Systematic Review

Exploring the Evolution of Big Data Technologies: A Systematic Literature Review of Trends, Challenges, and Future Directions

Tahani Ali Hakami [1], Yasser M. Alginahi [2] and Omar Sabri [3],*

[1] Department of Accounting and Finance, Jazan University, Jazan 82721, Saudi Arabia; thakami@jazanu.edu.sa
[2] Department of Computer Science, Adrian College, Adrian, MI 49221, USA; yalginahi@adrian.edu
[3] Zekelman School of Business & IT, St. Clair College, Windsor, ON N9A 6S4, Canada
* Correspondence: osabri@stclaircollege.ca

Abstract

This study examines the evolution and impact of Big Data technologies across sectors, emphasizing key algorithms, emerging trends, and organizational challenges in their adoption. Special attention is given to ethical concerns related to data privacy, security, and scalability, underscoring the importance of responsible governance frameworks. The review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines to ensure transparency and methodological rigor. A comprehensive literature search identified 83 peer-reviewed articles from high-indexed journals, and a complementary bibliometric analysis of 1108 Scopus-sourced articles (2015–2024) was conducted using R Biblioshiny. This dual-method approach offers both qualitative depth and quantitative insights into major trends, influential sources, and leading countries in Big Data research. Key findings reveal that real-time data processing and AI integration have significantly enhanced data management capabilities, supporting faster and more informed organizational decision-making. This study concludes by highlighting the importance of ethical governance and recommending future research on sector-specific adoption patterns and strategic frameworks that maximize Big Data's value while safeguarding privacy and trust.

Keywords: big data; big data technologies; big data application; big data trends; big data analytics

1. Introduction

In recent years, Big Data technologies have significantly impacted sectors such as healthcare, finance, marketing, and education [1]. The rapid increase in data generation, fueled by vast datasets and sophisticated analytics, has surpassed the capabilities of traditional data management tools [2]. As a result, there has been a rising demand for advanced Big Data Analytics (BDA) frameworks like Hadoop MapReduce and Spark, which are capable of efficiently processing large, diverse, and complex datasets [3]. These frameworks are crucial for key processes such as data storage, preprocessing, and analysis, especially in domains reliant on real-time data processing, AI integration, and edge computing [4]. According to Nagaraj et al. [5], the expansion of Big Data signifies its growing relevance across industries, particularly in sectors like healthcare, finance, and marketing, where it plays a crucial role in enhancing decision-making and strategic planning. Organizations in these sectors are increasingly adopting advanced technologies to facilitate data-driven decision-making, improve management and operations, and develop future strategies [6–10]. Emerging trends such as real-time analytics and industry-specific

applications are reshaping organizational strategies and addressing gaps in the existing literature [7]. This research contributes by identifying how these trends, along with AI integration and edge computing, influence organizational strategies and fill critical gaps in Big Data literature.

Despite the rapid growth of BDA, it is evident that much of the existing literature remains outdated, particularly in addressing the ethical implications of Big Data technologies [8]. Many scholars have pointed out that there is a limited amount of research in areas such as data privacy, security, and governance, which are becoming increasingly critical as organizations increasingly adopt data-driven approaches [5,9–11]. This study, therefore, seeks to address these pressing challenges by comparing traditional and modern frameworks, exploring sector-specific issues, and evaluating the ethical concerns surrounding Big Data technologies. By offering a novel perspective on these areas, this research contributes significant insights into the responsible and effective use of Big Data. Such studies are essential in the field, as they provide crucial guidance to both practitioners and scholars in navigating the complexities of Big Data implementation while also considering the associated risks and ethical concerns. This study not only sheds light on the importance of Big Data frameworks but also emphasizes the pressing need for ongoing research in addressing the ethical implications of their use, too. As Big Data technologies continue to evolve, it is crucial that future studies also focus on finding solutions that can mitigate risks and enhance the positive impact of these technologies across industries.

By providing a comparative analysis of traditional and modern Big Data frameworks, this study offers valuable insights into the evolution of Big Data technologies. In addition, by highlighting key sector-specific trends, it contributes significantly to both academic research and industry practice. The findings underscore the critical importance of emerging technologies in shaping organizational strategies, while also emphasizing the need for responsible data governance. Ultimately, this research enhances understanding of how Big Data can be leveraged to improve operational efficiency while ensuring ethical and effective use across various sectors.

After the Introduction section, the rest of this research will cover the following areas. The literature review will examine key themes and advancements in Big Data technologies. The methodology section will describe the research design and data analysis techniques. Next, the trends in Big Data technologies section will explore recent developments and emerging patterns. The challenges and opportunities in Big Data section will discuss obstacles and potential benefits of Big Data adoption. The social implications section will address the real-world applications and ethical considerations of Big Data. The visual bibliometric analysis section provides a mapping for the global dynamics of Big Data research. The discussion section will interpret the findings and their implications. The research limitations and future directions section will highlight this study's constraints and suggest areas for further research. Finally, the conclusion section will summarize the findings, contributing to the overall understanding of Big Data technologies.

2. Background on Big Data

The terms "big data" and "big data analytics" have varied interpretations, but they are generally understood as referring to data that possess unique characteristics compared to those in traditional data management systems [12]. Big Data refers to vast, complex datasets characterized by their volume, variety, and velocity, which present challenges to traditional tools used for data management [13,14]. In contrast, BDA involves applying advanced techniques to analyze this data, uncovering patterns, trends, and actionable insights [15]. While Big Data focuses on collection and storage, BDA extracts value from that data to inform decision-making. The defining characteristics of Big Data—volume, velocity, variety,

and sometimes veracity—are widely discussed in the literature [16]. These "four Vs" have become a common conceptual foundation for understanding Big Data technologies [17].

Big Data is typically characterized by four elements, often referred to as the 'Four Vs' [13,18,19]:

Volume: Data is produced in volumes that traditional methods can no longer handle, ranging from terabytes to petabytes. By 2020, a significant amount of data was being produced globally, with projections indicating continued exponential growth through 2027.

Velocity: Data is generated at speeds beyond the capabilities of traditional methods for detection and management. With the increase in data production, fast processing is required to keep up with the rapid influx of information.

Variety: Big Data includes a wide array of data types, including structured, semi-structured, and unstructured forms.

Veracity: The reliability and accuracy of data can impact the insights drawn from analysis.

The evolution of BDA technologies has been significant, originating before this decade to handle large, fast-growing datasets [2]. Initially, BDA tools focused on managing data size and access time, but modern applications address all four key priorities—volume, velocity, variety, and veracity—often simultaneously [20]. There are two general technological phases associated with Big Data: the early phase, which focused on handling basic data management problems, and the contemporary phase, where BDA technologies have advanced to handle more complex challenges [19]. Today, BDA technologies are employed across multiple sectors, including IT, medicine, space research, economics, and education [21]. This widespread use has driven the development of new Big Data ecosystems and conceptual foundations that define and support these technologies. According to Arena and Pau [7] and Jagatheesaperumal et al. [22], Big Data technologies aim to achieve the following four tasks: (1) listing technologies by age, whether old or new, (2) outlining various application scenarios, (3) comparing technologies and components between early and current phases, and (4) discussing key challenges from earlier periods and today. So, what is Big Data?

According to Nagaraj et al. [5], existing definitions of Big Data serve as a framework of tools for evaluating emerging datasets in the context of traditional data management systems. Although the concept of Big Data emerged in the late 1990s, it became an industry standard only after 2005 [23]. It encompasses a broad field of data processing, prompting companies to invest heavily in related technologies, including distributed systems and parallel processing. Companies such as Amazon Web Services (AWS) and Microsoft have been at the forefront, with services like Amazon Redshift, Amazon EMR, Azure Synapse Analytics, and Azure HDInsight all designed to efficiently manage and analyze large datasets [24]. Similarly, Google Cloud offers BigQuery and Dataflow, facilitating both batch and stream processing [25]. IBM also recognized the growing importance of Big Data, acquiring a data logging company in 1998 and subsequently launching a product tailored to the Big Data market in the mid-2000s [26].

2.1. Historical Development of Big Data Technologies

The development of Big Data technologies can be categorized into distinct phases, beginning with innovations in database systems at universities which were later adopted by businesses [27]. Early advancements in software for data warehousing, particularly in digital marketing, improved data organization and analysis capabilities. The rise of the internet led to the creation of web analytics technologies, which facilitated more detailed log file analysis and opened new possibilities for data insights [28]. However, the major breakthrough came with the introduction of cloud computing in the late 2000s, which

revolutionized data processing by enabling businesses to store and process vast datasets using scalable, on-demand cloud infrastructures [24].

Distributed file systems like the Hadoop Distributed File System (HDFS) and MapReduce, introduced in the early 2000s, played a pivotal role in managing massive datasets through batch processing, providing scalability and fault tolerance [29–31]. NoSQL databases like MongoDB and Cassandra emerged as solutions to manage unstructured data with flexible data models and horizontal scaling capabilities [32]. By the 2010s, real-time processing gained prominence with technologies such as Apache Kafka and Apache Flink, which enabled stream processing, addressing the latency issues inherent in batch processing [15]. The integration of artificial intelligence (AI) and machine learning (ML) into BDA further expanded the field, making predictive analytics a core component. Technologies like Kubernetes, which allows for scalable orchestration of containerized applications, have become essential in this context [33]. In the 2020s, frameworks like Apache Beam unified batch and stream processing, providing flexibility for both historical and real-time data [7]. Additionally, the advent of Data Mesh, which promotes decentralized data management architectures, has allowed for more specialized data ownership but has introduced new challenges in coordination between systems [34]. Emerging technologies like quantum computing (QC) also promise significant advances in the speed of data processing, although these technologies are still in experimental stages [13]. Table 1 summarizes the technological advancements across different periods, underscoring the evolution of Big Data technologies and their increasing ability to handle growing data demands.

Table 1. Technological advancements and characteristics across eras.

Era	Technologies	Characteristics
2000s	Distributed File Systems (HDFS, GFS) Apache Hadoop NoSQL Databases (MongoDB, Cassandra) Apache Spark	MapReduce, batch processing Scalability, fault tolerance Flexible data models, horizontal scaling In-memory processing, faster analytics
2010s	Apache Kafka Apache Flink ML and AI Kubernetes	Real-time streaming, event processing Stream processing, event time processing Data mining, predictive analytics Container orchestration, scalability
2020s	Apache Beam Kubernetes Operators Data Mesh Quantum Computing	Unified batch and stream processing Automation, manage stateful applications Decentralized data architecture Potential for processing massive datasets

The evolution of Big Data technologies has led to a notable surge in scientific research, as evidenced by the steady increase in publications from 2015 to 2024, as shown in Figure 1. Specifically, the number of articles began at 403 in 2015 and gradually rose to 558 in 2016 and 721 in 2017, indicating the growing academic and industrial interest in Big Data as sectors increasingly adopt data-driven solutions. Notably, by 2018, the number of publications had nearly reached 1000, with 990 articles published that year, and this trend continued to accelerate, peaking at 1701 articles in 2022. This sharp rise aligns with advancements in key areas such as real-time data processing, AI integration, and the broader adoption of Big Data technologies across multiple industries. However, in 2023 and 2024, the growth in publications appeared to stabilize, with 1328 and 1329 articles, respectively. This slight plateau suggests that the field may be reaching a point of maturity, with research now focusing more intensively on specific issues such as scalability, data privacy, and the ethical challenges associated with Big Data.
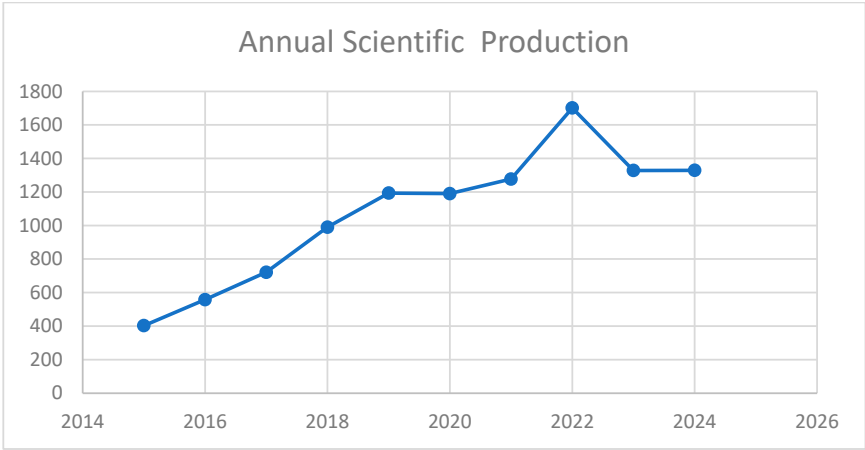
Figure 1. Annual scientific research production in Big Data technologies (2015–2024).

