World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Review Article)

Check for updates

Tackling security and privacy challenges in the realm of big data analytics

Janet Ngesa *

Kenya Agricultural and Livestock Research Organization, Nairobi, Kenya.

World Journal of Advanced Research and Reviews, 2024, 21(02), 552–576

Publication history: Received on 17 December 2023; revised on 06 February 2024; accepted on 09 February 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.21.2.0429

Abstract

As organizations increasingly harness the power of big data analytics to derive insights and drive decision-making, the paramount concerns of security and privacy have come to the forefront. This paper presents a comprehensive framework for addressing the multifaceted challenges of big data security and privacy. Drawing on a synthesis of cutting-edge technologies, encryption methods, and access control mechanisms, our approach aims to fortify the entire big data lifecycle. The paper delves into innovative strategies for secure data storage, transmission, and processing, ensuring that sensitive information is shielded from unauthorized access or malicious attacks. Additionally, the framework incorporates robust privacy-preserving techniques, including anonymization and differential privacy, to uphold individual confidentiality. Through a meticulous analysis of current trends, emerging threats, and regulatory landscapes, this paper not only provides theoretical insights but also practical guidelines for organizations seeking to navigate the intricate landscape of big data while safeguarding the integrity, security, and privacy of the vast datasets at their disposal.

Keywords: Big data; Privacy; Security; Analytics; Breaches

1. Introduction

In the era of unprecedented data proliferation, the advent of big data has revolutionized the landscape of information management, empowering organizations with unparalleled insights [1]-[4]. However, this data-driven paradigm shift comes hand-in-hand with profound challenges, particularly concerning the security and privacy of the vast and sensitive datasets involved [5]. As organizations increasingly rely on big data analytics to extract value and make informed decisions, the need for a robust and comprehensive framework to address the intricate issues of security and privacy becomes imperative. This paper seeks to delve into the multifaceted dimensions of big data security and privacy, offering a thorough exploration of the current landscape, emerging threats, and innovative solutions. The exponential growth of data, fueled by the Internet of Things (IoT), social media, and other sources, underscores the urgency of developing effective strategies to protect against unauthorized access, data breaches, and malicious activities [6]-[10]. In this context, the paper aims to provide a holistic understanding of the challenges posed by the sheer volume, velocity, and variety of big data, emphasizing the critical importance of safeguarding sensitive information throughout its lifecycle. Figure 1 shows some of the big data concepts and their brief descriptions.

* Corresponding author: Janet Ngesa

Figure 1 Big data concepts

According to [11], there is symbiotic relationship between big data security and privacy. As organizations collect and analyze vast datasets containing personally identifiable information (PII), there is a growing need to reconcile the tension between leveraging data for insights and preserving individual privacy rights. This paper acknowledges this delicate balance and explores innovative privacy-preserving techniques, such as anonymization and differential privacy, to ensure that individuals' confidentiality is maintained while still allowing organizations to harness the power of big data analytics. Through a synthesis of theoretical foundations and practical guidelines, this paper aims to equip researchers, practitioners, and decision-makers with the tools necessary to navigate the complex landscape of big data security and privacy in the contemporary data-driven ecosystem. By providing a comprehensive overview of the challenges and the need for a proactive approach, the paper aims to contribute to the discourse surrounding the responsible and secure utilization of big data for the benefit of organizations and society at large.

2. Big data technologies

Big data technologies encompass a diverse set of tools and frameworks designed to handle the challenges posed by large volumes, velocity, and variety of data [12]. These technologies enable the storage, processing, and analysis of massive datasets efficiently. Below is an extensive description of various big data technologies:

2.1. Hadoop

Apache Hadoop is a foundational framework for distributed storage and processing of large datasets. It consists of the Hadoop Distributed File System (HDFS) for storage and MapReduce for processing [13]. Hadoop allows data to be distributed across multiple nodes in a cluster, facilitating parallel processing.

2.2. Apache Spark

Apache Spark is an open-source, in-memory data processing engine that provides fast and flexible analytics. Spark supports batch processing, interactive queries, streaming, and machine learning [14]. Its in-memory processing capability enhances the performance of iterative algorithms [15] commonly used in machine learning.

2.3. NoSQL Databases

NoSQL databases, such as MongoDB, Cassandra, and Couchbase, are designed to handle unstructured or semi-structured data [16]. They provide scalable and flexible storage solutions, allowing organizations to store and retrieve data without the constraints of traditional relational databases.

2.4. Apache Flink

Apache Flink is a stream processing framework for big data analytics. It enables real-time processing and analytics on continuous data streams [17]-[19]. Flink's event-driven architecture is well-suited for applications requiring low-latency [20] processing, such as fraud detection and monitoring.

2.5. Apache Kafka

Apache Kafka is a distributed event streaming platform that enables the collection, storage, and real-time processing of streaming data [21]. Kafka serves as a reliable and scalable messaging system, allowing seamless communication between various components of a big data ecosystem.

2.6. Apache HBase

Apache HBase is a distributed, scalable, and consistent NoSQL database that runs on top of Hadoop Distributed File System (HDFS). It is particularly suitable for random, real-time read and write access to large datasets and is often used for applications requiring low-latency access to big data [22]-[24].

2.7. Apache Hive

Apache Hive is a data warehouse infrastructure built on top of Hadoop for querying [25] and managing large datasets. It provides a SQL-like language called HiveQL, which allows users to write queries to analyze data stored in Hadoop [26].

2.8. Apache Drill

Apache Drill is a distributed, schema-free SQL query engine for big data exploration [27]. It supports a wide range of data sources, including Hadoop, NoSQL databases, and cloud storage. Drill allows users to query and analyze data across multiple data formats and storage systems.

2.9. Distributed Storage Systems

Distributed storage systems like Amazon S3, Google Cloud Storage, and Azure Data Lake Storage offer scalable and cost-effective solutions for storing large volumes of data in the cloud [28], [29]. These systems provide high durability, availability, and the ability to scale storage as needed.

2.10. Machine Learning Frameworks

Frameworks like TensorFlow, PyTorch, and scikit-learn facilitate the implementation of machine learning algorithms [30] on big data. They allow data scientists and engineers to build, train, and deploy machine learning models at scale.

In a nutshell, these technologies collectively form the backbone of the big data ecosystem, empowering organizations to extract meaningful insights, perform advanced analytics, and derive value from massive and complex datasets.

3. Security and privacy issues in big data analytics

Security and privacy are paramount concerns in the realm of big data analytics due to the unprecedented scale, diversity, and sensitivity of the data involved [31]-[35]. Addressing these issues is crucial to ensure the responsible and ethical use of data. Below is an extensive discussion of security and privacy issues in big data analytics:

3.1. Data Breaches

Big data repositories are attractive targets for cybercriminals. A breach could expose sensitive information, leading to financial losses, reputational damage, and legal repercussions. Data breaches in big data analytics represent a critical and pervasive threat, where large-scale repositories of sensitive information become vulnerable to unauthorized access, exploitation, or theft [36]-[40]. These breaches pose substantial risks, ranging from financial losses and reputational damage to legal consequences. The vast and interconnected nature of big data environments amplifies the potential impact, as cybercriminals target multiple access points and exploit vulnerabilities in distributed systems. The consequences of a data breach in big data analytics extend beyond individual privacy concerns to encompass the compromise of intellectual property, confidential business strategies, and personally identifiable information [41]-[45]. Addressing this menace requires a comprehensive approach involving robust encryption protocols, stringent access controls, continuous monitoring, and proactive cybersecurity measures to safeguard the integrity and security of the voluminous and diverse datasets processed within the ambit of big data analytics.

3.2. Unauthorized Access

With multiple access points and users in a big data environment, there is a risk of unauthorized users gaining access to sensitive data, leading to misuse or theft. According to [46], unauthorized access in big data analytics poses a significant threat, encompassing the risk of unapproved entry into expansive datasets, potentially leading to data manipulation, theft, or misuse. With numerous access points and intricate data ecosystems, the challenge of preventing unauthorized entry becomes pronounced. Malicious actors may exploit vulnerabilities in the architecture, circumvent authentication protocols, or misuse insider privileges, jeopardizing the confidentiality and integrity of vast and diverse datasets [47]-[50]. The consequences of unauthorized access range from compromising individual privacy and sensitive business information to undermining the trust stakeholders place in data-driven decision-making processes. Counteracting this threat necessitates the implementation of robust authentication mechanisms, stringent access controls, regular security audits, and comprehensive user education to fortify the defenses of big data analytics systems against unauthorized intrusions.

3.3. Data Residency and Sovereignty

As data is often stored in distributed environments or the cloud, compliance with regional data protection regulations and ensuring data sovereignty become critical concerns. As explained in [51], data residency and sovereignty issues in big data analytics revolve around the challenges associated with storing and processing data across geographical boundaries while adhering to diverse and often stringent data protection regulations. As organizations increasingly leverage cloud-based solutions and distributed storage, the concern arises regarding compliance with regional data privacy laws and the safeguarding of data within specific jurisdictions. Ensuring data residency involves addressing questions of where data is physically stored and processed, while sovereignty pertains to respecting the legal and regulatory frameworks of the countries involved [52]-[55]. These considerations are crucial as non-compliance may lead to legal consequences, fines, and reputational damage. Consequently, navigating data residency and sovereignty in big data analytics necessitates careful selection of storage solutions, clear data governance policies, and ongoing adherence to evolving international and regional privacy regulations.

3.4. Inadequate Data Governance

Weak data governance can result in improper handling, storage, and sharing of data, leading to privacy violations and compromised data integrity. According to [56], inadequate data governance in big data analytics represents a significant challenge, encompassing deficiencies in the management, quality control, and security of vast and diverse datasets. The absence of robust data governance frameworks can lead to improper handling, storage, and sharing of information, raising concerns about data integrity, privacy violations, and compliance issues. In a big data environment, where data comes from various sources and is processed across distributed systems, the lack of standardized governance protocols heightens the risk of inaccurate analytics, compromising the reliability of insights [57]-[60]. To address these challenges, organizations must establish comprehensive data governance policies, enforce data quality standards, and conduct regular audits to ensure compliance with internal and external regulations, thereby fortifying the foundation of responsible and effective big data analytics practices.

3.5. De-identification Challenges

Anonymizing or de-identifying data to protect privacy can be challenging, especially when dealing with diverse datasets that can be re-identified when combined. De-identification challenges in big data analytics center around the intricate task of anonymizing or pseudonymizing data to protect individual privacy while maintaining data utility [61]-[65]. As big data often involves diverse datasets, the risk of re-identification through the combination of seemingly anonymous information becomes a persistent concern. Traditional de-identification methods may fall short in the face of evolving re-identification techniques and complex data interrelationships. Striking a balance between privacy preservation and data analysis requires constant adaptation of de-identification strategies, incorporating advanced techniques like differential privacy, which provides a rigorous mathematical framework for quantifying and controlling the privacy risks associated with data release [66]-[70]. The dynamic nature of big data and the need for comprehensive privacy protection mechanisms underscore the ongoing challenges in achieving effective de-identification in the context of large-scale analytics.

3.6. Algorithmic Bias and Discrimination

Biases in data or algorithms used for analytics can lead to discriminatory outcomes, infringing upon privacy and perpetuating social inequalities [71], [72]. Algorithmic bias and discrimination in big data analytics pose ethical challenges, reflecting the potential for biases encoded in algorithms to result in discriminatory outcomes, particularly against certain demographic groups. Big data analytics heavily relies on algorithms for decision-making, and if these algorithms are trained on biased datasets, they may perpetuate and amplify existing societal biases. Issues such as gender, race, or socioeconomic bias can emerge, leading to unfair treatment and reinforcing systemic inequalities. Addressing algorithmic bias requires thorough scrutiny of training datasets, continuous monitoring of model outputs, and efforts to enhance transparency and accountability in algorithmic decision-making processes [73]-[75]. As big data analytics increasingly influences critical decisions in various domains, mitigating bias becomes paramount to ensure fair, equitable, and ethical outcomes.

