Big data and data science in global governance: anticipating future needs and applications in the UN and beyond

Lanxin Li[1], Jiarou Wang[2], Xi Wang[3], Peng Peng[4], Jiaying Shen[5], Haining Zhu[6] and Ziyang Zhang[7]*

[1]Graduate School of Arts and Sciences, Columbia University, New York, NY, United States, [2]School of Humanities, Beijing University of Posts and Telecommunications, Beijing, China, [3]Asian Family Services, Auckland, New Zealand, [4]Graduate School of Arts and Sciences, Georgetown University, Washington, DC, United States, [5]Department of Media and Communications, London School of Economics, London, United Kingdom, [6]Department of Economics, Albert-Ludwigs-Universität Freiburg, Freiburg im Breisgau, Germany, [7]Sun Yat-sen University, Guangzhou, China

This paper explores the transformative potential of big data and data science in global governance, with particular emphasis on their application in international organizations addressing sustainable development challenges. Through comprehensive analysis of theoretical frameworks, current applications, and future directions, we examine how big data technologies enhance decision-making processes and operational efficiency in global governance frameworks, particularly within United Nations agencies and affiliated international organizations. The research identifies the "4Vs" of big data (Volume, Velocity, Variety, and Veracity) as fundamental characteristics reshaping governance approaches while highlighting innovative applications like UN Global Pulse, SDG tracking systems, and AI-driven predictive analytics in crisis prevention. We assess technical, ethical, and organizational challenges, including data quality inconsistencies, interoperability issues, privacy concerns, algorithmic bias, and resource constraints that impede the full integration of big data into governance systems. The paper proposes forward-looking strategies for infrastructure development, skills enhancement, and policy frameworks that can maximize big data's benefits while addressing ethical considerations and regulatory requirements. Our findings suggest that big data, when properly governed through international cooperation and ethical frameworks, can significantly enhance crisis response capabilities, improve resource allocation, and accelerate progress toward sustainable development goals. This research contributes to the evolving understanding of big data's role in addressing transnational challenges through improved monitoring systems, predictive capabilities, and evidence-based policy interventions.

KEYWORDS

big data, global governance, sustainable development goals, international organizations, data ethics, AI policy, data sovereignty and digital transformation

1 Introduction

1.1 Context and importance

Big data and data science offers new opportunities for managing global issues, increasing the effectiveness and speed of implementation (Hansen and Porter, 2017). These technologies gather information from diverse sources—from satellite imagery and remote sensors to social media—and translate them into actionable insights for addressing

global challenges. By enabling real-time analysis and predictive modeling, big data assists policymakers and governments in making plans and drawing solutions to current and future global challenges (Giest, 2017).

The United Nations Sustainable Development Goals (SDGs), which aim to eradicate poverty, eliminate hunger, expand access to clean water, and combat climate change among other goals, exemplify the types of complex global issues that can benefit from data-driven solutions. For instance, satellite observation coupled with modeling enables stakeholders to track the dynamics of deforestation and urbanization processes and evaluate agricultural productivity for some SDG indicators.

As the world becomes increasingly interconnected, traditional policy responses often struggle to keep pace with the scale, speed, and complexity of global phenomena such as pandemics, natural disasters, and military conflicts (Kuzio et al., 2022). In this context, big data provides a means to enhance evidence-based governance across national borders, facilitating smarter, more inclusive international responses. According to Sirbu et al. (2021), real-time data analytics are useful because they provide extra context concerning crises and tracking movement or trends. This contextual importance serves to justify the need for big data and data science in the future of global governance.

1.2 Research objectives and key questions

This paper explores the potential of big data and data science in global governance, with an emphasis on future needs and applications. It provides a comprehensive review of current applications of big data in international organizations; limitations in technological infrastructure, ethical guidelines, and political frameworks; solutions to address current challenges; and future directions for emerging technologies in multilateral institutions.

Main research questions of the paper include:

(1) How are international organizations currently applying big data to support global governance goals such as the SDGs, humanitarian relief, and crisis prevention?
(2) What technical and organizational infrastructures are required to implement big data systems at scale within international organizations?
(3) What ethical, legal, and political challenges arise from the cross-border use of big data for global governance?
(4) How can big data and AI enhance early warning systems for crises, from food insecurity to political conflict?
(5) What policy reforms, standards, and capacity-building strategies are necessary to ensure equitable and responsible use of big data in international governance?

Answering these questions, the paper aims to offer a theoretical and practical understanding of how big data can help address global challenges (Table 1).

1.3 Justification of scope and sampling strategy

To explore these questions, we focus on a purposive sample of international organizations and their projects that represent the diversity of big data-driven governance initiatives globally. This includes:

- **Multilateral organizations** such as the UN, World Bank, and WHO—selected for their central roles in shaping global development, health, and humanitarian agendas.
- **Key projects and initiatives utilizing big data**, such as UN Global Pulse, the Global Information and Early Warning System on Food and Agriculture (GIEWS), and WHO's NEXO weather stations—chosen because they exemplify cutting-edge integration of data science into institutional workflows and operate at a relatively large scale with demonstrable influence.
- **Specific big data applications,** spanning domains like climate change, food security monitoring, public health analytics, and conflict prediction—selected to demonstrate the breadth of use cases and the evolving toolkit of data governance.

This sampling strategy enables a nuanced exploration of both technical innovations and institutional pathways, ensuring the paper remains grounded in practical experiences while offering insights transferable across sectors and regions.

1.4 Structure overview

This paper adopts a systematic approach to examine the role of big data in global governance. The paper is divided into eight sections. Section 1 is an introduction that presents the topic under discussion, whereas Section 8 discusses the research findings and outlines recommendations concerning the policy implications and future research prospects (Figure 1). The rest of the sections presents different issues concerning big data and global governance:

Section 2: Background introduces the foundational concepts of big data and data science, including related technologies such as AI and Internet of Things, as well as emerging technological trends such as data democratization, cloud storage evolution, and visualization advances. It establishes the conceptual and historical groundwork necessary for understanding pathways for smarter, real-time decision-making.

Section 3: Current Applications examines how major international organizations (UN, World Bank, WHO) are deploying big data. Case studies include SDG monitoring through satellite imagery, AI-assisted social welfare targeting, and global health AI standards.

Section 4: Challenges and Risks analyzes the technical, ethical, and organizational barriers to big data adoption, including data quality, algorithmic bias, interoperability issues, and digital authoritarianism.

TABLE 1 Correspondence between research questions and article sections.

Research question	Primary sections	Secondary sections
RQ1: How are international organizations currently applying big data to support global governance goals?	**Section 3:** Current Applications of Big Data in International Organizations	**Section 6:** Emerging Applications and Future Developments
RQ2: What technical and organizational infrastructures are required?	**Section 5:** Anticipated Future Needs (*5.1 Infrastructure Needs*)	**Section 2:** Background on Big Data (*2.3 Technological Trends*)
RQ3: What ethical, legal, and political challenges arise?	**Section 4:** Potential Challenges and Risks	**Section 7:** Policy and Ethical Considerations
RQ4: How can big data and AI enhance early warning systems?	**Section 6:** Emerging Applications (*6.1 Predictive Analytics in Crisis Prevention*)	**Section 3.1:** UN Global Pulse Examples
RQ5: What policy reforms and capacity-building strategies are necessary?	**Section 7:** Policy and Ethical Considerations	**Section 5:** Anticipated Future Needs (*5.2 Skills Development, 5.3 Policy Frameworks*)

FIGURE 1
Map and structure of our research in big data in global governance.
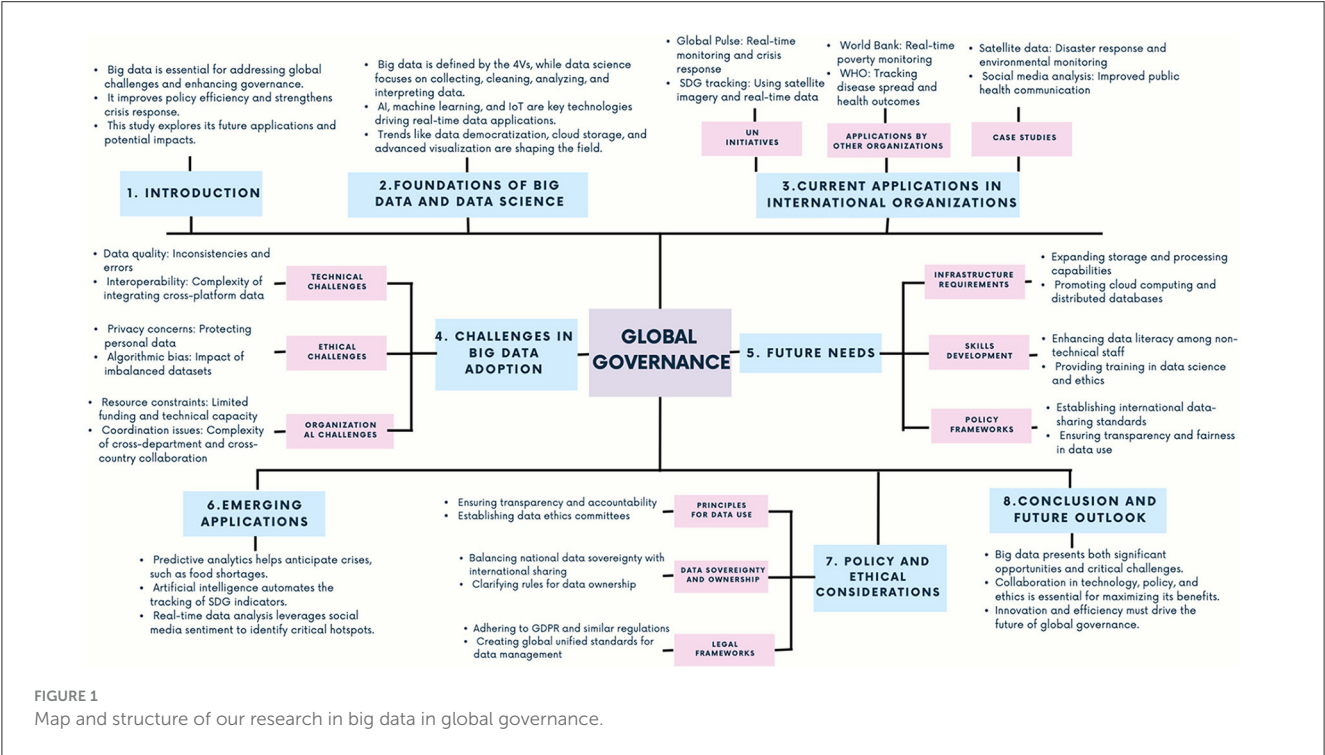
Section 5: Anticipated Needs outlines future requirements in infrastructure, training, and regulatory development for expanding data-driven governance. It identifies bottlenecks in cloud computing, data literacy, and policy coordination.

Section 6: Emerging Applications explores the latest emerging applications of big data within international organizations and identifies areas where further development and innovation are needed, particularly following progress in infrastructure, training, and regulatory development highlights new uses of AI and predictive analytics.
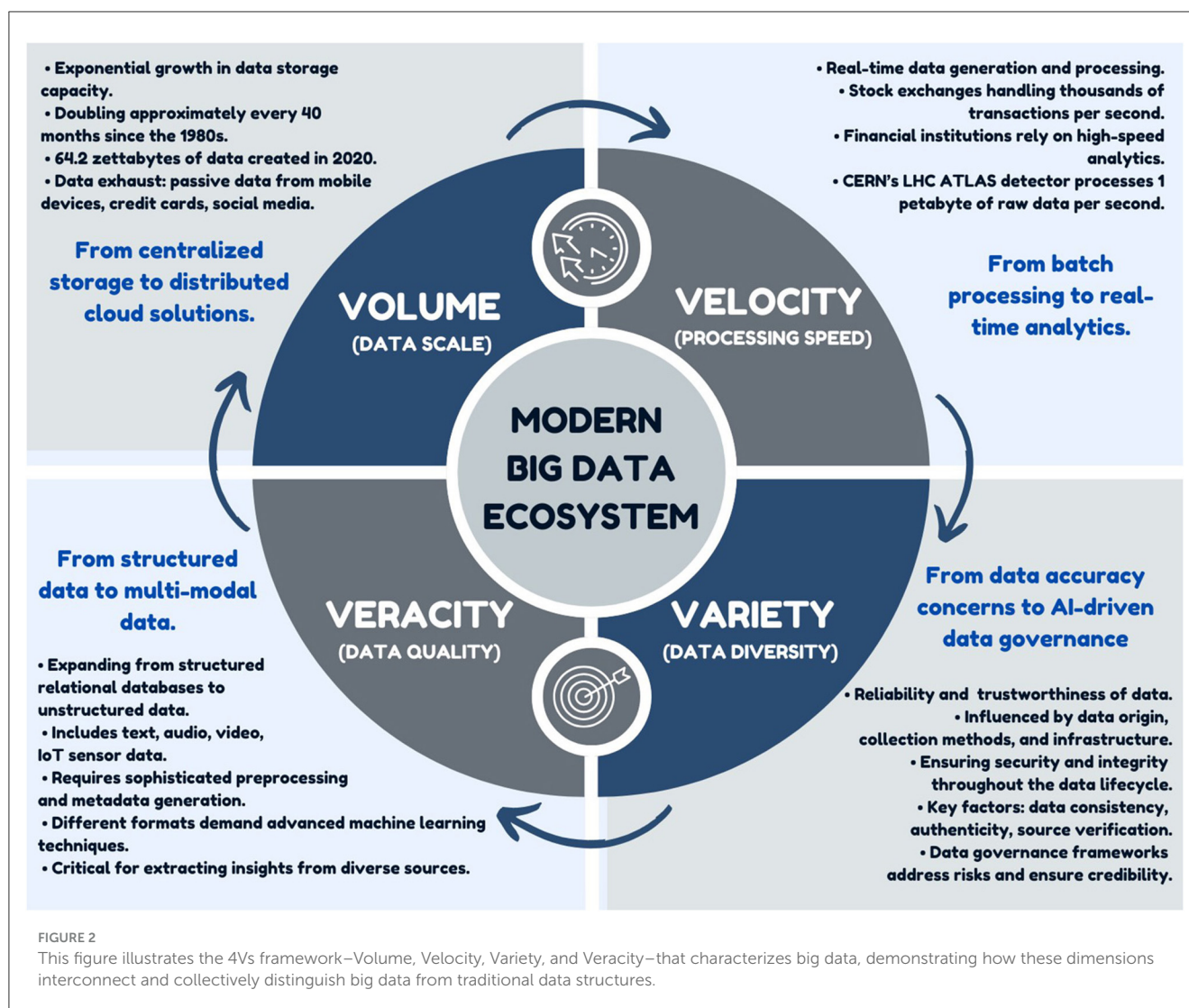
Section 7: Policy and Ethical Considerations evaluates normative frameworks governing data use, including data sovereignty, international guidelines and regulations, the General Data Protection Regulation (GDPR) compliance, and accountability mechanisms.