2.2. Key Components of Big Data Ecosystem

Over the past few years, BDA has emerged as a key catalyst for deriving valuable insights that accelerate decision-making and strategy formulation. Several prevalent systems and tools address the efficient data storage and processing needs [5]. Organizations often include numerous storage and data processing systems, such as databases, data warehousing platforms, in-memory databases, and enterprise application tools. However, these isolated and time-stamped views are insufficient for the expansive Big Data landscape [35]. Therefore, a comprehensive approach to storage, processing, analytics, and visualization tools is essential for effectively managing the entirety of BDA systems and leveraging accumulated assets. To handle such voluminous data, the Big Data ecosystem relies on large-capacity storage systems and NoSQL databases [32]. The efficiency of these storage systems is further enhanced by a range of processing frameworks. These frameworks incorporate multiple paradigms, such as batch and real-time processing, using a combination of tools and methods that leverage data science, data mining, and statistical principles [11,36]. In addition to these storage and processing components, data management tools provide the infrastructure for processing large datasets, utilizing various ML algorithms supplemented with programming models [37]. Aggregated operational information can then be visualized using business intelligence (BI) tools, providing valuable insights that enhance decision-making [38]. However, innovative technologies have since evolved in response to the ecosystem's need for robust capabilities to process and analyze large datasets. Table 2 summarizes the key components of the BDA ecosystem and their functions, highlighting the essential systems and tools that organizations can utilize.

Table 2. Key components of BDA ecosystem and their functions.

Component	Description	Examples
Data Storage Systems	Systems that provide large-capacity storage for managing voluminous data.	NoSQL databases, Data lakes
Data Processing Frameworks	Frameworks designed to efficiently process large datasets using various paradigms.	Hadoop, Apache Spark
Processing Paradigms	Methods employed for data processing, including batch and real-time processing.	Batch processing, Stream processing
Data Management Tools	Tools that facilitate the management and analysis of large datasets, including the application of ML algorithms.	ETL tools, data governance tools
Visualization Tools	BI tools that present data insights in accessible formats for decision-making.	Tableau, Power BI, QlikView
Analytics Techniques	Techniques leveraging data science, data mining, and statistical methods for extracting insights from data.	ML, predictive analytics

2.3. Overview of Key Big Data Algorithms and Their Characteristics

Several Big Data algorithms play pivotal roles in processing and analyzing large datasets, each tailored to specific analytical needs. MapReduce excels in managing distributed computing, making it highly effective in processing large datasets; however, it introduces programming complexities [36]. Hadoop offers strong storage and processing capabilities but demands significant resources as well as substantial configuration efforts [37]. On the other hand, Spark enables real-time data processing through in-memory computation, though it is memory-intensive, too [37]. For clustering, K-Means is effective for small datasets but is sensitive to the initial centroids, whereas Density-Based Spatial Clustering of Applications with Noise (DBSCAN) handles clusters of arbitrary shapes but struggles with varying densities. Apriori works for small datasets in market basket analysis but becomes inefficient with large data [39]. Random Forest handles high-dimensional data well, although it lacks interpretability [40], and Gradient Boosting delivers high accuracy, but it risks overfitting [36]. In dimensionality reduction, Principal Component Analysis (PCA) reduces dimensions while retaining variance, yet it assumes linearity, and Naive Bayes is efficient for large datasets but assumes feature independence. Neural Networks and Support Vector Machines (SVMs) excel in complex and high-dimensional tasks, though they demand substantial resources as well [39,41]. Lastly, Logistic Regression provides a simple solution for binary classification, but it is constrained by linearity assumptions, as noted by Liang et al. [39]. Table 3 provides a concise comparison of these algorithms, highlighting their strengths, limitations, and potential applications across various Big Data tasks.

Table 3. Comparative analysis of Big Data algorithms across different perspectives.

Algorithm	Source	Type	Applications	Scalability	Performance	Advantages	Challenges
MapReduce	Ejimofor & Okonkwo [36]	Distributed Computing	Big Data processing	Highly scalable (horizontal scaling)	Efficient for large datasets	Scalable and fault-tolerant	Complexity in programming model
Hadoop	Y. Li & Hei [37]	Distributed Framework	Data storage and processing	High scalability with distributed storage	Handles large volumes of data	Open-source and widely adopted	Requires significant resources
Spark	Y. Li & Hei [37]	Distributed Computing	Real-time data processing	Highly scalable (in-memory)	Faster due to in-memory processing	Supports various data processing tasks	Memory consumption can be high
K-Means	Liang et al. [39]	Clustering	Customer segmentation, image compression	Moderate scalability (depending on implementation)	Efficient for small to medium datasets	Simple and easy to implement	Sensitive to initial centroids
DBSCAN	Liang et al. [39]	Clustering	Spatial data analysis, anomaly detection	Moderate scalability (due to spatial indexing)	Can find arbitrarily shaped clusters	No need to specify number of clusters	Struggles with varying density clusters
Apriori	Liang et al. [39]	Association Rule Learning	Market basket analysis	Low scalability (explosive growth of candidate sets)	Effective for small datasets	Simple and interpretable	Inefficient for large datasets due to combinatorial explosion
Random Forest	Gao et al. [40]	Ensemble Learning	Classification and regression tasks	High scalability (parallelizable)	Robust against overfitting	Handles high dimensionality well	Can be less interpretable than simpler models

Table 3. *Cont.*

Algorithm	Source	Type	Applications	Scalability	Performance	Advantages	Challenges
Gradient Boosting (XGBoost)	Adankon et al. [42]; Liang et al. [39]	Ensemble Learning	Classification, regression, ranking	Highly scalable (with optimizations for distributed data)	High accuracy and efficiency	Handles missing values well	Sensitive to overfitting without tuning
Principal Component Analysis (PCA)	Liang et al. [39]	Dimensionality Reduction	Data visualization, noise reduction	Scalable with optimizations	Reduces dimensionality effectively	Enhances interpretability of data	Assumes linear relationships
Naive Bayes	Liang et al. [39]	Classification	Text classification, spam detection	High scalability (linear)	Fast and efficient	Works well with large datasets	Assumes feature independence
Deep Learning (Neural Networks)	Liang et al. [39]	Machine Learning	Image recognition, natural language processing	High scalability (with GPU and parallel processing)	High accuracy with large datasets	Learns complex patterns	Requires large amounts of data and computational power
Support Vector Machine (SVM)	Adankon et al. [42]	Classification	Image classification, bioinformatics	Low to moderate scalability (depends on dataset size)	Effective in high-dimensional spaces	Robust against overfitting	Memory-intensive for large datasets
Recurrent Neural Networks (RNN)	Liang et al. [39]	Deep Learning	Time series prediction, language modeling	High scalability (with optimizations like LSTM)	Captures temporal dependencies	Suitable for sequential data	Difficult to train and tune
Logistic Regression	Liang et al. [39]	Regression	Binary classification tasks	Highly scalable (linear scalability)	Simple and interpretable	Efficient for binary outcomes	Assumes linear relationship between features

3. Materials and Methods

This SLR was conducted in accordance with the PRISMA 2020 guidelines [43,44]. The review protocol was not registered.

3.1. Search Strategy

A comprehensive search was performed in the Scopus database to identify peer-reviewed journal articles published between 2015 and 2024, following the approach outlined by Gioia et al. [45] and Page et al. [43]. The search terms used included "big data," "big data analytics," and "big data technologies." The following filters were applied to refine the results:

- English language only
- Peer-reviewed journal articles
- Full-text availability
- Subject areas: computer science, engineering, information systems, business, and management
- Publication years: 2015–2024

3.2. Bibliometric Analysis

A bibliometric analysis was conducted using the Biblioshiny package in R, based on an initial dataset of 1108 articles. This analysis offered valuable insights into publication trends, prominent authors and sources, co-authorship networks, and frequently occurring thematic keywords. The complete study selection process is illustrated in the PRISMA 2020

flow diagram (Figure 2), while detailed inclusion and exclusion criteria are outlined in Supplementary Material File S1 [45,46]. Initially, the search yielded 1108 records; however, 752 non-qualifying sources—including conference proceedings, books, editorials, duplicates, and low-quality reviews—were excluded. This left 356 articles for title and abstract screening. Subsequently, 243 articles were removed due to irrelevance or failure to meet the inclusion criteria, and an additional 36 were excluded for being non-English. As a result, 77 full-text articles met the eligibility criteria. To enhance the comprehensiveness of the review, backward citation tracking identified six additional studies, bringing the total to 83 peer-reviewed journal articles as provided in Supplementary Material Table S1. Finally, to further align with the review's objectives, five more relevant studies were manually included, culminating in a final sample of 88 articles.
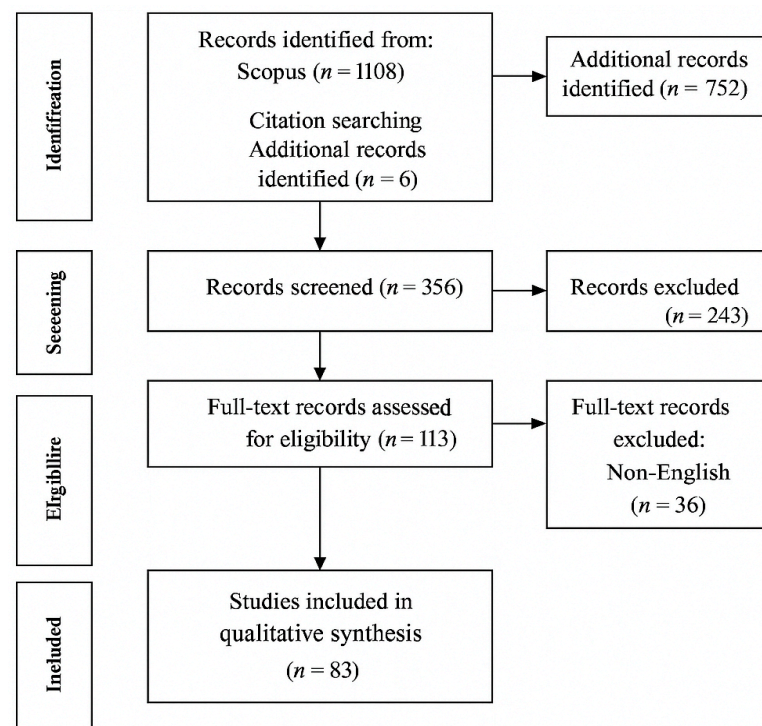

Figure 2. SLR process.

3.3. Study Selection

As suggested by Sarkis-Onofre et al. [47], all records and reports were independently screened by the same two reviewers, who evaluated each title, abstract, and full text against predefined eligibility criteria. Studies were included if they met the following criteria:

- Were peer-reviewed journal articles
- Focused on Big Data technologies or their applications
- Were published in English

Disagreements between reviewers were resolved through discussion and consensus.

3.4. Data Extraction

The same two reviewers independently extracted data from each included study using a structured data-extraction form. Discrepancies in extracted data were resolved by discussion. No automation tools were employed for this step. We extracted data on the following predefined outcomes:

- Adoption of Big Data technologies
- Application domains (e.g., healthcare, finance, logistics)

- Methodological approaches used in the studies
- Challenges reported in the adoption and use of Big Data
- Bibliometric indicators, such as citation counts

All results related to these outcome domains were collected across all studies, regardless of time point, measurement tool, or analysis type, ensuring a comprehensive synthesis. No additional variables—such as participant characteristics or funding sources—were collected, and no assumptions were made regarding missing or unclear data in those areas.

3.5. Risk of Bias and Confidence in Findings

Risk of bias was independently assessed for each included study by the same two reviewers. Discrepancies were resolved through discussion to ensure consistency. Formal tools such as GRADE were not applied, as this review focused on non-clinical studies. Instead, confidence in the findings was assessed conceptually, based on thematic saturation across studies, consistency in reported results, and bibliometric indicators such as citation frequency and co-authorship patterns [48].

4. Trends in Big Data Technologies

This section outlines key trends in Big Data, including real-time processing, AI, ML, edge computing, and cloud-based solutions. It also addresses data privacy, security, governance, and the rise of data democratization. Applications in healthcare, finance, smart cities, education, and marketing are discussed, each with unique challenges and opportunities. Table 4 summarizes these trends, and the top 10 sources will be listed at the end. The increasing volume and variety of data have made real-time processing essential in the Big Data landscape, enabling businesses to gain immediate insights and enhance operational efficiency [1]. Systems must now analyze data as it is generated, allowing for timely decision-making. High-velocity data, such as that from sensor networks, can predict equipment failures, thus providing businesses with critical insights [35]. Furthermore, ML and AI have evolved significantly, allowing for large-scale data analysis without task-specific programming, thus revolutionizing sectors like e-commerce and healthcare [49]. However, ethical concerns about data bias must also be addressed [50].