3.7. Privacy-preserving Analytics

Balancing the need for analytics with privacy protection is a delicate task. Performing analytics while safeguarding individual privacy rights poses technical challenges. Privacy-preserving challenges in big data analytics arise from the tension between extracting valuable insights from vast datasets and safeguarding individual privacy rights [76], [77]. As organizations strive to leverage the power of analytics, the risk of inadvertently exposing sensitive information looms large. Traditional approaches to privacy, such as anonymization, face challenges in the era of big data due to the potential for re-identification through sophisticated techniques. Implementing privacy-preserving technologies, like homomorphic encryption and secure multi-party computation, introduces computational complexities [78]-[80]. Striking a delicate balance between data utility and privacy protection remains a persistent challenge, necessitating ongoing advancements in techniques that enable meaningful analysis without compromising the confidentiality of personal information. Addressing these challenges is vital to building trust among stakeholders and ensuring responsible and ethical use of big data analytics in a privacy-conscious landscape.

3.8. Regulatory Compliance

Adhering to data protection laws and regulations, such as GDPR, HIPAA, or CCPA, is crucial but can be complex in the context of big data analytics. Regulatory compliance challenges in big data analytics stem from the complex and evolving landscape of data protection laws and industry regulations [81], [82]. As organizations harness the power of big data, navigating a myriad of compliance requirements, such as GDPR, HIPAA, or CCPA, becomes increasingly intricate. Ensuring that data processing activities align with the principles of consent, data minimization, and transparency poses a constant challenge, especially in the context of diverse and large-scale datasets. The global nature of data flows and storage exacerbates compliance complexities, as organizations must contend with varying regional and international regulations [83], [84]. Meeting regulatory standards necessitates ongoing efforts to stay informed about evolving legal frameworks, conduct thorough impact assessments, and implement mechanisms for obtaining and managing consent, all while adapting to the dynamic and multifaceted nature of big data analytics.

3.9. Insider Threats

Malicious activities or inadvertent errors by internal personnel can pose significant security risks, compromising the confidentiality and integrity of big data [85]. Insider threats present significant challenges in big data analytics, as the vast and interconnected nature of data ecosystems introduces vulnerabilities stemming from both malicious intent and inadvertent actions by internal personnel. Employees with privileged access may intentionally exploit their positions to compromise the confidentiality and integrity of large datasets, leading to data breaches or unauthorized access. Additionally, unintentional errors or negligence can pose serious risks to data security. The dynamic and distributed nature of big data environments exacerbates these challenges, requiring organizations to implement stringent access controls, conduct regular employee training programs, and employ advanced monitoring mechanisms to detect and mitigate insider threats effectively [86]-[90]. Navigating the complexities of insider threats in big data analytics demands a holistic approach that blends technological solutions with robust policies and ongoing vigilance to safeguard against potential breaches originating from within the organization.

3.10. Data Lifecycle Management

Incomplete or improper handling of data throughout its lifecycle, from creation to deletion, can result in security vulnerabilities and privacy infringements. Data lifecycle management in big data analytics presents challenges due to the sheer volume, variety, and velocity of data, necessitating careful consideration at every stage from creation to disposal [91], [92]. Handling the vast amount of information generated and collected requires efficient storage solutions, data cleansing, and transformation processes to ensure data quality and integrity. Figure 2 shows a typical data lifecycle management. As data moves through its lifecycle, issues of access control, privacy preservation, and regulatory compliance become paramount. Challenges also arise in determining the optimal retention periods for different types of data and managing the secure deletion of obsolete information [93], [94]. In the dynamic landscape of big data analytics, developing and adhering to comprehensive data lifecycle management strategies that align with organizational goals, industry regulations, and ethical considerations is essential to maximize the value of data while minimizing associated risks and inefficiencies.

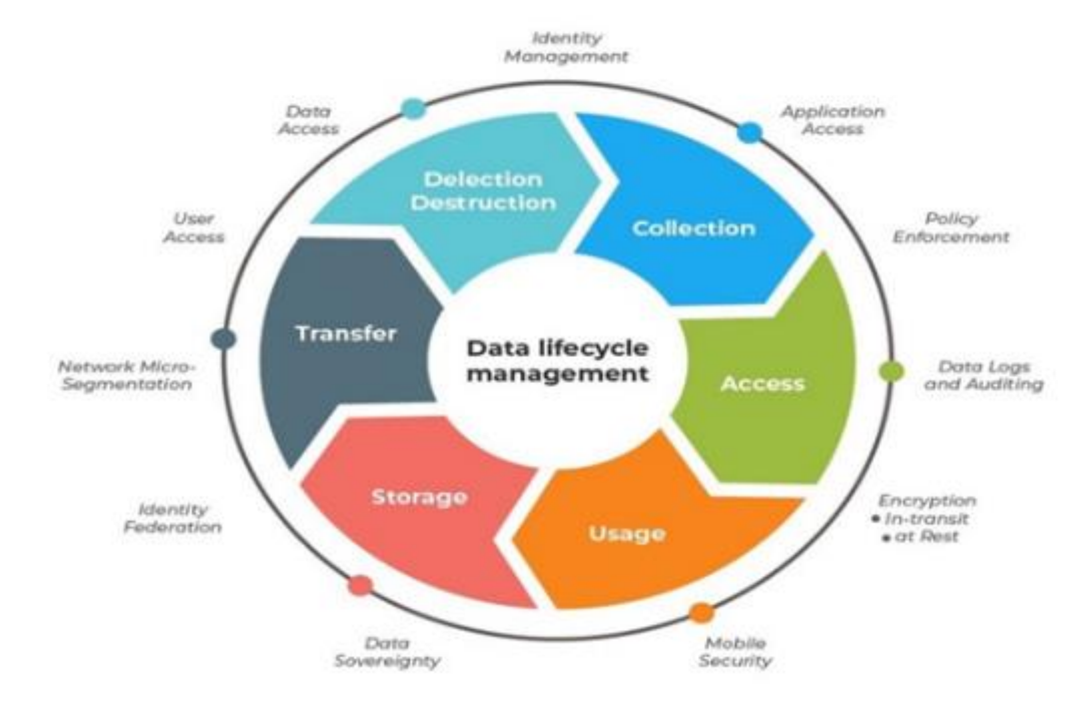

Figure 2 Data Lifecycle Management (DLM)

As shown in Figure 2, DLM involves the systematic management of data throughout its entire lifecycle, encompassing creation, storage, usage, and disposal. It includes processes for data classification, storage optimization, access controls, and archival, ensuring data is handled efficiently and in compliance with regulatory requirements. DLM aims to maximize the value of data while minimizing risks and costs, providing a structured framework for organizations to handle data from its inception to its eventual retirement.

4. Solutions for security and privacy issues in big data analytics

Addressing security and privacy issues in big data analytics requires a multifaceted approach, combining technological solutions, robust policies, and proactive measures to safeguard the integrity and confidentiality of vast and diverse datasets. According to [95], addressing security and privacy concerns in big data analytics necessitates a comprehensive set of solutions to safeguard sensitive information and ensure responsible data practices. Table 1 presents some of the solutions to security and privacy issues discussed in section 3 above.

Table 1 Solutions to big data analytics security and privacy issues

Threat	**Remedy**
Data Breaches	Implement robust encryption for data at rest and in transit, enforce access controls, and regularly conduct security audits and monitoring.
Unauthorized Access	Employ strong authentication mechanisms, implement role-based access controls, and conduct regular security training for personnel.
Data Residency and Sovereignty	Choose data storage solutions that comply with relevant regulations, implement geofencing, and establish clear data governance policies.

Inadequate Data Governance	Establish comprehensive data governance policies, enforce data quality standards, and conduct regular audits to ensure compliance.
De-identification Challenges	Implement advanced anonymization techniques, such as differential privacy, and stay informed about evolving re-identification risks.
Algorithmic Bias and Discrimination	Regularly audit algorithms for bias, promote transparency in algorithmic decision-making, and strive for diversity in dataset representation.
Privacy-preserving Analytics	Implement privacy-preserving techniques such as homomorphic encryption, secure multi-party computation, and federated learning to enable data analysis without exposing sensitive information.
Regulatory Compliance	Stay informed about relevant regulations, conduct impact assessments, and implement mechanisms for obtaining and managing consent.
Insider Threats	Implement employee training programs, enforce the principle of least privilege, and monitor user activities for anomalous behavior.
Data Lifecycle Management	Develop clear data lifecycle policies, including secure data disposal practices, and automate data retention and deletion processes.

Robust encryption techniques, both for data at rest and in transit, provide a foundational layer of protection against unauthorized access. Access controls, multi-factor authentication, and regular security audits bolster defenses, reducing the risk of insider threats and unauthorized data manipulation [96]-[100]. Privacy-preserving technologies, including advanced anonymization and differential privacy methods, enable organizations to glean meaningful insights while upholding individual privacy rights. Comprehensive data governance policies guide ethical data handling practices throughout the data lifecycle. Secure data transmission protocols and user education efforts further fortify security measures, while staying abreast of regulatory compliance ensures adherence to data protection laws. By embracing these solutions collectively, organizations can establish a resilient framework that not only mitigates risks but also fosters trust among stakeholders in the dynamic landscape of big data analytics. The sub-sections below describe some of these solutions in greater detail.

4.1. Encryption Techniques

Implementing robust encryption mechanisms for data at rest and in transit is fundamental. Techniques such as homomorphic encryption enable computations on encrypted data without exposing the raw information, ensuring privacy during processing. Encryption in big data involves the application of cryptographic techniques to secure sensitive information throughout its lifecycle, encompassing storage, transmission, and processing stages[101]-[105]. As big data environments deal with vast and diverse datasets, encryption serves as a fundamental safeguard against unauthorized access and potential breaches. In data storage, encryption transforms raw data into unreadable ciphertext, rendering it indecipherable without the appropriate decryption keys. During data transmission, secure communication protocols like SSL/TLS encrypt data flowing across networks, preventing eavesdropping and man-in-the-middle attacks. Encryption also plays a pivotal role in securing data processing by adopting techniques like homomorphic encryption, enabling computations on encrypted data without exposing the plaintext information [106], [107]. This holistic approach to encryption ensures the confidentiality and integrity of large-scale datasets, mitigating the risks associated with the dynamic landscape of big data analytics.

4.2. Access Controls and Authentication

Enforcing stringent access controls and multi-factor authentication mechanisms ensures that only authorized personnel can access and manipulate sensitive data. Access controls and authentication in big data are pivotal components of ensuring the security and integrity of vast and diverse datasets. Access controls involve defining and enforcing policies that determine which users or systems are granted permissions to access specific data or perform certain operations within the big data environment [108]-[110]. Authentication, on the other hand, verifies the identity of individuals or systems seeking access, typically through credentials like usernames and passwords, biometrics, or multi-factor authentication methods. In the context of big data analytics, robust access controls and authentication mechanisms are essential for mitigating the risks of unauthorized access, insider threats, and data manipulation [111], [112]. By implementing these measures, organizations can establish a secure framework that governs data access and usage, fostering a proactive defense against potential security breaches and ensuring the confidentiality of sensitive information throughout the analytics lifecycle.

4.3. Data Masking and Anonymization

Employing data masking and anonymization techniques helps in disguising specific information, preserving privacy while maintaining data utility. Data masking and anonymization in big data involve techniques that protect sensitive information by disguising or modifying identifiable attributes within datasets, thereby preserving privacy while maintaining data utility [113], [114]. Data masking entails replacing or scrambling specific data elements with fictitious or generalized values, ensuring that the masked dataset retains its analytical value without exposing sensitive details. Anonymization, on the other hand, involves removing or encrypting personally identifiable information to prevent re-identification. These privacy-enhancing measures are crucial in big data analytics, where diverse and expansive datasets often contain sensitive information [115]-[119]. By applying data masking and anonymization, organizations can comply with privacy regulations, mitigate the risk of data breaches, and facilitate the secure sharing of data for research or collaborative purposes without compromising individual privacy.

4.4. Secure Data Transmission Protocols

Utilizing secure communication protocols, such as HTTPS or SSL/TLS, ensures that data transmitted between systems is encrypted and protected from interception. Secure data transmission protocols in big data are fundamental to protecting information as it travels across networks or between distributed systems[120], [121]. These protocols, such as HTTPS (Hypertext Transfer Protocol Secure) or SSL/TLS (Secure Sockets Layer/Transport Layer Security), encrypt the data in transit, safeguarding it from unauthorized access, interception, or tampering. In the context of big data analytics, where large volumes of information flow between nodes and clusters, implementing secure transmission protocols is essential for maintaining the confidentiality and integrity of the data [122]-[125]. Figure 3 shows the SSL handshake procedures.

