Chelonian Conservation And Biology

Vol. 17No.2 (2022) | https://www.acgpublishing.com/ | ISSN - 1071-8443

DOI:doi.org/10.18011/2022.10(4).315.327

PRIVACY-PRESERVING TECHNIQUES IN AI-DRIVEN BIG DATA CYBER SECURITY FOR CLOUD

Anjan Kumar ReddyAyyadapu

anjanreddy8686@gmail.com

Abstract

Artificial intelligence (AI) and big data technologies are evolving at a rapid pace, completely changing the cybersecurity landscape in cloud environments. But this advancement also presents previously unheard-of difficulties, namely with regard to privacy issues when managing enormous volumes of sensitive data. This abstract investigates privacy-preserving methods in the context of cloud-based AI-driven big data cybersecurity. Strong privacy safeguards are essential as more and more businesses rely on cloud services to handle and keep their data. This study explores cutting-edge techniques and tools designed to protect user privacy while utilising AI to analyses large amounts of data for cybersecurity purposes. Thanks to digital technology, a wide range of organizations, including banks, supply chains, e-commerce, healthcare, and retail, are producing enormous amounts of data. Machines and people both add to data through internet-based records, shut circuit TV streaming, and different means. Online entertainment and cell phones make tremendous measures of data consistently. To aid in direction, the monstrous measures of data delivered by the various sources can be handled and analyzed. Data examination, in any case, is defenseless against privacy encroachment. Suggestion frameworks, which are habitually utilized by web-based business destinations like Amazon and Flipkart to propose things to clients based on their buying propensities, are one utilization of data examination that could bring about deduction assaults.

Keywords:*Privacy-Preserving,Techniques,Ai-Driven,Big Data,Cyber Security,Cloud*

1.INTRODUCTION

Big data technology and artificial intelligence (AI) have emerged as distinguishing features of the contemporary digital landscape, revolutionizing how businesses run and handle data. The combination of AI and large datasets in cybersecurity provides hitherto unseen possibilities for anomaly analysis, threat detection, and proactive defense strategies. These developments are made possible in large part by the cloud, which serves as a central hub for data processing and storage. This enables enterprises to access the computing capacity needed for complex AI-driven analysis. But while AI's potential for cybersecurity grows, a related worry about the privacy of sensitive data processed and stored in cloud environments has also surfaced. These ecosystems handle enormous amounts of data, including personally identifiable information (PII), confidential company information, and other delicate information that, in the event of a breach, might have dire repercussions. It has become increasingly difficult to protect people's and organizations' privacy in this complex network of linked AI and big data technologies.

The purpose of this study is to explore the field of privacy-preserving methods specifically designed for cloud computing-based AI-driven big data cybersecurity. Such solutions are necessary since it has been realized that standard cybersecurity measures might not be able to adequately handle the complex privacy challenges that AI algorithms that analyses large datasets face. It is crucial to find a balance between the use of potent AI technologies for cybersecurity and the preservation of user privacy as more and more businesses move their operations to cloud platforms. Among the creative solutions investigated in this study are cryptographic protocols like federated learning, secure multiparty computation, and homomorphic encryption. By enabling AI models to function on encrypted data, these methods seek to reduce the possibility that private information will be revealed during the analysis phase. Furthermore, in the context of big data cybersecurity, differential privacy techniques—which introduce controlled noise into datasets to protect individual privacy—are investigated as possible defenses. The ultimate objective of the project is to support the creation of a thorough framework that will allow organizations to reap the rewards of AI-driven big data cybersecurity while respecting the rights of individuals and entities to privacy. In addition to promoting a more safe and dependable cloud environment, addressing these privacy issues guarantees the ethical and responsible use of cutting-edge technology in the rapidly changing cybersecurity landscape.

1.1 AI and Big Data in Cybersecurity

Organizations' approaches to threat detection, analysis, and overall security policies have changed dramatically as a result of the integration of artificial intelligence (AI) and big data in cybersecurity.

Superior Danger Identification:

- Artificial Intelligence (AI) provides cybersecurity with sophisticated analytics capabilities that enable real-time threat assessment of intricate and dynamic threats.
- By using historical data, machine learning algorithms can identify trends that point to malicious activity, improving their capacity to identify cyber dangers that have never been seen before.