In addition, edge computing has gained traction, processing data closer to its source to reduce latency and improve bandwidth efficiency, particularly for Internet of Things (IoT) applications [51]. This trend complements cloud infrastructures and enhances data privacy and security [52]. Moreover, cloud-based solutions offer scalable platforms for efficient data management, as exemplified by Netflix's use of AWS for personalized recommendations [24]. Despite lowering barriers for businesses, these solutions raise significant security concerns, necessitating robust measures to safeguard data. Consequently, data privacy and security remain critical as organizations handle increasing volumes of sensitive information. Compliance with regulations like General Data Protection Regulation (GDPR) and implementing encryption and access controls are essential [52]. However, stringent security measures can also strain operational budgets [53]. Additionally, data lakes and data warehouses are key components of data management, offering flexible storage for unstructured data and optimized querying for structured data, respectively [54]. Nevertheless, effective governance is required to prevent data lakes from becoming disorganized [55]. Furthermore, data governance ensures integrity, quality, and security, while ethical guidelines prevent bias in AI implementations [41]. While these frameworks protect users, they may slow innovation due to regulatory pressures [56]. Finally, data democratization empowers non-technical users to make data-driven decisions, enhancing innovation [57]. Moreover, organizations must address security risks and misinterpretation through training and stringent measures [58].

Table 4. Emerging trends in Big Data technologies.

| Trend	Author/Year	Description	Example	Implication
Real-Time Data Processing	Dubuc et al. [15]; Jabbar et al. [7]; Mir [59]	Real-time data processing allows organizations to analyze and act on data instantly, as it is generated, improving operational efficiency and decision-making.	Apache Kafka and Flink used in banking for fraud detection.	Increases demand for infrastructure that can handle high-speed data, impacting industries like finance and healthcare.
AI and ML Integration	Kumar & Singh [21]	The integration of AI and ML into Big Data systems automates the analysis of large datasets, offering predictive insights and advanced decision-making capabilities.	IBM Watson in healthcare for disease diagnosis based on large datasets.	Enhances decision-making across industries but raises concerns about bias and the need for ethical guidelines in AI use.
Edge Computing	Rathore et al. [60]; Hamdan et al. [61]	Edge computing processes data closer to its source, reducing latency and improving bandwidth efficiency, especially in IoT applications and real-time systems.	Edge computing in autonomous vehicles for real-time decision-making.	Enables the growth of IoT applications, but challenges include security and infrastructure upgrades.
Cloud-Based Big Data Solutions	Tuli et al. [62]	Cloud solutions provide scalable and flexible platforms for Big Data processing, reducing the need for on-premise hardware and allowing organizations to scale on-demand.	Netflix using AWS for user preference analysis and content recommendations.	Accelerates the adoption of Big Data by reducing entry barriers for organizations, though security concerns remain significant.
Data Privacy and Security	Bansal et al. [63]; Yang et al. [64]	Ensuring data privacy and security is critical as the volume of data grows. Solutions include encryption, access controls, and compliance with regulations like GDPR.	Financial institutions using homomorphic encryption for secure computations on encrypted data.	Stricter regulations push industries to adopt stronger security measures, but compliance costs can be burdensome for businesses.
Data Lakes and Data Warehouses	Nambiar & Mundra [54]; Saddad et al. [17]	Data lakes store large volumes of raw, unstructured data, while data warehouses are optimized for querying structured data. Organizations often use both for comprehensive analytics.	Amazon S3 (data lake) and Redshift (data warehouse) for storing and analyzing different types of data.	A hybrid approach allows for more comprehensive analytics but requires careful data management to avoid inefficiencies.
Data Governance and Ethics	Kroll [56], Micheli et al. [41]	Data governance ensures the integrity, quality, and security of data across its lifecycle, while ethical guidelines prevent bias and ensure fairness in data and AI models.	Microsoft's AI ethics board ensuring fairness in AI systems.	Drives the need for transparent and responsible use of Big Data and AI but may slow innovation in highly regulated sectors.
Data Democratization	Wang et al. [57]	Data democratization enables non-technical users to access and analyze data, fostering innovation by empowering employees across an organization to make data-driven decisions.	Tableau and Power BI empowering business users to create their own reports.	Facilitates innovation by enabling widespread access to data but requires robust training and security measures to avoid misuse.

Recent research highlights the evolving landscape of Big Data, with a strong focus on privacy, security, and democratization. Strang and Sun [65] examine the state of pri-

vacy and security within the Big Data paradigm, emphasizing ongoing concerns about data protection amid rapid technological progress. Lefebvre et al. [66] contribute to this discussion by exploring data democratization, aiming to deepen our understanding of how accessible and usable data is becoming at various organizational levels. Building on this, Samarasinghe, Lokuge, and Snell [67] investigate the core principles of data democratization, shedding light on its potential to empower decision-making through wider data access. In the healthcare sector, Arshad et al. [68] analyze current trends and challenges in applying Big Data intelligence to enable transformative outcomes, particularly highlighting the need for secure and scalable solutions. Additionally, Big Data technologies are making a transformative impact across five vital sectors, including but not limited to healthcare, finance, smart cities, education, and marketing. Each sector uniquely utilizes Big Data, facing distinct challenges and opportunities as well as various implications, as shown in Table 5. Overall, these studies underline a shared priority: leveraging Big Data effectively while protecting ethical and privacy standards.

Table 5. Transformative impact of Big Data technologies across sectors.

Sector	Sources	Applications	Description	Challenges	Opportunities	Implications
Healthcare	Gomes et al. [6]; Buck et al. [69]	Predictive analytics, personalized medicine, disease surveillance, electronic health records (EHRs), telemedicine and remote monitoring	Integration of technology to improve patient care and streamline processes.	Data privacy, resistance to change, high costs of implementation.	Enhanced patient engagement, improved access to care, cost reduction.	Improved health outcomes, increased efficiency in healthcare delivery.
Financial	Agustí & Orta-Pérez [70]; Nneka Adaobi Ochuba et al. [71]; Rani et al. [72]; Karim et al. [73]	Fraud detection and risk management, algorithmic trading, customer segmentation and targeting, mobile banking, blockchain technology, and robo-advisors	Use of technology to enhance financial services and customer experience.	Cybersecurity threats, regulatory compliance, technology adoption.	Increased financial inclusion, reduced transaction costs, innovation.	Greater economic stability, improved access to financial services.
Smart Cities	Chang, [50]; Waterson et al. [74]	Traffic management and optimization, energy management, public safety and emergency response, IoT applications, smart transportation, and energy management systems	Integration of technology to enhance urban living and sustainability.	Infrastructure costs, data management, public acceptance.	Improved urban planning, enhanced quality of life for residents.	Sustainable urban development, increased efficiency in resource management.
Education	Ang et al. [28]; Hamad [75]; Ikegwu et al. [76]	Learning analytics, institutional performance assessment, online learning platforms, e-learning platforms, AI tutors, and virtual classrooms	Technology-enhanced learning environments to improve educational outcomes.	Digital divide, resistance from traditional educators, funding issues.	Broader access to education, personalized learning experiences.	Improved educational attainment, workforce readiness.
Marketing	Jabbar et al. [1]; Tran et al. [77]	Consumer behavior analysis, social media analytics, market segmentation, digital marketing, social media analytics, personalized advertising, and customer segmentation	Leveraging data analytics to target consumers effectively.	Data privacy concerns, rapid technological changes, market saturation.	Enhanced customer engagement, improved ROI on marketing campaigns.	Shift in consumer behavior, increased competition among brands.

Figure 3 presents the top 10 journals in Big Data research, ranked by the number of published articles. As illustrated, IEEE Access leads the list with a significant total of 551 articles, reflecting its prominence in the field. Following closely, the journal Applied Mathematics and Nonlinear Sciences contributes 339 articles, further emphasizing the journal's relevance in Big Data research. Additionally, other key contributors include the journal Computational Intelligence and Neuroscience, with 155 articles, and the Journal of Big Data, which has published 148 articles. Furthermore, journals such as Wireless Communications and Mobile Computing (141 articles), Future Generation Computer Systems (133 articles), and Multimedia Tools and Applications (127 articles) also showcase the expansive reach of Big Data across various fields. The journal Neurocomputing (125 articles), The Journal of Supercomputing (121 articles), and the journal Mobile Information Systems (118 articles) continue to highlight the interdisciplinary nature of this rapidly evolving research area. This diversity in publication sources underscores the broad and growing applications of Big Data across different scientific disciplines and industries.
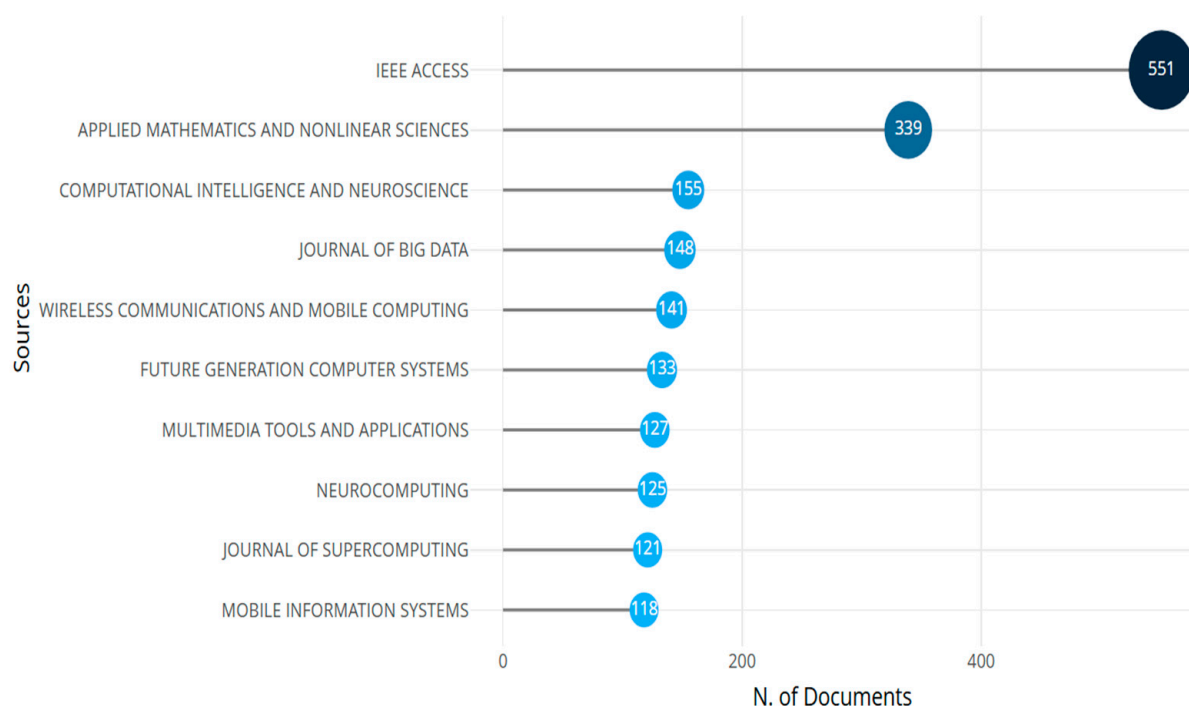
Figure 3. Top 10 journals in Big Data Research by article count (2015–2024).

5. Challenges and Opportunities in Big Data

This section addresses key aspects of Big Data management, focusing on three areas: data privacy and security, which covers concerns related to sensitive information and the necessary guidelines; ethical considerations, which examines privacy rights, consent, and potential discrimination; and scalability issues, which highlights the need for adaptable systems and cloud solutions to handle large datasets. As data usage evolves, it presents challenges such as security risks, ethical dilemmas, and scalability concerns while also providing opportunities for innovation, as shown in Table 6.

Table 6. Challenges and opportunities in Big Data management.

Aspect	Challenges	Opportunities
Data Privacy and Security	- Risk of data breaches due to sensitive information. - Need for stringent legal guidelines for data management. - Complexity in ensuring data integrity and protection against unauthorized access.	- Potential for developing privacyaware and ethically compliant solutions. - Opportunity to enhance data security measures.
Ethical Considerations	- Issues surrounding privacy rights and ownership of personal data. - Debate over adequacy of consent, especially for sensitive data. - Risk of discrimination through predictive analytics.	- Chance to prioritize ethical guidelines and build public trust. - Opportunity to adopt transparent practices in data strategies.
Scalability Issues	- Exponential data growth requires adaptable processing systems. - Traditional databases struggle with oversized datasets. - Need for effective management of extensive data centers.	- Opportunities for implementing distributed computing and cloud solutions. - Ability to focus on data access and distribution without physical infrastructure constraints.

