MZ Journals
Journal of Artificial Intelligence

Privacy Preservation in the Age of Big Data: Insights from Information Theory

Henrik Andersen, Freja Larsen
University of Copenhagen, Denmark

Abstract

Privacy preservation in the age of big data presents a multifaceted challenge, demanding innovative approaches grounded in information theory. As vast amounts of personal data are collected and analyzed, ensuring the confidentiality and integrity of sensitive information becomes paramount. Information theory offers valuable insights by quantifying the amount of information leaked during data processing and transmission, enabling the development of robust privacy-preserving mechanisms. Techniques such as differential privacy, homomorphic encryption, and secure multiparty computation emerge as promising solutions, leveraging mathematical principles to safeguard privacy without sacrificing utility. By embracing the principles of information theory, stakeholders can navigate the complexities of big data while upholding individuals' right to privacy in an increasingly data-driven world.

Keywords: Privacy Preservation, Information Theory, Data Processing, Differential Privacy

1. Introduction

The proliferation of big data has revolutionized the way organizations collect, analyze, and utilize vast amounts of information to derive insights and drive decision-making processes. However, this data-driven paradigm has also raised significant concerns regarding the privacy and security of personal information. As the volume, velocity, and variety of data continue to expand exponentially, ensuring the confidentiality and integrity of sensitive information has become increasingly challenging. Traditional approaches to privacy preservation often fall short in the face of these complexities, necessitating innovative solutions grounded in rigorous mathematical principles [1]. In this context, information theory emerges as a crucial framework for understanding and addressing the intricate challenges of privacy preservation in the age of big data. By quantifying the amount of information leaked during data processing and transmission, information theory offers valuable insights into designing robust privacy-preserving mechanisms that balance the need for data utility with the imperative of protecting individual privacy rights. This paper explores the role of information theory in elucidating key privacy preservation challenges, examining various techniques and mechanisms informed by information theory principles, and outlining future directions for research and innovation in this critical domain. Protection of Individual Rights: In the digital age, individuals generate and share massive amounts of personal data through various online activities, including social media interactions, online

purchases, and internet browsing [2]. Preserving privacy ensures that individuals maintain control over their personal information and have the right to determine how it is collected, used, and shared. Mitigation of Risks: Big data analytics offer tremendous opportunities for innovation and advancement in various fields, including healthcare, finance, and transportation. However, the vast amounts of data collected also pose significant risks if not adequately protected. Privacy breaches can result in identity theft, financial fraud, reputational damage, and other adverse consequences for individuals and organizations alike. Preservation of Democracy: Privacy preservation is integral to maintaining democratic principles and individual freedoms. In an era where data is increasingly used for political targeting and manipulation, ensuring privacy safeguards protects against the misuse of personal information for political gain or suppression of dissenting voices. Ethical Considerations: Respecting individuals' privacy is not only a legal requirement but also an ethical imperative [3]. Organizations have a moral obligation to prioritize privacy and ensure that data collection and usage practices align with ethical principles such as transparency, fairness, and accountability. Long-term Sustainability: Building a sustainable data ecosystem requires a balance between data-driven innovation and privacy protection. By implementing privacy-preserving measures, organizations can create a sustainable framework that fosters innovation while safeguarding privacy for current and future generations. In summary, privacy preservation in the age of big data is essential for upholding individual rights, fostering trust and confidence, mitigating risks, ensuring legal compliance, preserving democracy, adhering to ethical principles, and promoting long-term sustainability in the digital era [4].

The outline of this paper is as follows: Differential privacy preservation preliminaries are illustrated in section 2. Understanding Big Data and Privacy Preservation is explained in section 3. The Privacy-Preserving Techniques and Mechanisms are explained in section 4. The overall conclusion of the paper is given in section 5.

2. Preliminaries for differential privacy preservation

Differential privacy for securing the statistical database from several attacks. The differential privacy methods do not increase or decrease the output data about changes in individual information in the database [5]. Suppose Y is a random algorithm and Q is the possible outcome for two sets of P and R,

$$\Pr[Y[P]\varepsilon S] \leq \exp(\varepsilon) \times \Pr[Y[R]\varepsilon S] \quad (1)$$

Then, algorithm Y yields (ε) differential privacy (ε > 0).