Analysis of Anomalies:

- Combining AI with big data analytics makes it possible to find abnormalities in massive databases.
- Artificial intelligence (AI) algorithms can differentiate between typical behavior and odd patterns, which can assist organizations in spotting possible security lapses or unusual activity that can point to a cyberattack.

Preventive Defense Techniques:

- Proactive defense techniques can be implemented by cybersecurity systems powered by AI.
- Organizations can boost their overall security posture by taking preventive action by using predictive analytics to foresee possible security threats and vulnerabilities.

Conduction Examination:

- By continuously observing and learning from user actions and network behaviors, AI is able to do behavioral analysis.
- This makes it possible to spot behavioral anomalies, which aids in identifying insider threats and advanced cyberattacks that conventional security systems would miss.

1.2 Privacy Concerns in AI-Driven Big Data Cybersecurity

As AI-driven big data cybersecurity becomes more widely used, a number of privacy issues arise, most of which are related to the handling, processing, and storing of enormous volumes of sensitive data. Comprehending these issues is essential to guaranteeing that cybersecurity endeavors conform to privacy anticipations and legal mandates. The following are important specifics about privacy issues in big data cybersecurity driven by AI:

Exposition to Sensitive Data:

- Personally identifiable information (PII), confidential company information, and other sensitive information are frequently found in the massive datasets utilized for AI-driven cybersecurity.
- Severe privacy breaches may arise from unintentional or authorized disclosure of this data during analysis.

Discrimination and Algorithmic Bias:

- AI systems may unintentionally provide discriminatory results by aggravating or maintaining preexisting biases in the training set.
- When biases in the data cause people to be unfairly targeted or influenced by judgements made by AI algorithms, privacy problems arise**.**

Ownership of Data and Informed Consent:

- It's possible that users are unaware of the full scope of how their data is being used for cybersecurity driven by AI.
- As issues with data ownership and informed permission surface, it becomes more important than ever to have open lines of communication and explicit guidelines for using user data.

2. REVIEW OF LITERATURE

Almasoud et al.'s (2022) paper offers a thorough investigation of deep learning combined with image classification to strengthen Cyber-Physical Systems (CPS) in the medical field. The writers skillfully negotiate the intricacies of safe medical settings, emphasizing the critical function that deep learning performs in image categorization. The research adds to the changing

field of safe systems in a crucial area by addressing the nexus between technology and healthcare.

With an emphasis on 6G, Atlam, Azad, Altamimi, and Fadhel (2022) explore the possibilities for wireless communication in the future. Their research highlights the benefits of combining blockchain and AI, providing information on how these two technologies might work together to improve security and privacy in the developing 6G environment. This research makes a substantial contribution to the discussion on the security concerns of developing wireless networks by offering both a theoretical underpinning and practical implementations.

For Internet of Things (IoT) networks, Attkan and Ranga (2022) provide a thorough analysis covering classical, blockchain, and AI-based key security. Through its examination of the various security issues related to IoT, the article offers an invaluable resource for both scholars and industry professionals. Several security paradigms are included to deepen the conversation and provide a comprehensive roadmap to comprehending and resolving security issues in the rapidly changing IoT network landscape.

Bellagarda and Abu-Mahfouz (2022) provide an extensive overview of the convergence of artificial intelligence (AI) and distributed ledger technology (DLT). Their work offers a current assessment of this convergence's status, pointing out its main obstacles and suggesting its future paths. This survey, which was published in IEEE Access, is a useful tool for comprehending the changing environment at the intersection of DLT and AI. It provides insights that are relevant to both academics and business experts.

In the context of the Industrial Internet of Things (IIoT), Bugshan, Khalil, Rahman, Atiquzzaman, Yi, and Badsha (2022) add to the body of work with an emphasis on reliable and privacy-preserving federated deep learning. Their work tackles the crucial requirement for strong privacy and security safeguards in IIoT architectures. The suggested approach takes into account the unique needs of the industrial environment by prioritizing privacy and trust preservation in addition to the incorporation of federated deep learning.

3. PRIVACY THREATS IN DATA ANALYTICS

The ability to control access and conclude what data can be shared is known as privacy. Since the data is claimed by the data carrier, it represents a danger to individual privacy on the off chance that it is in the public domain. Interpersonal interaction applications, sites, versatile applications, online business locales, banks, emergency clinics, and different foundations can all be data holders. Guaranteeing the privacy of clients' data is the obligation of the data holder. Notwithstanding data in the public domain, clients themselves might add to data spills readily or unconsciously. For example, most of portable applications request to get to our contacts, documents, cameras, and other data. We agree to all agreements without perusing the privacy proclamation, which adds to data spillage.