5.1. Data Privacy and Security

Data privacy and security are critical concerns in Big Data, with the vast amounts of sensitive information collected increasing the risk of data breaches [71]. Legal regulations impose strict guidelines on data collection, usage, and management, prompting public discussions about the potential misuses of Big Data [78]. Organizations must integrate privacy-aware and ethically compliant solutions into their data management processes. Security challenges include maintaining data confidentiality, integrity, and availability [64]. As organizations gather extensive data from various stakeholders, they must monitor and manage these sources to uphold data integrity [79]. Effective strategies such as encryption, key management, and robust data governance are essential to protect Big Data from unauthorized access and ensure ethical consent for data usage.

5.2. Ethical Considerations

The ethical implications of Big Data technologies are significant, particularly regarding individuals' rights to privacy and ownership of personal data [79]. The debate over implied versus explicit consent, especially for sensitive information like genetic data, highlights the need for ethical guidelines [64]. Additionally, profiling individuals raises concerns about discrimination and unfair treatment based on predictive analytics [38]. Hence, organizations must prioritize transparency and accountability in their data strategies to maintain public trust [2]. Ignoring ethical obligations can lead to legal consequences and reputational damage, making "big data ethics" essential for fostering trust in decision-making [80].

5.3. Scalability Issues

Scalability is essential for Big Data systems, as the rapid growth of data requires adaptable processing capabilities [38]. Traditional databases often struggle with oversized datasets, prompting organizations to adopt scalable systems that can integrate new nodes

without sacrificing performance [28]. Distributed computing models are increasingly utilized for the parallel processing of large data volumes [35]. Cloud solutions are also gaining traction, allowing organizations to lease infrastructure and shift the burden of managing extensive data centers to vendors [81]. While scalability poses challenges, it also presents opportunities for innovation in data management, privacy, security, and ethics. By proactively addressing these issues, organizations can leverage Big Data for informed decision-making and competitive advantages. Table 6 summarizes the associated challenges and opportunities in Big Data management.

6. Societal Implications

The top 10 topics in Big Data, as highlighted in Table 7, show a diverse range of impacts. Big Data itself, with 6334 occurrences (17%), plays a central role as industries increasingly adopt data-driven strategies. This shift opens significant opportunities in data-driven governance and smart cities, enabling the optimization of public transport, carbon footprint reduction, and enhanced healthcare protocols [78]. Key advancements such as data mining (2039 occurrences, 5%), learning algorithms (1531 occurrences, 4%), and ML (1523 occurrences, 4%) are advancing decision-making processes, particularly in education and social networks [52]. Furthermore, deep learning (1503 occurrences, 4%) is set to revolutionize AI and autonomous systems by interpreting unstructured data. As datasets grow, clustering algorithms (1323 occurrences, 4%) and learning systems (1322 occurrences, 4%) will be essential for pattern recognition and continuous model adaptation. The efficiency of algorithms (1160 occurrences, 3%) and classification (1128 occurrences, 3%) will remain critical for managing large volumes of data in sectors such as healthcare and finance. Additionally, data handling (1078 occurrences, 3%) will continue evolving to ensure scalability, security, and real-time data processing. Hence, the continued development of these areas will be crucial to meet the demands of future Big Data applications.

Table 7. Key themes and their impact in the Big Data landscape.

	Terms	Frequency	Percentage
1	Big Data	6334	17.09%
2	data mining	2039	5.50%
3	learning algorithms	1531	4.13%
4	machine learning	1523	4.11%
5	deep learning	1503	4.05%
6	clustering algorithms	1323	3.57%
7	learning systems	1322	3.57%
8	algorithm	1160	3.13%
9	classification (of information)	1128	3.04%
10	data handling	1078	2.91%

However, while technological advancements are vital, societal factors must also be considered. Regulatory frameworks need to address ethical issues like data privacy, as emphasized by [58]. For instance, Europe's increased investment in public data governance highlights the need for public consensus and informed decision-making [80]. Public literacy on data privacy rights will be crucial, as well as ensuring that data power does not fall into the hands of a few private tech companies [6]. Hence, the balance between data governance and societal trust in Big Data is pivotal [22]. Ensuring strong regulations will help mitigate risks from tech monopolies and protect privacy, fostering public trust in AI-driven solutions. Empowering citizens to voice their opinions is vital, as is preventing private tech monopolies from controlling public discourse [6]. Redistributing data power might involve creating public monopolies or non-profit organizations [41]. This connection

between data governance and societal trust in Big Data-driven AI is crucial [22]. Allowing digital companies to retain vast data and influence societal impacts without strong regulations poses significant risks. Ultimately, the future presents two paths: one focused on technical advancements in data protection and the other on broader societal aspects of Big Data governance. Both paths carry potential benefits and challenges, especially concerning societal risks.

This paper highlights how organizations can strategically adopt Big Data technologies to drive innovation, optimize operations, and address market shifts. It emphasizes the importance of addressing key ethical concerns, such as data privacy, security, and scalability, through robust governance frameworks, enabling businesses to enhance decision-making and maintain a competitive edge. The findings also underscore the social implications of Big Data adoption, particularly in improving data management and responsiveness to societal needs. Responsible data governance promotes public trust and supports positive outcomes in sectors such as healthcare, education, and public services. Ultimately, the integration of Big Data technologies can enhance organizational effectiveness and contribute to societal well-being.

7. Mapping the Global Dynamics of Big Data Research: A Visual Bibliometric Analysis

This section presents a visual and thematic extension of the bibliometric analysis of Big Data technologies, leveraging advanced visualization tools to capture not only the volume and distribution of research output, but also the evolving intellectual, geographical, and institutional dynamics shaping the field.

Figure 4 offers a compelling visual summary of the global landscape of Big Data technologies research, mapping the relationship between research themes (author keywords), contributing countries, and publication outlets. On the right, we see familiar terms like "big data," "machine learning," and "deep learning" dominating the field; this is unsurprising, as these are the engines driving much of the innovation and scholarly interest in the area. These keywords represent the foundational concepts that most researchers are engaging with, forming the thematic backbone of the domain. At the left of the diagram are the countries making the largest contributions, with China, India, and the United States standing out as the primary research powerhouses. Their central positions and multiple connections suggest not only high publication volume but also a broad engagement across various subtopics. China, in particular, shows a strong association with "big data," much of which is channeled into publications like *IEEE Access*, visible on the right column of the plot. This aligns with earlier observations from the bibliometric data pointing to China's dominant citation count and global influence in the field. The center of the figure highlights the preferred journals for publishing Big Data research, with *IEEE Access*, *Applied Mathematics and Nonlinear Sciences*, and *Neurocomputing* being the most prominent. These outlets appear repeatedly across different themes and countries, emphasizing their role as central platforms for disseminating advancements in the field.

What makes Figure 4 particularly insightful is its ability to show how research priorities vary across regions. For example, while all three leading countries are active in core themes, the volume and journal preferences may reflect differences in institutional focus, funding structures, or even national strategies around AI and data innovation. Moreover, the visualization underscores the global and collaborative nature of Big Data research; despite regional differences, there is a clear convergence around certain journals and concepts.
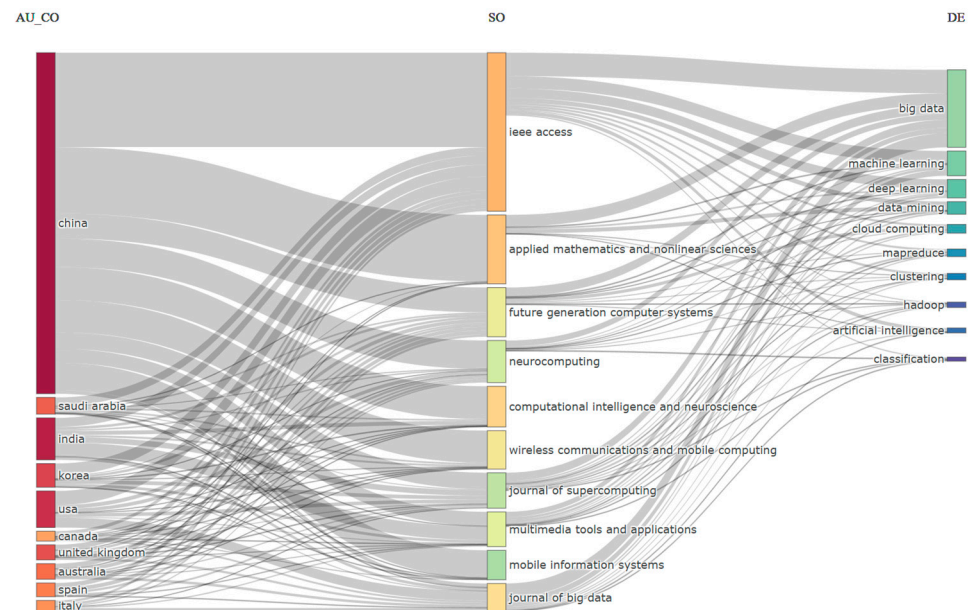
**Figure 4.** Research publications flow from countries to journals and research domains.

This three-field plot does not just map data; it tells a story of how different regions are shaping the field, what topics they are prioritizing, and where they are choosing to share their findings. It is a useful lens through which to understand both the geopolitical and intellectual dynamics of the Big Data research ecosystem.

A longitudinal view of cumulative publication trends across major academic sources in the field of Big Data technologies from 2015 to 2024 is presented in Figure 5. One of the most noticeable patterns is the consistent and dominant performance of *IEEE Access*, which steadily increases its publication output year after year, ultimately crossing 400 cumulative publications by 2024. This linear and sustained growth highlights its role as a central hub for disseminating research in Big Data and related technologies. The *Journal of Big Data* follows a slower, more measured trajectory, reaching around 150 publications by the end of the period, indicating a steady but more specialized role within the research ecosystem. The figure shows a sudden surge in contributions to *Applied Mathematics and Nonlinear Sciences* beginning around 2022. From nearly negligible levels, its publication count shoots up sharply, climbing past 300 publications by 2024; this is a clear signal of the journal's rapid rise in relevance, possibly due to increasing interest in mathematical and algorithmic approaches to Big Data problems. Meanwhile, both *Computational Intelligence and Neuroscience* and *Wireless Communications and Mobile Computing* show more modest publication trends, each converging near 150 cumulative outputs by 2024. The latter shows a gentle uptick in the 2020–2022 window, possibly reflecting a brief period of increased interest in wireless and mobile applications for Big Data.

Trends in publication volume reveal shifting research priorities and the rise of new platforms. While some journals remain consistently influential, others—like Applied Mathematics and Nonlinear Sciences—are gaining traction, reflecting a growing interest in interdisciplinary, data-driven research. Regional variations in output and journal preference hint at differing national strategies and institutional focuses. Still, the overall convergence around key themes highlights the global and collaborative nature of Big Data research.

