources to avoid waste. Priority will be given to the adoption of lightweight large-scale models and edge computing solutions, and lightweight models will be deployed at grassroots terminals for real-time data preprocessing to reduce computational pressure on central systems. Increase fiscal investment and use a special construction fund to promote public-private partnerships and the involvement of social capital in sharing construction and maintenance expenses, thus reducing the local fiscal burden. At the same time, we will promote the application of domestic computing technologies and models to improve technological self-sufficiency and reduce outsourcing costs.

6.2 Strengthening Data Foundation to Enhance Model Training Quality

Set up a regular data quality control mechanism and strictly enforce unified collection standards to ensure data quality at the source. Construct high-quality urban data annotation libraries and perform accurate data labelling to provide stable data support for large-scale model training. Promote the all-round integration of data by breaking down departmental data silos, establishing compulsory data-sharing evaluation mechanisms, and thus motivating authorities to open data access for multi-dimensional data convergence. For local urban scenarios, perform special fine-tuning and training of large models that integrate professional urban governance knowledge to improve model adaptability and inference accuracy, and reduce the problem of model hallucination.

6.3 Strengthen safety control to ensure reliable model operation

Construct an all-encompassing security protection system for large models at all times, strengthen safeguards in the process of training, deployment, inference and iteration, and prevent attacks, tampering and data loss. Introduce model interpretability features to improve the openness of inference and ensure that the reason for a decision can be traced back. Add security barriers to restrict the output of the model strictly and prevent the generation of false or non-compliant information. Adjust the Data Access Control and, in accordance with the principle of least privilege, establish a tiered permission structure to prevent data leaks or unauthorized access. Build a model for security monitoring and assessment, conduct periodic vulnerability analyses, and promptly strengthen protective measures.

6.4 Cultivating Professional Talent and Strengthening Technical Support

Strengthen the cultivation of interdisciplinary talents by setting up special programmes for artificial intelligence, big data and urban governance at the university level to provide targeted professional training. On-the-job training initiatives will be introduced to provide tailored training for large-scale model technology and data governance practices of urban management and data governance professionals to improve their practical skills. Recruit high-end technical teams, collaborate with research institutes and technology companies on research and development, model debugging and operational maintenance support, etc., to address technology implementation challenges. Design talent incentive mechanisms to attract high-quality professionals to participate in the development of smart cities and data governance.

6.5 Improve collaborative mechanisms and break down institutional barriers

Establish an all-weather smart city data governance coordination body to guide the planning, centralised construction and cooperation of reform in the data governance system. Clarify the distribution of powers and responsibilities among departments and stakeholders, set specific rules for data sharing and joint governance, improve the efficiency of governance through performance evaluation, and consider the interests of departments. Build a government-enterprise collaborative governance framework with multi-stakeholder participation, motivate enterprises to participate in the construction of a data governance platform, carry out technology R&D and scenario-based applications, and achieve complementary strengths and synergistic effects. Construct an all-weather, all-level data-sharing platform to support cross-level cooperation and all-encompassing data exchange among cities, districts, sub-districts and villages.

6.6 Improve compliance systems and strengthen regulatory constraints

Strengthen the legal system and regulatory environment for the use of large models and urban data governance, issue explicit instructions on data ownership, sharing norms, privacy protection and security responsibilities, and provide a compliance foundation for governance reform. An all-encompassing monitoring system will be established to observe the full life cycle of the data, including collection and processing, sharing and application, and model operation, and regulatory violations will be strictly penalized. A data compliance registration and security assessment system will be built to conduct regular compliance audits of the security of data and models to reduce risks and ensure lawful and orderly implementation of data governance.

7 Conclusion and Prospects

7.1 Research Conclusions

Multidimensional Data Governance in Smart Cities is the Foundation for Urban Digitalisation and Intelligent Development. The old model of linear and segment data governance has many defects and cannot meet the requirements of the new urban governance model. Based on the above core strengths - multimodal understanding, massive data processing, deep reasoning and autonomous iteration - large-scale models have become key technology drivers in the restructuring of smart city data governance ecosystems. Research has shown that these models can support all links in the data lifecycle management system systematically, from data collection and aggregation to cleaning/pre-processing, integration and sharing, analytical mining, security management, and application deployment. This way to build a closed-loop, intelligent and integrated data governance system that can address problems such as data silos, inefficient governance, unused value potential and increased security risks. By improving governance efficiency and quality and reducing operating costs simultaneously, these solutions enable cities to build precise, data-driven and predictive smart urban governance systems, thereby promoting high-quality development of smart urban ecosystems.