Along these lines, it's vital to illuminate cell phone clients about privacy issues and privacy concerns. Key privacy dangers include: (1) Separation; (2) Exposure; (3) Reconnaissance; and (4) Misuse and Individual Reverence.

3.1 Surveillance

Various organizations, like those in retail and online business, look at the buying examples of their clients with an end goal to foster extraordinary contributions and worth added administrations. Long range informal communication stages give thoughts for new acquaintances, objections to visit, people to follow, and other substance in light of feeling examination and assessment data. This is just attainable assuming they watch out for each exchange made by their clients. This represents a serious gamble to privacy in light of the fact that nobody agrees to being watched.

3.2 Disclosure

Envision an emergency clinic that has patient data on document, like Compress, orientation, age, and sickness.

The data proprietor has given delicate, by and by recognizable data to an outsider for investigation, guaranteeing that the data can't be connected to the person. An outsider data expert can plan this data onto freely open outer data sources, like evaluation data, and distinguish

people who are experiencing explicit problems. This is the technique by which a singular's confidential data might be uncovered, which is viewed as a significant intrusion of their privacy.

3.3 Discrimination

Separation is the predisposition or disparity that might emerge from the exposure of a singular's confidential data. For instance, factual assessment of political race results showed that inhabitants of one local area were eagerly against the party that comprised the organization. The public authority is presently allowed to dismissal or show inclination towards that local area.

3.4 Personal embracement and abuse

Any revelation of an individual's confidential data may possibly bring about misuse or individual embracement. For example, an individual may be watchfully taking prescription for a certain issue and regularly buying meds from a drug store. The clinical store may, as a component of their standard working technique, give telephone updates and offers pertaining to these drugs. This could bring about private acknowledgment and even abuse on the off chance that a relative notification it.

Data privacy will be influenced by data examination action. Regulations safeguarding privacy are being implemented in numerous countries. Another significant element adding to privacy breaks is obliviousness. For example, a ton of cell phone clients know nothing about the data that various applications take from their gadgets. Just 17% of cell phone clients know about privacy issues, as per prior examinations.

4. PRIVACY PRESERVATION METHODS

Although many privacy-preserving approaches have been developed, the majority of them rely on data anonymization. The ways for maintaining privacy are listed below.

- K anonymity
- L diversity
- T closeness
- Randomization
- Data distribution

- Cryptographic techniques
- Multidimensional Sensitivity Based Anonymization (MDSBA)

❖ **K anonymity**

Anonymization is the act of changing data before it is accommodated data investigation. In the event that the anonymized data is planned with outer data sources trying to de-recognize, it will bring about K vague records since de-ID isn't feasible. The homogeneity assault and the foundation information attack are the two sorts of assaults that can target K obscurity. To ensure anonymization, a couple of the techniques utilized are In secret and Mondrian. Table 1 shows the patient data with K secrecy applied. Data before anonymization is shown in the table.

To ensure three vague records if an endeavor is made to distinguish a particular individual's data, the K namelessness calculation is applied, with a worth of 3. The two credits — Postal division and age — that are shown in Table 1 are dependent upon K obscurity. Table 2 shows the result of applying anonymization to the age and postal district ascribes.

Table 2: Prior to being anonymized, patient data

ID	Patient ID	Age	Diagnosis
1	32561	32	Cardiac problem
2	41251	14	Cardiac problem
3	32541	25	Cardiac problem
4	53262	36	Skin allergy
5	52012	44	Cardiac problem
6	43625	21	Cancer
7	30125	36	Cardiac problem
8	42361	10	Cancer
9	53621	15	Cance

This dataset contains patient data that includes age, related diagnoses, and unique identifiers (Patient IDs). The patients' health issues are varied, with heart issues taking the stage. Many people—patients with IDs 32561, 41251, 32541, 52012, and 30125, in particular—have been diagnosed with heart problems, suggesting that cardiovascular health problems may be common in this community. Moreover, cancer diagnoses have been made for patients with IDs 43625, 42361, and 53621, indicating the existence of oncological diseases in the dataset. Remarkably, the dataset also contains an entry (Patient ID 53262) with the diagnostic "Skin allergy," demonstrating the variety of medical conditions among the patients who were registered. The identified health issues within this group may warrant additional analysis and investigation into trends, risk factors, and possible interventions. This dataset may provide a foundation for such investigations. Improved patient outcomes and more focused healthcare initiatives can result from an understanding of these patterns.