Figure 3 SSL Handshake process

These measures not only mitigate the risks of eavesdropping and man-in-the-middle attacks but also ensure that sensitive information remains protected during the dynamic processes of data exchange and communication within the extensive big data ecosystem.

4.5. Regular Security Audits and Monitoring

Conducting regular security audits and implementing continuous monitoring systems help identify and address vulnerabilities or suspicious activities promptly. Regular security audits and monitoring in big data are crucial components of maintaining a robust cyber-security posture [126], [127]. Figure 4 presents some of the concepts in security audit and monitoring.

Figure 4 Security Audits and Monitoring

Security audits involve systematic examinations of the big data infrastructure, configurations, and access controls to identify vulnerabilities, mis-configurations, or potential weaknesses. Continuous monitoring, on the other hand, involves real-time scrutiny of network activities, user behaviors, and system logs to detect and respond promptly to any suspicious or anomalous activities. In the context of big data analytics, where vast and complex datasets are processed, regular audits and monitoring provide insights into potential security threats, unauthorized access, or data breaches [128]-[130]. By proactively identifying and addressing security issues through these measures, organizations can enhance their ability to safeguard sensitive information and maintain the integrity of their big data environments.

4.6. Privacy-preserving Technologies

Leveraging advanced privacy-preserving technologies, such as differential privacy and secure multi-party computation, allows organizations to perform analytics on sensitive data without exposing individual-level details [131]. Privacy-preserving technologies in big data encompass a range of methods and tools designed to extract valuable insights from datasets while safeguarding individual privacy. Techniques like homomorphic encryption enable computations on encrypted data without revealing the underlying information, ensuring confidentiality during data processing. Differential privacy introduces noise or randomness to query responses, preventing the identification of specific individuals in the dataset. Additionally, secure multi-party computation allows parties to jointly compute functions over their inputs while keeping those inputs private [132]-[135]. These technologies are indispensable in the context of big data analytics, where the sheer scale and diversity of datasets necessitate innovative approaches to protect sensitive information. By adopting privacy-preserving technologies, organizations can strike a balance between deriving meaningful analytics and respecting the privacy rights of individuals within the evolving landscape of big data.

4.7. Comprehensive Data Governance

Establishing clear data governance policies, including data classification, access management, and data lifecycle management, helps ensure that data is handled responsibly and in compliance with regulations [136], [137]. Comprehensive data governance in big data involves the establishment of structured policies and frameworks to manage, protect, and ensure the quality of data throughout its entire lifecycle. This approach encompasses data classification, defining ownership and stewardship roles, and implementing access controls to govern who can access and modify specific datasets. It also involves developing clear data quality standards, metadata management, and establishing procedures for data retention and disposal. In big data analytics, where diverse and voluminous datasets are processed, comprehensive data governance is crucial for maintaining data integrity, complying with regulatory requirements, and fostering a culture of responsible and ethical data management [138]-[140]. By embracing a comprehensive data governance framework, organizations can enhance the reliability, security, and ethical use of data, ensuring that it remains a valuable asset in the analytics process.

4.8. User Education and Awareness

Conducting regular training programs to educate employees and stakeholders about security best practices and the importance of privacy protection fosters a security-conscious culture within the organization [141]. User education and awareness in big data involve fostering a culture of understanding and responsibility among individuals interacting with and managing large datasets. This includes educating users, ranging from data scientists and analysts to non-technical stakeholders, about the potential security and privacy implications of their actions within a big data environment [142]-[144]. Awareness programs cover topics such as data handling best practices, the importance of adhering to access controls, recognizing phishing attempts, and understanding the ethical considerations in data usage. By equipping users with the knowledge and awareness of potential risks, organizations can significantly reduce the likelihood of inadvertent security breaches, insider threats [145], and data misuse, contributing to a more secure and privacy-respecting big data ecosystem.

4.9. Regulatory Compliance Adherence

Staying informed about and adhering to data protection laws and industry regulations ensures that the organization remains compliant with regional and international standards. Adhering to regulatory compliance in big data involves navigating and satisfying the complex web of data protection laws, industry regulations, and privacy standards that govern the collection, storage, and processing of data[146], [147]. This includes compliance with frameworks such as GDPR, HIPAA, or CCPA, depending on the nature of the data and the geographic locations involved. Ensuring regulatory compliance in big data analytics requires organizations to implement robust data governance practices, enforce privacy-preserving technologies, and conduct thorough impact assessments to identify and mitigate potential risks [148]. Staying informed about evolving regulations, obtaining and managing user consent, and establishing transparent data practices are essential components of regulatory compliance in the dynamic and expansive landscape of big data analytics.

4.10. Ethical Use and Responsible AI Practices

Promoting ethical considerations in big data analytics, including fairness, transparency, and accountability, helps mitigate algorithmic biases and ensures responsible use of data. Ethical use and responsible AI practices in big data involve ensuring that the collection, processing, and utilization of data align with principles of fairness, transparency, and accountability [149], [150]. This encompasses scrutinizing algorithms for biases, addressing potential discrimination, and promoting transparency in decision-making processes. Organizations engaged in big data analytics must prioritize responsible AI practices, considering the societal impact of their data-driven initiatives. Striving for inclusivity and diversity in dataset representation, actively addressing biases, and regularly auditing algorithms for ethical implications are crucial components [151], [152]. By prioritizing ethical considerations, organizations can contribute to the development of trustworthy and socially responsible applications of big data analytics, fostering public trust and minimizing the risks of unintended consequences associated with algorithmic decision-making. Table 2 presents a summary of these solutions, including their strengths.

Table 2 Big data analytics security solutions and their strengths

Solution	Strength
Encryption Techniques	Protects sensitive data from unauthorized access and enhances confidentiality, especially when data is stored or transmitted across networks.
Access Controls and Authentication	Mitigates the risk of unauthorized access, limiting the potential for data breaches and insider threats.
Data Masking and Anonymization	Allows for meaningful analysis while minimizing the risk of re-identification and protecting individual privacy.
Secure Data Transmission Protocols	Safeguards against eavesdropping and unauthorized access during data transmission, maintaining the integrity and confidentiality of information.
Regular Security Audits and Monitoring	Enhances situational awareness, enabling organizations to proactively respond to potential security threats and ensure the ongoing effectiveness of security measures.
Privacy-preserving Technologies	Enables organizations to derive insights while respecting individual privacy rights and complying with data protection regulations.

Comprehensive Data Governance	Provides a framework for consistent and ethical data handling, reducing the risk of privacy violations and ensuring data quality and integrity.
User Education and Awareness	Minimizes the likelihood of human errors, mitigates insider threats, and enhances overall security awareness.
Regulatory Compliance Adherence	Mitigates legal risks, avoids regulatory penalties, and builds trust with customers and stakeholders.
Ethical Use and Responsible AI Practices	Fosters trust among users, minimizes the risk of discrimination, and aligns organizational practices with ethical standards.

Evidently, a holistic approach that integrates these solutions is essential for effectively addressing security and privacy challenges in big data analytics. Organizations must continually adapt and enhance their strategies to keep pace with evolving technologies, regulations, and threat landscapes, thereby ensuring responsible and secure use of large-scale data analytics.

5. Research gaps

Identifying and understanding research gaps in security and privacy issues in big data analytics is crucial for advancing the field and addressing emerging challenges. Several research gaps persist in the domain of security and privacy issues in big data analytics. First, there is a need for scalable privacy-preserving techniques that can efficiently handle the massive volumes of data processed in big data environments without compromising computational performance [153]-[155]. Additionally, research should focus on developing adaptive security measures capable of dynamically responding to the evolving threat landscape. Ethical considerations surrounding the use of artificial intelligence and algorithmic fairness in big data analytics require deeper exploration, with a focus on frameworks for assessing and mitigating biases. User-centric privacy solutions, including enhanced user consent mechanisms and data ownership models, demand more attention. The interoperability of diverse privacy-enhancing technologies within complex big data ecosystems remains a critical research area. Security and privacy challenges specific to edge computing, behavioral aspects of insider threats, and the integration of human factors into security measures represent additional gaps that require comprehensive investigation [156]-[160]. Finally, addressing cross-border data governance and regulatory compliance complexities is crucial for organizations operating in global contexts. Bridging these research gaps will contribute to the development of more resilient and ethical security and privacy solutions in the ever-evolving landscape of big data analytics. The sub-sections below describe some of these gaps in detail.

5.1. Privacy-preserving Analytics Scalability

As the scale of data continues to grow, ensuring privacy becomes increasingly challenging. Scalability in this context involves developing systems and algorithms that can handle massive datasets without compromising privacy. This entails implementing techniques such as differential privacy, secure multiparty computation, and homomorphic encryption to enable meaningful analysis while protecting sensitive information [161]-[165]. Achieving privacy-preserving analytics scalability is crucial for addressing the ethical and legal concerns associated with handling vast amounts of personal data in big data analytics, ensuring that data-driven insights can be derived responsibly and in compliance with privacy regulations. While privacy-preserving techniques like homomorphic encryption and differential privacy show promise, scalability remains a significant challenge. Research is needed to develop scalable solutions that can handle the massive volumes of data processed in big data analytics without sacrificing computational efficiency.

5.2. Dynamic Threat Landscape and Adaptive Security

As organizations increasingly rely on big data analytics to extract valuable insights, they become susceptible to a variety of cyber threats, including sophisticated attacks and evolving tactics. Adaptive security in this context involves implementing dynamic and responsive measures to counteract these threats effectively. This may include real-time monitoring, machine learning algorithms for anomaly detection, and adaptive access controls [166]-[170]. By continuously adapting security measures to the changing threat landscape, organizations can enhance their resilience against cyber threats in the realm of big data analytics, ensuring the confidentiality, integrity, and availability of critical data and insights. The evolving threat landscape demands research on adaptive security measures capable of dynamically responding to emerging cyber threats. Investigating real-time threat intelligence and self-learning security systems tailored to big data environments is essential for staying ahead of evolving security risks.

5.3. Ethical Use of AI and Algorithmic Fairness

The ethical use of AI and algorithmic fairness in big data analytics is paramount to ensure equitable and unbiased outcomes in decision-making processes. As big data analytics increasingly influences various aspects of society, from healthcare to finance, there is a growing concern about the potential reinforcement of existing biases and discrimination in algorithmic decision models. Addressing this requires a commitment to ethical principles, transparency, and the incorporation of fairness considerations throughout the data analytics pipeline. This involves careful examination and mitigation of biases in data collection, preprocessing, and model training stages [171]-[175]. Implementing measures such as explainability in algorithms, regular audits, and diverse representation in dataset curation can contribute to promoting fairness and mitigating unintended consequences, fostering responsible and ethical use of AI in the context of big data analytics. The ethical implications of AI algorithms used in big data analytics, including issues of bias and fairness, require further exploration. Research should focus on developing frameworks for assessing and mitigating algorithmic biases, ensuring fair and equitable outcomes in decision-making processes.

5.4. User-Centric Privacy Solutions

User-centric privacy solutions in big data analytics prioritize the protection of individuals' privacy while still enabling meaningful data analysis. These solutions focus on empowering users with greater control over their personal information, ensuring transparency, and obtaining informed consent. Implementing privacy-preserving techniques such as anonymization, pseudonymization, and secure data aggregation helps minimize the risk of unauthorized disclosure of sensitive details. Additionally, incorporating user-friendly interfaces for privacy settings and providing clear communication about data usage practices fosters a more transparent and user-centric approach [176]-[180]. Striking a balance between data utility and privacy concerns is essential to building trust among users, promoting ethical data handling practices, and ensuring that big data analytics can deliver valuable insights without compromising individual privacy rights. Research should delve into user-centric privacy solutions that empower individuals to have greater control over their data. This includes exploring technologies and mechanisms for enhanced user consent, data ownership, and transparency in data collection and processing practices.

5.5. Interoperability of Privacy-enhancing Technologies

The interoperability of privacy-enhancing technologies (PETs) in big data analytics is crucial for fostering a cohesive and effective approach to safeguarding privacy across diverse systems and applications. As organizations deploy various PETs like homomorphic encryption, differential privacy, and secure multi-party computation, ensuring seamless integration and communication between these technologies becomes paramount. Interoperability allows for the creation of a comprehensive privacy infrastructure, enabling different components to work together harmoniously while preserving data confidentiality [181]-[185]. This ensures that privacy measures can be uniformly applied throughout the big data analytics ecosystem, promoting standardized practices and facilitating collaborative efforts in addressing privacy concerns across platforms and applications. Such interoperability is instrumental in building a robust and consistent framework for privacy protection in the evolving landscape of big data analytics. Investigating the interoperability of various privacy-preserving technologies is vital to create a cohesive and standardized approach. Research should explore how different privacy solutions can seamlessly work together within complex big data ecosystems while maintaining efficiency and effectiveness.