Laplace mechanism

The Laplace and exponential are the basic concepts behind differential privacy preservation. Moreover, the randomization is examined in both concepts through the sensitivity analysis. For a query, Q the sensitivity is expressed as,

$$\Delta Q = \max \| Q(P) - Q(R) \|_1 \quad (2)$$

It also uses the Laplace transform for the numeric output, adding independent noise. For Q: dataset P W range, the method M provides ε-differential privacy by the following equation.

$$M(P) = Q(P) + Lap\ \Delta Q \qquad (3)$$

Exponential mechanism

The exponential mechanism provides randomized results for non-numeric queries merged with the score function to examine the quality of the output S. Assume that c(P; θ) is the score function of the dataset P, which evaluates the output quality θ λ and ΔQ is the sensitivity of θ. The exponential mechanism satisfies ε - differential privacy by the following equation.

$$M(P) = \quad return\ \theta\ \alpha \exp \frac{eq(P;\theta)}{2\Delta Q} \qquad (4)$$

2.2.Differential privacy preservation in deep learning for big data

Data protection and security are the most important aspects of the big data platform, and if the data are not secured with security measures, the data can easily be compromised. To be able to deal with the huge amount of data, the DL methods are developed [6]. The data protection method offers some control measures over the data displayed. In addition, the advantages of DL approaches are not only limited to data analysis but also know image classification, speech, and text analysis. Thus, the need for differential data protection in DL increases exponentially, whereby data protection is guaranteed without loss of information. Numerous attempts have been made to provide data protection in DL. The basic concept of DL-based data protection in Big Data is shown in Figure 3.

The privacy protection method stops the information available to unknown sources. Privacy practices regulate a person's data and ensure users' data. The security protocols protect the data from malicious attacks and information misuse. Below are some of the methods that are used to maintain privacy [7].

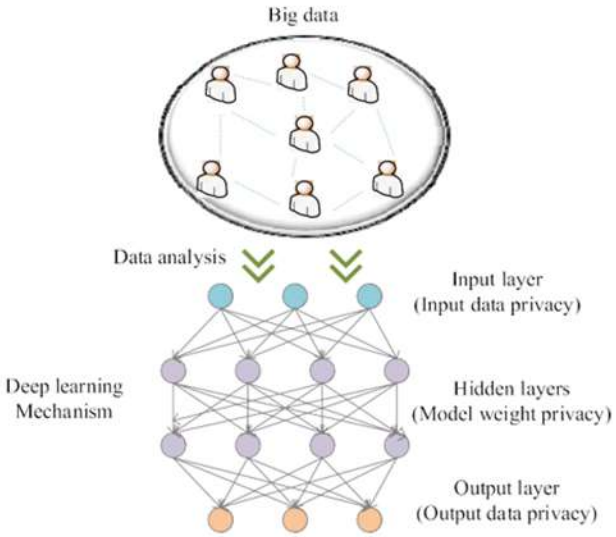

Figure 1: Privacy preservation of deep learning in big data.

Foundations of Differential Privacy: Differential privacy, a widely recognized privacy-preserving framework, is grounded in information theory principles. It quantifies the privacy guarantee provided by a data analysis algorithm, ensuring that the presence or absence of any individual's data does not significantly affect the outcome [8]. Information theory underpins the mathematical definitions and analysis of differential privacy, enabling rigorous evaluation and comparison of different privacy-preserving mechanisms. Evaluation of Privacy-Preserving Mechanisms: Information theory provides a framework for evaluating the efficacy of various privacy-preserving mechanisms, such as encryption techniques, anonymization methods, and data obfuscation strategies. By quantifying the level of privacy protection provided by these mechanisms, organizations can make informed decisions about their implementation and deployment in real-world scenarios. Design of Secure Communication Protocols: Information theory informs the design of secure communication protocols that protect data confidentiality and integrity during transmission over insecure channels. Encryption schemes, error-correcting codes, and authentication protocols leverage information theory concepts to ensure secure communication and prevent unauthorized access to sensitive information. Optimization of Data Utilization and Privacy Preservation Trade-offs: Information theory enables organizations to optimize the trade-off between data utilization and privacy preservation [9]. By quantifying the utility of data and the privacy loss associated with its use, organizations can devise strategies to maximize the value derived from data analytics while minimizing the risk of privacy breaches. Development of Privacy-Preserving Data Analysis Techniques: Information theory inspires the development of innovative data analysis techniques that preserve privacy without sacrificing utility. Techniques such as secure multiparty computation, homomorphic encryption, and privacy-preserving machine learning algorithms leverage information theory principles to enable collaborative data analysis while protecting the privacy of individual data contributors. In summary, information theory serves as a foundational framework for addressing privacy preservation challenges by quantifying