The cumulative number of published articles by key countries in the domain of Big Data technologies between 2015 and 2024 is provided in Figure 6, revealing distinct national trajectories and underlying shifts in global research leadership. China clearly dominates the landscape, showing an exceptionally steep and sustained growth curve. While the dataset

contains a total of 1108 documents, the country-level frequencies appear larger because they are calculated from author affiliations; a single article may be counted for multiple countries and, in some cases, multiple times, for the same country if several co-authors share that affiliation. Within this framework, China's output approaches 10,000 counts by 2024, underscoring its strategic emphasis on data-intensive research and signaling its preeminent position in the global knowledge economy. The United States, traditionally a leader in scientific output, also exhibits a steady rise, reaching around 2500 counts by 2023 and continuing gradually thereafter. India's trajectory is particularly noteworthy; from 2018 onward it accelerates sharply, surpassing the U.S. around 2023 and exceeding 2500 counts by 2024. This shift reflects the growing role of emerging economies in driving global scientific progress. In comparison, the United Kingdom and South Korea display more modest trends, with the UK surpassing 500 counts by 2024 and South Korea remaining below 250. Although smaller in scale, these figures still highlight consistent engagement and contribution to the field. We note that countries with larger research populations naturally produce more papers; therefore, these cumulative counts reflect both total research activity and the underlying author base, and further normalized analyses could provide additional insights into per-capita or efficiency-adjusted contributions.
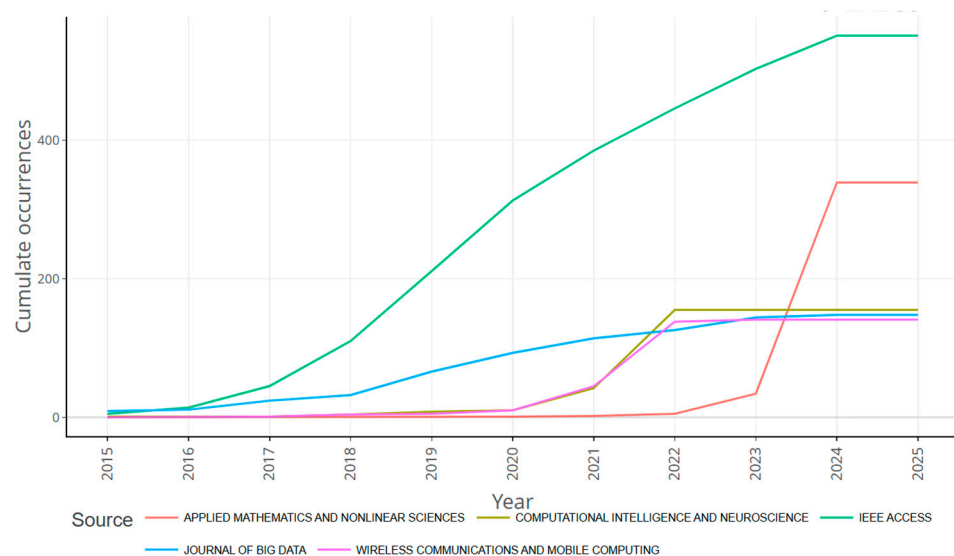
Figure 5. Sources' production over time.
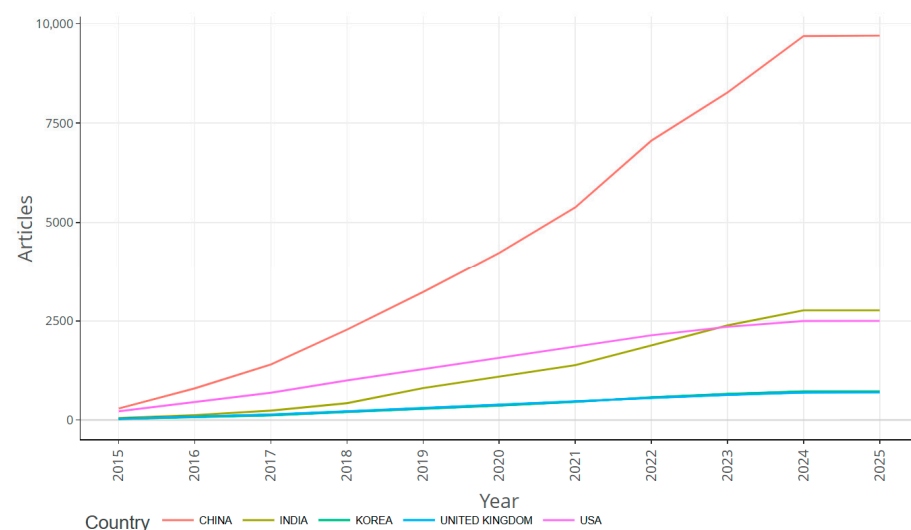
Figure 6. Countries' production over time.

Collectively, Figure 6 highlights the dynamic and evolving geography of Big Data research, where China's explosive growth and India's rapid rise are reshaping the global academic landscape. It also reflects broader geopolitical and policy-driven influences on research funding, institutional priorities, and technological development, particularly in the areas of AI, data science, and analytics.

An analysis of the most-cited countries reveals important trends in the global research landscape on cloud-based technologies and internal auditing. As presented in Table 8, China stands out with the highest total citations (72,962), reflecting its dominant research output in the field. However, the relatively modest average citations per article (17.30) suggest that while China produces a large volume of publications, the impact of individual studies may be less pronounced. In contrast, the United Kingdom demonstrates the highest average article citation rate (54.40), indicating a strong emphasis on research quality and influence despite a smaller overall output. The United States exhibits a balanced profile, with a high total citation count (30,653) and an average of 45.50 citations per article, highlighting its consistent contribution to both volume and scholarly impact. Other countries such as Australia (37.30), Canada (31.30), and Spain (27.80) also show strong per-article citation performance, underscoring the high quality of their research. Meanwhile, emerging contributors like India and Korea are gaining visibility, although their average citation rates remain moderate. These patterns, detailed in Table 8, illustrate a diverse and evolving international research ecosystem where both output quantity and citation quality play critical roles in shaping scholarly influence.

Table 8. Most-cited countries.

| Country	TC	Average Article Citations
China	72,962	17.30
USA	30,653	45.50
India	15,466	16.50
United Kingdom	11,202	54.40
Spain	6011	27.80
Korea	5855	17.50
Australia	5453	37.30
Italy	4301	22.50
Germany	3591	27.40
Canada	3323	31.30

Figure 7 presents a co-occurrence network of keywords commonly found in Big Data research, helping to visualize how different concepts are connected within the literature. At the center of the map is the term "big data," depicted as the largest node, with strong connections to terms like "data mining," "machine learning," "deep learning," and "artificial intelligence." These links emphasize the core technologies and methods central to Big Data. Surrounding nodes such as "data analytics," "information management," and "Internet of Things" reflect the diverse applications and domains where Big Data is actively explored. The thickness and number of lines between nodes represent the frequency of co-occurrence, highlighting the interconnected and interdisciplinary nature of Big Data research. Overall, this network provides a clear snapshot of the field's main themes, demonstrating how technical tools and practical applications are closely intertwined.
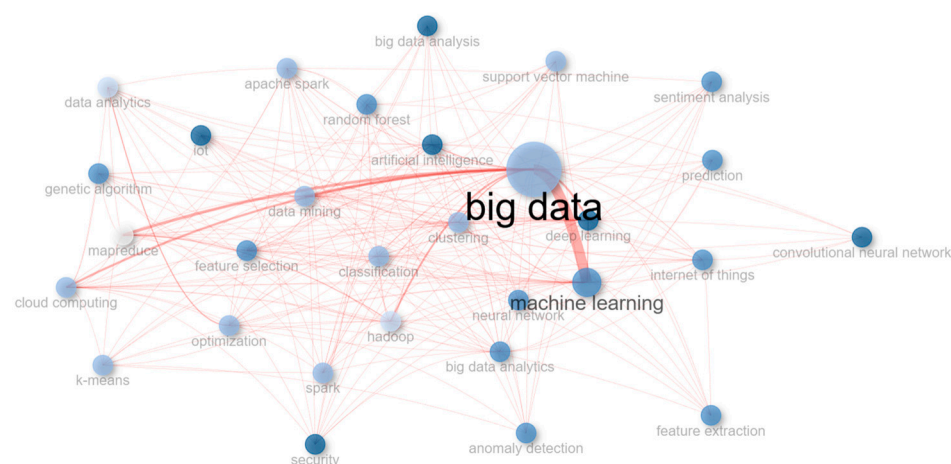
Figure 7. Co-occurrence network.

Figure 8 illustrates the temporal evolution and persistence of core themes in Big Data research between 2015 and 2024. Each bubble represents the frequency of a term within a given year, with larger bubbles signifying higher prominence in the literature. The horizontal line segments extending from each bubble indicate the period during which a term remains visible in the academic discourse. This visualization highlights both the emergence of new concepts (e.g., deep learning, Hadoop, information management) and the sustained dominance of foundational themes such as big data, data mining, machine learning, and clustering algorithms.
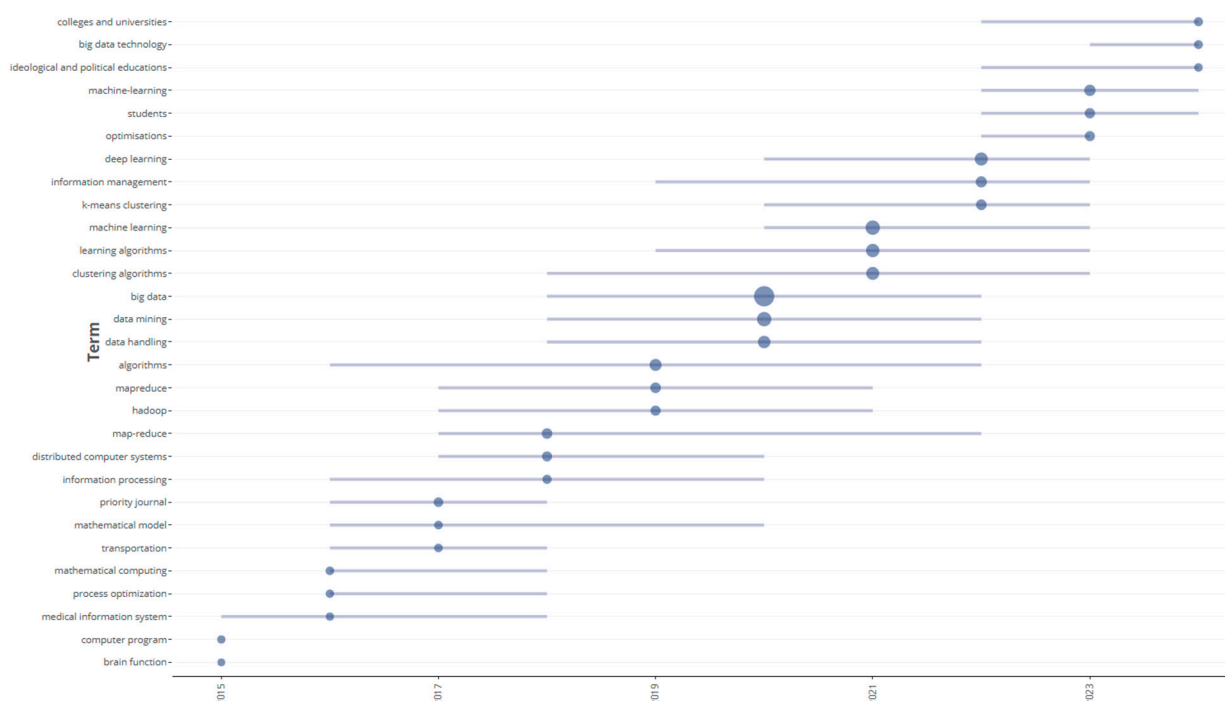
Figure 8. Trend topics in Big Data publications (2015–2024).

The chart reveals two key patterns. First, enduring terms like big data and machine learning demonstrate longitudinal relevance, showing consistent frequency growth across the decade. Second, the appearance of more specialized or technical terms (e.g., Hadoop, distributed computer systems, medical information systems) reflects diversification of the Big Data research agenda into applied domains. This temporal mapping thereby underscores how the field is not static but continually adapting, with older paradigms being

complemented rather than replaced by newer approaches. In addition, Table 9 presents the raw frequency distribution of the most frequently occurring terms in Big Data research between 2015 and 2024. The terms "big data" (7320 occurrences), "machine learning" (2101), "data mining" (2084), "learning algorithms" (1529), "deep learning" (1371), and "clustering algorithms" (1330) dominate the research landscape, reflecting both methodological emphases and evolving thematic priorities in the field. The data were extracted from a comprehensive corpus of publications, and the frequency counts serve as a foundational layer for identifying emerging research priorities.

Table 9. Word cloud of core topics in Big Data research (2015–2024).

Terms	Frequency
Big Data	7320
machine learning	2101
data mining	2084
learning algorithms	1529
deep learning	1371
clustering algorithms	1330
learning systems	1319
algorithm	1181
classification (of information)	1128
data handling	1079

From an analytical standpoint, this table was included to provide a transparent, unprocessed view of the raw term distribution prior to normalization or ranking (as presented in Table 7). Such a presentation is particularly important in academic research because it enables readers to directly assess the magnitude of specific terms without the abstraction of percentages or composite measures. Moreover, frequency-based word clouds have become an accepted exploratory tool in bibliometric studies, as they allow scholars to visually and quantitatively capture the prominence of concepts.

By providing Table 9 alongside subsequent analytical tables, this study emphasizes both exploratory mapping (via raw counts) and interpretive synthesis (via ranked percentages and impact measures). This dual approach not only increases methodological rigor but also demonstrates transparency in how key themes were derived, thereby strengthening the scholarly contribution of this study. Figure 8 provides a dynamic complement to the static frequency tables (Tables 7 and 9). While the tables quantify the relative importance of terms, the timeline chart adds a crucial historical perspective, showing how concepts rise and fall in scholarly attention. Together, these visual and tabular analyses provide a holistic understanding of both the structural hierarchy and the temporal trajectory of Big Data themes.