At the same time, large-scale models are enabling the restructuring of the multidimensional data governance chain in smart cities, but they also face numerous problems, such as high computational costs, poor data quality, security risks of models, lack of talent, institutional obstacles, and incomplete compliance frameworks. The following all-around measures will be taken to address the above issues: coordinate computing resources, strengthen the data foundation, improve security control, nurture specialized talents, enhance collaborative mechanisms, and optimize the compliance system. Through the above measures, we will be able to apply the technology successfully.

7.2 Future Outlook

With the continuous progress of large-scale models and reduced computational costs, and the development of strong data governance systems for various data, their integration into smart city data governance platforms will be deepened, and the scope of application will be expanded. In the future, lightweight, specialised and localised urban models will be widespread, and edge devices will boost real-time governance by carrying out second-level processing and accurate predictive analysis in all parts of society. The data governance process will be more intelligent and closed-loop, and fully autonomous operation without human intervention will be achieved. The value of data will be enhanced to drive all-round improvements in urban governance, building a new kind of modern city that is efficient, safe, livable and intelligent.

At the same time, with the progress of market-oriented development for data elements, large models will help promote the compliant circulation and market-based trading of urban data, further activating data resources and creating new business forms and models. In the future, we will continue to advance technological research, optimise governance systems, balance security and development, and so on, to better empower the construction of smart cities and serve the overall goals of urban governance modernisation and digital China development.

Funding

This work was supported by the Scientific Research Project of the Department of Education in Jilin Province. Project Number: JJKH20261526KJ

About the Author

Yang Zhang was born in 1986 in Jilin City, Jilin Province, China, and in 2016 he obtained a Master of Science degree in Software Engineering from Inner Mongolia University. He is now a member of the faculty at the School of Information Engineering, Jilin Vocational College of Industry and Technology, and his research interests lie in big data analysis and artificial intelligence.

References

[1] Bozkurt, Y., Rossmann, A., Pervez, Z., et al. (2025). Assessing data governance models for smart cities: Benchmarking data governance models on the basis of European urban requirements. Sustainable Cities and Society, 130, 106528. doi: 10.1016/j.scs.2025.106528.

[2] Gilman, E., Bugiotti, F., Khalid, A., et al. (2024). Data governance requirements for distributed compliance in IoT. ACM Transactions on Intelligent Systems and Technology, 15(6), Article 87, 1–34. doi: 10.1145/3653689.

[3] Bibri, S. E., Huang, J., Krogstie, J. (2024). Artificial intelligence of things for synergizing smarter eco-city brain, metabolism, and platform: Pioneering data-driven environmental governance. Sustainable Cities and Society, 108, 105516. doi: 10.1016/j.scs.2024.105516.

[4] Bibri, S. E., Huang, J. (2025). Data-driven smart eco-cities and sustainable integrated districts: A best-evidence synthesis approach to an extensive literature review. Environmental Science and Ecotechnology, 26, 100591. doi: 10.1016/j.ese.2025. 100591.

[5] Bibri, S. E., Krogstie, J., Kaboli, A., et al. (2023). Smarter eco-cities and their leading-edge artificial intelligence of things solutions for environmental sustainability: A comprehensive systematic review. Environmental Science and Ecotechnology, 19, 100330. doi: 10.1016/j.ese.2023.100330.

[6] Lartey, D., Law, K. M. Y. (2025). Incorporating generative artificial intelligence with geographic information science to understand human-environment interactions. Landscape and Urban Planning, 258, 105337. doi: 10.1016/j.landurbplan.2025.105337.

[7] Zheng, Y., Xu, F., Lin, Y., et al. (2025). Urban science in the era of artificial intelligence. Nature Computational Science, 5(9), 727–736. doi: 10.1038/s43588-025-00846-1.

[8] Li, Z., Xia, L., Ren, X., et al. (2025). Urban computing in the era of large language models. ACM Transactions on Intelligent Systems and Technology, 16(6), Article 146, 1–43. doi: 10.1145/3768163.