Table 3: After utilising age and zip code anonymization

Sno	**Zip**	**Age**	**Disease**
1	262	1	Cardiac problem
2	362	3	Cardiac problem
3	414	2	Cardiac problem
4	536	>50	Skin allergy
5	458	>50	Cardiac problem
6	369	>50	Cancer
7	585	4	Cardiac problem
8	695	4	Cancer
9	584	4	Cancer

This dataset includes data on various individuals, denoted by their zip codes (Zip), ages, and associated diagnosed ailments, in addition to their serial numbers (Sno). The dataset is divided into two halves, each of which has unique entries for different patients. The first set of patients is

denoted by serial numbers 1 through 9 and has a wide variety of medical issues. Notably, people with ages ranging from 1 to over 50 who reside in zip codes 262, 362, 414, 458, and 585 have been diagnosed with cardiac issues. Furthermore, patients with skin allergies and cancer diagnoses are found in zip codes 536, 369, 695, and 584. These results point to a combination of oncological and cardiovascular health problems in this group.The second portion of the dataset is a mirror image of the first, offering identical but redundant details about the same patients. The inclusion of patients whose ages are designated as ">50" represents a broad category for those over 50, which may indicate that they are a population more vulnerable to specific medical conditions. This dataset provides a starting point for additional research into the prevalence of particular diseases in various age groups and geographical areas. In order to address the health requirements of the population under consideration, healthcare professionals, policymakers, and researchers can better adapt interventions, resource allocation, and preventive measures by having a better understanding of these patterns.

- **L diversity**

Anonymization is the demonstration of changing data before it is obliged data examination. If the anonymized data is arranged with external data sources attempting to de-remember, it will achieve K ambiguous records since de-ID isn't attainable. The homogeneity attack and the establishment data assault are the two kinds of attacks that can target K lack of clarity. To guarantee anonymization, several the techniques used are stealthily and Mondrian. Table 1 shows the patient data with K mystery applied. Data before anonymization is displayed in the table.

To guarantee three obscure records in the event that an undertaking is made to recognize a specific person's data, the K anonymity estimation is applied, with a value of 3. The two credits — Postal division and age — that are displayed in Table 1 are reliant upon K lack of definition. Table 2 shows the consequence of applying anonymization to the age and postal locale credits.

- **T closeness**

An improvement to L variety is the T closeness measure, which characterizes an equivalency class as having "T closeness" on the off chance that all of the equivalency classes have T closeness and the distance between the delicate property dispersions in the class is under an edge. Each quality's T vicinity can be processed according to the delicate property.

Table 4 shows that even with John's age of 27, it will be hard to decide if he falls into the low-pay classification and whether he has a cardiovascular condition. Despite the fact that embracing T vicinity may not necessarily in all cases bring about the right circulation of data, it might guarantee quality divulgence.

4. RESULTS AND DISCUSSION

4.1 Randomization technique

The act of adding noise to data, known as randomization, is often carried out using a probability distribution. Surveys and sentiment analysis are two areas where randomization is used. It is not necessary for randomization to be aware of other records in the data. It can be used in the pre-processing and data collection phases. Randomization has no overhead related to anonymization. However, as our experiment, which is detailed below, shows, applying randomization on huge datasets is not feasible due to time complexity and data utility.

Ten thousand entries were put into the Hadoop Distributed File System from an employee database, and a Map Reduce Job was used to process the data. We've tried categorizing the staff members according to age groupings and salaries. After completing the Map Reduce process, the following observations were made in order to apply randomization: 5k records were randomly added to create a database of 15k records.

- As the volume of data rose, more Mappers and Reducers were employed.
- The outcomes before and after randomization differed significantly.
- A few records that are outliers are susceptible to adversarial attack and are not affected by randomization.
- Randomization may not be appropriate for maintaining privacy when it comes to attribute sharing, as privacy preservation at the expense of data utility is not valued.