Figure 9 illustrates the global collaboration network in Big Data research, emphasizing both the frequency and intensity of international co-authorship. Darker shading reflects higher levels of national research output, while the red connecting lines denote collaborative linkages across countries. The findings highlight that while China and the United States remain dominant hubs—acting as central nodes in dense global networks extending across continents—other regions also demonstrate substantial engagement.
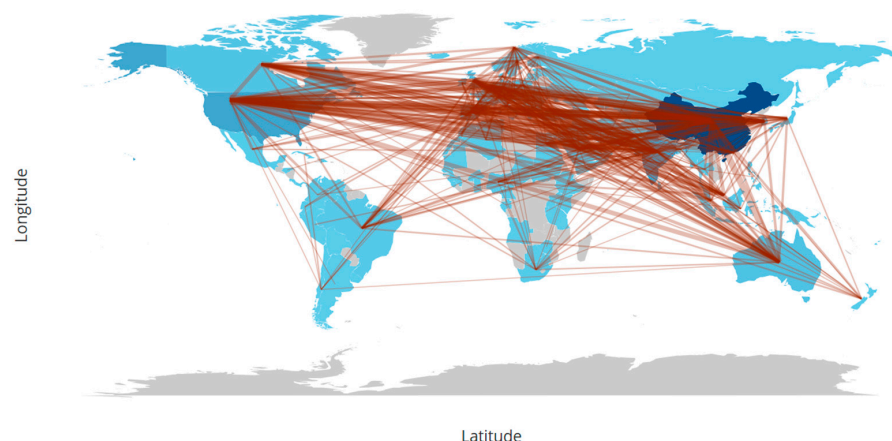
Figure 9. Countries' collaboration world map.

In particular, New Zealand emerges as a critical collaborative focal point, maintaining equally strong partnerships with a wide set of countries, including Australia, Canada, China, France, Hong Kong, India, Japan, Pakistan, Romania, Singapore, Spain, the United Arab Emirates, United Kingdom, and United States of America, each with a frequency of 171.48, indicating not only consistent engagement but also a uniform investment of collaborative effort across all these partnerships, Table 10. This consistent distribution reflects a structurally balanced network configuration where ties are uniformly strong rather than concentrated. In network theory terms, New Zealand exemplifies a distributed network model, avoiding the hierarchical centralization characteristic of dominant hubs like China or the United States. In such distributed networks, nodes are more equally weighted, reducing dependency on a few central actors and facilitating robust, resilient collaboration across multiple regions.

Table 10. Example uniform global collaborations (New Zealand).

From	To	Frequency
Australia	New Zealand	171.4849235
Canada	New Zealand	171.4849235
China	New Zealand	171.4849235
France	New Zealand	171.4849235
Hong Kong	New Zealand	171.4849235
India	New Zealand	171.4849235
Japan	New Zealand	171.4849235
Pakistan	New Zealand	171.4849235
Romania	New Zealand	171.4849235
Singapore	New Zealand	171.4849235
Spain	New Zealand	171.4849235
United Arab Emirates	New Zealand	171.4849235
United Kingdom	New Zealand	171.4849235
USA	New Zealand	171.4849235

This contrast is particularly instructive; whereas China and the United States reinforce their leadership through high-density, hierarchical networks that centralize influence, New Zealand's collaboration profile suggests a balanced, non-hierarchical structure, fostering inclusivity and reciprocal exchanges. From a theoretical perspective, this pattern aligns with concepts of network equality, where the uniformity of tie strength mitigates structural inequalities and enables a wide diffusion of knowledge and resources. Such a network

not only broadens the geographic and institutional scope of Big Data research but also enhances its capacity for innovation by connecting diverse epistemic communities.

At the same time, the visualization underscores the international character of Big Data research while also pointing to persistent disparities in global participation. Large regions of Africa and South America remain weakly connected, signaling underutilized potential and structural barriers in access to funding, infrastructure, and networks. Addressing these gaps through open collaboration frameworks, capacity-building initiatives, and inclusive policies will be essential for ensuring that Big Data evolves not merely as a field of technological progress but as a genuinely distributed and cooperative global enterprise in which network theory principles of balance, redundancy, and resilience are actively realized.

The visual bibliometric analysis presented here underscores the rapid expansion and global dispersion of Big Data research. The central themes of "machine learning," "deep learning," and "data mining" dominate the discourse, while journals such as *IEEE Access* and *Applied Mathematics and Nonlinear Sciences* emerge as key publication hubs. China leads in both publication volume and citations, with India showing a sharp rise in recent years, suggesting a shift in the global center of research gravity. Network visualizations reveal the interdisciplinary and interconnected nature of Big Data, as well as emerging research priorities and regional focuses. Importantly, the collaboration map highlights both the strength of existing international partnerships and the need to foster greater inclusion from underrepresented regions such as parts of Africa and South America.

Overall, this section provides a holistic and visually driven understanding of how Big Data research is evolving geographically, thematically, and institutionally, offering valuable context for future research direction and policy development.

8. Discussion

This study, which applies a bibliometric lens to 83 high-indexed journal articles, illuminates the dynamic evolution of Big Data technologies across sectors. The growing adoption of these technologies—particularly in conjunction with AI and ML—is reshaping organizational landscapes by enhancing operational efficiency and decision-making processes [49]. However, as implementation deepens, ethical concerns such as data privacy, security, and governance emerge as central challenges [51,64]. This calls for not just technical competence but also responsible stewardship from both researchers and practitioners.

The analysis supports a departure from the traditional "Four Vs" of Big Data—volume, velocity, variety, and veracity—with newer dimensions emphasizing ethical accountability and integration into complex sociotechnical ecosystems [27]. This shift highlights the need for holistic strategies that prioritize not only performance but also integrity, especially as organizations contend with societal expectations around transparency and fairness.

8.1. Insights from SLR

The SLR encompassed 83 peer-reviewed articles across diverse sectors, technologies, and geographies, providing a robust lens into the multifaceted landscape of BDT. The analysis revealed a consistent thematic progression from infrastructure-centric solutions to more nuanced concerns around ethics, human factors, real-time processing, and sector-specific transformations.

A striking observation across the reviewed literature is the rapid vertical integration of Big Data within critical domains like healthcare, education, marketing, manufacturing, and security. For instance, studies by Gomes et al. [6] and Kumar and Singh [21] emphasized the role of BDA in personalized medicine, disease surveillance, and electronic health records, highlighting transformative outcomes in patient care. Similarly, Arshad et al. [68] focused on Big Data intelligence in healthcare, underlining scalability and data privacy as emerging

concerns. In the education sector, researchers like Ang et al. [28], Ikegwu et al. [76], and Foffano et al. [82] explored the application of educational data analytics and computational intelligence, suggesting BDA's potential to enhance institutional performance, personalize learning, and support emergency remote teaching, especially during crises like COVID-19 [83].

In the business and management disciplines, studies by Ranjan and Foropon [11], Adewusi et al. [38], and Ochuba et al. [71] illustrate how organizations are using BDA for strategic planning, competitive intelligence, and customer segmentation. These findings align with Pawar and Paluri [10], who explored logistics and supply chain optimization through data-driven decision-making, confirming BDA's operational value across value chains. Likewise, Korherr and Kanbach [9] introduced a taxonomy of human-related capabilities that impact BDA implementation success, shifting attention to organizational readiness and workforce analytics.

A major technological trend observed is the acceleration of real-time data processing and the convergence with AI and ML, as detailed by Dubuc et al. [15], Jabbar et al. [7], and Jagatheesaperumal et al. [22]. These studies demonstrate how AI-enhanced BDA systems are enabling intelligent automation in sectors ranging from marketing and manufacturing to Industry 4.0 environments. Additionally, the inclusion of QC Agarwal and Alam [13] and hardware acceleration Sklyarov et al. [31] in select papers signals an emerging frontier aimed at solving high-volume, low-latency processing bottlenecks.

Data governance and ethics emerged as core challenges, particularly in studies focused on high-stakes applications. Micheli et al. [41] and Char et al. [79] addressed the ethical dimensions of AI and BDA in medical and geospatial contexts, raising concerns about consent, transparency, and algorithmic bias. These concerns were echoed by Kroll [56] and Favaretto et al. [80], who emphasized the need for more robust, context-specific ethical frameworks to regulate data use across both public and private domains.

Furthermore, the integration of cloud computing and edge computing was highlighted as critical for managing distributed data environments. Studies such as those by Hofmann et al. [24], Oliveira et al. [51], and Tuli et al. [62] illustrate how organizations are transitioning to hybrid architectures that blend Amazon Web Services (AWS), Google Cloud Platform (GCP), fog, and edge computing to reduce latency and improve scalability. These technologies are particularly vital for IoT-heavy sectors like smart cities, healthcare, and industrial automation.

Lastly, regional diversity in BDT adoption and innovation is notable. Contributions from countries such as India, China, Saudi Arabia, and Brazil reflect a democratization of Big Data research, with emerging economies addressing local challenges through tailored applications. For example, Amanullah et al. [1] investigated IoT security using deep learning in the Middle East and South Asia, while Tariq et al. [3] applied intelligent analytics to understand human behavior across multinational contexts.

In summary, the SLR underscores that Big Data Technologies are no longer confined to technical backends but are deeply embedded in organizational strategy, public policy, and ethical discourse. The integration of AI, the expansion of real-time and decentralized architectures, and a growing call for ethical oversight point to a maturing discipline with far-reaching implications.

8.2. Insights from the Bibliometric Analysis

The bibliometric analysis, based on 1108 Scopus-indexed articles from 2015 to 2024, complements the SLR by offering a data-driven view of the global BDT research ecosystem. The results reveal not only the thematic evolution of the field but also notable trends in regional output, citation influence, and interdisciplinary convergence. Together, these findings illustrate the field's maturation and diversification over the past decade.

A key finding from the analysis is the exponential growth in publication volume between 2015 and 2022, followed by a plateau in 2023–2024. This trajectory suggests a shift from foundational exploration toward consolidation and specialization. Early studies primarily focused on infrastructure and scalability challenges, while more recent publications reflect emerging interests in AI ethics, data democratization, and cross-sector applications. Keyword co-occurrence maps confirm this shift; terms like ML, deep learning, data governance, and edge computing dominate the intellectual landscape, suggesting a blending of data science, AI, and domain-specific innovation.

The analysis also sheds light on geographic research dynamics. China leads in publication volume with nearly 10,000 cumulative articles by 2024, demonstrating its strategic investment in data-intensive research. However, countries like the United Kingdom and Australia exhibit higher average citations per article (54.4 and 37.3, respectively), indicating a focus on high-impact scholarship. The United States maintains a balanced profile, with a strong presence in both volume and citation metrics. Meanwhile, India's recent surge—surpassing the U.S. in article output by 2023—signals the rise of emerging economies as serious contributors to the global BDT knowledge base [84].

Institutional and journal contributions further illuminate the field's evolution. *IEEE Access* emerges as the most prolific source, consistently publishing on a wide range of BDT topics including IoT, real-time analytics, and intelligent systems. Meanwhile, the sudden rise of *Applied Mathematics and Nonlinear Sciences* since 2022 reflects a growing emphasis on algorithmic modeling and mathematical approaches to Big Data problems. The diversity of journals, ranging from *Neurocomputing* to *Multimedia Tools and Applications*, underscores BDT's interdisciplinary reach across health informatics, cloud systems, marketing analytics, and energy management [85].

From a thematic perspective, the co-occurrence network and keyword frequency charts show how technical concepts (clustering algorithms, data mining, classification, dimensionality reduction) are deeply intertwined with application-driven themes (smart cities, healthcare analytics, cybersecurity). This reflects a dual movement: one toward technical depth and another toward practical integration. The rise of "data democratization" as a recurring theme, highlighted in studies from Wang et al. [57], Lefebvre et al. [66], and Samarasinghe et al. [67], also points to a social shift in data accessibility and decision-making power beyond data specialists [86].

The three-field plot and global collaboration maps highlight the highly international and collaborative nature of BDT research. Countries such as the United States, China, and India are not only high-output contributors but also key nodes in international co-authorship networks. However, the analysis also reveals regional disparities; Africa and South America remain underrepresented, suggesting potential for expanded research funding, infrastructure support, and North–South collaboration. The dominance of a few publication venues (e.g., *IEEE Access*) and countries (e.g., China, USA) also raises questions about knowledge centralization and visibility for scholars in developing regions [87].

In conclusion, the bibliometric analysis paints a vivid picture of a field that is expanding in scope, diversifying in methods, and shifting geographically. While foundational research laid the groundwork in infrastructure and processing, contemporary studies are increasingly engaging with ethical governance, societal applications, and algorithmic fairness. The global diffusion of research efforts—though uneven—indicates a future where Big Data technologies are shaped not only by computational advances but also by cultural, ethical, and geopolitical contexts. These trends reinforce the need for interdisciplinary collaboration and inclusive research practices to address complex, data-driven global challenges [88].