5.6. Security and Privacy in Edge Computing

Owing to data processing occurring closer to the source, often at the edge devices themselves, there's a need to implement robust security measures to protect sensitive information. Figure 5 shows an edge computing architecture deployed in many organizations. Edge computing introduces new challenges such as the potential exposure of data at the edge, making it susceptible to various cyber threats. Privacy concerns arise as personal and sensitive data may be processed locally. Implementing strong encryption, secure communication protocols, and access controls becomes essential to safeguard data integrity and confidentiality [186]-[190]. Furthermore, privacy-preserving techniques must be integrated into edge analytics to ensure responsible handling of information, considering regulatory requirements and user expectations. Balancing security and privacy in edge computing for big data analytics is pivotal to harness the benefits of distributed processing while mitigating the associated risks. With the rise of edge computing in processing data closer to the source, understanding and addressing security and privacy challenges specific to edge environments is critical.

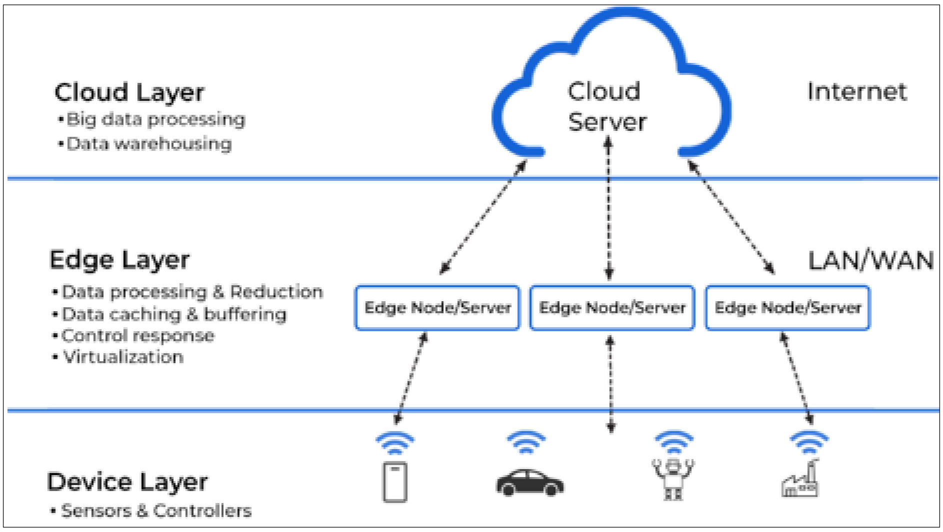

Figure 5 Edge computing architecture

Research should explore novel approaches for securing data at the edge, considering factors like resource constraints and distributed computing models.

5.7. Behavioral Aspects of Insider Threats

The behavioral aspects of insider threats in big data analytics delve into the human elements that contribute to potential security risks from within an organization. Insider threats involve individuals with authorized access who, either intentionally or unintentionally, compromise data confidentiality, integrity, or availability. In the context of big data analytics, understanding the behavioral patterns of employees, contractors, or collaborators is crucial. This includes recognizing signs of disgruntlement, identifying abnormal data access patterns, or monitoring changes in work behavior that may indicate a potential insider threat [191]-[195]. Behavioral analytics, incorporating machine learning and anomaly detection, can play a pivotal role in predicting, detecting, and responding to insider threats in a proactive manner, thereby enhancing the overall security posture of big data analytics environments. Combining technological solutions with an awareness of human behavior is essential to mitigate the risks associated with insider threats in the realm of big data analytics. While there is recognition of insider threats, research should delve deeper into the behavioral aspects of insiders with privileged access in big data analytics environments. Understanding the motivations, patterns, and early indicators of malicious or unintentional insider activities can inform more effective prevention and detection strategies.

5.8. Regulatory Compliance and Cross-Border Data Governance

Regulatory compliance and cross-border data governance are critical considerations in big data analytics, given the global nature of data flows and the diversity of data protection laws. Organizations engaged in big data analytics must navigate a complex landscape of regulations, such as the General Data Protection Regulation (GDPR), the Health Insurance Portability and Accountability Act (HIPAA), and other regional or industry-specific mandates. Ensuring compliance involves understanding and adhering to the principles of data minimization, purpose limitation, and transparency. Cross-border data governance is particularly challenging, requiring strategies for managing data sovereignty issues and addressing the varying legal frameworks across jurisdictions [192]-[199]. Implementing robust data governance policies, encryption protocols, and secure data transfer mechanisms are essential to meet regulatory requirements and build trust among users and stakeholders, thereby fostering responsible and lawful big data analytics practices on a global scale. Research is needed to navigate the complexities of cross-border data governance and ensure compliance with diverse and evolving privacy regulations. Addressing challenges related to conflicting legal frameworks, data sovereignty, and international data transfers is crucial for organizations operating in global contexts.

5.9. Adversarial Machine Learning in Big Data

Adversarial machine learning in big data analytics refers to the vulnerabilities and challenges posed by deliberate attempts to manipulate or deceive machine learning models. Adversarial attacks can take various forms, including injecting malicious data, exploiting model vulnerabilities, or employing sophisticated techniques to generate misleading inputs. Figure 6 presents a typical adversarial machine learning attack. As shown in Figure 6, adversarial machine learning involves the deliberate manipulation of machine learning models by exploiting vulnerabilities, injecting misleading data, or employing deceptive techniques to compromise the accuracy and reliability of analytical outcomes. In the vast landscape of big data analytics, where intricate models process massive datasets, adversarial attacks pose significant challenges. Threats can manifest as crafted inputs designed to mislead models, leading to erroneous predictions or compromising the confidentiality of sensitive information. Addressing Adversarial Machine Learning in Big Data requires deploying advanced defenses like adversarial training, robust anomaly detection, and continuous monitoring to fortify machine learning systems against intentional manipulations, ensuring the trustworthiness and security of data-driven insights.

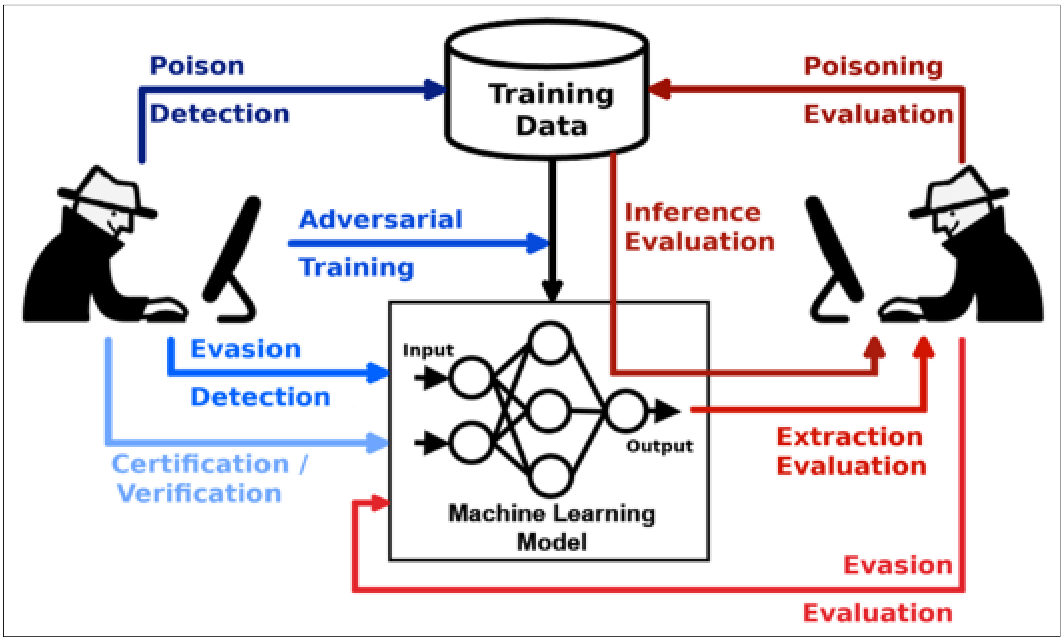

Figure 6 Adversarial machine learning

In the context of big data analytics, where large datasets are used to train complex models, the potential impact of adversarial attacks is amplified. Adversarial machine learning aims to compromise the accuracy and reliability of predictive models, which can have serious consequences in critical domains such as finance, healthcare, and security [200]-[205]. Mitigating adversarial threats in big data analytics involves deploying robust defenses such as adversarial training, anomaly detection, and continuous model monitoring to enhance the resilience of machine learning systems against intentional manipulations, ensuring the trustworthiness of analytical results in the face of adversarial challenges. Investigating the vulnerability of machine learning models to adversarial attacks within big data analytics is essential. Research should focus on developing resilient models and algorithms that can withstand intentional manipulation attempts aimed at undermining the integrity and effectiveness of analytics outcomes.

5.10. Integration of Human Factors in Security and Privacy

The integration of human factors in security and privacy within big data analytics recognizes the pivotal role of individuals in the overall data protection framework. Understanding and addressing human behaviors, perceptions, and cognitive biases are essential for designing effective security and privacy measures. In big data analytics, where the human element is deeply involved in data handling and decision-making, considerations such as user awareness, training, and the design of user interfaces become crucial. Incorporating a human-centric approach involves educating users about security best practices, fostering a culture of privacy awareness, and ensuring that security protocols are intuitive and user-friendly [206]-[210]. Additionally, it involves recognizing the impact of human factors on potential vulnerabilities, such as unintentional data leaks or social engineering attacks. By integrating human factors into the design and implementation of security and privacy measures, organizations can enhance the overall resilience of big data analytics systems while promoting a collaborative and informed approach to safeguarding sensitive information. Considering the role of human factors in security and privacy is often overlooked. Research should explore the human element in the context of big data analytics, including user behaviors, awareness, and perceptions, to design more effective security and privacy solutions that align with user expectations.

Addressing these research gaps is essential to fortify the foundations of security and privacy in big data analytics, enabling the development of innovative and effective solutions that meet the evolving challenges of the digital landscape.

6. Conclusion

This paper has extensively examined the multifaceted landscape of security and privacy issues in big data analytics, revealing the intricate challenges and complexities inherent in the processing and analysis of vast and diverse datasets. The research has underscored the critical importance of addressing these issues to ensure the responsible and ethical use of big data, balancing the imperative for valuable insights with the need to protect individual privacy and maintain data integrity. The identified solutions, ranging from encryption techniques and access controls to privacy-preserving technologies and user-centric approaches, provide a roadmap for organizations to fortify their defenses against evolving threats. However, as the digital landscape continues to evolve, with new technologies and regulatory frameworks emerging, it is evident that the journey towards securing and preserving privacy in big data analytics is an ongoing and dynamic process. Bridging the research gaps identified in this study will be imperative for staying ahead of the curve and fostering an environment where data-driven innovations can flourish responsibly. Ultimately, a concerted effort from researchers, practitioners, and policymakers is required to navigate these challenges and establish a secure and privacy-respecting foundation for the future of big data analytics.

Compliance with ethical standards

Disclosure of conflict of interest

The author declares that she has no any conflict of interest.

References

[1] Soto-Acosta P. Navigating Uncertainty: Post-Pandemic Issues on Digital Transformation. Information Systems Management. 2024 Jan 2, 41(1):20-6.

[2] Panori A, Kakderi C, Komninos N. Transformation of smart city public services through AI and big data analytics: towards universal cross-sector solutions. InHandbook of Research on Artificial Intelligence, Innovation and Entrepreneurship 2023 Feb 14 (pp. 292-307). Edward Elgar Publishing.

[3] Asikpo NA. Impact of Digital Transformation on Financial Reporting in the 21st Century. International Journal of Comparative Studies and Smart Education. 2024 Jan 6, 1(1):34-45.

[4] Silitonga D, Rohmayanti SA, Aripin Z, Kuswandi D, Sulistyo AB. Edge Computing in E-commerce Business: Economic Impacts and Advantages of Scalable Information Systems. EAI Endorsed Transactions on Scalable Information Systems. 2024, 11(1).