information leakage, guiding the design and evaluation of privacy-preserving mechanisms, optimizing data utilization and privacy trade-offs, and inspiring the development of innovative data analysis techniques that uphold privacy while enabling data-driven innovation.

3. Understanding Big Data and Privacy Preservation

Understanding big data and privacy preservation is essential in navigating the complexities of modern data-driven environments while safeguarding individuals' privacy rights. Big data refers to the massive volumes of structured and unstructured data generated from various sources, including social media, sensors, transactions, and online activities. This data is characterized by its volume, velocity, and variety, presenting both opportunities and challenges for organizations seeking to extract valuable insights. Privacy preservation in the context of big data involves protecting individuals' personal information from unauthorized access, use, and disclosure. It encompasses various principles and practices aimed at ensuring that individuals maintain control over their data and are not subject to privacy violations. Key aspects of privacy preservation in the age of big data include Data Collection and Consent: Organizations must obtain informed consent from individuals before collecting their data [10]. This involves communicating the purposes for which data will be used, the types of data collected, and any potential risks or implications for privacy. Transparent data collection practices empower individuals to make informed decisions about sharing their information. Data Minimization: Data minimization principles advocate for collecting and retaining only the minimum amount of personal data necessary to achieve specific purposes. By limiting the collection and retention of unnecessary data, organizations can reduce the risk of privacy breaches and mitigate the potential impact of data misuse. Security Measures: Robust security measures, including encryption, access controls, and data encryption, are essential for protecting personal data against unauthorized access, theft, or tampering. Secure data storage and transmission practices ensure that sensitive information remains confidential and integrity intact throughout its lifecycle. Privacy by Design and Default: Privacy by design and default principles advocate for embedding privacy considerations into the design and implementation of data systems and processes from the outset. By prioritizing privacy at the design stage, organizations can proactively identify and address potential privacy risks, leading to more robust and privacy-preserving solutions. Accountability and Transparency: Organizations should demonstrate accountability for their data practices by being transparent about how they collect, use, and protect personal data. Transparency builds trust with individuals and stakeholders and allows for greater scrutiny and oversight of data-handling processes [11]. In summary, understanding big data and privacy preservation requires a comprehensive approach that encompasses responsible data collection, anonymization, data minimization, security measures, privacy by design, regulatory compliance, and transparency. By adopting privacy-preserving practices and principles, organizations can leverage the benefits of big data analytics while respecting individuals' privacy rights and fostering trust in the data ecosystem.

3.2.Privacy preservation model based on the Data Lake Concept

We propose a novel privacy preservation model based on the Data Lake concept to hold a variety of data from diverse sources. A data lake is a repository to holds data from diverse sources in their raw format. Data ingestion from a variety of sources can be done using Apache Flume and an intelligent algorithm based on machine learning can be applied to identify sensitive attributes dynamically [12]. The algorithm will be trained with existing data sets with known sensitive attributes and rigorous training of the model will help in predicting the sensitive attributes in a given data set. The accuracy of the model can be improved by adding more layers of training leading to deep learning techniques. Advanced computing techniques like Apache Spark can be used in implementing privacy-preserving algorithms which is a distributed massive parallel computing with in-memory processing to ensure very fast processing. The proposed model is shown in Fig. 3.