[9] Zhang, S., Fu, D., Liang, W., et al. (2024). TrafficGPT: Viewing, processing and interacting with traffic foundation models. Transport Policy, 150, 95–105. doi: 10.1016/j.tranpol.2024.03.006.

[10] Chen, Y., Zhang, H., Li, C., et al. (2025). Large model empowered smart city mobility. Frontiers of Engineering Management, 12(1), 201–207. doi: 10.1007/s42524-025-4213-0.

[11] Xu, H., Omitaomu, F., Sabri, S., et al. (2024). Leveraging generative AI for urban digital twins. Urban Informatics, 3(1), 29. doi: 10.1007/s44212-024-00060-w.

[12] Zou, X., Yan, Y., Hao, X., et al. (2025). Deep multimodal fusion: A survey. Information Fusion, 113, 102606. doi: 10.1016/j.inffus.2024.102606.

[13] Dritsas, E., Trigka, M. (2025). Big data and Internet of Things applications in smart cities. Internet of Things, 34, 101770. doi: 10.1016/j.iot.2025.101770.

[14] Bhatia, M., Kumar, V. (2025). Artificial intelligence in digital twin technology: Scientometric insights on architecture, applications, and tools. IEEE Internet of Things Journal, 12(16), 32653–32675. doi: 10.1109/JIOT.2025.3575476.

[15] Lnenicka, M., Kysela, T., Horák, O. (2026). Building security and resilience: A guide to implementing effective cybersecurity and data protection measures in smart cities. Smart and Sustainable Built Environment, 15(2), 908–937. doi: 10.1108/SASBE-09-2024-0363.

[16] Rajamäe Soosaar, K., Nikiforova, A. (2025). Assessing the role of open government data in achieving the Sustainable Development Goals: A case study of the European Union. Computer Law & Security Review, 56, 106099. doi: 10.1016/j.clsr.2024.106099.

[17] Liva, G., Micheli, M., Schade, S., et al. (2023). City data ecosystems between theory and practice: A qualitative exploratory study in seven European cities. Data & Policy, 5, e17. doi: 10.1017/dap.2023.13.

[18] Lnenicka, M., Nikiforova, A., Luterek, M., et al. (2024). Understanding the development, use, and impact of the city data ecosystem: A systematic literature review and guidelines for future research. Telematics and Informatics, 94, 102190. doi: 10.1016/j.tele.2024.102190.

[19] Lnenicka, M., Nikiforova, A., Wang, D., et al. (2025). Investigating the user experience of smart city portals: A cross-country comparison of France, Germany, Italy, Spain, and the United Kingdom. Telematics and Informatics, 100, 102284. doi: 10.1016/j.tele.2025.102284.

[20] Lnenicka, M., Nikiforova, A., Clarinval, A., et al. (2024). Advancing urban sustainability: The role of smart city data ecosystems in managing open data initiatives. Cities, 148, 104851. doi: 10.1016/j.cities.2024.104851.

[21] Lnenicka, M., Nikiforova, A., Luterek, M., et al. (2024). Benchmarking open data efforts through indices and rankings: Assessing development and contexts of use. Government Information Quarterly, 41(1), 101898. doi: 10.1016/j.giq.2023.101898.

[22] Chang, Y., Wang, X., Wang, J., et al. (2024). A survey on evaluation of large language models. ACM Transactions on Intelligent Systems and Technology, 15(3), Article 39, 1–45. doi: 10.1145/3641289.

[23] Das, B. C., Amini, M. H., Wu, Y. (2025). Security and privacy challenges of large language models: A survey. ACM Computing Surveys, 57(6), 1–39. doi: 10.1145/3712001.

[24] Yao, Y., Duan, J., Xu, K., et al. (2024). A survey on large language model security and privacy: The good, the bad, and the ugly. High-Confidence Computing, 4(2), 100211. doi: 10.1016/j.hcc.2024.100211.

[25] Stahl, B. C., Eke, D. (2024). The ethics of ChatGPT: Exploring the ethical issues of an emerging technology. International Journal of Information Management, 74, 102700. doi: 10.1016/j.ijinfomgt.2023.102700.