[5] Al-Chaab W, Abduljabbar ZA, Abood EW, Nyangaresi VO, Mohammed HM, Ma J. Secure and Low-Complexity Medical Image Exchange Based on Compressive Sensing and LSB Audio Steganography. Informatica. 2023 May 31, 47(6).

[6] Tariq U, Ahmed I, Bashir AK, Shaukat K. A Critical Cybersecurity Analysis and Future Research Directions for the Internet of Things: A Comprehensive Review. Sensors. 2023 Apr 19, 23(8):4117.

[7] Sharma P, Barua S. From data breach to data shield: the crucial role of big data analytics in modern cybersecurity strategies. International Journal of Information and Cybersecurity. 2023 Sep 5, 7(9):31-59.

[8] Ahmed S, Khan M. Securing the Internet of Things (IoT): A Comprehensive Study on the Intersection of Cybersecurity, Privacy, and Connectivity in the IoT Ecosystem. AI, IoT and the Fourth Industrial Revolution Review. 2023 Sep 16, 13(9):1-7.

[9] Nag A, Hassan MM, Das A, Sinha A, Chand N, Kar A, Sharma V, Alkhayyat A. Exploring the applications and security threats of Internet of Thing in the cloud computing paradigm: A comprehensive study on the cloud of things. Transactions on Emerging Telecommunications Technologies.:e4897.

[10] Nyangaresi VO. Extended Chebyshev Chaotic Map Based Message Verification Protocol for Wireless Surveillance Systems. InComputer Vision and Robotics: Proceedings of CVR 2022 2023 Apr 28 (pp. 503-516). Singapore: Springer Nature Singapore.

[11] Abba Ari AA, Ngangmo OK, Titouna C, Thiare O, Mohamadou A, Gueroui AM. Enabling privacy and security in Cloud of Things: Architecture, applications, security & privacy challenges. Applied Computing and Informatics. 2024 Jan 5, 20(1/2):119-41.

[12] Jaramillo M, Pavón W, Jaramillo L. Adaptive Forecasting in Energy Consumption: A Bibliometric Analysis and Review. Data. 2024 Jan 11, 9(1):13.

[13] Gupta U, Sharma R. Apache Hadoop framework for big data analytics using AI. InArtificial Intelligence and Blockchain in Industry 4.0 2024 (pp. 130-140). CRC Press.

[14] Azeem M, Abualsoud BM, Priyadarshana D. Mobile Big Data Analytics Using Deep Learning and Apache Spark. Mesopotamian Journal of Big Data. 2023 Feb 7, 2023:18-30.

[15] Nyangaresi VO, El-Omari NK, Nyakina JN. Efficient Feature Selection and ML Algorithm for Accurate Diagnostics. Journal of Computer Science Research. 2022 Jan 25, 4(1):10-9.

[16] Sen PS, Mukherjee N. An ontology-based approach to designing a NoSQL database for semi-structured and unstructured health data. Cluster Computing. 2023 Apr 8:1-8.

[17] Deepthi BG, Rani KS, Krishna PV, Saritha V. An efficient architecture for processing real-time traffic data streams using apache flink. Multimedia Tools and Applications. 2023 Sep 30:1-7.

[18] Almeida A, Brás S, Sargento S, Pinto FC. Time series big data: a survey on data stream frameworks, analysis and algorithms. Journal of Big Data. 2023 May 28, 10(1):83.

[19] Tantalaki N, Souravlas S, Roumeliotis M. A review on big data real-time stream processing and its scheduling techniques. International Journal of Parallel, Emergent and Distributed Systems. 2020 Sep 2, 35(5):571-601.

[20] Nyangaresi VO, Al Sibahee MA, Abduljabbar ZA, Alhassani A, Abduljaleel IQ, Abood EW. Intelligent Target Cell Selection Algorithm for Low Latency 5G Networks. InAdvances in Computational Intelligence and Communication: Selected Papers from the 2nd EAI International Conference on Computational Intelligence and Communications (CICom 2021) 2022 Dec 14 (pp. 79-97). Cham: Springer International Publishing.

[21] Raptis TP, Cicconetti C, Passarella A. Efficient topic partitioning of Apache Kafka for high-reliability real-time data streaming applications. Future Generation Computer Systems. 2024 Jan 8.

[22] Yang L, He W, Qiang X, Zheng J, Huang F. Research on remote sensing image storage management and a fast visualization system based on cloud computing technology. Multimedia Tools and Applications. 2024 Jan 2:1-26.

[23] Manchanda A. Computational Intelligence for Big Data Analysis. InComputational Science and Its Applications 2024 Jan 9 (pp. 199-230). Apple Academic Press.

[24] Szafir MF. Digital Transformation Enabled by Big Data. IEEE Technology and Engineering Management Society Body of Knowledge (TEMSBOK). 2023 Dec 8:297-329.

[25] Hussain MA, Hussien ZA, Abduljabbar ZA, Ma J, Al Sibahee MA, Hussain SA, Nyangaresi VO, Jiao X. Provably throttling SQLI using an enciphering query and secure matching. Egyptian Informatics Journal. 2022 Dec 1, 23(4):145-62.

[26] Horvath K, Abid MR, Merino T, Zimmerman R, Peker Y, Khan S. Cloud-Based Infrastructure and DevOps for Energy Fault Detection in Smart Buildings. Computers. 2024 Jan 16, 13(1):23.

[27] Wawrzoniak M, Moro G, Fraga Barcelos Paulus Bruno R, Klimovic A, Alonso G. Off-the-shelf Data Analytics on Serverless. InCIDR'24: 14th Annual Conference on Innovative Data Systems Research 2024 Jan 14.

[28] Darius PS, Sowjanya K, Manju VN, Saha S, Mitra P, Majumder P, Suneetha J, Prabhu SM. From Data to Insights: A Review of Cloud-Based Big Data Tools and Technologies. Big Data Computing. 2024:86-110.

[29] Kumari A, Patra MK, Sahoo B. Data Controlling and Security Issues in Cloud: A Step Towards Serverless. InPerspectives on Social Welfare Applications’ Optimization and Enhanced Computer Applications 2023 (pp. 105-124). IGI Global.

[30] Nyangaresi VO. Target Tracking Area Selection and Handover Security in Cellular Networks: A Machine Learning Approach. InProceedings of Third International Conference on Sustainable Expert Systems: ICSES 2022 2023 Feb 23 (pp. 797-816). Singapore: Springer Nature Singapore.

[31] Shukla S, Bisht K, Tiwari K, Bashir S. Comparative Study of the Global Data Economy. InData Economy in the Digital Age 2023 Nov 15 (pp. 63-86). Singapore: Springer Nature Singapore.

[32] Debbarma R. The changing landscape of privacy laws in the age of big data and surveillance. Rivista Italiana di Filosofia Analitica Junior. 2023 Sep 10, 14(2):1740-52.

[33] Ahmed A, Xi R, Hou M, Shah SA, Hameed S. Harnessing big data analytics for healthcare: A comprehensive review of frameworks, implications, applications, and impacts. IEEE Access. 2023 Oct 10.

[34] Rosário AT, Dias JC. How has data-driven marketing evolved: Challenges and opportunities with emerging technologies. International Journal of Information Management Data Insights. 2023 Nov 1, 3(2):100203.

[35] Umran SM, Lu S, Abduljabbar ZA, Nyangaresi VO. Multi-chain blockchain based secure data-sharing framework for industrial IoTs smart devices in petroleum industry. Internet of Things. 2023 Dec 1, 24:100969.

[36] Shukla S, George JP, Tiwari K, Kureethara JV. Data Ethics and Challenges. Springer, 2022 Mar 31.

[37] Rao NT, Bhattacharyya D, Joshua ES. An extensive discussion on utilization of data security and big data models for resolving healthcare problems. InMulti-chaos, fractal and multi-fractional artificial intelligence of different complex systems 2022 Jan 1 (pp. 311-324). Academic Press.

[38] Djenna A, Harous S, Saidouni DE. Internet of things meet internet of threats: New concern cyber security issues of critical cyber infrastructure. Applied Sciences. 2021 May 17, 11(10):4580.

[39] Rasool RU, Ahmad HF, Rafique W, Qayyum A, Qadir J. Security and privacy of internet of medical things: A contemporary review in the age of surveillance, botnets, and adversarial ML. Journal of Network and Computer Applications. 2022 May 1, 201:103332.

[40] Nyangaresi VO. Provably secure authentication protocol for traffic exchanges in unmanned aerial vehicles. High-Confidence Computing. 2023 Sep 15:100154.

[41] Ghosh J, Sinha VK. Big Data Analytics in Industry 4.0 in Legal Perspective: Past, Present and Future. InDigital Transformation: Industry 4.0 to Society 5.0 2024 Jan 30 (pp. 151-177). Singapore: Springer Nature Singapore.

[42] Wilkinson D, Christie A, Tarr AA, Tarr JA. Big Data, Artificial Intelligence and Insurance. InThe Global Insurance Market and Change 2024 (pp. 22-46). Informa Law from Routledge.

[43] Dong Q, Wu Y, Lin H, Sun Z, Liang R. Fostering green innovation for corporate competitive advantages in big data era: The role of institutional benefits. Technology Analysis & Strategic Management. 2024 Feb 1, 36(2):181-94.

[44] Durst S, Hinteregger C, Zieba M. The effect of environmental turbulence on cyber security risk management and organizational resilience. Computers & Security. 2024 Feb 1, 137:103591.

[45] Al-Sulami ZA, Abduljabbar ZA, Nyangaresi VO, Ma J. Knowledge Management and its Role in the Development of a Smart University in Iraq. TEM Journal. 2023 Aug 1, 12(3):1582.

[46] Ahmadi S. A Comprehensive Study on Integration of Big Data and AI in Financial Industry and Its Effect on Present and Future Opportunities. International Journal of Current Science Research and Review. 2024 Jan 6, 7(01).

[47] Faaique M. Overview of Big Data Analytics in Modern Astronomy. International Journal of Mathematics, Statistics, and Computer Science. 2024, 2:96-113.

[48] Rajeshkumar K, Dhanasekaran S, Vasudevan V. A novel three-factor authentication and optimal mapreduce frameworks for secure medical big data transmission over the cloud with shaxecc. Multimedia Tools and Applications. 2024 Jan 26:1-29.

[49] Sakthivel G, Madhubala P. Advanced set containment deep learned Rabin certificateless signcryption for secured transmission with big data in cloud. Concurrency and Computation: Practice and Experience. 2024 Jan 10, 36(1):e7883.

[50] Nyangaresi VO, Moundounga AR. Secure data exchange scheme for smart grids. In2021 IEEE 6th International Forum on Research and Technology for Society and Industry (RTSI) 2021 Sep 6 (pp. 312-316). IEEE.

[51] Glasze G, Cattaruzza A, Douzet F, Dammann F, Bertran MG, Bômont C, Braun M, Danet D, Desforges A, Géry A, Grumbach S. Contested spatialities of digital sovereignty. Geopolitics. 2023 Mar 15, 28(2):919-58.

[52] Arner DW, Castellano GG, Selga EK. The transnational data governance problem. Berkeley Tech. LJ. 2022, 37:623.

[53] Kotsev A, Minghini M, Tomas R, Cetl V, Lutz M. From spatial data infrastructures to data spaces—A technological perspective on the evolution of European SDIs. ISPRS International Journal of Geo-Information. 2020 Mar 16, 9(3):176.

[54] Coche E, Kolk A, Ocelík V. Unravelling cross-country regulatory intricacies of data governance: the relevance of legal insights for digitalization and international business. Journal of International Business Policy. 2023 Oct 6:1-6.

[55] Zhang H, Ma J, Qiu Z, Yao J, Sibahee MA, Abduljabbar ZA, Nyangaresi VO. Multi-GPU Parallel Pipeline Rendering with Splitting Frame. InComputer Graphics International Conference 2023 Aug 28 (pp. 223-235). Cham: Springer Nature Switzerland.

[56] Oladoyinbo TO, Olabanji SO, Olaniyi OO, Adebiyi OO, Okunleye OJ, Alao AI. Exploring the challenges of artificial intelligence in data integrity and its influence on social dynamics. Asian Journal of Advanced Research and Reports. 2024 Jan 13, 18(2):1-23.