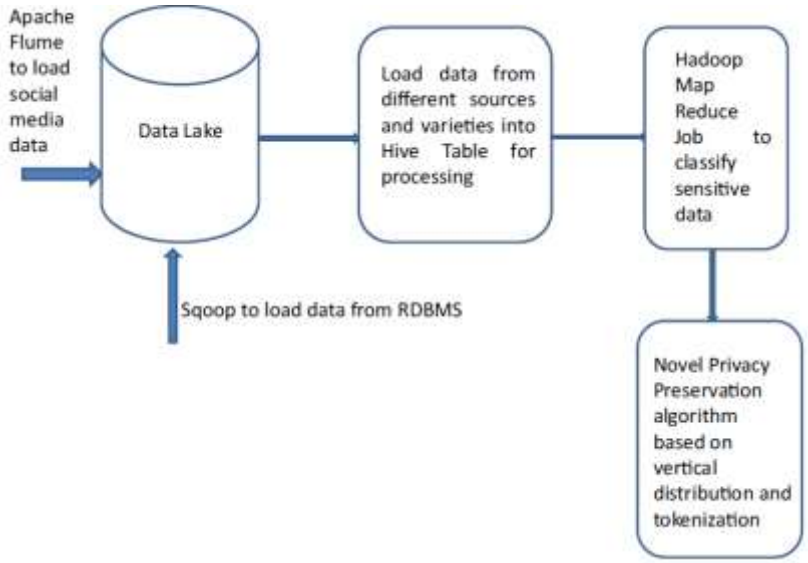

Figure 2: A Novel privacy preservation model based on vertical distribution and tokenization

Data analytics is done on the data collected from various sources. If an e-commerce site would like to perform data analytics, they needs transactional data, website logs, and customer opinions through social media pages. A Data lake is used to collect data from different sources. Apache Flume is used to ingest data from social media sites, and website logs into Hadoop Distributed File System(HDFS). Using SQOOP relational data can be loaded into HDFS. In Data lake the data can remain in its native form which is either structured or unstructured. When data has to be processed, it can be transformed into HIVE tables. A Hadoop map reduces jobs using machine learning that can be executed on the data to classify sensitive attributes [13]. The data can be vertically distributed to separate the sensitive attributes from the rest of the data and apply tokenization to map the vertically distributed data. The data without any sensitive attributes can be published for data analytics.

Big data refers to extremely large and complex datasets that cannot be effectively managed, processed, or analyzed using traditional data processing tools or methods. These datasets are

characterized by the following key attributes: Volume: Big data involves massive volumes of data, often ranging from terabytes to petabytes and beyond. These datasets may originate from various sources, including social media platforms, sensors, mobile devices, and transactional systems, generating an unprecedented amount of information. Velocity: Big data is generated at an incredibly high velocity, with data streams being produced rapidly and continuously. This real-time or near-real-time data flow requires rapid processing and analysis to extract valuable insights and respond to events as they occur. Variety: Big data encompasses diverse types of data, including structured, semi-structured, and unstructured data [14]. Structured data follows a predefined format and is typically stored in databases, while semi-structured and unstructured data, such as text documents, images, videos, and social media posts, lack a predefined schema and require advanced processing techniques for analysis. Variability: Big data exhibits variability in its format, quality, and structure. Data may arrive in inconsistent formats, with varying levels of completeness and accuracy, posing challenges for integration and analysis. Managing this variability requires flexible data processing and cleansing techniques to ensure data quality and consistency. Veracity: Veracity refers to the trustworthiness and reliability of big data, encompassing issues such as data accuracy, integrity, and provenance. Big data sources may contain errors, inconsistencies, or biases, necessitating rigorous data validation and quality assurance measures to ensure the reliability of analytical insights derived from the data. Value: Despite the challenges posed by its volume, velocity, variety, and veracity, big data holds immense value for organizations seeking to gain insights, make data-driven decisions, and drive innovation [15]. By effectively capturing, processing, and analyzing big data, organizations can uncover valuable patterns, trends, and correlations that inform strategic initiatives and improve business outcomes. In summary, big data is characterized by its large volume, high velocity, diverse variety, variability, veracity, and potential value. Understanding these characteristics is essential for organizations seeking to harness the power of big data to gain competitive advantages, drive innovation, and address complex business challenges.