9. Research Limitations and Future Directions

While this study provides valuable insights through bibliometric analysis, several limitations must be acknowledged. The use of only 83 samples limits statistical power and may restrict the diversity of perspectives captured. As such, the findings may not fully reflect the broader trends across disciplines, geographical regions, or industry sectors.

Furthermore, bibliometric methods inherently reflect patterns found in published literature, which may be subject to publication bias—favoring frequently cited or English-language sources over equally relevant but less visible work. These biases can distort impact metrics and skew interpretation.

The defined timeframe of 2015–2024, though methodologically sound, may also exclude emerging technologies that have not yet gained prominence in scholarly databases. Fields like AI, biotechnology, and QC may be underrepresented in this window, limiting future-oriented conclusions. To enhance robustness and relevance, future research should do the following:

- Expand the sample size for greater statistical reliability.
- Include empirical investigations across varied sectors and regions.
- Explore trends beyond 2024 to capture cutting-edge advancements.
- Address technological challenges such as scalability, interoperability, and ethical governance—especially as Big Data, blockchain, and QC continue to evolve.

Ultimately, achieving meaningful progress will require interdisciplinary collaboration, bringing together expertise in computer science, statistics, data analytics, and domain-specific disciplines to ensure responsible innovation and comprehensive understanding.

10. Conclusions

This study is structured to provide a logical and clear flow of ideas, ensuring that the research question, methodology, and findings are easily digestible. It begins with a detailed literature review, followed by a structured methodology that employs bibliometric analysis, guiding the reader through the complexities of Big Data technologies. Furthermore, the focus on real-time processing, AI integration, and sector-specific applications creates a cohesive narrative that effectively links this study's theoretical framework to its practical implications. Each section builds upon the previous one, culminating in a discussion that not only synthesizes the findings but also provides actionable insights for both researchers and practitioners.

Big Data has had a profound impact since its inception, significantly influencing various sectors and reshaping organizational strategies. As new technologies emerge, Big Data has become increasingly actionable, driving advancements in decision-making through innovations such as real-time processing, AI, and edge computing. These technological advancements have greatly improved data management, enabling organizations to adapt more swiftly to market changes. However, the widespread adoption of Big Data also raises critical ethical concerns, particularly regarding data privacy, security, and scalability. Therefore, addressing these issues requires robust governance frameworks to maintain stakeholder trust and ensure responsible data practices.

In this context, this study investigates the evolution of Big Data technologies and their implications across various sectors by examining 83 papers from high-indexed, reputable journals. Additionally, a bibliometric analysis covering the period from 2015 to 2024 was employed, meticulously reviewing 1108 documents sourced from 544 publications. This comprehensive approach enhances our understanding of future trends in Big Data technologies and their broader implications. Consequently, the analysis of these documents provided valuable insights into the evolving landscape of Big Data and its future trajectory, identifying key themes that reflect significant trends and challenges in the field.

Moreover, to ensure relevance and quality, a rigorous article selection process was utilized—guided by the PRISMA 2020 framework—with strict inclusion and exclusion criteria. There is also a growing need to explore strategies that unlock the value of Big Data while addressing ethical concerns. For instance, identifying more efficient methods to transmit Big Data insights to decision-makers, customers, and citizens is crucial for maximizing its benefits. Recognizing the diverse perspectives on ethics and privacy from the private sector, scientific community, public sector, and citizens is essential for guiding the practical development of Big Data and analytics. By integrating these viewpoints, future research can foster a more inclusive and effective approach to leveraging Big Data.

References

1. Amanullah, M.A.; Habeeb, R.A.A.; Nasaruddin, F.H.; Gani, A.; Ahmed, E.; Nainar, A.S.M.; Akim, N.M.; Imran, M. Deep learning and big data technologies for IoT security. *Comput. Commun.* **2020**, *151*, 495–517. [CrossRef]
2. Arena, F.; Pau, G. An overview of big data analysis. *Bull. Electr. Eng. Inform.* **2020**, *9*, 1646–1653. [CrossRef]
3. Tariq, M.U.; Babar, M.; Poulin, M.; Khattak, A.S.; Alshehri, M.D.; Kaleem, S. Human Behavior Analysis Using Intelligent Big Data Analytics. *Front. Psychol.* **2021**, *12*, 686610. [CrossRef]
4. Mishra, D.; Luo, Z.; Jiang, S.; Papadopoulos, T.; Dubey, R. A bibliographic study on big data: Concepts, trends and challenges. *Bus. Process Manag. J.* **2017**, *23*, 555–573. [CrossRef]
5. Nagaraj, K.; Sharvani, G.S.; Sridhar, A. Emerging trend of big data analytics in bioinformatics: A literature review. *Int. J. Bioinform. Res. Appl.* **2018**, *14*, 144–205. [CrossRef]
6. Gomes, M.A.S.; Kovaleski, J.L.; Pagani, R.N.; da Silva, V.L.; Pasquini, T.C.d.S. Transforming healthcare with big data analytics: Technologies, techniques and prospects. *J. Med. Eng. Technol.* **2023**, *47*, 1–11. [CrossRef] [PubMed]
7. Jabbar, A.; Akhtar, P.; Dani, S. Real-time big data processing for instantaneous marketing decisions: A problematization approach. *Ind. Mark. Manag.* **2020**, *90*, 558–569. [CrossRef]
8. Cremin, C.J.; Dash, S.; Huang, X. Big data: Historic advances and emerging trends in biomedical research. *Curr. Res. Biotechnol.* **2022**, *4*, 138–151. [CrossRef]
9. Korherr, P.; Kanbach, D. Human-related capabilities in big data analytics: A taxonomy of human factors with impact on firm performance. *Rev. Manag. Sci.* **2023**, *17*, 1943–1970. [CrossRef]
10. Pawar, P.V.; Paluri, R.A. Big Data Analytics in Logistics and Supply Chain Management: A Review of Literature. *Vision* **2022**, 1–20. [CrossRef]
11. Ranjan, J.; Foropon, C. Big Data Analytics in Building the Competitive Intelligence of Organizations. *Int. J. Inf. Manag.* **2021**, *56*, 102231. [CrossRef]
12. Ajah, I.A.; Nweke, H.F. Big data and business analytics: Trends, platforms, success factors and applications. *Big Data Cogn. Comput.* **2019**, *3*, 32. [CrossRef]

13. Agarwal, P.; Alam, M. Exploring Quantum Computing to Revolutionize Big Data Analytics for Various Industrial Sectors. In *Big Data Analytics*; Auerbach Publications: Boca Raton, FL, USA, 2021. [CrossRef]

14. Agrawal, R.; Wankhede, V.A.; Kumar, A.; Luthra, S.; Huisingh, D. Big data analytics and sustainable tourism: A comprehensive review and network based analysis for potential future research. *Int. J. Inf. Manag. Data Insights* **2022**, *2*, 100122. [CrossRef]

15. Dubuc, T.; Stahl, F.; Roesch, E.B. Mapping the Big Data Landscape: Technologies, Platforms and Paradigms for Real-Time Analytics of Data Streams. *IEEE Access* **2021**, *9*, 15351–15374. [CrossRef]

16. Morawiec, P.; Sołtysik-Piorunkiewicz, A. Cloud Computing, Big Data, and Blockchain Technology Adoption in ERP Implementation Methodology. *Sustainability* **2022**, *14*, 3714. [CrossRef]

17. Saddad, E.; El-Bastawissy, A.; Mokhtar, H.M.O.; Hazman, M. Lake data warehouse architecture for big data solutions. *Int. J. Adv. Comput. Sci. Appl.* **2020**, *11*, 417–424. [CrossRef]

18. Bazzaz Abkenar, S.; Haghi Kashani, M.; Mahdipour, E.; Jameii, S.M. Big data analytics meets social media: A systematic review of techniques, open issues, and future directions. *Telemat. Inform.* **2021**, *57*, 101517. [CrossRef]

19. Naeem, M.; Jamal, T.; Diaz-Martinez, J.; Butt, S.A.; Montesano, N.; Tariq, M.I.; De-la-Hoz-Franco, E.; De-La-Hoz-Valdiris, E. Trends and Future Perspective Challenges in Big Data. In *Smart Innovation, Systems and Technologies*; Springer: Singapore, 2022; Volume 253. [CrossRef]

20. Lu, Y. Artificial intelligence: A survey on evolution, models, applications and future trends. *J. Manag. Anal.* **2019**, *6*, 1–29. [CrossRef]

21. Kumar, S.; Singh, M. Big data analytics for healthcare industry: Impact, applications, and tools. *Big Data Min. Anal.* **2019**, *2*, 48–57. [CrossRef]

22. Jagatheesaperumal, S.K.; Rahouti, M.; Ahmad, K.; Al-Fuqaha, A.; Guizani, M. The Duo of Artificial Intelligence and Big Data for Industry 4.0: Applications, Techniques, Challenges, and Future Research Directions. *IEEE Internet Things J.* **2022**, *9*, 12861–12885. [CrossRef]

23. Swazan, I.S.; Das, D. Bangladesh's Emergence as a Ready-Made Garment Export Leader: An Examination of the Competitive Advantages of the Garment Industry. *Int. J. Glob. Bus. Compet.* **2022**, *17*, 162–174. [CrossRef]

24. Hofmann, W.; Lang, S.; Reichardt, P.; Reggelin, T. A brief introduction to deploy Amazon Web Services for online discrete-event simulation. *Procedia Comput. Sci.* **2022**, *200*, 386–393. [CrossRef]

25. Sukhdeve, D.S.R.; Sukhdeve, S.S. *Google Cloud Platform for Data Science*; Apress: Berkeley, CA, USA, 2023. [CrossRef]

26. Azeem, M.; Haleem, A.; Bahl, S.; Javaid, M.; Suman, R.; Nandan, D. Big data applications to take up major challenges across manufacturing industries: A brief review. *Mater. Today Proc.* **2022**, *49*, 339–348. [CrossRef]

27. Botvin, M.; Hershkovitz, A.; Forkosh-Baruch, A. Data-driven decision-making in emergency remote teaching. *Educ. Inf. Technol.* **2023**, *28*, 489–506. [CrossRef] [PubMed]

28. Ang, K.L.M.; Ge, F.L.; Seng, K.P. Big Educational Data Analytics: Survey, Architecture and Challenges. *IEEE Access* **2020**, *8*, 116392–116414. [CrossRef]

29. Dean, J.; Ghemawat, S. MapReduce: Simplified data processing on large clusters. *Commun. ACM* **2008**, *51*, 107–113. [CrossRef]

30. Khezr, S.N.; Navimipour, N.J. MapReduce and Its Applications, Challenges, and Architecture: A Comprehensive Review and Directions for Future Research. *J. Grid Comput.* **2017**, *15*, 295–321. [CrossRef]

31. Sklyarov, V.; Skliarova, I.; Utepbergenov, I. Hardware Accelerators for Data Processing in High-Performance Computing Systems. In Proceedings of the 15th IEEE International Conference on Application of Information and Communication Technologies, AICT 2021, Baku, Azerbaijan, 13–15 October 2021. [CrossRef]

32. Bhogal, J.; Choksi, I. Handling Big Data Using NoSQL. In Proceedings of the 2015 IEEE 29th International Conference on Advanced Information Networking and Applications Workshops, WAINA 2015, Gwangju, Republic of Korea, 24–27 March 2015. [CrossRef]

33. Felstaine, E.; Hermoni, O. Machine Learning, Containers, Cloud Natives, and Microservices. In *Artificial Intelligence for Autonomous Networks*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2018. [CrossRef]

34. Lucas-Noll, J.; Lleixà-Fortuño, M.; Queralt-Tomas, L.; Panisello-Tafalla, A.; Carles-Lavila, M.; Clua-Espuny, J.L. Organization and costs of stroke care in outpatient settings: Systematic review. *Aten. Primaria* **2023**, *55*, 102578. [CrossRef]

35. Bilal, M.; Oyedele, L.O.; Qadir, J.; Munir, K.; Ajayi, S.O.; Akinade, O.O.; Owolabi, H.A.; Alaka, H.A.; Pasha, M. Big Data in the construction industry: A review of present status, opportunities, and future trends. *Adv. Eng. Inform.* **2016**, *30*, 500–521. [CrossRef]

36. Ejimofor, I.A.U.; Okonkwo, O.O.R. Development of a Knowledge Discovery System in Big Data Mining Environment. *Int. Res. J. Innov. Eng. Technol.* **2021**, *5*, 65–70. [CrossRef]