[57] Kumbhare A, Thakur PK, Patnaik BR, Midiyam K. Blockchain's Data Integrity and Reliability. InBuilding Secure Business Models Through Blockchain Technology: Tactics, Methods, Limitations, and Performance 2023 (pp. 231-250). IGI Global.

[58] Thulare T, Herselman M, Botha A. A scoping review on identifying aspects of data integrity in health information systems for South Africa. In2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD) 2020 Aug 6 (pp. 1-8). IEEE.

[59] Saxena UR, Alam T. Provisioning trust-oriented role-based access control for maintaining data integrity in cloud. International Journal of System Assurance Engineering and Management. 2023 Dec, 14(6):2559-78.

[60] Nyangaresi VO. Masked Symmetric Key Encrypted Verification Codes for Secure Authentication in Smart Grid Networks. In2022 4th Global Power, Energy and Communication Conference (GPECOM) 2022 Jun 14 (pp. 427-432). IEEE.

[61] Oh J, Lee K. Data De-identification Framework. Computers, Materials & Continua. 2023 Feb 1, 74(2).

[62] Şahin Y, DOGRU İ. An Enterprise Data Privacy Governance Model: Security-Centric Multi-Model Data Anonymization. International Journal of Engineering Research and Development. 2023, 15(2):574-83.

[63] Sampaio S, Sousa PR, Martins C, Ferreira A, Antunes L, Cruz-Correia R. Collecting, Processing and Secondary Using Personal and (Pseudo) Anonymized Data in Smart Cities. Applied Sciences. 2023 Mar 16, 13(6):3830.

[64] Chevrier R, Foufi V, Gaudet-Blavignac C, Robert A, Lovis C. Use and understanding of anonymization and de-identification in the biomedical literature: scoping review. Journal of medical Internet research. 2019 May 31, 21(5):e13484.

[65] Kumar S, Chinthaginjala R, Anbazhagan R, Nyangaresi VO, Pau G, Varma PS. Submarine Acoustic Target Strength Modelling at High-Frequency Asymptotic Scattering. IEEE Access. 2024 Jan 1.

[66] Maity A, More R, Kambli G, Ambadekar S. Preserving Privacy in Video Analytics: A Comprehensive Review of Face De-Identification and Background Blurring Techniques. Authorea Preprints. 2023 Dec 7.

[67] Ye M, Shen W, Zhang J, Yang Y, Du B. SecureReID: Privacy-Preserving Anonymization for Person Re-Identification. IEEE Transactions on Information Forensics and Security. 2024 Jan 19.

[68] Thantilage RD, Le-Khac NA, Kechadi MT. Healthcare data security and privacy in Data Warehouse architectures. Informatics in Medicine Unlocked. 2023 May 12:101270.

[69] Neves F, Souza R, Sousa J, Bonfim M, Garcia V. Data privacy in the Internet of Things based on anonymization: A review. Journal of Computer Security. 2023(Preprint):1-31.

[70] Nyangaresi VO, Ogundoyin SO. Certificate based authentication scheme for smart homes. In2021 3rd Global Power, Energy and Communication Conference (GPECOM) 2021 Oct 5 (pp. 202-207). IEEE.

[71] Arowosegbe JO. Data bias, intelligent systems and criminal justice outcomes. International Journal of Law and Information Technology. 2023 Mar 1, 31(1):22-45.

[72] Kuiler EW, McNeely CL. Panopticon implications of ethical AI: equity, disparity, and inequality in healthcare. InAI Assurance 2023 Jan 1 (pp. 429-451). Academic Press.

[73] Behrendt H, Loh W. Informed consent and algorithmic discrimination–is giving away your data the new vulnerable?. Review of Social Economy. 2022 Jan 2, 80(1):58-84.

[74] Mühlhoff R. Predictive privacy: towards an applied ethics of data analytics. Ethics and Information Technology. 2021 Dec, 23(4):675-90.

[75] Al Sibahee MA, Nyangaresi VO, Abduljabbar ZA, Luo C, Zhang J, Ma J. Two-Factor Privacy Preserving Protocol for Efficient Authentication in Internet of Vehicles Networks. IEEE Internet of Things Journal. 2023 Dec 7.

[76] Ahmed SF, Alam MS, Afrin S, Rafa SJ, Taher SB, Kabir M, Muyeen SM, Gandomi AH. Towards a secure 5G-enabled Internet of Things: A survey on requirements, privacy, security, challenges, and opportunities. IEEE Access. 2024 Jan 10.

[77] Dhinakaran D, Sankar SM, Selvaraj D, Raja SE. Privacy-Preserving Data in IoT-based Cloud Systems: A Comprehensive Survey with AI Integration. arXiv preprint arXiv:2401.00794. 2024 Jan 1.

[78] Habbal A, Ali MK, Abuzaraida MA. Artificial Intelligence Trust, Risk and Security Management (AI TRiSM): Frameworks, applications, challenges and future research directions. Expert Systems with Applications. 2024 Apr 15, 240:122442.

[79] Ajani SN, Khobragade P, Dhone M, Ganguly B, Shelke N, Parati N. Advancements in Computing: Emerging Trends in Computational Science with Next-Generation Computing. International Journal of Intelligent Systems and Applications in Engineering. 2024, 12(7s):546-59.

[80] Nyangaresi VO, Morsy MA. Towards privacy preservation in internet of drones. In2021 IEEE 6th International Forum on Research and Technology for Society and Industry (RTSI) 2021 Sep 6 (pp. 306-311). IEEE.

[81] Winter JS, Davidson E. Harmonizing regulatory regimes for the governance of patient-generated health data. Telecommunications Policy. 2022 Jun 1, 46(5):102285.

[82] Park SE. Technological convergence: Regulatory, digital privacy, and data security issues. Congressional Research Service, Tech. Rep. 2019 May 30.

[83] Winter JS, Davidson EJ. Harmonizing regulatory spheres to overcome challenges for governance of patient-generated health data in the age of artificial intelligence and big data. InTPRC48: The 48th research conference on communication, information and internet Policy 2020 Dec 15.

[84] Saeed MM, Saeed RA, Ahmed ZE. Data Security and Privacy in the Age of AI and Digital Twins. InDigital Twin Technology and AI Implementations in Future-Focused Businesses 2024 (pp. 99-124). IGI Global.

[85] Abduljabbar ZA, Abduljaleel IQ, Ma J, Al Sibahee MA, Nyangaresi VO, Honi DG, Abdulsada AI, Jiao X. Provably secure and fast color image encryption algorithm based on s-boxes and hyperchaotic map. IEEE Access. 2022 Feb 11, 10:26257-70.

[86] Hasan M, Hoque A, Le T. Big data-driven banking operations: Opportunities, challenges, and data security perspectives. FinTech. 2023 Jul 19, 2(3):484-509.

[87] Sundaram A, Abdel-Khalik HS, Ashy O. A data analytical approach for assessing the efficacy of Operational Technology active defenses against insider threats. Progress in Nuclear Energy. 2020 Jun 1, 124:103339.

[88] Lavanya P, Shankar Sriram VS. Detection of insider threats using deep learning: a review. Computational Intelligence in Data Mining: Proceedings of ICCIDM 2021. 2022 May 7:41-57.

[89] Telo J. Smart City Security Threats and Countermeasures in the Context of Emerging Technologies. International Journal of Intelligent Automation and Computing. 2023 Feb 27, 6(1):31-45.

[90] Nyangaresi VO. Privacy Preserving Three-factor Authentication Protocol for Secure Message Forwarding in Wireless Body Area Networks. Ad Hoc Networks. 2023 Apr 1, 142:103117.

[91] Namose MS, Hiwarkar T. Exploring Unconventional Sources in Big Data: A Comprehensive Data Lifecycle for Social and Economic Analysis. International Journal of Intelligent Systems and Applications in Engineering. 2024, 12(1s):809-20.

[92] Munappy AR, Bosch J, Olsson HH, Arpteg A, Brinne B. Data management for production quality deep learning models: Challenges and solutions. Journal of Systems and Software. 2022 Sep 1, 191:111359.

[93] Abughazala M. Architecting Data-Intensive Applications: From Data Architecture Design to Its Quality Assurance. arXiv preprint arXiv:2401.12011. 2024 Jan 22.

[94] Nikolakopoulos A, Julian Segui M, Pellicer AB, Kefalogiannis M, Gizelis CA, Marinakis A, Nestorakis K, Varvarigou T. BigDaM: Efficient Big Data Management and Interoperability Middleware for Seaports as Critical Infrastructures. Computers. 2023 Oct 27, 12(11):218.

[95] Abood EW, Abdullah AM, Al Sibahe MA, Abduljabbar ZA, Nyangaresi VO, Kalafy SA, Ghrabta MJ. Audio steganography with enhanced LSB method for securing encrypted text with bit cycling. Bulletin of Electrical Engineering and Informatics. 2022 Feb 1, 11(1):185-94.

[96] Aldboush HH, Ferdous M. Building Trust in Fintech: An Analysis of Ethical and Privacy Considerations in the Intersection of Big Data, AI, and Customer Trust. International Journal of Financial Studies. 2023 Jul 10, 11(3):90.

[97] Wylde V, Prakash E, Hewage C, Platts J. The Use of AI in Managing Big Data Analysis Demands: Status and Future Directions. Artificial Intelligence and National Security. 2022 May 5:47-67.

[98] Villegas-Ch W, García-Ortiz J. Toward a Comprehensive Framework for Ensuring Security and Privacy in Artificial Intelligence. Electronics. 2023 Sep 7, 12(18):3786.

[99] Cheryl BK, Ng BK. Protecting the unprotected consumer data in internet of things: Current scenario of data governance in Malaysia. Sustainability. 2022 Aug 10, 14(16):9893.

[100] Nyangaresi VO. A Formally Verified Authentication Scheme for mmWave Heterogeneous Networks. Inthe 6th International Conference on Combinatorics, Cryptography, Computer Science and Computation (605-612) 2021.

[101] Sutar S, Jose K, Gaikwad V, Mishra V, Wankhede D, Karnik M. Enhancing Data Management: An Integrated Solution for Database Backup, Recovery, Conversion, and Encryption Capabilities. International Journal of Intelligent Systems and Applications in Engineering. 2024, 12(6s):720-34.

[102] Bande V, Raju BD, Rao KP, Joshi S, Bajaj SH, Sarala V. Designing Confidential Cloud Computing for Multi-Dimensional Threats and Safeguarding Data Security in a Robust Framework. International Journal of Intelligent Systems and Applications in Engineering. 2024 Jan 11, 12(11s):246-55.

[103] Ahmed SF, Alam MS, Afrin S, Rafa SJ, Rafa N, Gandomi AH. Insights into Internet of Medical Things (IoMT): Data fusion, security issues and potential solutions. Information Fusion. 2024 Feb 1, 102:102060.

[104] Weippl E, Schrittwieser S. Introduction to Security and Privacy. Hannes Werthner· Carlo Ghezzi· Jeff Kramer· Julian Nida-Rümelin· Bashar Nuseibeh· Erich Prem·. 2024:397.

[105] Al Sibahee MA, Abdulsada AI, Abduljabbar ZA, Ma J, Nyangaresi VO, Umran SM. Lightweight, Secure, Similar-Document Retrieval over Encrypted Data. Applied Sciences. 2021 Jan, 11(24):12040.

[106] Goswami C, Tamil Selvi P, Sreenivas V, Seetha J, Kiran A, Talasila V, Maithili K. Securing healthcare big data in industry 4.0: cryptography encryption with hybrid optimization algorithm for IoT applications. Optical and Quantum Electronics. 2024 Mar, 56(3):366.

[107] Attuluri S, Ramesh M. Discrete Particle Swarm Optimization based Data Encryption and Distribution to Mass Cloud Storage system. International Journal of Intelligent Systems and Applications in Engineering. 2024, 12(5s):17-25.

[108] Olabanji SO, Olaniyi OO, Adigwe CS, Okunleye OJ, Oladoyinbo TO. AI for Identity and Access Management (IAM) in the Cloud: Exploring the Potential of Artificial Intelligence to Improve User Authentication, Authorization, and Access Control within Cloud-Based Systems. Asian Journal of Research in Computer Science. 2024 Jan 25, 17(3):38-56.