2 Background on big data and data science

2.1 Definitions and core concepts

2.1.1 Big data definition

Big data has emerged as a transformative concept in the digital age, driven by the exponential growth of data generated through various digital activities. The McKinsey Foundation defines big data as "data sets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" (Manyika et al., 2011). The inherent value of big data lies in its ability to reveal patterns through the interconnections of data points concerning individuals, groups, or the underlying structure of information itself (Vargas-Solar et al., 2016).

FIGURE 2
This figure illustrates the 4Vs framework—Volume, Velocity, Variety, and Veracity—that characterizes big data, demonstrating how these dimensions interconnect and collectively distinguish big data from traditional data structures.

The concept of big data is also commonly characterized by the 4Vs framework, which effectively offers a comprehensive understanding of its nature (Kulkarni et al., 2016; Huang et al., 2015) (Figure 2).

2.1.1.1 Volume: scale and size of data

Volume represents the unprecedented scale and magnitude of data in the modern era. The world's data storage capacity has exhibited exponential growth, doubling approximately every 40 months since the 1980s. The COVID-19 pandemic catalyzed an unprecedented surge in information demand, resulting in the creation of 64.2 zettabytes of data in 2020—a remarkable 314% increase from 2015 levels. A significant portion of this consists of "data exhaust", representing passively collected data from daily interactions with digital products and services, including mobile devices, credit cards and social media platforms. It is precisely because people generate data through a variety of activities, including creating documents, downloading media and using applications. This has led to the demand for sophisticated storage solutions and enhanced data processing capabilities.

2.1.1.2 Velocity: speed of data generation and processing

Velocity encompasses both the rapid pace of data generation and the critical need for swift processing capabilities in the big data era. Real-time systems, particularly in financial trading, depend heavily on high-velocity data processing. Stock exchanges, for example, handle thousands of transactions per second, while financial institutions conduct real-time market data analysis for informed decision-making. In industrial applications, sensor networks continuously generate high-speed data streams, exemplified by manufacturing plant sensors monitoring various metrics. A notable example is the LHC ATLAS detector at CERN, which employs approximately 80 readout channels and possesses up to 1 petabyte of unfiltered data per second, subsequently refined to around 100 megabytes per second. This sophisticated system is engineered to record up to 40 million collision events every second (Demchenko et al., 2024).

2.1.1.3 Variety: different types and sources of data

Variety addresses the diverse spectrum of data types in the contemporary landscape. With traditional data structures were primarily confined to structured data in relational databases, the

big data era has introduced numerous forms of unstructured data, including text, audio, and video formats. These diverse data types require sophisticated preprocessing techniques to extract meaningful insights and generate appropriate metadata (Demchenko et al., 2024).

2.1.1.4 Veracity: data quality and reliability

The veracity dimension of Big Data encompasses dual aspects: data consistency determined by statistical reliability, and data trustworthiness, influenced by factors including data origin, collection methodologies and infrastructure integrity. Big Data veracity guarantees the trustworthiness, authenticity of used data while protecting against unauthorized access and modification. Data security must be maintained throughout its lifecycle, from collection from trusted sources to processing on verified computing facilities and secure storage systems. Essential considerations in maintaining data veracity include data and linked data integrity, data authenticity and trusted origin, identification of both data and source, computing and storage platform trustworthiness, availability and timeliness, as well as accountability and reputation management (Demchenko et al., 2024).

2.1.2 Historical evolution of the concept

The evolution of big data represents a gradual progression aligned with information technology advancement. Initially, limited data storage and processing capabilities restricted the understanding of large-scale data management. With the advancement of computer technology, the amount of generated and stored data gradually increased, and people began to recognize the potential value of data. The 1990s witnessed the emergence of data warehousing and business intelligence technologies, providing enterprises with enhanced data management and analysis tools. During this period, the scale and complexity of data grew steadily, though not yet reaching big data proportions. The early twenty-first century, marked by Internet proliferation and e-commerce growth, saw an exponential increase in data generation rates and scale. The emergence of search engines, social media platforms and similar technologies led to unprecedented accumulation of user behavior data, focusing attention on large-scale data processing and analysis. In 2012, the US government launched the "Big Data Research and Development Initiative", elevating big data to the national strategic level. This move attracted extensive attention from various organizations and countries around the world, and the relevant infrastructure, industrial applications and theoretical systems of big data have been continuously developed and improved. Since then, big data has gradually transformed from a single technical concept into a new element, strategy and mindset (Data Strategy Key Laboratory, 2017).

2.1.3 Current understanding of big data in global governance

Big data has transformed global governance by automating governance processes, reshaping interactions among nations, and empowering non-state actors. It transcends national boundaries, establishes new connection patterns, and employs algorithms and automation to influence governance structures. "Big data relies on the globalization operation of new media and strengthens, extends, obscures and confounds power in new ways" (Hansen and Porter, 2017), presenting a governance model distinct from traditional sovereignty. Cross-border data flows have spurred competition for control, with countries and enterprises establishing new "boundaries" to manage data access, altering traditional governance patterns. However, challenges persist, including privacy concerns, "data nationalism," and unequal access to critical data hinder global cooperation, particularly affecting marginalized populations (Castro, 2013).

Despite these challenges, big data offers opportunities to address global issues. The UN Global Pulse Initiative exemplifies this potential through real-time food price monitoring, providing governments with timely decision-making tools for crises prevention (UN Global Pulse, 2024). Additionally, big data analysis reveals previously hidden social inequalities, such as the marginalization of women in informal sectors, offering insights for more inclusive policy-making (United Nations, 2023). Through the integration of algorithms, automation, and cross-border collaboration, big data continues to redefine governance, amplifying the roles of technology platforms, private companies, and non-human actors (Barnett et al., 2021).

2.1.4 Data science components

Data Science Components refer to the fundamental components that make up workflows and systems in the field of data science. These components work together to help data scientists extract valuable insights from data, construct models, and facilitate data-driven decision-making.

2.1.4.1 Data collection methodologies

Data collection represents the initial phase of any data science project. Sources vary widely, including databases (both relational such as MySQL, and non-relational databases like MongoDB), file systems (CSV files, JSON files, etc.), web crawlers (extracting data from web pages), and sensors (data collected by Internet of Things devices), etc. For example, an e-commerce companies might obtain user purchase records from its sales databases, browsing behavior data from website log files, and competitor pricing information through web crawlers establishing a comprehensive foundation for subsequent analysis (Wang et al., 2019).

2.1.4.2 Data cleaning techniques

Raw data often has various problems, such as missing values, duplicate values, incorrect values, and inconsistent data formats. Data cleaning addresses these problems through systematic approaches. For missing values, methods such as deleting records with missing values or filling with the mean or median can be used; for duplicate values, they need to be identified and deleted; for incorrect values, they may need to be corrected based on business rules; for inconsistent data formats, such as different date formats, they need to be unified. For example, in a medical data set, there may be some incorrect inputs in the patient's age field (such as a negative age), which need to be corrected. At the same time, if there are spelling errors or duplicate records in the patient's name, they also need to be cleaned (Kulkarni et al., 2016).

2.1.4.3 Analysis approaches

Analysis primarily involves preliminary data examination to understand distribution patterns and feature relationships. This includes calculating fundamental statistics (including means, medians, and standard deviations) and generating visual representations (such as histograms, scatter plots, and box plots). Through data exploration, outliers and skewed distributions in the data can be discovered. For example, when analyzing telecom customer data, by drawing a histogram of the monthly call duration, the distribution of the customer call duration can be understood, and whether there are a few customers with extremely long call durations (possibly business users) can be known, thus providing clues for subsequent customer classification.

2.1.4.4 Interpretation frameworks

Interpretation frameworks comprise methodologies and strategies for understanding and communicating analysis results and model outputs, helping stakeholders (such as business decision-makers, customers, etc.) understand the significance and value of data science work.

For machine learning models, especially complex ones (such as deep learning models), it is important to explain their working principles and output results. For example, for an image classification model based on deep learning, feature importance analysis (such as calculating the importance scores of features) can be used to explain how the model classifies according to different features of the image (such as color, shape, etc.). For decision tree models, the decision-making process of the model can be explained by showing the structure of the decision tree (nodes and branches).

2.1.5 Application of data science in international organizations

The application of data science in international organizations demonstrates its vital role in decision-making, policy monitoring, and fostering global collaboration. These organizations leverage statistical data across domains such as trade, unemployment, and public health to provide critical insights to the global community. These analytical tools not only measure development progress but also define international challenges, from identifying "least developed countries" to establishing global debt benchmarks. Additionally, quantitative research plays a pivotal role in driving public policy reforms through international oversight. The statistical operations within international organizations involve complex collaborations among intergovernmental bodies, expert committees, and secretariats, achieved through sophisticated cross-national coordination.

Moreover, the quantitative approaches of these organizations are often viewed as authoritative tools, valued for their neutrality and independence from domestic political conflicts, while promoting policy innovation and cross-border cooperation. Data science, through the analysis of statistical methodologies and data collection techniques, offers a unique perspective on the historical evolution and impact of international organizations, as exemplified by UNESCO's adaptation of its statistical framework to align with neoliberal policy objectives (Cussó and Piguet, 2023).

In practical applications, data science supports humanitarian aid, health monitoring, and the achievement of Sustainable Development Goals (SDGs) (Kirkpatrick and Vacarelu, 2018). The United Nations optimizes resource allocation through population displacement data in conflict zones, WHO employs data modeling for epidemic trends, and UNICEF uses data analysis to enhance service coverage for vulnerable groups. Similarly, the International Organization for Migration monitors migration pathways and evaluates policy impacts, while platforms like OCHA's Relief Web integrate data from multiple sources to aid disaster response and information dissemination.

Across disciplines such as economics and agriculture, data science facilitates investigation of complex issues such as war-induced displacement and human rights violations. With the integration of big data and artificial intelligence, data science has become indispensable for managing global affairs and improving service delivery in international organizations, providing robust support for global governance and policy design (Cussó and Piguet, 2023; Dixit and Gill, 2024).

2.2 Technologies related to big data

2.2.1 AI and machine learning

The integration of AI and Machine Learning algorithms with Big Data enables advanced analytical capabilities, including predictive insights generation, decision-making processes automation, and support for innovative applications spanning recommendation systems, fraud detection, and personalized marketing strategies (Demchenko et al., 2024,).

In the domain of artificial intelligence and machine learning, there are three primary methodologies–supervised learning, unsupervised learning, and reinforcement learning—addressing distinct aspects of data-driven problem-solving in different ways.

2.2.1.1 Supervised learning

Supervised learning involves the training of models using labeled datasets to establish input-output mappings. This methodology finds extensive application in classification tasks (such as spam detection) and regression problems (such as house price prediction), where the goal is to optimize the model's performance by minimizing prediction errors. Notable supervised learning algorithms include XGBoost, LightGBM, and k-Nearest Neighbors, widely used across finance, healthcare, and e-commerce sectors. Additionally, supervised learning contributes significantly to feature engineering and time series forecasting through specialized algorithms including Object2Vec and DeepAR.

2.2.1.2 Unsupervised learning

Unsupervised learning focuses on processing unlabeled data to discover inherent patterns and structures within the data. This approach aims to identify data characteristics through various techniques like clustering, dimensionality reduction, and anomaly detection. Prominent algorithms in the category include Principal Component Analysis (PCA) for dimensionality reduction, K-Means for clustering, and Random Cut Forest (RCF) for anomaly detection. These techniques are particularly effective in customer segmentation, genetic data analysis, and fraud detection. Unsupervised learning is well-suited for handling complex high-dimensional data, and methods such as IP Insights (used for

recognizing network patterns) have proven valuable in specialized fields (Demchenko et al., 2024).

2.2.1.3 Reinforcement learning

Reinforcement learning, in contrast to the previous two methods, does not rely on labeled data but instead learns through trial and error by interacting with the environment. The methodology emphasizes cumulative reward optimization through action adjustment based on environmental feedback. In reinforcement learning, an agent takes actions in an environment and refines its strategy over time to maximize long-term benefits. Contemporary algorithms such as Deep Q-Networks (DQN) and policy gradient methods are primarily applied in areas like robotics, game AI, and autonomous driving. Reinforcement learning emphasizes dynamic decision-making processes and adapts strategies to complex environments, driving innovation in a variety of applications.

2.2.2 Applications of AI and machine learning in global governance

AI and machine learning demonstrate significant impact across multiple domains of global governance. In the economic field, in the face of severe disruptions to the global food supply chain caused by abnormal events such as trade conflicts, natural disasters, and pandemics, AI methodologies (including machine learning, reinforcement learning, and deep learning) can identify regular, irregular, and contextual components, enhancing understanding events outcomes and guiding decision-making guidance for suppliers, farmers, processors, wholesalers, retailers, and policymakers to promote welfare improvement (Batarseh and Gopinath, 2020).

In the technological field, particularly network development, the evolution from 5G to next-generation 6G networks increasingly relies on intelligent network orchestration and management. AI will play a key role in the emerging 6G paradigm. Moreover, as a system or machine that simulates human thinking, although most AI systems have no perception ability, they can think and calculate like humans through algorithms and cooperate with technologies such as machine learning (Yarali, 2023).

In addition, in international migration governance, AI technology is widely applied in various scenarios such as the prediction management of migration flows, automated decision-making, identity recognition, machine learning and matching, sentiment analysis, border monitoring, and robotics. While technological iterations drive governance concepts and model transformation, they simultaneously impact governance systems through capability disparities among governance subjects, fairness considerations, and normative frameworks (Chen and Wu, 2021).

2.2.3 Internet of Things

The Internet of Things (IoT) is also closely related to big data, as it generates vast volumes of real-time data from connected devices that require advanced analytics to process and interpret. The IoT architecture represents an ecosystem comprising interconnected physical objects accessible via Internet protocols. This architecture embodies the continuous integration between digital and physical realms, incorporating various technologies to form the Internet of Things (Margaret Mary et al., 2021). The purpose of designing the IoT architecture is to meet system requirements in specific application fields. Implementing an IoT system based on a given architecture to achieve required characteristics is essential.

The main IoT architecture categories include software IoT architecture, hardware IoT architecture, and general IoT architecture. Research suggests an end-to-end IoT architecture based on a five-layer model with various enabling technologies (Hafdi, 2019).

The IoT system contains several key components. Devices (Things), as the core element, evolve in aspects like user interface, appearance, performance, energy consumption, and security. New IoT devices explore applications with enhanced intelligence, incorporating various sensors and actuators, microcontroller architectures (such as ARM Cortex-M), input/output interfaces, programming models, and real-time operating systems (Firouzi et al., 2020).

2.2.4 Real-world applications of IoT in governance

IoT applications in governance continue to expand, transforming social governance. In smart city initiatives, IoT devices monitor environmental indicators like air and water quality in real time, providing data-driven support for decision-making. Wearable devices collect health data for personalized healthcare services. In traffic management, IoT optimizes signal control through real-time monitoring, while smart parking systems improve efficiency by helping drivers quickly locate available parking spaces (Nimkar and Khanapurkar, 2021). Water quality monitoring supports ecological goals through scientific management solutions (Lan and Fan, 2002). In agriculture, IoT sensors monitor soil conditions for precision farming while ensuring food safety through product traceability. Industrial applications improve production efficiency through equipment monitoring and intelligent management. In transportation, IoT enhances logistics by tracking shipments in real-time (Verma et al., 2022).