4. Privacy-Preserving Techniques and Mechanisms

Privacy-preserving techniques and mechanisms play a crucial role in safeguarding individuals' privacy rights while enabling the analysis and utilization of big data. Several advanced techniques have been developed to address privacy concerns in various data processing scenarios. Some key privacy-preserving techniques and mechanisms include Differential Privacy: Differential privacy is a rigorous privacy-preserving framework that aims to protect individuals' sensitive information while allowing for meaningful data analysis. It achieves this by adding noise to query responses or data records in a way that obscures individual contributions to the dataset, thus ensuring plausible deniability. Differential privacy provides strong privacy guarantees by quantifying the maximum impact of any single individual's data on the outcome of a computation, thereby protecting against privacy breaches even when adversaries have access to auxiliary information. Homomorphic Encryption: Homomorphic encryption allows for computations to be performed directly on encrypted data without decrypting it first. This enables secure data processing while

preserving the confidentiality of sensitive information. Homomorphic encryption schemes come in different forms, such as partially homomorphic encryption, fully homomorphic encryption, and leveled homomorphic encryption, each offering varying degrees of computational efficiency and functionality. Secure Multiparty Computation (SMPC): SMPC enables multiple parties to jointly compute a function over their private inputs while keeping these inputs confidential. This is achieved through cryptographic protocols that allow parties to collaborate without revealing their sensitive data to each other or any external observer. SMPC protocols ensure that computation results are accurate and valid while protecting the privacy of individual inputs. Privacy-Preserving Data Masking and Perturbation: Data masking and perturbation techniques involve modifying or obfuscating sensitive data to prevent unauthorized disclosure while maintaining its utility for analysis. This includes methods such as k-anonymity, l-diversity, and t-closeness, which aim to anonymize datasets by generalizing or suppressing identifying attributes to ensure individuals' identities cannot be readily inferred. Privacy-Preserving Machine Learning: Privacy-preserving machine learning techniques aim to train models on sensitive data while preserving the privacy of individual training examples. This includes approaches such as federated learning, where models are trained collaboratively across decentralized devices or data sources without exchanging raw data, and secure enclaves, where computations are performed within secure hardware environments to protect sensitive information. Data Obfuscation and De-identification: Data obfuscation techniques involve obscuring or redacting sensitive information in datasets to prevent the identification or re-identification of individuals. This includes methods such as data masking, tokenization, and anonymization, which replace or suppress identifying attributes while retaining the utility of non-sensitive data for analysis. By leveraging these privacy-preserving techniques and mechanisms, organizations can protect individuals' privacy rights while harnessing the value of big data for analysis, decision-making, and innovation. However, it's essential to carefully evaluate and select the most appropriate techniques based on the specific privacy requirements, data characteristics, and regulatory considerations of each use case.

Table 1 illustrates Privacy Preservation in the Age of Big Data, viewed through the lens of Information Theory, which focuses on quantifying and mitigating the risks associated with data processing and transmission. Information Theory provides a mathematical framework to assess the amount of information leakage during data analysis, enabling the development of robust privacy-preserving mechanisms. Techniques such as differential privacy, homomorphic encryption, and secure multiparty computation are applied to ensure that an individual's privacy rights are upheld while enabling valuable data analysis. By leveraging Information Theory insights, organizations can strike a balance between data utility and privacy protection in the era of big data, fostering trust and confidence in data-driven decision-making processes.