[109] Ramani K. Impact of Big Data on Security: Big Data Security Issues and Defense Schemes. InCloud Security: Concepts, Methodologies, Tools, and Applications 2019 (pp. 2014-2038). IGI Global.

[110] Nyangaresi VO, Petrovic N. Efficient PUF based authentication protocol for internet of drones. In2021 International Telecommunications Conference (ITC-Egypt) 2021 Jul 13 (pp. 1-4). IEEE.

[111] Zahid R, Altaf A, Ahmad T, Iqbal F, Vera YA, Flores MA, Ashraf I. Secure data management life cycle for government big-data ecosystem: Design and development perspective. Systems. 2023 Jul 25, 11(8):380.

[112] Aboukadri S, Ouaddah A, Mezrioui A. Machine Learning in Identity and Access Management Systems: Survey and Deep Dive. Computers & Security. 2024 Jan 23:103729.

[113] Monteiro S, Oliveira D, António J, Sá F, Wanzeller C, Martins P, Abbasi M. Data Anonymization: Techniques and Models. InInternational Conference on Marketing and Technologies 2022 Dec 1 (pp. 73-84). Singapore: Springer Nature Singapore.

[114] Harini S, Dharshini R, Agalya N, Priya RL, Nair A. Analysis of Data Anonymization Techniques in Biometric Authentication System. Automated Secure Computing for Next-Generation Systems. 2024 May 3:205-22.

[115] Abood EW, Hussien ZA, Kawi HA, Abduljabbar ZA, Nyangaresi VO, Ma J, Al Sibahee MA, Kalafy A, Ahmad S. Provably secure and efficient audio compression based on compressive sensing. International Journal of Electrical & Computer Engineering (2088-8708). 2023 Feb 1, 13(1).

[116] Maurya A, Joshi M. Exploring Privacy-Preserving Strategies: A Comprehensive Analysis of Group-Based Anonymization and Hybrid ECC Encryption Algorithm for Effective Performance Evaluation in Data Security. International Journal of Intelligent Systems and Applications in Engineering. 2024 Jan 29, 12(13s):517-27.

[117] Torra V, Navarro-Arribas G. Attribute disclosure risk for k-anonymity: the case of numerical data. International Journal of Information Security. 2023 Dec, 22(6):2015-24.

[118] Aufschläger R, Folz J, März E, Guggumos J, Heigl M, Buchner B, Schramm M. Anonymization Procedures for Tabular Data: An Explanatory Technical and Legal Synthesis. Information. 2023 Sep 1, 14(9):487.

[119] Liu P, Song Y, Zou Q, Tang J, Fernandez J. Environmental press of urban greenspace pre-post COVID on older adults: A big data study in metropolitan Atlanta. Cities. 2024 Feb 1, 145:104733.

[120] Nyangaresi VO, Alsamhi SH. Towards secure traffic signaling in smart grids. In2021 3rd Global Power, Energy and Communication Conference (GPECOM) 2021 Oct 5 (pp. 196-201). IEEE.

[121] Mohammad AS, Pradhan MR. Machine learning with big data analytics for cloud security. Computers & Electrical Engineering. 2021 Dec 1, 96:107527.

[122] Lin Q, Li X, Cai K, Prakash M, Paulraj D. Secure Internet of medical Things (IoMT) based on ECMQV-MAC authentication protocol and EKMC-SCP blockchain networking. Information Sciences. 2024 Jan 1, 654:119783.

[123] Priyadarshini SB, Dash SK, Sahani A, Mishra BK, Nath MP. An Introduction to Security in Internet of Things (IoT) and Big Data. A Roadmap for Enabling Industry 4.0 by Artificial Intelligence. 2022 Dec 16:169-200.

[124] Liu L, Li J, Lv J, Wang J, Zhao S, Lu Q. Privacy-Preserving and Secure Industrial Big Data Analytics: A Survey and the Research Framework. IEEE Internet of Things Journal. 2024 Jan 15.

[125] Abduljaleel IQ, Abduljabbar ZA, Al Sibahee MA, Ghrabat MJ, Ma J, Nyangaresi VO. A Lightweight Hybrid Scheme for Hiding Text Messages in Colour Images Using LSB, Lah Transform and Chaotic Techniques. Journal of Sensor and Actuator Networks. 2022 Dec, 11(4):66.

[126] Yu B, Hu J. Construction of Data Security System. InInternational Conference on Big Data 2023 Sep 23 (pp. 218-229). Cham: Springer Nature Switzerland.

[127] Patel S, Shah M. A Comprehensive Study on Implementing Big Data in the Auditing Industry. Annals of Data Science. 2023 Jun, 10(3):657-77.

[128] Rahman SF, Irwansyah I. The role of big data in audit quality and fraud disclosure. InProceeding International Conference on Accounting and Finance 2024 Jan 25 (pp. 467-476).

[129] Minh TN. The role of big data analytics in auditing financial statements with the improvement of audit quality in Vietnam. Onomázein. 2023 Nov 10(62 (2023): December):348-57.

[130] Nyangaresi VO, Mohammad Z. Session Key Agreement Protocol for Secure D2D Communication. InThe Fifth International Conference on Safety and Security with IoT: SaSeIoT 2021 2022 Jun 12 (pp. 81-99). Cham: Springer International Publishing.

[131] Lee CC, Gheisari M, Shayegan MJ, Ahvanooey MT, Liu Y. Privacy-Preserving Techniques in Cloud/Fog and Internet of Things. Cryptography. 2023 Oct 16, 7(4):51.

[132] Miao J, Wang Z, Wu Z, Ning X, Tiwari P. A blockchain-enabled privacy-preserving authentication management protocol for Internet of Medical Things. Expert Systems with Applications. 2024 Mar 1, 237:121329.

[133] Hastings M, Falk BH, Tsoukalas G. Privacy-preserving network analytics. Management Science. 2023 Sep, 69(9):5482-500.

[134] Ntizikira E, Lei W, Alblehai F, Saleem K, Lodhi MA. Secure and privacy-preserving intrusion detection and prevention in the internet of unmanned aerial vehicles. Sensors. 2023 Sep 25, 23(19):8077.

[135] Mutlaq KA, Nyangaresi VO, Omar MA, Abduljabbar ZA. Symmetric Key Based Scheme for Verification Token Generation in Internet of Things Communication Environment. InApplied Cryptography in Computer and Communications: Second EAI International Conference, AC3 2022, Virtual Event, May 14-15, 2022, Proceedings 2022 Oct 6 (pp. 46-64). Cham: Springer Nature Switzerland.

[136] Zeng M, Xu Y, Wu H, Ma J, Gao J. Sustainable Insights for Energy Big Data Governance in China: Full Life Cycle Curation from the Ecosystem Perspective. Sustainability. 2022 May 16, 14(10):6013.

[137] Yallop AC, Gică OA, Moisescu OI, Coroș MM, Séraphin H. The digital traveller: implications for data ethics and data governance in tourism and hospitality. Journal of Consumer Marketing. 2023 Feb 7, 40(2):155-70.

[138] Akour I, Al Kurdi B, Nuseir MT, Alzoubi HM, Alshurideh MT, AlHamad AQ. Modelling Big Data Management for the Finance Sector Using Artificial Intelligence. InCyber Security Impact on Digitalization and Business Intelligence: Big Cyber Security for Information Management: Opportunities and Challenges 2024 Jan 4 (pp. 25-37). Cham: Springer International Publishing.

[139] Yang X, Huang K, Yang D, Zhao W, Zhou X. Biomedical Big Data Technologies, Applications, and Challenges for Precision Medicine: A Review. Global Challenges. 2024 Jan, 8(1):2300163.

[140] Nyangaresi VO, Ma J. A Formally Verified Message Validation Protocol for Intelligent IoT E-Health Systems. In2022 IEEE World Conference on Applied Intelligence and Computing (AIC) 2022 Jun 17 (pp. 416-422). IEEE.

[141] Hasan R, Kamal MM, Daowd A, Eldabi T, Koliousis I, Papadopoulos T. Critical analysis of the impact of big data analytics on supply chain operations. Production Planning & Control. 2024 Jan 2, 35(1):46-70.

[142] Hammami S, Durrah O, El-Maghraby L, Jaboob M, Kasim S, Baalwi K. Understanding how big data awareness affects healthcare institution performance in Oman. InArtificial Intelligence, Big Data, Blockchain and 5G for the Digital Transformation of the Healthcare Industry 2024 Jan 1 (pp. 271-297). Academic Press.

[143] Farouk FM, Siew EG, Yusof SH. Overcoming resistance to change in a big data analytics implementation case study. Journal of Information Technology Teaching Cases. 2024 Jan 10:20438869231226395.

[144] Hemmati A, Arzanagh HM, Rahmani AM. A taxonomy and survey of big data in social media. Concurrency and Computation: Practice and Experience. 2024 Jan 10, 36(1):e7875.

[145] Abduljabbar ZA, Nyangaresi VO, Ma J, Al Sibahee MA, Khalefa MS, Honi DG. MAC-Based Symmetric Key Protocol for Secure Traffic Forwarding in Drones. InFuture Access Enablers for Ubiquitous and Intelligent Infrastructures: 6th EAI International Conference, FABULOUS 2022, Virtual Event, May 4, 2022, Proceedings 2022 Sep 18 (pp. 16-36). Cham: Springer International Publishing.

[146] Kardas P, Aguilar-Palacio I, Almada M, Cahir C, Costa E, Giardini A, Malo S, Massot Mesquida M, Menditto E, Midão L, Parra-Calderón CL. The need to develop standard measures of patient adherence for big data. Journal of Medical Internet Research. 2020 Aug 27, 22(8):e18150.

[147] James M. The Ethical and Legal Implications of Using Big Data and Artificial Intelligence for Public Relations Campaigns in the United States. International Journal of Communication and Public Relation. 2024 Jan 19, 9(1):38-52.

[148] Regulwar GB, Mahalle A, Pawar R, Shamkuwar SK, Kakde PR, Tiwari S. Big Data Collection, Filtering, and Extraction of Features. InBig Data Analytics Techniques for Market Intelligence 2024 (pp. 136-158). IGI Global.

[149] Ahmad A. Ethical implications of artificial intelligence in accounting: A framework for responsible ai adoption in multinational corporations in Jordan. International Journal of Data and Network Science. 2024, 8(1):401-14.

[150] Nyangaresi VO. Provably Secure Pseudonyms based Authentication Protocol for Wearable Ubiquitous Computing Environment. In2022 International Conference on Inventive Computation Technologies (ICICT) 2022 Jul 20 (pp. 1-6). IEEE.

[151] Lemke C, Monett D, Mikoleit M. Digital ethics in data-driven organizations and AI ethics as application example. InApply Data Science: Introduction, Applications and Projects 2023 Jan 1 (pp. 31-48). Wiesbaden: Springer Fachmedien Wiesbaden.

[152] Fosso Wamba S, Queiroz MM. Responsible artificial intelligence as a secret ingredient for digital health: Bibliometric analysis, insights, and research directions. Information Systems Frontiers. 2023 Dec, 25(6):2123-38.

[153] Shanthi R, Babu MD, Kousika N, Vijayaraj C, Choubey SB, Sambooranalaxmi S. Advanced Privacy-Preserving Framework Using Homomorphic Encryption and Adaptive Privacy Parameters for Scalable Big Data Analysis. International Journal of Intelligent Systems and Applications in Engineering. 2024 Jan 11, 12(11s):160-5.

[154] Zhao B, Chen WN, Li X, Liu X, Pei Q, Zhang J. When Evolutionary Computation Meets Privacy. IEEE Computational Intelligence Magazine. 2024 Jan 8, 19(1):66-74.

[155] Hussien ZA, Abdulmalik HA, Hussain MA, Nyangaresi VO, Ma J, Abduljabbar ZA, Abduljaleel IQ. Lightweight Integrity Preserving Scheme for Secure Data Exchange in Cloud-Based IoT Systems. Applied Sciences. 2023 Jan, 13(2):691.

[156] Dhinakaran D, Selvaraj D, Dharini N, Raja SE, Priya CS. Towards a Novel Privacy-Preserving Distributed Multiparty Data Outsourcing Scheme for Cloud Computing with Quantum Key Distribution. International Journal of Intelligent Systems and Applications in Engineering. 2024, 12(2):286-300.

[157] Gupta M, Dwivedi RK. Fortified MapReduce Layer: Elevating Security and Privacy in Big Data. EAI Endorsed Transactions on Scalable Information Systems. 2023 Oct 2, 10(6).