2.3 Technological trends

As countries and organizations increasingly depend on data-driven strategies and digital infrastructures, several technological trends have emerged. Among them, data democratization, the evolution of cloud storage, and advancements in data visualization are transforming how information is accessed, processed, and interpreted. Understanding these trends offers critical insight into both current capabilities and future trajectories across industries.

2.3.1 Data democratization

Data Democratization (DD) enables broader employee access to data understanding and usage within organizations. Its core features include wider data access with security controls, self-service analytics tools, enhanced data literacy training,

collaborative knowledge sharing, and promoting data value recognition (Lefebvre et al., 2021).

This concept has seen notable implementation in digital-native companies such as Airbnb, Uber, and Netflix, where data-driven thinking is deeply embedded in organizational culture. In contrast, traditional enterprises, despite investments in IT infrastructure, often grapple with cultural inertia—stemming from limited managerial awareness and underestimated employee data capabilities (Lefebvre et al., 2021).

Research on data democratization remains focused on specific areas without a comprehensive theoretical framework. Key dimensions include data accessibility, analytical skill development, and knowledge sharing, with applications in healthcare and urban planning.

Looking ahead, data democratization continues to evolve with exponential data growth projected for the next 5 years. The advancement in data analytics enables the extraction of meaningful context from external data sources for better business decisions. Future trends show that technology, especially AI and machine learning, will accelerate the process by automatically identifying patterns, generating insights, and handling data, thus lowering the usage threshold. For enterprises, benefits include improved decision-making, enhanced cohesion, and deeper customer understanding. Challenges remain in organizational silos, specialist dependence, and data visualization limitations (Lefebvre et al., 2021).

2.3.2 Cloud storage evolution

Building upon the data access revolution, the evolution of cloud storage marks another foundational shift in digital infrastructure. Originally celebrated for overcoming traditional limitations in data capacity and accessibility, cloud platforms like Google Drive, OneDrive, and Dropbox have become integral to storing and sharing information efficiently (Rawat, 2020).

Cloud storage technology continues to evolve with cutting-edge developments. AIOps platforms leverage machine learning, and analytics to enhance IT operations through proactive insights (Levin et al., 2019). In electrical automation, cloud storage facilitates efficient data processing for electrical automation control, combining work data from production and manufacturing with big data storage to guide work processes and inform operational decisions (Yu, 2020). Meanwhile, in the security system sector, the integration of cloud computing and cloud storage technologies has become increasingly profound, significantly enhancing the performance and reliability of security systems to meet the growing demands for safety (Zhou, 2018).

However, this progress brings security challenges. Data encryption faces challenges with duplicate detection, while convergent encryption remains vulnerable to attacks. Client-side deduplication technology presents security risks despite storage efficiency benefits (Park et al., 2015). Key management issues also persist, particularly in secure deduplication systems as user numbers increase (Babu and Babu, 2016).

To counteract these concerns, flexible distributed schemes can enhance data accuracy and security. Fuzzy authorization schemes can also improve access control through encryption and third-party auditor mechanisms, reducing storage consumption while maintaining public auditing capabilities (Puranik et al., 2016).

2.3.3 Visualization advances

As more organizations gain access to data and the storage infrastructure becomes more powerful, the ability to interpret and communicate complex information becomes crucial. This is where data visualization technologies have made their mark, offering dynamic and intuitive ways to translate raw data into meaningful insights.

The latest tools and techniques for big data visualization have significantly advanced, offering powerful functionalities tailored to diverse analytical scenarios. Tableau connects to various data sources, providing intuitive interfaces for complex analyses. RapidMiner excels in data mining and machine learning visualization. R Studio, with packages like ggplot2 enables detailed statistical visualization for fields such as finance and market analysis (Okechukwu, 2022).

Current visualization technologies can handle large-scale datasets efficiency, supporting diverse data types, and interactive exploration. Tools like Tableau and RapidMiner efficiently process millions of records in real-time, managing structured and unstructured data while enabling collaborative decision-making through platforms like Tableau Public (Okechukwu, 2022).

Future developments will further enhance usability and impact. Technologically, they will become smarter, capable of identifying patterns, suggesting suitable chart types, and highlighting key data points, making analysis more efficient and user-friendly. Enhanced interactivity, such as gesture controls and voice commands, will enable users to explore data more intuitively (Xie et al., 2022). Moreover, the integration of AI, VR, and AR will deliver richer, more immersive visualization experiences (Sharma et al., 2021).

The applications of these tools will also expand significantly. In business, they will support decision-making by visualizing sales or customer data to guide precise strategies (Hu and Hao, 2020). In healthcare, they will help doctors analyze patient records and imaging data, improving diagnosis accuracy and enabling personalized treatments (Nazir et al., 2019). In journalism, data visualization will become a powerful storytelling tool, presenting trends and impacts through intuitive charts (Liang, 2017).

2.4 Summary

A solid understanding of big data, data science, and their associated technologies is foundational for navigating today's digital transformation. Key trends such as data democratization, cloud storage evolution, and advancements in data visualization are revolutionizing how institutions collect, share, and analyze information. These shifts not only enable more efficient data use but also pave the way for smarter, real-time decision-making.

As this section has shown the fundamentals of these technologies, the following sections will explore the concrete applications of big data within international organizations

such as the UN system—highlighting how big data is being harnessed to support sustainable development goals and enhance humanitarian response.

3 Current applications of big data in international organizations

This section directly addresses our first research question by examining concrete applications of big data across major international organizations. We analyze how the UN, World Bank, and WHO are leveraging big data technologies to advance SDGs, enhance humanitarian relief efforts, and strengthen crisis prevention mechanisms.

3.1 UN initiatives in big data applications

3.1.1 UN global pulse

UN Global Pulse was launched in 2009 to explore how big data and data science can support the Sustainable Development Goals (SDGs) and humanitarian action. The initiative harnesses emerging technologies to understand global challenges, enabling real-time decision-making and improving UN programs efficiency (United Nations, 2023; UN Global Pulse, 2023). Its core objectives include: harnessing big data for real-time monitoring of SDG progress (UN Global Pulse, 2023); using mobile phone data, satellite imagery, and social media to improve humanitarian aid effectiveness (Legner, 2021); and strengthening partnerships between international organizations, governments, and technology companies (Bandola-Gill et al., 2022). Global Pulse also partners with technology firms like Microsoft, Google, and IBM for data analytics capabilities, machine learning, and cloud storage (Soares, 2012). Furthermore, collaborations with the World Bank, WHO, and UNICEF ensure alignment with global development agendas (Aguerre et al., 2024).

Global Pulse is coordinated by the UN Secretary-General's office in New York, with regional hubs in Jakarta (Indonesia), Kampala (Uganda), and Helsinki (Finland), focusing on food security, migration, and urbanization (UN Global Pulse, 2023). These hubs collaborate with stakeholders to apply data analytics to local challenges (Hennen et al., 2023).

Pulse Lab Jakarta, a regional hub jointly operated by UN Global Pulse and the Government of Indonesia, serves the broader Asia-Pacific region as a data innovation facility. It has been instrumental in utilizing mobile data and satellite imagery for disaster risk assessment and urban development. The lab analyzes real-time data to track population movement during floods and guide resource allocation (UN Global Pulse, 2023; United Nations, 2023). This project underscores the importance of data-driven urban resilience (Aguerre et al., 2024).

Similarly, Pulse Lab Kampala, the UN Global Pulse regional hub based in Uganda, focuses on food security by integrating mobile phone data with weather forecasts. During the 2020 drought, Pulse Lab Uganda's predictions helped NGOs allocate food aid more efficiently (United Nations, 2023; Hennen et al., 2023).

3.1.1.1 Project analysis

Leveraging social media data from platforms like Twitter and Facebook, Global Pulse has developed systems for monitoring public sentiment and identifying emerging crises, including Ebola outbreak monitoring and public health messaging (UN Global Pulse, 2013; United Nations, 2023). The initiative utilizes Natural Language Processing (NLP), sentiment analysis, and machine learning to analyze unstructured social media data, helping track trends and gauge public sentiment in real time (Legner, 2021). The data includes social media posts, user-generated content, and geospatial information. These are analyzed to track crises, public opinion, and information diffusion (United Nations, 2023).

Sentiment analysis and geospatial mapping are core techniques used to identify regions of high activity or concern. These analyses enable the UN to monitor global events in real time and optimize resource allocation during disasters (UN Global Pulse, 2023). Meanwhile, geolocation helps map crisis impact, while text analysis provides insights into public health, migration trends, and social unrest (Soares, 2012).

With social media analytics, Global Pulse's projects have successfully improved response times and resource allocation, as demonstrated during the Nepal earthquake and Uganda food security initiatives (Legner, 2021; Bandola-Gill et al., 2022). During the 2015 Nepal earthquake, social media data helped track aftershocks and coordinate rescue efforts more effectively (United Nations, 2023; UN Global Pulse, 2023). In the Ebola outbreak, it helped monitor misinformation for better public health communication, while during Hurricane Maria in Puerto Rico, social media analysis directed aid to the hardest-hit areas (Legner, 2021).

Despite successes, challenges persist in data privacy, data quality, and cross-border data sharing. Global Pulse employs anonymization techniques to protect privacy but concerns about data sovereignty remain significant hurdles (Aguerre et al., 2024). Integrating datasets from countries with different data infrastructure presents logistical challenges (Soares, 2012).

Key lessons from Global Pulse's projects include collaborating with local actors to ensure data relevance and accuracy and the importance of data interoperability to facilitate cross-sector collaboration (Hennen et al., 2023). Moreover, ethical data collection and usage frameworks must be carefully designed to balance development goals with privacy and security concerns (Aguerre et al., 2024).

3.1.1.2 SDG tracking and monitoring

Big data also plays an important role in tracking SDG progress, and the UN has developed several platforms to support this effort. The primary platform is the Global SDG Indicator Platform by UN Statistics Division (UNSD), which consolidates data from national governments and international organizations. The SDG Tracker by Our World in Data also provides real-time visualizations of SDG indicators (UN Big Data, 2025)

SDG monitoring combines traditional statistics (census data, national surveys) with alternative sources like satellite imagery and mobile data. Poverty and health indicators are monitored through household surveys, while environmental indicators rely on remote

sensing data (Legner, 2021). Big data and sensor networks enable real-time monitoring.

Countries report through the Voluntary National Reviews (VNRs), at the UN High-Level Political Forum. These are complemented by the SDG Progress Report, which aggregates data to assess global progress (United Nations, 2023).

UNSD works with national statistical offices to ensure standardized data collection. Many countries have established national SDG monitoring frameworks, though challenges remain in aligning diverse data sources, particularly in low-resource settings (Aguerre et al., 2024).

3.1.1.3 Applications of satellite imagery in SDG monitoring

For SDG 2 (Zero Hunger), satellite imagery monitors agricultural productivity and food security. Remote sensing tracks crop yields, drought conditions, and soil moisture. FAO uses this data to monitor global food production and assess climate impacts on food systems (Bandola-Gill et al., 2022).

Satellite data can also track urban development in rapidly growing cities, supporting the tracking of SDG 11 (Sustainable Cities and Communities). The Global Urban Monitoring Initiative observes urban growth patterns and environmental impact (Aguerre et al., 2024).

Additionally, satellite plays a vital role in monitoring SDG 13 (Climate Action). The European Space Agency (ESA) and NASA monitor global temperature changes, sea level rise, and glacier melt. Sentinel satellites track changes in vegetation, ice caps, and water bodies (Hennen et al., 2023).

Furthermore, satellite imagery provides near-instantaneous data on disaster impact analysis. After natural disasters like hurricanes, floods, or earthquakes, satellite data helps responders quickly identify the scale of destruction and humanitarian needs (United Nations, 2023). As mentioned earlier, after the 2015 Nepal earthquake, satellite data helped assess infrastructure damage and guide humanitarian aid deployment. This technology is essential for ensuring that relief efforts are targeted efficiently, saving lives and resources (UN Global Pulse, 2023).

3.1.1.4 Future development

Future developments in big data applications in the UN should focus on enhancing data accuracy. The integration of 5G networks, edge computing, and blockchain technologies may revolutionize remote data collection. New satellite constellations will provide more frequent imagery, improving SDG monitoring capabilities (Aguerre et al., 2024).

Equally important is addressing the challenge of interoperability. Current limitations include data interoperability between platforms and standardization of collection methods. Data gaps exist between satellite observations and ground-level information, particularly in rural or conflict areas. Ensuring technology sustainability in low-resource settings remains critical for inclusive and equitable data-driven development (Bandola-Gill et al., 2022; Soares, 2012) (Figure 3).

3.2 Current applications of big data in World Bank

The integration of big data and artificial intelligence (AI) enables decision-makers to address complex global challenges with precision and efficiency. At the World Bank, AI is applied both internally and externally, from project design to policy advice (Okahashi and Blanco, 2020). Lutz Lersch, the Bank's Senior AI Officer, emphasizes the Bank's responsibility to explore tools like ChatGPT and Google's Gemini while ensuring responsible implementation (Saldinger, 2023). Presented below are key applications of big data and AI at the World Bank.

3.2.1 Project development

The World Bank's Environmental and Social Framework (ESF) includes 10 E & S Standards for sustainable development projects. The AI-powered ESF Risk Assessment Toolkit leverages big data to generate baseline information and flags environmental and social risks. These reports help project teams facilitate strategic dialogues, assist consultants, and support client preparation work.