37. Li, Y.; Hei, X. Performance optimization of computing task scheduling based on the Hadoop big data platform. *Neural Comput. Appl.* **2022**, *37*, 8181–8192. [CrossRef]

38. Adewusi, A.O.; Okoli, U.I.; Adaga, E.; Olorunsogo, T.; Asuzu, O.F.; Daraojimba, D.O. Business Intelligence in the Era of Big Data: A Review of Analytical Tools and Competitive Advantage. *Comput. Sci. IT Res. J.* **2024**, *5*, 415–431. [CrossRef]

39. Liang, H.; Li, J.; Wu, H.; Li, L.; Zhou, X.; Jiang, X. Mammographic Classification of Breast Cancer Microcalcifications through Extreme Gradient Boosting. *Electronics* **2022**, *11*, 2435. [CrossRef]

40. Gao, Q.; Jin, X.; Xia, E.; Wu, X.; Gu, L.; Yan, H.; Xia, Y.; Li, S. Identification of Orphan Genes in Unbalanced Datasets Based on Ensemble Learning. *Front. Genet.* **2020**, *11*, 820. [CrossRef]

41. Micheli, M.; Gevaert, C.M.; Carman, M.; Craglia, M.; Daemen, E.; Ibrahim, R.E.; Kotsev, A.; Mohamed-Ghouse, Z.; Schade, S.; Schneider, I.; et al. AI ethics and data governance in the geospatial domain of Digital Earth. *Big Data Soc.* **2022**, *9*, 1–5. [CrossRef]

42. Adankon, M.M.; Cheriet, M.; Biem, A. Semisupervised least squares support vector machine. *IEEE Trans. Neural Netw.* **2009**, *20*, 1858–1870. [CrossRef] [PubMed]

43. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*. [CrossRef]

44. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; Antes, G.; Atkins, D.; Barbour, V.; Barrowman, N.; Berlin, J.A.; Clark, J.; et al. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med.* **2009**, *6*, e1000097. [CrossRef] [PubMed]

45. Gioia, D.A.; Corley, K.G.; Hamilton, A.L. Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology. *Organ. Res. Methods* **2013**, *16*, 15–31. [CrossRef]

46. Korner, M.E.H.; Lambán, M.P.; Albajez, J.A.; Santolaria, J.; Corrales, L.D.C.N.; Royo, J. Systematic literature review: Integration of additive manufacturing and industry 4.0. *Metals* **2020**, *10*, 1061. [CrossRef]

47. Sarkis-Onofre, R.; Catalá-López, F.; Aromataris, E.; Lockwood, C. How to properly use the PRISMA Statement. *Syst. Rev.* **2021**, *10*, 117. [CrossRef]

48. Tahamtan, I.; Safipour Afshar, A.; Ahamdzadeh, K. Factors affecting number of citations: A comprehensive review of the literature. *Scientometrics* **2016**, *107*, 1195–1225. [CrossRef]

49. Kuang, L.; Liu, H.; Ren, Y.; Luo, K.; Shi, M.; Su, J.; Li, X. Application and development trend of artificial intelligence in petroleum exploration and development. *Pet. Explor. Dev.* **2021**, *48*, 1–14. [CrossRef]

50. Chang, V. An ethical framework for big data and smart cities. *Technol. Forecast. Soc. Change* **2021**, *165*, 120559. [CrossRef]

51. Oliveira, F.; Costa, D.G.; Assis, F.; Silva, I. Internet of Intelligent Things: A convergence of embedded systems, edge computing and machine learning. *Internet Things* **2024**, *26*, 101153. [CrossRef]

52. Sun, Z.; Strang, K.D.; Pambel, F. Privacy and security in the big data paradigm. *J. Comput. Inf. Syst.* **2020**, *60*, 146–155. [CrossRef]

53. Al-Ghabra, N. Toward Sustainable Smart Cities: Concepts & Challenges. *Archit. Plan. J.* **2022**, *28*, 3. [CrossRef]

54. Nambiar, A.; Mundra, D. An Overview of Data Warehouse and Data Lake in Modern Enterprise Data Management. *Big Data Cogn. Comput.* **2022**, *6*, 132. [CrossRef]

55. Azzabi, S.; Alfughi, Z.; Ouda, A. Data Lakes: A Survey of Concepts and Architectures. *Computers* **2024**, *13*, 183. [CrossRef]

56. Kroll, J.A. Data Science Data Governance [AI Ethics]. *IEEE Secur. Priv.* **2018**, *16*, 61–70. [CrossRef]

57. Wang, Y.; Blobel, B.; Yang, B. Reinforcing Health Data Sharing through Data Democratization. *J. Pers. Med.* **2022**, *12*, 1380. [CrossRef] [PubMed]

58. Marinakis, V.; Koutsellis, T.; Nikas, A.; Doukas, H. Ai and data democratisation for intelligent energy management. *Energies* **2021**, *14*, 4341. [CrossRef]

59. Mir, A.A. Optimizing Mobile Cloud Computing Architectures for Real-Time Big Data Analytics in Healthcare Applications: Enhancing Patient Outcomes through Scalable and Efficient Processing Models. *Integr. J. Sci. Technol.* **2024**, *1*, 1–11.

60. Rathore, M.M.; Shah, S.A.; Shukla, D.; Bentafat, E.; Bakiras, S. The Role of AI, Machine Learning, and Big Data in Digital Twinning: A Systematic Literature Review, Challenges, and Opportunities. *IEEE Access* **2021**, *9*, 32030–32052. [CrossRef]

61. Hamdan, S.; Ayyash, M.; Almajali, S. Edge-computing architectures for internet of things applications: A survey. *Sensors* **2020**, *20*, 6441. [CrossRef]

62. Tuli, S.; Mirhakimi, F.; Pallewatta, S.; Zawad, S.; Casale, G.; Javadi, B.; Yan, F.; Buyya, R.; Jennings, N.R. AI augmented Edge and Fog computing: Trends and challenges. *J. Netw. Comput. Appl.* **2023**, *216*, 103648. [CrossRef]

63. Bansal, M.; Chana, I.; Clarke, S. A Survey on IoT Big Data: Current Status, 13 V's Challenges, and Future Directions. *ACM Comput. Surv.* **2021**, *53*, 1–59. [CrossRef]

64. Yang, P.; Xiong, N.; Ren, J. Data Security and Privacy Protection for Cloud Storage: A Survey. *IEEE Access* **2020**, *8*, 131723–131740. [CrossRef]

65. Strang, K.D.; Sun, Z. Big Data Paradigm: What is the Status of Privacy and Security? *Ann. Data Sci.* **2017**, *4*, 1–17. [CrossRef]

66. Lefebvre, H.; Legner, C.; Fadler, M. Data democratization: Toward a deeper understanding. In Proceedings of the 42nd International Conference on Information Systems, ICIS 2021 TREOs: "Building Sustainability and Resilience with IS: A Call for Action", Austin, TX, USA, 12–15 December 2021.

67. Samarasinghe, S.S.U.; Lokuge, S.; Snell, L. Exploring Tenets of Data Democratization. *arXiv* **2022**, arXiv:2206.12051. [CrossRef]

68. Arshad, H.; Tayyab, M.; Bilal, M.; Akhtar, S.; Abdullahi, A.M. Trends and Challenges in harnessing big data intelligence for health care transformation. In *Artificial Intelligence for Intelligent Systems*; CRC Press: Boca Raton, FL, USA, 2024; pp. 220–240.

69. Buck, D.; Tucker, S.; Roe, B.; Hughes, J.; Challis, D. Hospital admissions and place of death of residents of care homes receiving specialist healthcare services: A systematic review without meta-analysis. *J. Adv. Nurs.* **2022**, *78*, 666–697. [CrossRef]

70. Agustí, M.A.; Orta-Pérez, M. Big data and artificial intelligence in the fields of accounting and auditing: A bibliometric analysis. *Span. J. Financ. Account./Rev. Española De Financ. Contab.* **2023**, *52*, 412–438. [CrossRef]

71. Ochuba, N.A.; Amoo, O.O.; Okafor, E.S.; Akinrinola, O.; Usman, F.O. Strategies for Leveraging Big Data and Analytics for Business Development: A Comprehensive Review Across Sectors. *Comput. Sci. IT Res. J.* **2024**, *5*, 562–575. [CrossRef]

72. Rani, S.; Bhambri, P.; Kataria, A. Integration of IoT, Big Data, and Cloud Computing Technologies: Trend of the Era. In *Big Data, Cloud Computing and IoT: Tools and Applications*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2023. [CrossRef]

73. Karim, A.; Siddiqa, A.; Safdar, Z.; Razzaq, M.; Gillani, S.A.; Tahir, H.; Kiran, S.; Ahmed, E.; Imran, M. Big data management in participatory sensing: Issues, trends and future directions. *Futur. Gener. Comput. Syst.* **2020**, *107*, 942–955. [CrossRef]

74. Waterson, P.; Carman, E.M.; Manser, T.; Hammer, A. Hospital Survey on Patient Safety Culture (HSPSC): A systematic review of the psychometric properties of 62 international studies. *BMJ Open* **2019**, *9*, e026896. [CrossRef] [PubMed]

75. Hamad, R.; Elser, H.; Tran, D.C.; Rehkopf, D.H.; Goodman, S.N. How and why studies disagree about the effects of education on health: A systematic review and meta-analysis of studies of compulsory schooling laws. *Soc. Sci. Med.* **2018**, *212*, 168–178. [CrossRef] [PubMed]

76. Ikegwu, A.C.; Nweke, H.F.; Anikwe, C.V. Recent trends in computational intelligence for educational big data analysis. *Iran J. Comput. Sci.* **2024**, *7*, 103–129. [CrossRef]

77. Tran, H.; Saleem, K.; Lim, M.; Chow, E.P.F.; Fairley, C.K.; Terris-Prestholt, F.; Ong, J.J. Global estimates for the lifetime cost of managing HIV. *AIDS* **2021**, *35*, 1273–1281. [CrossRef]

78. Deepa, N.; Pham, Q.V.; Nguyen, D.C.; Bhattacharya, S.; Prabadevi, B.; Gadekallu, T.R.; Maddikunta, P.K.R.; Fang, F.; Pathirana, P.N. A survey on blockchain for big data: Approaches, opportunities, and future directions. *Futur. Gener. Comput. Syst.* **2022**, *131*, 209–226. [CrossRef]

79. Char, D.S.; Abràmoff, M.D.; Feudtner, C. Identifying Ethical Considerations for Machine Learning Healthcare Applications. *Am. J. Bioeth.* **2020**, *20*, 7–17. [CrossRef]

80. Favaretto, M.; De Clercq, E.; Gaab, J.; Elger, B.S. First do no harm: An exploration of researchers' ethics of conduct in Big Data behavioral studies. *PLoS ONE* **2020**, *15*, e0241865. [CrossRef]

81. Sandhu, A.K. Big Data with Cloud Computing: Discussions and Challenges. *Big Data Min. Anal.* **2022**, *5*, 32–40. [CrossRef]

82. Foffano, F.; Scantamburlo, T.; Cortés, A. Investing in AI for social good: An analysis of European national strategies. *AI Soc.* **2023**, *38*, 479–500. [CrossRef]

83. Cui, Y.; Ma, Z.; Wang, L.; Yang, A.; Liu, Q.; Kong, S.; Wang, H. A survey on big data-enabled innovative online education systems during the COVID-19 pandemic. *J. Innov. Knowl.* **2023**, *8*, 100295. [CrossRef]

84. Market Research Future. Data Analytics Market Size, Share | Growth Analysis 2030. Available online: https://www.marketresearchfuture.com/reports/data-analytics-market-1689 (accessed on 24 December 2024).

85. Almunawar, M.N.; Anshari, M. Digital enabler and value integration: Revealing the expansion engine of digital marketplace. *Technol. Anal. Strateg. Manag.* **2022**, *34*, 847–857. [CrossRef]

86. Gartner Inc. What's New in Artificial Intelligence From the 2023 Gartner Hype Cycle. *Gartner Articles.* Available online: https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2023-gartner-hype-cycle (accessed on 27 December 2024).

87. Pradhan, S.K.; Heyn, H.M.; Knauss, E. Identifying and managing data quality requirements: A design science study in the field of automated driving. *Softw. Qual. J.* **2024**, *32*, 313–360. [CrossRef]

88. Eke, D.; Stahl, B. Ethics in the Governance of Data and Digital Technology: An Analysis of European Data Regulations and Policies. *Digit. Soc.* **2024**, *3*, 11. [CrossRef]