Table 4:T closeness privacy protection method

Sno	Zip	Age	Salary	Disease
1	262	1	5362	Cardiac problem
2	362	3	6251	Cancer

3	414	2	7145	Skin allergy
4	536	>50	8256	Skin allergy
5	458	>50	9362	Cardiac problem
6	369	>50	4582	Flu

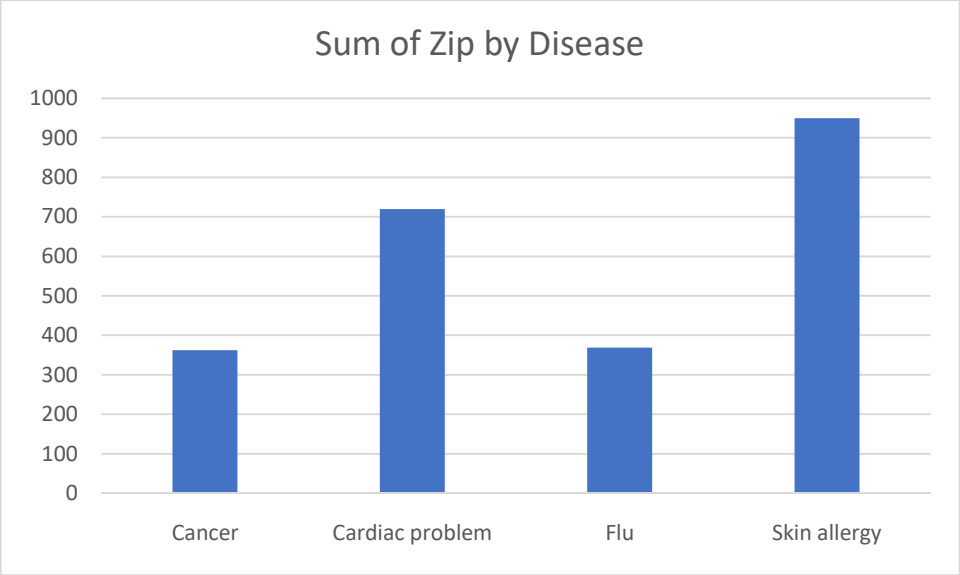

Figure 1:T closeness privacy protection method

The dataset offered includes data on a variety of people who live in various zip codes (Zip) and are recognized by serial numbers (Sno). Every person is identified by their age, stated income, and associated medical condition. One noteworthy aspect of this dataset is the wage data, which gives the analysis a new perspective. Patients with serial numbers 1 through 6 represent a spectrum of ages, with a focus on those who are over 50 (designated as ">50"). It is interesting to note that people in this age range have been diagnosed with a wide range of illnesses, such as the flu, cancer, and heart issues.

The dataset highlights the possible correlation between age, income, and the frequency of specific illnesses. The salary range reported by patients with heart difficulties is 5362 to 9362, demonstrating a variation in income levels within this health category. In a similar vein, those with skin allergies or cancer diagnoses also exhibit a range in reported salaries. This dataset

provides a chance to explore the connections between age, socioeconomic status, and the probability of particular health problems. In order to create focused interventions and healthcare policies that take into consideration the complex nature of health disparities within this group, healthcare practitioners and policymakers may find it essential to comprehend these relationships.

4.2 Data distribution technique

The data in this technique is dispersed over numerous sites. Data distribution can be accomplished in two ways:

- Horizontal distribution of data
- Vertical distribution of data

4.2.1 Horizontal distribution:

An even conveyance, as portrayed happens when data is scattered all through various spots that share indistinguishable elements.

Just when certain total capabilities or cycles are to be performed on the data without really sharing the data may even dispersion of data be applied. All for example, a retail store might utilize examination programming that figures on total data to survey deals across its branches. Notwithstanding, imparting the data to an outsider examiner might be important for the data holder as a component of the data examination process, which could disregard their privacy. While strategies for order and grouping can be utilized with disseminated data, privacy isn't ensured. Consequences of total capabilities might aid one party in finding data kept with different gatherings assuming the data is scattered all through different destinations that are claimed by different associations. We guess that all taking part destinations will be honest with each other in such conditions.

4.2.2 Vertical distribution of data:

The conveyance of individual explicit data across different spots under the guardianship of different associations is alluded to as an upward circulation. For example, throughout a criminal examination, cops might want to know explicit data about a suspect, like their own, monetary,

and wellbeing narratives. It's conceivable that not this data is all on one site. This sort of dissemination, known as upward circulation, has few an individual's traits at each spot. There is a gamble of privacy break when data from these sites should be consolidated for investigation purposes.