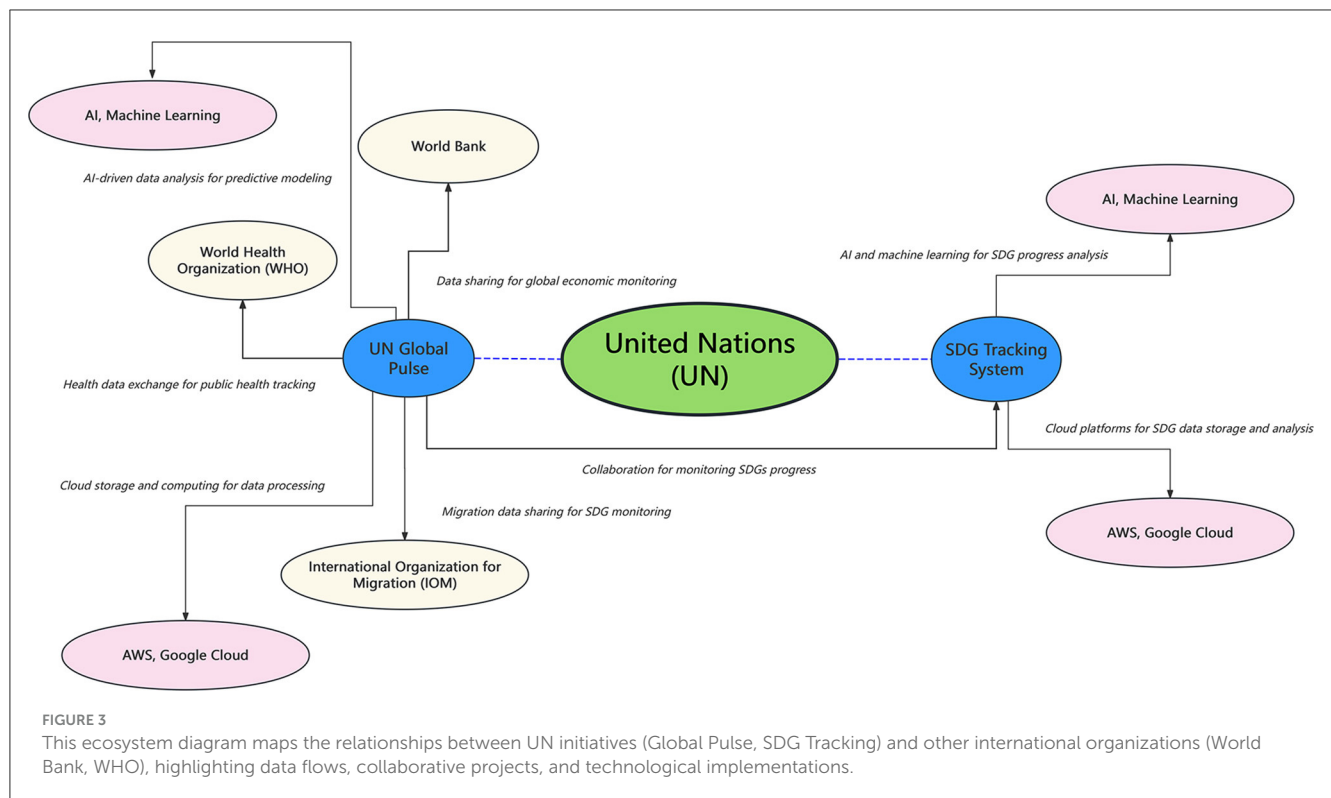### 3.2.2 Poverty reduction efforts

Since 2018, the World Bank Group has implemented nearly 45 projects including MALENA, Mai, and Impact AI (IFC Opens Public Access to MALENA: An AI-Powered Accelerator for Sustainable Investments). Impact AI enhances decision-making by utilizing big data to provide actionable solutions and evidence for development professionals and policymakers. The MALENA platform processes ESG data to optimize investments in poverty-stricken regions (IFC Opens Public Access to MALENA: An AI-Powered Accelerator for Sustainable Investments).

The Bank's big data and AI integration extends to real-time economic monitoring through satellite imagery and financial data (Real-Time Welfare Monitoring for Development Impact). A notable example is the $70 million Novissi cash transfer initiative in Togo during COVID-19. Using AI and satellite imagery, the program identified disadvantaged regions and leveraged mobile phone metadata to assess consumption patterns for 70% of the population, reaching over 57,000 beneficiaries between 2020 and 2021 (Prioritizing the Poorest and Most Vulnerable in West Africa).

For food insecurity. the Development Impact Group AI uses natural language processing on news articles as an early warning system. This approach improves crisis forecast accuracy by up to 50% for up to 1 year ahead, enabling proactive measures especially in data-limited areas (Fraiberger; Development Impact Group).

3.2.3 Climate finance and environmental sustainability

The integration of big data and AI in climate-smart decision-making offers governments seeking a promising tool to navigate climate resilience and sustainable development (UNDP, 2025). As climate change accelerates, governments face the dual challenge of

FIGURE 3
This ecosystem diagram maps the relationships between UN initiatives (Global Pulse, SDG Tracking) and other international organizations (World Bank, WHO), highlighting data flows, collaborative projects, and technological implementations.

protecting infrastructure from rising sea levels, extreme heat, and storm surges while ensuring investments contribute to emissions reduction. A key challenge is the gap between the vast climate data availability and governments' capacity to process and act on it. The World Bank highlights AI's potential to bridge this gap by analyzing large datasets and presenting actionable insights in user-friendly formats, enhancing climate-smart decision-making (Peixoto et al., 2023).

AI enables governments to move beyond simple data collection to proactive, context-specific decisions that advance climate resilience. For example, the Green Economy Diagnostic (GED) under the Climate-Smart Development Initiative integrates datasets- air quality, extreme weather patterns, and economic performance, to guide sub-national regions in climate-conscious investments. AI identifies trends, such as temperature and air quality variations, offering concrete recommendations on emissions standards and infrastructure resilience. As AI advances, it could further optimize resource allocation for climate mitigation and adaptation while fostering stakeholder collaboration.

However, AI's effectiveness in climate governance depends on multidisciplinary teams translating data into actionable solutions. Despite AI's ability to process large datasets, human and institutional barriers must be addressed. Governments need to invest in skilled teams for AI tool development and deployment, requiring substantial funding and support. Additionally, a "demos-driven" approach—focusing on practical results rather than theoretical frameworks—is crucial for refining AI applications. Hands-on experimentation and iterative learning allow AI tools to evolve in alignment with climate and development goals (Peixoto et al., 2023).

3.2.3.1 Case study: digital tools to safeguard the climate resilience of smallholder farmers

The Digital Farm initiative, funded by the World Bank's Trust Fund for Statistical Capacity Building III, empowered smallholder farmers in East Africa with real-time climate data for climate-smart agriculture (2018–2020). In partnership with the International Centre for Tropical Agriculture (CIAT), Climate Edge, and local cooperatives, the project trained lead farmers and youth agents in using digital tools such as a digital record-keeping app and NEXO weather stations (World Bank).

NEXO weather stations, using IoT M2M SIM cards, trained bioclimatic variables affecting coffee and tea crops. Integrated with satellite and farm-level data, the initiative provided farmer-friendly dashboards for decision-making. Addressing challenges like data literacy and connectivity, the project emphasized human-centered design to ensure usability.

Outcomes included training over 4,000 smallholders, installing 20 weather stations, and integrating climate data into FarmDirect's app, improving profitability and resilience. This success highlights the need for ongoing investment in digital literacy and data accessibility for smallholder farmers.

3.2.4 Social norms and behavior monitoring

The World Bank recognizes big data's role in addressing harmful social norms. The Development Impact Group (DIME) AI, supported by a Bill & Melinda Gates Foundation grant, enhances research and policy in India, Kenya, and Nigeria. This initiative conducts media consumed by Adolescent Girls and Young Women (AGYW), develops monitoring tools, and evaluates

media's influence on social norms. In order to improve accessibility, the group is utilizing machine learning to automate content analysis and monitoring tools. Additionally, policy dialogues will be conducted to maximize the effectiveness of these efforts, particularly within major entertainment hubs (Development Impact Group).

A key component is the development of Hate Speech Detection (HSD) models to combat harmful online content. Traditional HSD models, often trained on U.S. data, struggle with regional dialects, and biased datasets can distort moderation performance. To address this, the World Bank and academic partners developed culturally tailored HSD models, improving accuracy with human oversight. A study found that human reviewers monitoring just 1% of flagged tweets enabled AI to moderate 60% of hateful content effectively.

Understanding social media usage patterns helps policymakers address interethnic violence and societal trends. Leveraging detection models, the World Bank provides real-time policy tools. In Nigeria, a targeted intervention used the best-performing HSD model to identify users sharing hate content and delivered ads with prosocial messages. Early findings show a 15% reduction in hate content among those exposed to the ads, likely due to decreased sharing within their networks (Tonneau et al., 2024).

3.3 World Health Organization (WHO) applications of big data

WHO's Director General, Tedros Adhanom Ghebreyesus, highlights AI's growing role in digital healthcare, particularly in diagnosis, clinical care, drug development, disease monitoring, outbreak management, and overall health system organization (Harnessing Artificial Intelligence for Health) (Taylor, 2023).

WHO's AI strategy is built on three pillars: setting standards, governance, and policies for evidence-based AI in health; fostering pooled investments and global expert collaboration, and implementing sustainable AI programs at the national level (Artificial Intelligence for Health).

In collaboration with the International Telecommunication Union (ITU), WHO established the ITU/WHO Focus Group on "AI for Health" (FG-AI4H) in July 2018. The initiative aims to develop international evaluation standards for AI solutions in healthcare, focusing on machine learning, medicine, regulation, public health, statistics, and ethics. This multidisciplinary effort consolidates national expertise to create universal benchmarks for AI in health.

FG-AI4H operates through iterative, collaborative efforts, with bi-monthly meetings producing deliverables on AI ethics, regulatory best practices, software lifecycle specifications, data management, AI training, evaluation criteria, and AI adoption strategies. Specialized working groups generate documents on specific use cases and guidelines for AI deployment in healthcare settings (Kan et al., 2024).

A key outcome of the FG-AI4H is the development of guidelines and specifications designed for replicability. These guidelines support AI evaluation and contribute to an online platform for benchmarking AI applications. Leveraging global expertise, the Group ensures AI solutions are developed, tested, and implemented with consistency, transparency, and ethical consideration.

Several factors have contributed to FG-AI4H's progress, including active participation from national and regional health regulators, public health agencies, medical professionals, AI developers, and ITU and WHO inter-agency collaborations. However, challenges remain, such as ensuring cross-border coordination and balancing innovation with regulation. Despite these constraints, the Focus Group's collaborative approach and commitment to standardized, replicable AI evaluation frameworks promise to drive sustainable and scalable advancements in healthcare AI worldwide (Figure 3).

3.4 Summary

Recent years have witnessed a significant expansion in the application of big data across international organizations, particularly within the United Nations (UN), the World Bank, and the World Health Organization (WHO). The UN's Global Pulse initiative exemplifies the integration of mobile data, satellite imagery, and social media analytics to enhance humanitarian action, monitor public sentiment during crises, and track progress toward the Sustainable Development Goals (SDGs). At the World Bank, artificial intelligence and big data have been incorporated across sectors, from poverty alleviation programs to climate resilience initiatives and hate speech monitoring. Similarly, the WHO focuses on AI for healthcare and develops global standards for AI applications in healthcare.

While this section highlights the concrete applications of big data in advancing sustainable development, disaster resilience, public health, and global governance, the next section will critically examine the potential challenges and risks associated with integrating big data into global governance frameworks, including privacy concerns, infrastructural and human resource limitations, ethical dilemmas, and the potential misuse of data.

4 Potential challenges and risks in adopting big data for global governance

In the current era of increasing informatization and rapid technological development, the trend of both modern science and industrial development is gradually being driven by data. The utilization of Big Data extends beyond pioneering technology into most market sectors, enhancing overall operational efficacy (Figure 4). For organizations of global repute, the efficacious design and development of integrated Big Data systems is pivotal to the adoption of Big Data for global governance (Demchenko et al., 2024). However, the adoption of big data for governance is not without substantial challenges. This section critically examines the technical, ethical, organizational, and political risks associated with integrating big data into global governance frameworks.
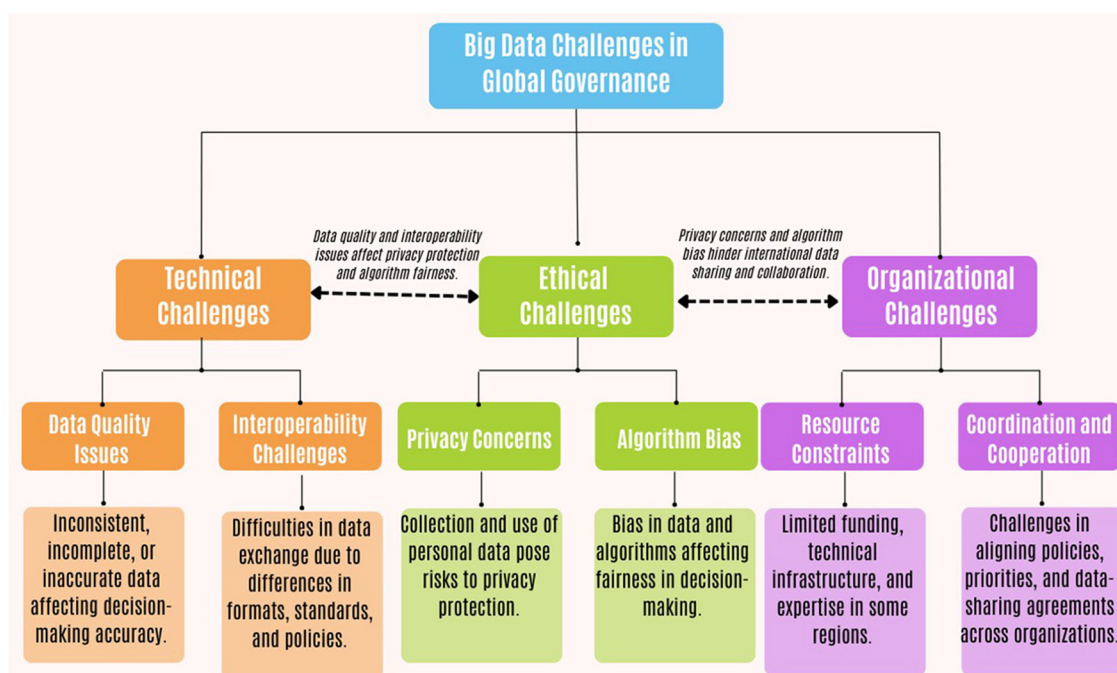
FIGURE 4
This framework visualizes the three core challenge dimensions (technical, ethical, and organizational) that impede big data implementation in global governance.

4.1 Technical challenges

The establishment of a big data platform to serve international organizations, such as the United Nations, is paramount for realizing of the Sustainable Development Goals (SDGs). Such platforms can provide real-time information for monitoring SDG progress and facilitate data sharing between member states (United Nations, 2023). However, new paths often come with challenges, particularly in managing data quality and overcoming interoperability limitations.

4.1.1 Data quality issues

Data quality varies for multiple reasons, with inconsistent and low-quality data impacting decision-making accuracy. According to the Journal of Accountancy, 1–5% of error rates can be attributed to human input, including spelling errors, incompleteness, and duplicate entries. The absence of standardized data management can result in inconsistencies and fragmentation of data sets (Elahi, 2022), as well as the fragmentation of data sets. This is due to the fact that data is stored in isolated data repositories maintained by departments and organizations, which can create data silos and impede the harmonization and sharing of information (Tsidulko, 2024). In addition, the integration of multiple data resources may result in discrepancies due to the presence of different quality standards when merging data from diverse systems or external sources (Batini and Scannapieco, 2016). Concurrently, the inadequate incorporation of third-party data represents a significant contributing factor to the prevalence of data quality concerns within organizational contexts. This is due to the absence

of oversight and control over the fundamental data sources (David, 2022). As technology advances, there is a possibility that older systems may not be compatible with the latest data formats, which could also result in the generation of unclean data. Common data quality issues under international governance are often similar to those faced by most organizations. According to the United Nations Data Quality Assurance Framework (UN DQAF) factsheet (International Monetary Fund, 2003), the issues described above need to be taken seriously.

High quality data must possess both accuracy and validity to reflect real situation accurately. Data integrity is crucial, as missing data can affect decision outcomes and lead to bias. Data consistency must be maintained, especially in cross-database systems. Moreover, timeliness is essential as only recent data can reliably reflect current situations.