Table 1: Privacy Preservation in the Age of Big Data from an Information Theory Perspective

Aspect	Description
Definition of Big Data	Big data refers to large and complex datasets characterized by volume, velocity, variety, and veracity.
Privacy Challenges	Challenges include protecting sensitive information, ensuring data anonymity, and mitigating privacy risks.
Role of Information Theory	Information theory provides mathematical frameworks to quantify and address privacy risks in big data environments.
Techniques and Mechanisms	Differential privacy, homomorphic encryption, and secure multiparty computation are key techniques for privacy preservation.
Applications	Real-world applications include healthcare analytics, financial fraud detection, smart city initiatives, and genomic research.
Future Directions	Future directions involve advancements in privacy-preserving techniques, interdisciplinary collaboration, and user-centric privacy solutions.
Opportunities	Opportunities include developing privacy-enhancing technologies, addressing ethical and societal implications, and establishing global data governance frameworks.

Differential privacy mechanisms and algorithms are foundational tools for preserving privacy in the analysis of sensitive data. These techniques ensure that the inclusion or exclusion of any individual's data does not significantly affect the outcome of a computation, providing strong privacy guarantees while enabling valuable data analysis. Some key differential privacy mechanisms and algorithms include the Laplace Mechanism: The Laplace mechanism adds noise sampled from a Laplace distribution to the output of a query, ensuring differential privacy. The amount of noise added is calibrated based on the sensitivity of the query, which quantifies how much the output changes when a single individual's data is modified. The Laplace mechanism provides ε-differential privacy, where ε represents the privacy budget or maximum allowable privacy loss. Exponential Mechanism: The exponential mechanism selects outputs from a set of candidate outcomes probabilistically, with probabilities determined by the utility or quality of each outcome and adjusted based on their sensitivity to changes in individuals' data. By selecting outputs in a privacy-aware manner, the exponential mechanism ensures differential privacy while maximizing the expected utility of the output. Differentially Private Stochastic Gradient Descent (DP-SGD): DP-SGD is a differential privacy algorithm commonly used in training machine learning models on sensitive data. It perturbs the gradients computed during model training with

noise sampled from a suitable distribution, such as the Gaussian or Laplace distribution, to ensure differential privacy. DP-SGD allows for the training of accurate models while protecting the privacy of individual training examples. Local Differential Privacy (LDP): Local differential privacy is a variant of differential privacy where individual data contributors perturb their data locally before sharing it with a data aggregator or analyst. This approach allows individuals to protect their privacy without relying on a trusted data curator. LDP mechanisms include strategies such as adding noise to data before transmission or applying randomized response techniques locally. Secure Multiparty Computation (SMPC): SMPC protocols enable multiple parties to jointly compute a function over their private inputs while preserving differential privacy. By leveraging cryptographic techniques, SMPC ensures that individual inputs remain private throughout the computation process, allowing parties to collaborate on data analysis tasks without sharing sensitive information. Differential Privacy in Database Systems: Differential privacy can be integrated into database systems to provide privacy guarantees for query responses. Techniques such as noisy aggregation, query rewriting, and query sampling are used to enforce differential privacy in database queries while maintaining data usability and query accuracy. These differential privacy mechanisms and algorithms form the foundation of privacy-preserving data analysis techniques, enabling organizations to analyze sensitive data while protecting individuals' privacy rights. When implementing differential privacy, it's essential to carefully calibrate privacy parameters, such as ε, and consider the trade-offs between privacy guarantees and data utility to ensure effective privacy protection without sacrificing analytical accuracy.

Future Directions: The future direction of privacy preservation in the age of big data, informed by insights from information theory, holds promise for innovative solutions that balance data utility with individual privacy rights. Advancements in information theory will continue to refine our understanding of privacy risks and guide the development of more robust privacy-preserving mechanisms. Differential privacy, homomorphic encryption, and secure multiparty computation will evolve to offer stronger privacy guarantees while minimizing the impact on data analysis tasks. Interdisciplinary collaboration among researchers, policymakers, and industry stakeholders will be essential for addressing complex privacy challenges and ensuring that privacy-preserving solutions align with legal, ethical, and societal norms. Furthermore, user-centric approaches and transparent data governance frameworks will empower individuals to control their data and make informed decisions about its use. By embracing these future directions, stakeholders can navigate the complexities of big data environments while upholding privacy as a fundamental right in the digital era.