[158] Kabdjou J, Tagne EF, Rawat DB, Acosta J, Kamhoua C. A pipeline approach for privacy preservation against poisoning attacks in a Mobile Edge Computing environment. Ad Hoc Networks. 2024 Mar 1, 154:103385.

[159] Wu G, Chen X, Gao Z, Zhang H, Yu S, Shen S. Privacy-preserving offloading scheme in multi-access mobile edge computing based on MADRL. Journal of Parallel and Distributed Computing. 2024 Jan 1, 183:104775.

[160] Nyakomitta SP, Omollo V. Biometric-Based Authentication Model for E-Card Payment Technology. IOSR Journal of Computer Engineering (IOSRJCE). 2014, 16(5):137-44.

[161] Aziz R, Banerjee S, Bouzefrane S, Le Vinh T. Exploring homomorphic encryption and differential privacy techniques towards secure federated learning paradigm. Future internet. 2023 Sep 13, 15(9):310.

[162] Rao S. Enhancing Cloud Data Privacy with a Scalable Hybrid Approach: HE-DP-SMC. J. Electrical Systems. 2023, 19(4):350-75.

[163] Jin W, Yao Y, Han S, Joe-Wong C, Ravi S, Avestimehr S, He C. FedML-HE: An Efficient Homomorphic-Encryption-Based Privacy-Preserving Federated Learning System. arXiv preprint arXiv:2303.10837. 2023 Mar 20.

[164] AbaOud M, Almuqrin M, Khan MF. Advancing Federated Learning through Novel Mechanism for Privacy Preservation in Healthcare Applications. IEEE Access. 2023 Aug 2.

[165] Nyangaresi VO. Lightweight anonymous authentication protocol for resource-constrained smart home devices based on elliptic curve cryptography. Journal of Systems Architecture. 2022 Dec 1, 133:102763.

[166] Safitra MF, Lubis M, Fakhrurroja H. Counterattacking cyber threats: A framework for the future of cybersecurity. Sustainability. 2023 Sep 6, 15(18):13369.

[167] Repetto M. Adaptive monitoring, detection, and response for agile digital service chains. Computers & Security. 2023 Jun 18:103343.

[168] Sun N, Ding M, Jiang J, Xu W, Mo X, Tai Y, Zhang J. Cyber Threat Intelligence Mining for Proactive Cybersecurity Defense: A Survey and New Perspectives. IEEE Communications Surveys & Tutorials. 2023 May 5.

[169] Rangaraju S. Secure by Intelligence: Enhancing Products with AI-Driven Security Measures. EPH-International Journal of Science And Engineering. 2023 Dec 1, 9(3):36-41.

[170] Ghrabat MJ, Hussien ZA, Khalefa MS, Abduljabba ZA, Nyangaresi VO, Al Sibahee MA, Abood EW. Fully automated model on breast cancer classification using deep learning classifiers. Indonesian Journal of Electrical Engineering and Computer Science. 2022 Oct, 28(1):183-91.

[171] Zhang J, Shu Y, Yu H. Fairness in design: a framework for facilitating ethical artificial intelligence designs. International Journal of Crowd Science. 2023 Mar, 7(1):32-9.

[172] Pfeiffer J, Gutschow J, Haas C, Möslein F, Maspfuhl O, Borgers F, Alpsancar S. Algorithmic Fairness in AI: An Interdisciplinary View. Business & Information Systems Engineering. 2023 Apr, 65(2):209-22.

[173] Filippi CG, Stein JM, Wang Z, Bakas S, Liu Y, Chang PD, Lui Y, Hess C, Barboriak DP, Flanders AE, Wintermark M. Ethical considerations and fairness in the use of artificial intelligence for neuroradiology. American Journal of Neuroradiology. 2023 Nov 1, 44(11):1242-8.

[174] Memarian B, Doleck T. Fairness, Accountability, Transparency, and Ethics (FATE) in Artificial Intelligence (AI), and higher education: A systematic review. Computers and Education: Artificial Intelligence. 2023 Jun 26:100152.

[175] Nyangaresi VO, Ahmad M, Alkhayyat A, Feng W. Artificial neural network and symmetric key cryptography based verification protocol for 5G enabled Internet of Things. Expert Systems. 2022 Dec, 39(10):e13126.

[176] Rivadeneira JE, Silva JS, Colomo-Palacios R, Rodrigues A, Boavida F. User-centric privacy preserving models for a new era of the Internet of Things. Journal of Network and Computer Applications. 2023 Jul 5:103695.

[177] Villarán C, Beltrán M. User-Centric Privacy for Identity Federations Based on a Recommendation System. Electronics. 2022 Apr 14, 11(8):1238.

[178] Wickramasinghe CI. Best-Practice-Based Framework for User-Centric Privacy-Preserving Solutions in Smart Home Environments. InInternational Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services 2022 Nov 14 (pp. 101-120). Cham: Springer Nature Switzerland.

[179] Lola J, Serrão C, Casal J. Towards Transparent and Secure IoT: Improving the Security and Privacy through a User-Centric Rules-Based System. Electronics. 2023 Jun 8, 12(12):2589.

[180] Abduljabbar ZA, Nyangaresi VO, Jasim HM, Ma J, Hussain MA, Hussien ZA, Aldarwish AJ. Elliptic Curve Cryptography-Based Scheme for Secure Signaling and Data Exchanges in Precision Agriculture. Sustainability. 2023 Jun 28, 15(13):10264.

[181] Lemieux VL, Werner J. Protecting Privacy in Digital Records: The Potential of Privacy-Enhancing Technologies. ACM Journal on Computing and Cultural Heritage. 2024 Jan 8, 16(4):1-8.

[182] Becher S, Gerl A, Meier B, Bölz F. Big picture on privacy enhancing technologies in e-health: a holistic personal privacy workflow. Information. 2020 Jul 8, 11(7):356.

[183] Bluemke E, Collins T, Garfinkel B, Trask A. Exploring the Relevance of Data Privacy-Enhancing Technologies for AI Governance Use Cases. arXiv preprint arXiv:2303.08956. 2023 Mar 15.

[184] Schmidt K, Munilla Garrido G, Mühle A, Meinel C. Mitigating Sovereign Data Exchange Challenges: A Mapping to Apply Privacy-and Authenticity-Enhancing Technologies. InInternational Conference on Trust and Privacy in Digital Business 2022 Aug 24 (pp. 50-65). Cham: Springer International Publishing.

[185] Nyangaresi VO. Lightweight key agreement and authentication protocol for smart homes. In2021 IEEE AFRICON 2021 Sep 13 (pp. 1-6). IEEE.

[186] Wang C, Yuan Z, Zhou P, Xu Z, Li R, Wu DO. The Security and Privacy of Mobile Edge Computing: An Artificial Intelligence Perspective. IEEE Internet of Things Journal. 2023 Aug 11.

[187] Osman L, Taiwo O, Elashry A, Ezugwu AE. Intelligent Edge Computing for IoT: Enhancing Security and Privacy. Journal of Intelligent Systems & Internet of Things. 2023 Jan 1, 8(1).

[188] Sindjoung ML, Velempini M, Djamegni CT. A data security and privacy scheme for user quality of experience in a Mobile Edge Computing-based network. Array. 2023 Sep 1, 19:100304.

[189] Kumar A, Upadhyay A, Mishra N, Nath S, Yadav KR, Sharma G. Privacy and security concerns in edge computing-based smart cities. InRobotics and AI for Cybersecurity and Critical Infrastructure in Smart Cities 2022 Mar 29 (pp. 89-110). Cham: Springer International Publishing.

[190] Alsamhi SH, Shvetsov AV, Kumar S, Shvetsova SV, Alhartomi MA, Hawbani A, Rajput NS, Srivastava S, Saif A, Nyangaresi VO. UAV computing-assisted search and rescue mission framework for disaster and harsh environment mitigation. Drones. 2022 Jun 22, 6(7):154.

[191] Marbut AR, Harms PD. Fiends and fools: a narrative review and neo-socioanalytic perspective on personality and insider threats. Journal of Business and Psychology. 2023 May 9:1-8.

[192] Al-Harrasi A, Shaikh AK, Al-Badi A. Towards protecting organisations' data by preventing data theft by malicious insiders. International Journal of Organizational Analysis. 2023 Apr 10, 31(3):875-88.

[193] Mehmood M, Amin R, Muslam MM, Xie J, Aldabbas H. Privilege Escalation Attack Detection and Mitigation in Cloud using Machine Learning. IEEE Access. 2023 May 8.

[194] Dwivedi R, Nerur S, Mangalaraj G. Predicting Insider Breaches Using Employee Reviews. Journal of Computer Information Systems. 2023 Jun 21:1-5.

[195] Nyangaresi VO. Terminal independent security token derivation scheme for ultra-dense IoT networks. Array. 2022 Sep 1, 15:100210.

[196] Li S. Towards Digital Money Interoperability: Data Governance Coordination for Cross-border Payments. Houston Journal of International Law. 2023, 45(1).

[197] Guan Z. Difficulties and Solutions for Industrial Data Security and Compliance Governance. InInternational Conference on Cloud Computing 2023 Dec 17 (pp. 66-75). Cham: Springer Nature Switzerland.

[198] Li W, Yang D. Decentralized but Coordinated: Probing Polycentricity in EU Data Protection Cross-border Enforcement. InGlobal Digital Data Governance (pp. 105-124). Routledge.

[199] Jones CA. Critical Success Factors for Data Governance of Cross-border e-Trade Data Among ASEAN Member States. Journal of Asian Economic Integration. 2021 Apr, 3(1):38-60.

[200] Yenurkar GK, Mal S, Nyangaresi VO, Hedau A, Hatwar P, Rajurkar S, Khobragade J. Multifactor data analysis to forecast an individual's severity over novel COVID-19 pandemic using extreme gradient boosting and random forest classifier algorithms. Engineering Reports. 2023:e12678.

[201] Bountakas P, Zarras A, Lekidis A, Xenakis C. Defense strategies for Adversarial Machine Learning: A survey. Computer Science Review. 2023 Aug 1, 49:100573.

[202] He K, Kim DD, Asghar MR. Adversarial machine learning for network intrusion detection systems: a comprehensive survey. IEEE Communications Surveys & Tutorials. 2023 Jan 3.

[203] Ayoub S, Gulzar Y, Rustamov J, Jabbari A, Reegu FA, Turaev S. Adversarial Approaches to Tackle Imbalanced Data in Machine Learning. Sustainability. 2023 Apr 24, 15(9):7097.

[204] Alotaibi A, Rassam MA. Adversarial machine learning attacks against intrusion detection systems: A survey on strategies and defense. Future Internet. 2023 Jan 31, 15(2):62.

[205] Honi DG, Ali AH, Abduljabbar ZA, Ma J, Nyangaresi VO, Mutlaq KA, Umran SM. Towards Fast Edge Detection Approach for Industrial Products. In2022 IEEE 21st International Conference on Ubiquitous Computing and Communications (IUCC/CIT/DSCI/SmartCNS) 2022 Dec 19 (pp. 239-244). IEEE.

[206] Ishaq M, Kifayat K, Zafar M. A Survey on Human Factors in Cyberspace: A New Dimension of Privacy Threats. In2023 International Conference on Communication, Computing and Digital Systems (C-CODE) 2023 May 17 (pp. 1-6). IEEE.

[207] Nurgalieva L, Frik A, Doherty G. A Narrative Review of Factors Affecting the Implementation of Privacy and Security Practices in Software Development. ACM Computing Surveys. 2023 Jul 17, 55(14s):1-27.

[208] Pottebaum J, Rossel J, Somorovsky J, Acar Y, Fahr R, Cabarcos PA, Bodden E, Gräßler I. Re-Envisioning Industrial Control Systems Security by Considering Human Factors as a Core Element of Defense-in-Depth. In2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW) 2023 Jul 3 (pp. 379-385). IEEE.

[209] Fourné M, Wermke D, Fahl S, Acar Y. A Viewpoint on Human Factors in Software Supply Chain Security: A Research Agenda. IEEE Security & Privacy. 2023 Nov 13, 21(6):59-63.

[210] Zarei E, Khan F, Abbassi R. How to account artificial intelligence in human factor analysis of complex systems?. Process safety and environmental protection. 2023 Jan 28.