4.1.2 Interoperability challenges

Beyond internal data quality, the challenge of interoperability across systems, sectors, and national borders is critical. Cultural, linguistic, and technical differences exacerbate difficulties in establishing standardized governance frameworks (Sargiotis, 2024). Interoperability challenges arise when data systems across organizations, regions, and countries need to communicate, or integrate data. Different countries use varying data formats, including basic units of measurement and date formats, which can lead to incompatibility during data exchange. Overcoming these barriers requires the development of internationally agreed-upon data standards and robust protocols for ensuring compatibility across diverse platforms.

4.2 Ethical issues

As global governance increasingly relies on big data analytics, ethical concerns regarding privacy, transparency, and algorithmic fairness have also come to the forefront.

4.2.1 Privacy concerns

The vast amount of personal information collected through big data systems poses significant privacy challenges. Individuals often have little insight into how their data is used or whether sufficient protections are in place, particularly when data is transferred across jurisdictions with differing legal standards (Zuboff, 2019). Additionally, in the global governance context, data comes from different countries with varying privacy standards and regulations, further complicating efforts to protect personal information.

4.2.2 Algorithm bias

Algorithmic bias remains another significant ethical risk. Complex algorithms based on big data analytics can obscure the calculation process, leading to questions about impartiality. Without transparency, affected populations cannot challenge erroneous or discriminatory outcomes.

Algorithmic bias can be categorized into three types: data source bias, sampling bias, and algorithmic bias. Data source bias occurs when training data is concentrated in certain regions, while sampling bias results from unbalanced category distribution in datasets. Algorithmic bias stems from the design itself and can affect decision-making outcomes. These biases can reinforce social inequities or distort policy interventions.

For example, during the pandemic, the World Health Organization (WHO) used big data to calculate spread rates and predict infection numbers (Rossouw and Greyling, 2024). However, better data availability in Europe and the United States led to algorithmic bias, affecting resource distribution in developing regions (Bayati et al., 2022). Addressing such biases requires deliberate efforts to diversify datasets, audit algorithmic processes, and ensure inclusive model development practices.

4.3 Organizational inefficiencies and resource shortages

Beyond technological and ethical considerations, one of the most significant barriers to integrating big data into global governance lies in organizational inefficiencies and resource shortages.

The successful application of big data requires seamless coordination among a wide range of stakeholders, including international organizations, national governments, and private entities. However, this coordination is often hindered by competing interests, fragmented management systems, and communication barriers between countries (Kuzio et al., 2022).

Additionally, the UN system itself can still be ineffective even after integrating big data into global governance. This is due to longstanding structural issues of the system, including deep-seated divides between the Global North and the Global South, which shape disparities in resources, influence, and data access. Furthermore, overlapping jurisdictions among UN agencies, competing mandates, fragmented accountability structures, and entrenched bureaucratic processes create barriers to effective coordination and decision-making (Weiss, 2016). As mentioned earlier, many UN bodies—such as UNDP and WHO—collect and store data independently, using incompatible formats and systems. This lack of interoperability also limits the ability to conduct integrated analyses. Despite the potential for data-driven innovation, these systemic inefficiencies often hinder the UN's ability to act cohesively and responsively across complex global issues.

Furthermore, establishing big data infrastructure demands substantial upfront investment and specialized human resources. Many developing countries lack the technical infrastructure and data science expertise necessary to implement large-scale data initiatives, often depending on developed countries to supply resources and foster capacity-building partnerships.

Therefore, while big data holds great promise for enhancing global governance, its full integration is currently obstructed by institutional fragmentation, coordination failures, and uneven resource distribution across countries and agencies.

4.4 Misuse of data and risks of surveillance

Lastly, big data is not a neutral technological force, but one that is entangled with political power structures, capable of either increasing authoritarian surveillance or enhancing democratic transparency.

When misused, data streams such as facial recognition, social media activity, geolocation data, and satellite imagery can be harnessed to monitor dissent, track individual mobility, and enforce population control, thereby supporting what scholars call *digital authoritarianism*. Conversely, when strong legal safeguards, institutional transparency, and meaningful public participation are embedded within data governance frameworks, big data can become a powerful tool for the governments to identify underserved populations, better allocate resources, plan infrastructure, inform equitable policy interventions, and monitor sustainable development.

A research conducted by Feldstein (2019) revealed that at least 75 out of 176 countries are actively employing artificial intelligence technologies for surveillance purposes. These include 51% of advanced democracies, 37% of closed autocracies, and 41% of both competitive authoritarian regimes and electoral democracies. While the adoption of AI surveillance tools does not automatically imply the abuse of data, the breadth of deployment across regime types raises serious concerns about the potential for repression, data misuse, and rights violations. These developments underscore the urgent need for clear ethical standards and robust regulatory oversight in global data governance.

4.5 Summary

Technical challenges such as data quality and interoperability, ethical dilemmas around privacy and bias, organizational

inefficiencies, resource disparities, and the threat of political misuse must all be carefully navigated.

Moving forward, the development of digital infrastructure, investment in human capital, the establishment of robust governance frameworks, and a strong commitment to ethical principles are essential to ensuring that big data serves as a tool for inclusive, equitable, and democratic global development.

5 Anticipated future needs for big data in international organizations

Building on the current applications discussed, this section addresses our second research question by identifying the essential technical and organizational infrastructures required for scaling big data systems in international organizations.

5.1 Infrastructure needs

With the exponential growth of real-time data—from satellite imagery and sensor networks to social media and mobile devices—international organizations must anticipate and proactively address emerging needs for infrastructure, human capital, and policy frameworks.

5.1.1 Data storage solutions

Many international organizations, including the UN, depend on cloud storage solutions to manage Big Data. As more data is generated through high-resolution satellite imagery, sensor networks, social media analytics, and mobile data, existing infrastructure may not suffice (Singh and Sidhu, 2023). Future systems must scale to accommodate exabytes of information. Hybrid models involving cloud storage and on-premises infrastructure are likely to become more widespread (Soares, 2012; Aguerre et al., 2024). Technologies like quantum computing and distributed databases will also be crucial for storing vast datasets efficiently.

Furthermore, future data storage systems should feature low-latency access, robust encryption, and advanced compression techniques to ensure quick data processing while maintaining security (Hennen et al., 2023).

5.1.2 Data processing capabilities

The growing demand for real-time analytics, particularly during crises and humanitarian emergencies, calls for significant increases in computing power. Technologies such as Graphics Processing Units (GPUs) and edge computing will be essential to manage the growing volume of data (Singh and Sidhu, 2023). Cloud computing and edge computing solutions will also play key roles in ensuring swift data processing (Legner, 2021).

A reliable data processing system also requires distributed architectures that can operate across multiple regions and handle redundancy. For global organizations that rely on real-time data—such as the UN's SDG monitoring systems—maintaining continuous service is crucial. Building in automatic failover

capabilities and geo-redundancy will ensure data resilience (Soares, 2012).

5.1.3 Cloud solutions

As organizations increasingly migrate to cloud environments, cybersecurity becomes a top priority. Strong encryption, strict access control protocols, and compliance with data privacy frameworks—such as the General Data Protection Regulation (GDPR)—are essential safeguards (United Nations, 2023).

Equally important is ensuring that cloud-based platforms remain accessible to authorized users across borders while ensuring data security. This is crucial for international cooperation and timely decision-making during global crises (Soares, 2012).

However, the transition to cloud solutions presents several challenges, particularly when migrating from legacy systems. Key issues include data migration complexities, ensuring data interoperability, and training staff on new technologies. Organizations must carefully manage the change process while meeting local legal requirements for data protection (Legner, 2021).

5.2 Skills development

As technical infrastructure evolves, parallel investments in human capacity are critical. Bridging the global data literacy gap is essential for the equitable and effective use of big data in international governance.

5.2.1 Data literacy programs

A significant barrier to the effective use of big data is the lack of data literacy among policymakers and staff. Many individuals lack skills to interpret complex data from sources like satellite imagery, and social media analytics. According to a report by the (UN Global Pulse, 2023), this gap hinders the capacity to leverage data-driven insights for achieving the SDGs.

Addressing this requires systematic training needs assessments to identify skill gaps at all levels of the organization. This includes assessing staff capabilities, familiarity with data science tools, and capacity to apply data insights to policy-making. The assessment should consider institutional needs, focusing on types of data typically used and specific sector demands (Hennen et al., 2023).

Data Literacy Programs should focus on both fundamental data skills and advanced analytical capabilities. For entry-level learners, programs should introduce basic concepts in data management, and data visualization tools. For senior professionals, programs should emphasize data-driven decision-making and predictive analytics (Legner, 2021).

5.2.2 Specialized training

Advanced technical roles require training in programming languages such as Python, R, and SQL, as well as skills in machine learning, geospatial analysis, and ethical data use (Soares, 2012).

Course modules should cover foundational data science, advanced analytics, data visualization and communication,

and data governance frameworks, with a strong emphasis on ethics and equity. Training initiatives must be inclusive and scalable, ensuring that personnel in both developed and developing countries have equal access to learning opportunities. Additionally, courses should be designed around real-world case studies and development challenges, equipping professionals with practical tools to apply in the field (Singh and Sidhu, 2023) (Figure 5).

5.3 Policy frameworks

As the scale and sensitivity of data use increase, robust policy frameworks must ensure ethical, secure, and interoperable applications.

5.3.1 Ethical guidelines

Ethical guidelines for Big Data usage are essential to ensure data is collected, processed, and applied responsibly. Key principles include privacy, transparency, data sovereignty, fairness, and beneficence (Aguerre et al., 2024; Soares, 2012).

Organizations should establish ethics committees and develop data ethics code of conduct. Training programs should ensure staff understand ethical standards (Legner, 2021).

Enforcement of ethical standards requires clear sanctions for violations, such as disciplinary actions, and effective whistleblower protections to encourage reporting unethical practices. Regular audits and the use of automated systems to monitor compliance will ensure ethical breaches are detected and addressed promptly (Hennen et al., 2023).

Ethical guidelines must also be regularly reviewed and updated to address emerging challenges in Big Data, such as AI ethics and machine learning transparency. Feedback from stakeholders, including local communities and tech experts, should be incorporated to ensure the guidelines stay relevant and adaptable to new technologies (United Nations, 2023; Soares, 2012).

5.3.2 International standards

Although several global data governance frameworks exist— including the General Data Protection Regulation (GDPR), OECD principles, and ISO/IEC 27001 for information safety— gaps remain, particularly in addressing emerging issues like AI, blockchain, data sovereignty, and cross-border data flows (Aguerre et al., 2024).

The lack of harmonization between different regional frameworks creates challenges for global organizations. A unified global framework is needed to standardize data protection, privacy, and sharing rules. Such a framework would facilitate international cooperation, streamline data-sharing agreements, and enhance trust in global governance systems (Hennen et al., 2023).

Implementing this vision will require overcoming significant political resistance and regulatory fragmentation. Balancing innovation and regulation remains critical for global data governance (Legner, 2021; Soares, 2012).

5.4 Summary

In sum, unlocking the full potential of big data in international governance depends on scalable infrastructure, skilled human resources, sound policy frameworks, and strong ethical safeguards.

The following section will explore the latest emerging applications of big data within international organizations and identifies areas where further development and innovation are needed, particularly following progress in infrastructure, capacity-building, and regulatory alignment.

6 Emerging applications and future developments of big data in international organizations

With the rapid advancement of data science, emerging techniques such as predictive analytics, machine learning, and Natural Language Processing (NLP) are increasingly transforming the landscape of modern monitoring systems in international affairs. These cutting-edge technologies are improving early warning systems and automating the tracking of progress toward the Sustainable Development Goals (SDGs).
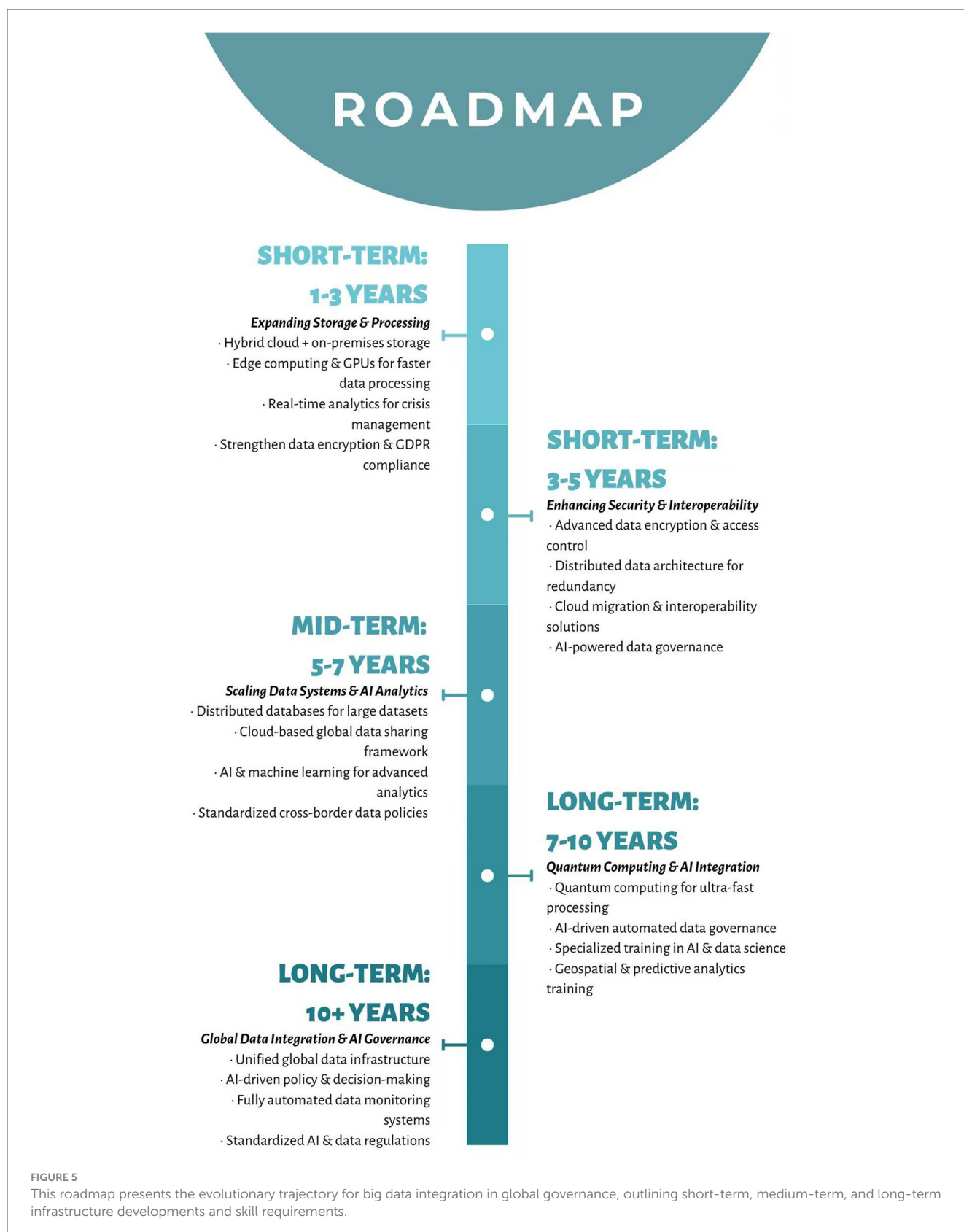
6.1. Predictive analytics in crisis prevention

This section specifically addresses our fourth research question by examining how AI and machine learning technologies are revolutionizing early warning systems for various types of crises.