5. Conclusion

In conclusion, the application of information theory insights to privacy preservation in the age of big data marks a crucial step forward in addressing the challenges posed by the vast collection and analysis of personal information. By leveraging concepts such as differential privacy, homomorphic encryption, and secure multiparty computation, stakeholders can develop robust mechanisms to safeguard confidentiality and integrity while maintaining data utility. These

techniques offer a promising avenue for balancing the benefits of data-driven innovation with the protection of individual privacy rights. However, ongoing research and collaboration are essential to continually refine and adapt these approaches to evolving threats and privacy concerns in the dynamic landscape of big data. Ultimately, by integrating information theory principles into privacy-preserving practices, we can foster trust, transparency, and accountability in the responsible use of data in our increasingly interconnected world.

Reference

[1] B. Jiang, M. Seif, R. Tandon, and M. Li, "Context-aware local information privacy," *IEEE Transactions on Information Forensics and Security,* vol. 16, pp. 3694-3708, 2021.

[2] J. S. Davis and O. Osoba, "Improving privacy preservation policy in the modern information age," *Health and Technology,* vol. 9, pp. 65-75, 2019.

[3] T. Wang, Z. Zheng, M. H. Rehmani, S. Yao, and Z. Huo, "Privacy preservation in big data from the communication perspective—A survey," *IEEE Communications Surveys & Tutorials,* vol. 21, no. 1, pp. 753-778, 2018.

[4] J. Puaschunder, "Towards a utility theory of privacy and information sharing and the introduction of hyper-hyperbolic discounting in the digital big data age," in *Research anthology on privatizing and securing data*: IGI Global, 2021, pp. 68-111.

[5] B. Jiang, M. Li, and R. Tandon, "Context-aware data aggregation with localized information privacy," in *2018 IEEE Conference on Communications and Network Security (CNS)*, 2018: IEEE, pp. 1-9.

[6] L. Xu, C. Jiang, J. Wang, J. Yuan, and Y. Ren, "Information security in big data: privacy and data mining," *Ieee Access,* vol. 2, pp. 1149-1176, 2014.

[7] M. G. Raeini and M. Nojoumian, "Privacy-preserving big data analytics: from theory to practice," in *Security, Privacy, and Anonymity in Computation, Communication, and Storage: SpaCCS 2019 International Workshops, Atlanta, GA, USA, July 14–17, 2019, Proceedings 12*, 2019: Springer, pp. 45-59.

[8] B. Jiang, M. Li, and R. Tandon, "Local information privacy and its application to privacy-preserving data aggregation," *IEEE Transactions on Dependable and Secure Computing,* vol. 19, no. 3, pp. 1918-1935, 2020.

[9] J. M. Puaschunder, "Dignity and utility of privacy and information sharing in the digital big data age," *International Journal of Commerce and Management Research,* vol. 5, no. 4, pp. 62-70, 2018.

[10] W. Zhang, B. Jiang, M. Li, and X. Lin, "Privacy-preserving aggregate mobility data release: An information-theoretic deep reinforcement learning approach," *IEEE Transactions on Information Forensics and Security,* vol. 17, pp. 849-864, 2022.

[11] P. Ram Mohan Rao, S. Murali Krishna, and A. Siva Kumar, "Privacy preservation techniques in big data analytics: a survey," *Journal of Big Data,* vol. 5, no. 1, p. 33, 2018.

[12] O. Tene and J. Polonetsky, "Big data for all: Privacy and user control in the age of analytics," *Nw. J. Tech. & Intell. Prop.,* vol. 11, p. 239, 2012.

[13] O. Choudhury *et al.*, "Anonymizing data for privacy-preserving federated learning," *arXiv preprint arXiv:2002.09096*, 2020.

[14] X. Wu, T. Wu, M. Khan, Q. Ni, and W. Dou, "Game theory based correlated privacy-preserving analysis in big data," *IEEE Transactions on Big Data*, vol. 7, no. 4, pp. 643-656, 2017.

[15] J. Zhao, Y. Chen, and W. Zhang, "Differential privacy preservation in deep learning: Challenges, opportunities, and solutions," *IEEE Access*, vol. 7, pp. 48901-48911, 2019.