According to the UN Office for Disaster Risk Reduction, an Early Warning System (EWS) is a comprehensive system that integrates hazard monitoring, forecasting, prediction, disaster risk assessment, communication, and preparedness efforts (UN Office for Disaster Risk Reduction). Traditional EWS often face limitations including reliance on outdated data, and slow analysis processes. However, the incorporation of big data and predictive analytics offers promising solutions.

6.1.1 Food insecurity early warning systems

International organizations such as the UN are increasingly utilizing big data and predictive analytics to enhance their early warning systems. A notable example is the Global Information and Early Warning System on Food and Agriculture (GIEWS) established by the Food and Agriculture Organization (FAO) of the UN in the early 1970 (FAO, 2025). GIEWS serves as a vital tool for monitoring and providing early warnings of potential food crises at national, regional, and global levels. The system integrates conventional data including the "Country Cereal Balance Sheet" database and the "Food Price Monitoring and Analysis (FPMA) Tool," which provides monthly domestic retail and wholesale price data for major food items in 126 countries, as well as weekly or monthly prices for 88 internationally traded foods (FAO, 2025). In addition to conventional data, GIEWS has also incorporated remote sensing technology to gain big earth data on

FIGURE 5
This roadmap presents the evolutionary trajectory for big data integration in global governance, outlining short-term, medium-term, and long-term infrastructure developments and skill requirements.

water availability and vegetation health during cropping seasons. In 2014, GIEWS introduced the Agricultural Stress Index System (ASIS), designed to identify croplands at risk of water deficits or drought. ASIS generates pre-processed, publish-ready maps and zonal statistics every 10 days, delivered through the FAO GIEWS Earth Observation website (FAO, 2025). These timely updates provide critical information that supports early intervention efforts, helping to mitigate the impacts of food insecurity.

Nevertheless, there is potential for improvements. Currently, GIEWS primarily relies on a narrow set of data. For comprehensive understanding, it should expand to include climate data, demographic patterns, income levels and livestock health, pest outbreaks, and data on political stability. For instance, migration patterns, pest invasions, and poverty rates all influence food security. Incorporating this multidimensional data is essential for improving GIEWS' statistical models and predictions.

GIEWS could benefit from increased use of machine learning and predictive analytics. Machine learning algorithms, capable of analyzing vast amounts of complex, multi-dimensional historical data, can uncover patterns and trends that may not be apparent through traditional, labor-intensive methods. With advanced AI support, reports could shift from monthly or 10-day intervals to daily updates, enabling more responsive crisis prevention.

A promising approach is scenario development, which involves analyzing potential future events and their impacts. The Famine Early Warning Systems Network (FEWS NET) established by the United States Agency for International Development in 1985, uses this methodology to project food insecurity in Africa and other vulnerable regions. FEWS NET's approach recognizes that while precise predictions are impossible, scenario development can construct "most likely" outcomes based on available data (FEWS NET, 2018). Research indicates FEWS NET's projections achieve 84% accuracy, though accuracy decreases with higher levels of food insecurity (Backer and Billing, 2021).

Additionally, integrating remote sensing technology with predictive analytics could enable GIEWS to detect food crises at the local community level, rather than just nationally. This precision would allow for more targeted intervention plans.

Through enhanced predictive analytics, statistical models, and big data sources, systems like GIEWS can provide more timely and precise warnings of potential food crises months in advance, helping governments and communities develop proactive response strategies.

6.1.2 Political instability early warning systems

When issuing early warnings for military and political conflicts, the UN primarily relies on expert analysis rather than AI models. The UN Human Rights Office of the High Commissioner (OHCHR) identifies risks of human rights violations, including breaches of economic, social, cultural, civil, and political rights. These violations can serve as both long-term drivers of conflict and immediate triggers for crises (OHCHR, 2025). Similarly, the UN Department of Political and Peacebuilding Affairs (DPPA) provides analysis of global political developments and potential conflicts. Currently, no conflict prevention early warning systems in international organizations utilize machine learning or predictive analytics.

However, predictive analytics holds significant potential to enhance conflict early warning efforts. Although military and political conflicts may appear random, they are often influenced by underlying factors that predictive analytics may reveal. By analyzing socio-economic data and social media activity, these tools can identify patterns and emerging risks, enabling more accurate conflict forecasting.

Some regional organizations have already embraced data-based early warning systems. Joint Research Centre, the European Commission's science and knowledge service, developed the Global Conflict Risk Index (GCRI) in 2014, which assesses conflict risks using 22 variables from social, economic, security, political, geographical, environmental, and demographic dimensions. These 22 quantitative indicators are democracy, state capacity, repression, corruption, recent internal conflicts, years since last conflict, neighboring conflict, homicide rate, female empowerment, ethnic exclusion, transnational ethnic ties, food security, unemployment, GDP per capita, income inequality, trade openness, oil exports, droughts, temperature change, population, youth bulge, and child mortality (European Commission, 2025). By analyzing these variables, the GCRI supports the EU's conflict prevention capacities and decision-making on long-term conflict risks. Similarly, the Violence & Impacts Early-Warning System (VIEWS), led by Uppsala University and Peace Research Institute Oslo, generates monthly conflict forecasts up to 3 years in advance.

Natural Language Processing (NLP) can enhance conflict prevention through real-time social media monitoring. As a branch of AI, NLP enables computers to comprehend, generate, and manipulate human language (Eppright, 2025). It can analyze unstructured data to identify patterns, sentiments, or phrases indicating escalating tensions. For instance, an NLP system can read and process real-time texts, classify each text into one representative real-world category, and assign a citizen satisfaction value to each event (Hodorog et al., 2022). NLP can also analyze social media patterns and trends in hate speech in terms of time, geography, and actors, to understand the potential offline implications and threats of this hate speech [UN Office on Genocide Prevention and the Responsibility to Protect (UNOSAPG), 2024]. The UN High-Level Committee on Management (HLCM) adopted Personal Data Protection and Privacy Principles in 2018. These include only utilizing personal data for a specific permitted purpose; only retaining data for as long as required to serve that purpose; recognizing that deidentification measures such as not recording social media users' account names are often inadequate to fully anonymize data; treating all social media data as personal, among others (UN HLCM, 2018). Additionally, transparency in how data is collected and used, as well as oversight by independent bodies, can help build public trust while maintaining the effectiveness of these systems.

While the UN's reliance on human rights and political affairs experts remains essential, integrating predictive analytics, machine learning, and NLP could strengthen these efforts. Research indicates that predictive models can offer greater forecasting power than traditional indicators of current risks (Perry, 2013). By combining expert insights with advanced analytics, the UN could develop more comprehensive early warning systems to prevent conflicts before they escalate.

6.2 AI and machine learning for SDG monitoring

The UN's 17 Sustainable Development Goals (SDGs) provide a comprehensive global framework to address pressing issues such as

poverty, health care, education, gender equality, economic growth, and climate change. However, monitoring progress toward these goals faces several challenges.

6.2.1 Current challenges in SDG monitoring

One significant challenge is data collection. Since SDG monitoring relies on member states to collect and report data, countries with lower statistical capacity often struggle to gather national data. A 2019 World Bank report revealed that data were only available for "just over 50% of all the indicators and for just 19% of what is needed for comprehensively tracking progress across countries and over time" (Dang and Serajuddin, 2019). The number of countries missing data varies from just four for SDG 15 (Life on Land) to 96 for SDG 13 (Climate Action) (Dang and Serajuddin, 2019).

Time lag between data collection, analysis, and reporting also hampers effective monitoring. Traditional methods rely heavily on manual surveys and census reports, which are often outdated by compilation time. In some cases, even fundamental data such as GDP estimates are reported with significant time delays (Dang and Serajuddin, 2019). Such delay may distort the true picture of progress, leading to inaccurate interpretations of trends and achievements in sustainable development.

Additionally, inconsistent methodologies and incomplete datasets can lead to discrepancies in reporting. Research has shown that the data used to assess SDG progress can stem from various sources, yielding different numbers. Even two household surveys implemented by a national statistical office to collect data on the same employment characteristics may not produce the same statistics (Dang and Serajuddin, 2019). These inconsistencies may also undermine the SDG monitoring system and impede data-driven decision-making.

6.2.2 AI solutions

Emerging technologies like remote sensing, deep learning, and Natural language processing (NLP) offer transformative potential for automating data collection, enhancing analytical precision, and expediting timely reporting.

AI-powered tools can revolutionize data collection by reducing reliance on manual methods. Satellite imagery, drones, and remote sensors equipped with AI can gather real-time data on environmental and human-induced changes. These technologies support SDGs. In agriculture and forestry, satellite imagery can monitor crop yields, assess irrigation needs, and detect deforestation. For climate and environmental monitoring, AI tracks greenhouse gas emissions, temperature fluctuations, and air pollution. In urban planning, satellite imagery identifies land use changes, monitors industrial activity, and oversees infrastructure. Beyond these areas, remote sensing technologies can even be used to infer socioeconomic data. By detecting nightlights, consumption activities, wealth assets, building density, large stadiums, and other important infrastructure, satellites can estimate poverty rates and gauge economic development across regions (Ahn et al., 2023). These methods have proven effective in tracking socioeconomic status in areas such as North and

South America (Google Cloud, 2025a,b), Sub-Saharan Africa (Jean et al., 2016; Ayush et al., 2020), and Southeast Asia (Tingzon et al., 2019; Han et al., 2020). With the power of AI and remote sensing technologies, it will become much easier to collect up-to-date and accurate data across a variety of indicators, where traditional data collection methods often face significant barriers.

Manual object detection from satellite imagery is labor-intensive and costly, whereas deep learning, a subset of machine learning, employs artificial neural networks inspired by the human brain to recognize patterns and detect objects in images (Google Cloud, 2025a,b). Among all deep learning models, Convolutional Neural Networks (CNNs), for instance, has demonstrated exceptional performance in various remote sensing applications, including image segmentation, object detection, and classification (Adegun et al., 2023). CNNs are designed to learn spatial hierarchies of features automatically and adaptively from input images. They consist of multiple layers, each responsible for detecting different features such as edges, textures, or more complex structures. This layered approach enables CNNs to capture intricate patterns in satellite imagery, which is critical for precise analysis. A 2019 study in *Nature*'s *Scientific Reports* demonstrated that CNNs could accurately map vegetation species with 84% accuracy, significantly reducing the time required compared to manual methods (Kattenborn et al., 2019). These models enhance scalability and enable near real-time satellite image analysis while maintaining high accuracy.

Traditional SDG monitoring often struggles to integrate diverse datasets, including economic indicators, health records, satellite imagery, and social media analytics. Machine learning excels in synthesizing these large-scale datasets, offering a cohesive view of development trends.

Additionally, NLP enhances qualitative data analysis—such as national reports, policy documents, and news articles—when monitoring progress toward the 17 SDGs. As introduced before, NLP enables computers to comprehend, generate, and manipulate human language. NLP algorithms categorize information by SDG targets, summarize findings and translate documents into multiple languages and adapt them to various reading levels, ensuring that SDG-related materials from different countries are accessible to a broader audience (Fouch, 2024).

By automating data collection and analysis, AI-powered tools facilitate efficient, accurate SDG monitoring, enabling organizations to process complex datasets and generate near real-time insights.

6.2.3 Implementation framework

To integrate AI into SDG monitoring, the UN must invest in technological infrastructure, including high-performance computing, cloud storage, and reliable connectivity. Open data platforms should promote collaboration among stakeholders. Partnerships with tech companies and research institutions can provide access to advanced AI tools and expertise.

Successful deployment requires financial, human, and institutional resources. Investments are needed for AI software, hardware, and data systems. Training programs should build

technical capacity among UN staff and member states. Institutional reforms may be necessary to embed AI into SDG frameworks.

A phased approach is recommended. Short-term pilot projects can test AI applications in areas such as environmental monitoring and health data analysis. Medium-term goals should scale successful initiatives to other SDG and regions. Long-term objectives should establish an integrated AI-driven SDG monitoring system across all member states.

To measure effectiveness of AI integration, success metrics should assess data coverage, timeliness, and accuracy, as well as the quality of actionable insights. Annual evaluations should refine strategies and ensure continued improvements.

6.3 Conclusion

Emerging techniques in data science, including predictive analytics, machine learning, deep learning, and NLP, are revolutionizing monitoring systems in international affairs. These technologies enhance existing early warning systems, automate SDG progress tracking, and improve data-driven decision-making. By leveraging AI innovations, organizations can continue to improve responses to global challenges and drive sustainable development.

7 Policy and ethical considerations

Big data and data science present significant ethical and legal concerns despite offering substantial opportunities for solving global challenges (Crampton, 2015). Addressing these issues requires global application of big data while adhering to international standards, protecting individual rights, and ensuring fairness. This section examines ethical principles, data ownership and sovereignty, and legal frameworks governing big data worldwide.

7.1 Ethical data usage principles

7.1.1 Transparency and accountability

Transparency is crucial in international big data applications. Every stakeholder must understand the processes of data collection, processing, analysis, and use (Schubert and Barrett, 2024). Transparency fosters trust and minimizes concerns about data misuse or confidentiality breaches (Government Digital Service, 2020). For example, clearly explaining data analysis methods help reduce biases against vulnerable groups.

Accountability mechanisms and governance frameworks are essential for oversight and risk mitigation in organizations using big data (Singhal et al., 2024). Measures such as third-party audits, regular reporting, and data ethics committees help ensure compliance with international ethical standards (Vedder and Naudts, 2017). Accountability should align with global governance protocols, particularly those of the UN and its affiliated institutions.

7.1.2 Informed consent and privacy

Protecting individual privacy is a fundamental ethical principle in big data use. Informed consent ensures individuals understand and agree to how their data will be collected and used (Andreotta, 2024). However, obtaining consent globally is challenging, especially in regions with low digital literacy (Kayaalp, 2018). Strong privacy protections are necessary to prevent data breaches or misuse. The Information Commissioner's Office (2012) recommends anonymization, secure storage, and encryption to safeguard personal data.

For organizations like the UN, privacy considerations must account for cultural perspectives on data protection (Yilma, 2019). Implementing privacy-by-design principles in data systems helps address cultural differences while maintaining ethical integrity in data collection and use.

7.2 Data sovereignty and ownership

7.2.1 Data ownership

Data ownership in global governance is complex. while governments or private entities control data at the national level, international ownership remains ambiguous (Aaronson, 2021). For example, satellite imagery or humanitarian data collected during crises often involves multiple stakeholders, including private companies and NGOs. Clear ownership guidelines are essential to ensure proper data usage, sharing, and security responsibilities.

Collective ownership models treat data as a global public good, particularly for issues like climate change, public health, poverty reduction (Pradeep and Muytenbaeva, 2023). Establishing clear ownership policies helps prevent monopolization by a few entities and ensures equitable data access for international institutions.

7.2.2 Jurisdictional challenges

Data sovereignty presents significant challenges in transnational business and governance. Nations have varying regulations on data collection, storage, and processing (AllahRakha, 2024). For instance, GDPR imposes strict rules on data transfers outside the EU, complicating international cooperation. Such jurisdictional differences can hinder global initiatives, such as tracking Sustainable Development Goals (Obendiek, 2023).

International organizations must advocate for standardized data policies to facilitate cooperation. Implementing treaties that balance national sovereignty with global data flows can reduce regulatory conflicts and improve data accessibility.

7.3 Legal frameworks and international policies

7.3.1 GDPR and other regulations

GDPR sets global benchmarks for data protection and privacy, influencing policies worldwide. It emphasizes principles like data minimization, purpose limitation, and accountability. Compliance with GDPR ensures organizations including the UN, use big

data responsibly while protecting individual rights. Other major regulations include the US Cloud Act and China's Cybersecurity Law (Du, 2018; Schwartz and Peifer, 2019), which address national security and civil rights concerns. This diversity of regulations underscores the need for robust international legal frameworks to govern big data application (Garcia, 2020).

7.3.2 Shaping ethical standards for global leadership

International collaboration is essential for harmonizing big data policies. Organizations like the UN should lead efforts to establish global guidelines on data governance, drawing from frameworks like GDPR (Yilma, 2019; Hassani et al., 2021). These guidelines should cover data transfer, protection, and dispute resolution. Additionally, ethical considerations in emerging technologies like AI and machine learning must be addressed to prevent biases and protect privacy.

7.3.3 Case studies in legal framework application

Numerous cases show how big data can be applied while following ethical and legal standards at all levels. For instance, the WHO utilizes big data and AI in pandemic analysis while adhering to GDPR, ensuring both efficiency and data privacy (WHO Europe Regional Office, 2021). Similarly, the UN Global Pulse initiative promotes data accountability through transparency and governance measures (UN Global Pulse and Pulse Lab Kampala, 2018). Establishing sound policies and governance frameworks enhances ethical big data usage in international organizations.

8 Conclusion and future outlook

8.1 Summary of key insights

This study has systematically examined the transformative role of big data and data science in global governance, addressing five critical research questions about current applications, infrastructure needs, challenges, early warning systems, and policy requirements.

8.2 Answers to research questions

RQ1: How are international organizations currently applying big data?
Our analysis reveals that international organizations are extensively leveraging big data across multiple domains. The UN Global Pulse initiative uses social media analytics, mobile data, and satellite imagery for real-time crisis monitoring and SDG tracking. The World Bank employs AI-powered tools like MALENA and the Novissi cash transfer program to enhance poverty reduction efforts and climate resilience. WHO has established the FG-AI4H to develop global standards for healthcare AI applications. These applications demonstrate significant progress in humanitarian response, sustainable development monitoring, and public health management.

RQ2: What technical and organizational infrastructures are required?
The research identifies three critical infrastructure components: (1) Scalable data storage solutions including hybrid cloud models and quantum computing capabilities; (2) High-performance processing systems with GPU support and edge computing for real-time analytics; and (3) Interoperable platforms that can bridge different data formats and national systems. Organizationally, successful implementation requires cross-sector coordination mechanisms, dedicated data governance units, and sustained financial investment in digital infrastructure.

RQ3: What ethical, legal, and political challenges arise?
The study reveals multiple layers of challenges: Technical issues include data quality inconsistencies and interoperability barriers across jurisdictions. Ethical concerns center on privacy protection, algorithmic bias, and the need for transparent accountability mechanisms. Political risks include the potential for surveillance misuse and digital authoritarianism, with 75 of 176 countries using AI for surveillance purposes. Legal challenges stem from fragmented regulatory frameworks, with GDPR, national data sovereignty laws, and varying privacy standards creating complex compliance requirements.

RQ4: How can big data and AI enhance early warning systems?
Our findings demonstrate substantial potential for AI-enhanced early warning systems. Machine learning algorithms can improve food insecurity predictions by up to 50% through multi-dimensional data analysis. Natural Language Processing enables real-time social media monitoring for conflict prevention, while satellite imagery combined with deep learning provides near-instantaneous disaster impact assessments. However, current systems like GIEWS could benefit from expanded data sources and increased use of predictive analytics for daily rather than monthly updates.

RQ5: What policy reforms and capacity-building strategies are necessary?
The research identifies several critical requirements: (1) Development of unified global data governance frameworks that balance innovation with privacy protection; (2) Comprehensive data literacy programs targeting both technical specialists and policy makers; (3) International standards for cross-border data sharing that respect sovereignty while enabling collaboration; (4) Investment in capacity-building partnerships between developed and developing nations; and (5) Establishment of ethics committees and regular audit mechanisms to ensure responsible data use.

8.3 Future directions and policy recommendations

Based on these findings, we recommend a three-phase implementation strategy:

Short-term (1–2 years):

- Establish interoperability standards for existing systems.
- Launch pilot data literacy programs in key agencies.
- Create cross-organizational data-sharing protocols.

Medium-term (3–5 years):

- Develop unified global data governance frameworks.
- Scale successful pilot programs across regions.
- Implement AI transparency requirements.

Long-term (5+ years):

- Achieve universal access to big data technologies.
- Establish a Global Data Governance Alliance.
- Integrate predictive analytics into all major governance systems.

The success of big data in global governance ultimately depends on balancing technological innovation with ethical responsibility, ensuring that these powerful tools serve humanity's collective interests while respecting individual rights and cultural diversity.

Author contributions

LL: Conceptualization, Formal analysis, Investigation, Validation, Writing – original draft, Writing – review & editing. JW: Conceptualization, Investigation, Software, Writing – original draft, Writing – review & editing. XW: Conceptualization, Software, Writing – original draft, Writing – review & editing. PP: Conceptualization, Formal analysis, Project administration, Writing – original draft, Writing – review & editing. JS: Conceptualization, Formal analysis, Resources, Validation, Writing – original draft, Writing – review & editing. HZ: Visualization, Writing – original draft, Writing – review & editing. ZZ: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Writing – original draft, Writing – review & editing.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

References

Aaronson, S. A. (2021). *Data is Disruptive: How Data Sovereignty is Challenging Data Governance*. Available online at: https://www.wita.org/wp-content/uploads/2021/08/Data-is-disruptive-Hinrich-Foundation-white-paper-Susan-Aaronson-August-2021.pdf (Accessed February 11, 2025).

Adegun, A. A., Viriri, S., and Tapamo, J. R. (2023). Review of deep learning methods for remote sensing satellite images classification: experimental survey and comparative analysis. *J. Big Data* 10, 1–20. doi: 10.1186/s40537-023-00772-x

Aguerre, C., Galperin, H., and Lobato, R. (2024). *Global Digital Data Governance: Polycentric Perspectives*. Boca Raton, FL: Taylor and Francis.

Ahn, D., Yang, J., Cha, M., Yang, H., Kim, J., Park, S., et al. (2023). A human-machine collaborative approach measures economic development using satellite imagery. *Nat. Commun.* 14, 1–10. doi: 10.1038/s41467-023-42122-8

AllahRakha, N. (2024). Rethinking digital borders to address jurisdiction and governance in the global digital economy. *Int. J. Law Policy* 2, 1–8. doi: 10.59022/ijlp.124

Andreotta, A. J. (2024). *Rethinking Informed Consent in the Big Data Age*. Abingdon: Taylor and Francis.

Ayush, K., Uzkent, B., Burke, M., Lobell, D., and Ermon, S. (2020). "Generating interpretable poverty maps using object detection in satellite images," in *Proceedings of the 29th International Joint Conference on Artificial Intelligence* (New York City, NY: ACM), 4410–4416.

Babitha, M. P., and Babu, K. (2016). Enhancing the secure cloud storage in cloud environments. *In. J. Technol. Res. Eng.* 4, 2347–4718.

Backer, D., and Billing, T. (2021). Validating famine early warning systems network projections of food security in Africa, 2009–2020. *Global Food Secur.* 29, 1–10. doi: 10.1016/j.gfs.2021.100510

Bandola-Gill, J., Grek, S., and Tichenor, M. (2022). *Governing the Sustainable Development Goals: Quantification in Global Public Policy*. Cham: Springer. doi: 10.1007/978-3-031-03938-6

Barnett, M. N., Pevehouse, J. C., and Raustiala, K. (2021). "Introduction: the modes of global governance," in *Global Governance in a World of Change* (Cambridge: Cambridge University Press), 1–10.

Batarseh, F. A., and Gopinath, M. (2020). "Panel: Economic Policy and Governance during Pandemics Using AI," in *The 2020 AAAI Fall Symposium Series: Artificial Intelligence in Government and Public Sector, Technical Report FSS-20* (Blacksburg, VA: Commonwealth Cyber Initiative, Virginia Tech), 2–3.

Batini, C., and Scannapieco, M. (2016). "Data quality issues in data integration systems," in *Data and Information Quality* (Berlin: Springer), 279–307.

Bayati, M., Noroozi, R., Ghanbari-Jahromi, M., and Jalali, F. S. (2022). Inequality in the distribution of COVID-19 vaccine: a systematic review. *Int. J. Equity Health* 11, 1–15. doi: 10.1186/s12939-022-01729-x

Castro, D. (2013). *The False Promise of Data Nationalism*. Washington, DC: The Information Technology and Innovation Foundation.

Chen, C., and Wu, R. (2021). The application of artificial intelligence technology in governing international migration and its effects. *Overseas Chin. Hist. Stud.* 2021, 60–69. Available online at: https://qikan.cqvip.com/Qikan/Article/Detail?id=7105427752

Crampton, A. (2015). *Decolonizing Social Work "Best Practices" through a Philosophy of Impermanence*. Social and Cultural Sciences Faculty Research and Publications. 159. Available online at: https://epublications.marquette.edu/socs_fac/159

Cussó, R., and Piguet, L. (2023). "Statistics and quantification," in *International Organizations and Research Methods: An Introduction*, ed. F. Badache, et al. (Ann Arbor, MI: University of Michigan Press), 174–181.

Dang, H.-A. H., and Serajuddin, U. (2019). *Tracking the Sustainable Development Goals: Emerging Measurement Challenges and Further Reflections*. Washington, DC: World Bank Group.

Data Strategy Key Laboratory (2017). Big data concepts and development. *Chin. Scientif. Technol. Terms* 4, 43–50 (Annotation: This Chinese journal article, published by the Data Strategy Laboratory, provides a comprehensive overview of the concept and development of big data, tracing its historical evolution and strategic importance in global contexts. It is relevant to the discussion of big data's historical progression in Section 2.1, 6). doi: 10.3969/j.issn.1673-8578.2017.04.009

David, M. (2022). *Where Do Data Quality Issues Come From? Datafold*. Available online at: https://www.datafold.com/blog/where-do-data-quality-issues-come-from (Accessed February 11, 2025).

Demchenko, Y., Cuadrado-Gallego, J. J., Chertov, O., and Aleksandrova, M. (2024). *Big Data Infrastructure Technologies for Data Analytics: Scaling Data Science Applications for Continuous Growth* (Berlin: Springer).

Dixit, S., and Gill, I. (2024). *AI: The New Wingman of Development*. Washington, D.C.: The World Bank Group.

Du, Y. Y. S. (2018). *The Impact of the GDPR and China's Data Protection Regime Towards Chinese Cloud Service Providers with Regards to Cross-Border Data Transfers*. Tilburg: Tilburg University.

Elahi, E. (2022). *The Impact of Poor Data Quality: Risks, Challenges, and Solutions*. Data Ladder. Available online at: https://dataladder.com/the-impact-of-poor-data-quality-risks-challenges-and-solutions/ (Accessed February 11, 2025).

Eppright, C. (2025). *What Is Natural Language Processing (NLP)?* Oracle. Available online at: https://www.oracle.com/artificial-intelligence/what-is-natural-language-processing/ (Accessed February 11, 2025).

European Commission (2025). *Global Conflict Risk Index*. Available online at: https://drmkc.jrc.ec.europa.eu/initiatives-services/global-conflict-risk-index#documents/1435/list (Accessed February 11, 2025).

FAO (2025). *Global Information and Early Warning System*. Available online at: https://www.fao.org/giews/en/ (Accessed February 11, 2025).

Feldstein, S. (2019). "The global expansion of AI surveillance," in *Carnegie Endowment for International Peace*. Available online at: http://carnegieendowment.org/research/2019/09/the-global-expansion-of-ai-surveillance (Accessed July 30, 2025).

FEWS NET (2018). *Scenario Development for Food Security Early Warning: Guidance Document Number 1*. Famine Early Warning Systems Network. Available online at: https://fews.net/sites/default/files/documents/reports/Guidance_Document_Scenario_Development_2018.pdf (Accessed February 11, 2025).

Firouzi, F., Farahani, B., and Bojnordi, M. N. (2020). "The smart 'things' in IoT," in *Intelligent Internet of Things*, eds. F. Firouzi, K. Chakrabarty, and S. Nassif (Cham: Springer), 51–95.

Fouch, E. W. (2024). *Using Natural Language Processing to Make the United Nations 2030 Connect Platform More Accessible*. Multistakeholder Forum on Science, Technology and Innovation for the SDGs. Available online at: https://sdgs.un.org/tfm/STIForum2024 (Accessed February 11, 2025).

Garcia, M. (2020). Big data applications in global economics. *J. Financ. Stud.* 12, 45–67.

Giest, S. (2017). Big data for policymaking: fad or fasttrack? *Policy Sci.* 50, 367–382. doi: 10.1007/s11077-017-9293-1

Google Cloud (2025a). *What Is Big Data?*. Available online at: https://cloud.google.com/learn/what-is-big-data (Accessed February 11, 2025).

Google Cloud (2025b). *What Is Deep Learning?*. Available online at: https://cloud.google.com/discover/what-is-deep-learning (Accessed February 11, 2025).

Government Digital Service (2020). "Data ethics frameworks," in *Information - Wissenschaft and Praxis (Vol. 72)*. Crown. Available online at: https://assets.publishing.service.gov.uk/media/5f74a4958fa8f5188dad0e99/Data_Ethics_Framework_2020.pdf (Accessed July 30, 2025).

Hafdi, K., (2019). Overview on Internet of Things (IoT) architectures, enabling technologies and challenges. *J. Comput.* 14, 557–570. doi: 10.17706/jcp.14.9.557-570

Han, S., Ahn, D., Cha, H., Yang, J., Park, S., and Cha, M. (2020). Lightweight and robust representation of economic scales from satellite *imagery. AAAI Conf. Artif. Intell.* 34, 5379–5386. doi: 10.1609/aaai.v34i01.5379

Hansen, H. K., and Porter, T. (2017). What do big data do in global governance? *Global Govern.* 23, 2–34. doi: 10.1163/19426720-02301004

Hassani, H., Huang, X., MacFeely, S., and Entezarian, M. R. (2021). Big Data and the United Nations Sustainable Development Goals (UN SDGs) at a glance. *Big Data Cogn. Comput.* 5:28. doi: 10.3390/bdcc5030028

Hennen, L., Hahn, J., Ladikas, M., Lindner, R., Peissl, W., and Est, R. v. (2023). *Technology Assessment in a Globalized World: Facing the Challenges of Transnational Technology Governance*. Cham: Springer.

Hodorog, A., Petri, I., and Rezgui, Y. (2022). Machine learning and natural language processing of social media data for event detection in smart cities. *Sustain. Cities Soc.* 85, 1–10. doi: 10.1016/j.scs.2022.104026

Hu, W., and Hao, F. (2020). Analysis of the application status and future trends of enterprise financial data visualization in the era of big data. *China Market*, 15, 187–195. (Annotation: This Chinese article analyzes data visualization in business, relevant to visualization trends.)

Huang, T., Lan, L., Fang, X., An, P., Min, J., and Wang, F. (2015). Promises and challenges of big data computing in health sciences. *Big Data Res.* 2, 2–11. doi: 10.1016/j.bdr.2015.02.002

Information Commissioner's Office (2012). *Anonymisation: Managing Data Protection Risk ? Code of Practice*, November 2012.

International Monetary Fund (2003). *Data Quality Assessment Framework*. Department of Economic and Social Affairs Statistics. Available online at: https://unstats.un.org/unsd/dnss/docs-nqaf/IMF-dqrs_factsheet.pdf (Accessed February 11, 2025).

Jean, N., Burke, M., Xie, M., Alampay Davis, W. M., Lobell, D. B., and Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science* 353, 790–794. doi: 10.1126/science.aaf7894

Kan, J. Y., Feng, Y. X., Xu, M., Yao, Y. N., Sun, R. D., and Xu, Y. (2024). Decoding environmental impact with image-based CO_2 emission analytics. *Carb. Neutral.* 3:27. doi: 10.1007/s43979-024-00103-w

Kattenborn, T., Eichel, J., and Fassnacht, F. E. (2019). Convolutional neural networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution UAV imagery. *Scientif. Rep.* 9, 1–10. doi: 10.1038/s41598-019-53797-9

Kayaalp, M. (2018). Patient privacy in the era of big data. *Balkan Med. J.* 35, 8–17. doi: 10.4274/balkanmedj.2017.0966

Kirkpatrick, R., and Vacarelu, F. (2018). A decade of leveraging big data for sustainable development. *UN Chronicle* 55, 26–31. doi: 10.18356/5f4fe2e2-en

Kulkarni, P., Joshi, P., and Dey, A. K. (2016). *Big Data Analytics*. Delhi :PHI Learning Pvt. Ltd.

Kuzio, J., Ahmadi, M., Migaud, M. R., Kim, K. C., Wang, Y. F., and Bullock, J. (2022). Building better global data governance. *Data Policy* 8, 1–15. doi: 10.1017/dap.2022.17

Lan, J., and Fan, F. (2002). Design and application of water environment internet of things. *Inform. Comput.* 162, 2–11. (Annotation: This Chinese article discusses IoT applications in water quality monitoring, relevant to governance for ecological management.) doi: 10.1016/j.bdr.2015.02.002

Lefebvre, H., Legner, C., and Fadler, M. (2021). "Data democratization: toward a deeper understanding," in *Proceedings of the Forty-Second International Conference on Information Systems*. Available online at: https://www.researchgate.net/publication/354906721 (Accessed February 11, 2025).

Legner, C. (2021). Data Democratization: Toward a Deeper Understanding. ResearchGate (Accessed February 11, 2025).

Levin, A., Garion, S., Kolodner, E. K., Lorenz, D. H., Barabash, K., Kugler, M., et al. (2019). "AIOps for a cloud object storage service," in *2019 IEEE International Congress on Big Data*, 165–169.

Liang, J. (2017). The application prospects of data visualization in news reporting. *Literat. Art Life* 281, 94–95. doi: 10.3969/j.issn.1005-5312.2017.10.224

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al. (2011). *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute. Available online at: https://archive.org/details/mgi_big_data_full_report/page/n9/mode/2up?q$=$+$beyond$+the+$abilityandview=theater (Accessed February 11, 2025).

Margaret Mary, T., Sangamithra, A., and Gopalakrishnan, R. (2021). "Architecture of IoT and challenges," in *Cases on Edge Computing and Analytics*, 31–54.

Nazir, S., Khan, M. N., Anwar, S., Adnan, A., Asadi, S., Shahzad, S., et al. (2019). Big data visualization in cardiology—a systematic review and future directions. *IEEE Access* 7, 115945–115958. doi: 10.1109/ACCESS.2019.2936133

Nimkar, S. S., and Khanapurkar, M. M. (2021). "Edge computing for IoT: a use case in smart city governance," in *2021 International Conference on Computational Intelligence and Computing Applications (ICCICA)*, 1–5. Available online at: https://www.semanticscholar.org/paper/Edge-Computing-for-IoT:-A-Use-Case-in-Smart-City-Nimkar-Khanapurkar/22d58f0d60a5986dba3d34112f20bdaa9f24c9fe (Accessed February 11, 2025).

Obendiek, A. S. (2023). *Data Governance: Value Orders and Jurisdictional Conflicts*. Oxford: Oxford University Press.

OHCHR (2025). *OHCHR: Prevention and Early Warning. United Nations Human Rights Office of the High Commissioner*. Available online at: http://www.ohchr.org/en/prevention-and-early-warning (Accessed February 11, 2025).

Okahashi, A., and Blanco, C. (2020). *How the World Bank Is Using AI and Machine Learning for Better Governance*. World Bank Blogs. Available online at: https://blogs.worldbank.org/en/governance/how-world-bank-using-ai-and-machine-learning-better-governance (Accessed January 3, 2025).

Okechukwu, O. C. (2022). "Big data visualization tools and techniques," in *Research Anthology on Big Data Analytics, Architectures, and Applications* (Hershey, PA: IGI Global), 465–492.

Park, C., Hong, S., Jung, S., and Lee, D. (2015). "Privacy preserving source based deduplication in cloud storage," in *Conference on Information Security and Cryptology*. Available online at: https://www.semanticscholar.org/paper/Privacy-Preserving-Source-Based-Deduplication-In-Park-Hong/f8d0c3ccf56cf43abf6ff0f00f31eab8c4f1826b (Accessed February 11, 2025).

Peixoto, T. C., Sjoblom, D., and Mustapha, S. (2023). *Leveraging Data and Artificial Intelligence for Climate-Smart Decision Making in Government*. World Bank Blogs. Available online at: https://blogs.worldbank.org/en/governance/leveraging-data-and-artificial-intelligence-climate-smart-decision-making-government (Accessed January 3, 2025).

Perry, C. (2013). Machine learning and conflict prediction: a use case. *Stabil. Int. J. Secur. Dev.* 2:56. doi: 10.5334/sta.cr

Pradeep, A., and Muytenbaeva, Z. (2023). "Leveraging data science to advance the united nations sustainable development goals," in *2023 7th International Multi-Topic ICT Conference (IMTIC)*, 1–6.

Puranik, J., Giri, D., and Dubey, S. (2016). *Security in Data Storage in Cloud Computing*. Available online at: https://www.semanticscholar.org/paper/Security-in-Data-Storage-in-Cloud-Computing-Puranik-Giri/e473fffbb8ee02144855f165cc428e5ed7ba0e9b (Accessed February 11, 2025).

Rawat, H. (2020). Cloud storage drive forensics. *4n6 J.* 2:5. doi: 10.46293/4n6/2020.02.02.05

Rossouw, S., and Greyling, T. (2024). "Big data, big data analytics, and policymaking during a global pandemic," in *Big Data Governance*, ed. J. Kuzio et al. (Berlin: Springer), 109–128.

Saldinger, A. (2023). *Devex Invested: Inside the World Bank's AI Strategy*. Devex. Available online at: https://www.devex.com/news/devex-invested-inside-the-world-bank-s-ai-strategy-107999 (Accessed February 11, 2025).

Sargiotis, D. (2024). "Overcoming challenges in data governance: strategies for success," in *Data Democratization for Business Impact: How to Democratize Data Usage for Greater Business Impact*, ed. D. Sargiotis and P. Ballon (Berlin: Springer), 167–181.

Schubert, K. D., and Barrett, D. (2024). "Data governance, privacy, and ethics," in *Human Privacy in Virtual and Physical Worlds: Multidisciplinary Perspectives*, eds. M. C. Lacity and L. Coon (Switzerland: Springer Nature), 87–110.

Schwartz, P., and Peifer, K-. N. (2019). Data localization under the CLOUD act and the GDPR. *Comput. Law Rev. Int.* 20, 1–10. doi: 10.9785/cri-2019-200102

Sharma, L., Anand, S., Sharma, N., and Routry, S. K. (2021). "Visualization of big data with augmented reality," in *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 928–932.

Singh, R., and Sidhu, H. (2023). *Machine Learning and Sustainable Manufacturing*. Abingdon: Taylor and Francis.

Singhal, A., Neveditsin, N., Tanveer, H., and Mago, V. (2024). Toward fairness, accountability, transparency, and ethics in AI for social media and health care: scoping review. *JMIR Med. Inform.* 12:e50048. doi: 10.2196/50048

Sîrbu, A., Pedreschi, D., Pollacci, L., Pratesi, F., Andrienko, G., Andrienko, N., et al. (2021). Human migration: the big data perspective. *Int. J. Data Sci. Analyt.* 11, 341–360. doi: 10.1007/s41060-020-00213-5

Soares, S. (2012). *Big Data Governance: An Emerging Imperative*. Boise: Mc Press.

Taylor, E. (2023). The role of big data in disease surveillance. *Public Health Analyt. J.* 15, 150–160.

Tingzon, I., Orden, A., Go, K. T., Sy, S., Sekara, V., García-Herranz, M., et al. (2019). Mapping poverty in the philippines using machine learning, satellite imagery, and crowd-sourced geospatial information. *ISPRS* 42, 425–431. doi: 10.5194/isprs-archives-XLII-4-W19-425-2019

Tonneau, M., Castro, P. V. Q. d., Lasri, K., Farouq, I., Orozco-Olvera, V., Fraiberger, S. P., et al. (2024). "NaijaHate: evaluating hate speech detection on nigerian twitter using representative data," in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 9020–9040

Tsidulko, J. (2024). *Break down Silos to Realize the Full Benefits of Your Data*. Oracle. Available online at: https://www.oracle.com/database/data-silos/ (Accessed February 11, 2025).

UN Big Data (2025). *Big Data Methods for SDG Indicators*. Available online at: https://unstats.un.org/bigdata/task-teams/sdgs/indicators.cshtml (Accessed February 11, 2025).

UN Global Pulse (2013). *Information Sheet: United Nations Global Pulse*. United Nations. Available online at: https://www.un.org/millenniumgoals/pdf/GP%20Backgrounder-General2013_Sept2013.pdf (Accessed February 11, 2025).

UN Global Pulse (2023). *UN Global Pulse Annual Report*. United Nations. Available online at: https://www.unglobalpulse.org (Accessed February 11, 2025).

UN Global Pulse (2024). *Secretary-General Describes 'Global Pulse' Initiative as Quick Way to Check Needs of Vulnerable Groups, Enabling Faster, Better Response*. United Nations. Available online at: https://press.un.org/en/2011/sgsm13929.doc.htm (Accessed February 11, 2025).

UN Global Pulse and Pulse Lab Kampala (2018). *Experimenting with Big Data and Artificial Intelligence to Support Peace and Security*. New York, NY: UN Global Pulse.

UN HLCM (2018). *Personal Data Protection and Privacy Principles*. Available online at: https://unsceb.org/sites/default/files/imported_files/UN-Principles-on-Personal-Data-Protection-Privacy-2018_0.pdf (Accessed July 30, 2025).

UN Office on Genocide Prevention and the Responsibility to Protect (UNOSAPG) (2024). *A Comprehensive Methodology for Monitoring Social Media to Address and Counter Online Hate Speech*. Available online at: https://www.un.org/sites/un2.un.org/files/comprehensive_methodology_monitoring_social_media.pdf (Accessed July 30, 2025).

UNDP (2025). *OSDG Initiative Recognized in Top 100 AI Projects for Advancing Sustainable Development Goals*. Available online at: https://www.undp.org/news/osdg-initiative-recognized-top-100-ai-projects-advancing-sustainable-development-goals (Accessed February 11, 2025).

United Nations (2023). *UN Global Pulse Annual Report*. Available online at: https://www.unglobalpulse.org (Accessed July 30, 2025).

Vargas-Solar, G., Barhamgi, M., Risse, T., and Mouhib, I. (2016). "Big continuous data: dealing with velocity by composing event streams," in *Big Data Concepts, Theories, and Applications*, ed. S. Yu and S. Guo (Berlin: Springer International Publishing), 7. Available online at: https://archive.org/details/big-data-collection-pdf/3319277618%20Big%20Data%20Concepts%2C%20Theories%2C%20And%20Applications%20%28Springer%2C%202016%29/page/1/mode/2up (Accessed February 11, 2025).

Vedder, A., and Naudts, L. (2017). Accountability for the use of algorithms in a big data environment. *Int. Rev. Law Comput. Technol.* 31, 1–19. doi: 10.1080/13600869.2017.1298547

Verma, A., Verma, P., Farhaoui, Y., and Lv, Z. (2022). *Emerging Real-World Applications of Internet of Things* (Boca Raton, FL: CRC Press).

Wang, Y., Wang, Y., Sivrikaya, F., Albayrak, S., and Anelli, V. W. (2019). Call for papers: special issue on data science for next-generation recommender systems. *Int. J. Data Sci. Analyt.* 16, 135–145. doi: 10.1007/s41060-023-00404-w

Weiss, T. G. (2016). *What's Wrong with the United Nations and How to Fix It* (3rd Edn.) Hoboken, NJ: John Wiley and Sons.

WHO Europe Regional Office (2021). *The Protection of Personal Data in Health Information Systems- Principles and Processes for Public Health*. Geneva: World Health Organization.

Xie, T., Zhang, M., Wang, Z., Yang, L., and Zhang, D. (2022). "Research on intelligent analysis and visualization of big data based on ELK," in *Conference on Mechatronics and Computer Technology Engineering*.

Yarali, A. (2023). "Artificial intelligence and machine learning in the era of 5G and 6G technology," in *From 5G to 6G: Technologies, Architecture, AI, and Security* (Hoboken, NJ: Wiley-IEEE Press), 65–72.

Yilma, K. M. (2019). The United Nations data privacy system and its limits. *Int. Rev. Law Comput. Technol.* 33, 224–248. doi: 10.1080/13600869.2018.14 26305

Yu, L. (2020). The application of cloud storage technology in electrical automation control. *Secur. Technol.* 15. (Annotation: This Chinese article explores cloud storage applications in electrical automation, relevant to big data infrastructure.)

Zhou, N. (2018). The application of cloud computing and cloud storage in security systems. *Comput. Knowl. Technol.* 51–52, 26–29. (Annotation: This Chinese journal article traces the evolution of big data, providing historical context for its strategic importance.) Available online at: http://m.qikan.cqvip.com/Article/ArticleDetail?id=7108409787

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* New York: PublicAffairs.