It is extremely challenging to maintain privacy while performing data investigation on in an upward direction appropriated data, in light of the fact that the traits are scattered across numerous locales under the guardianship of different gatherings. This is particularly obvious if the datasets are shared. For example, to look into the blamed's character, the police researching official might have to obtain data about him from his bank, work, and wellbeing organization. During this system, the charged may uncover to the examining official certain private and delicate realities that could make them feel manhandled or humiliated. Complete records that are not needed for examination can't be anonymized. Data circulation intently looks like encryption draws near, yet it won't ensure privacy assurance.

4.3 Cryptographic techniques

Prior to delivering the data for examination, the data proprietor might encode it. In any case, it is truly challenging to scramble colossal measures of data utilizing customary encryption techniques, and they ought to possibly be utilized while gathering data. Differential privacy approaches have proactively been utilized in circumstances where a few total data calculations are performed without the sources of info being shared. A capability F(x, y) will be determined, for example, in the event that x and y are two data things, to obtain some total data from both x and y without really sharing x and y. This is relevant in circumstances when x and y are kept with discrete gatherings, like in vertical circulation. Differential privacy, nonetheless, can't be utilized in the event that the data is kept in a solitary area and is overseen by a solitary association. Albeit utilized, an alternate related technique known as secure multiparty calculation has demonstrated to be lacking for maintaining secrecy. Assuming that encryption is utilized during data investigation, the handiness of the data will diminish. Not exclusively is tus encryption hard to set up, yet it likewise decreases the helpfulness of the data.

4.4 Multidimensional Sensitivity Based Anonymization (MDSBA)

The conventional techniques of anonymization, known as hierarchical speculation and base up speculation, were utilized on very much addressed organized data records. Yet, it's undeniably challenging to utilize similar for enormous scope data sets, which creates some issues with versatility and data misfortune. An improved sort of anonymization that has demonstrated to be more powerful than customary anonymization strategies is called multi-faceted responsiveness based anonymization.

A superior anonymization strategy that can be utilized on tremendous data sets with less data misfortune and predefined semi identifiers is called multi-faceted responsiveness based anonymization. Huge data sets have been taken care of utilizing the Apache Guide Diminish structure as a feature of this technique. The data in a customary Hadoop Conveyed Record Framework is separated into 64-or 128-MB blocks, which are then scattered among a few hubs without considering the items in the blocks.

Involving channels in the Apache Pig programming language, the data is separated into unmistakable sacks as per the semi identifiers' likelihood dispersion as a feature of the Complex Responsiveness Based Anonymization approach.

Base up speculation is utilized in Multi-layered Responsiveness Based Anonymization, however just for a subset of characteristics with explicit class values, where class means a delicate property. Dispersion of data was made actually when contrasted with the conventional block strategy. Utilizing Apache Pig, four semi identifiers were utilized to anonymize the data.

Assuming that the sack incorporates not many ascribes, it tends to be shielded from foundation information assault since the data is upward apportioned into various gatherings. This approach additionally makes it trying to plan the data with outside sources to uncover any actually recognizable data.

Apache Pig was utilized for the execution of this methodology. Since Apache Pig is a prearranging language, less work has gone into its turn of events. Nonetheless, on the grounds that every Apache Pig script should ultimately be changed into a Guide Lessen work, the coding proficiency of Apache Pig is relatively lower than that of Guide Decrease occupations. Huge scope data is more qualified for Multi-faceted Responsiveness Based Anonymization, yet just

when the data is very still. For streaming data, Multi-faceted Awareness Based Anonymization isn't appropriate.

4.5 Discussion

The properties of different privacy preservation strategies, such as data type, data utility, attribute preservation, and complexity, have all been explored. Table 5 presents a comparison of several privacy preservation approaches.

Table 5:A comparison of privacy-preserving methods

Features	Privacy preservation techniques	Anonymization techniques	Cryptographic techniques	Data distribution	Randomization	MDSBA	Suitability for unstructured data	Attribute preservation	Damage to data utility	Very complex to apply	Accuracy of results of data analytics
×	×	×	×	✓	×	×	✓	×	×	×	×
×	×	×	✓	×	×	×	✓	×	✓	×	×
×	×	×	✓	×	×	×	✓	×	✓		×
✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓

Creating Privacy-Preserving Mechanisms That Are Versatile Enough to Handle Both Structured and Unstructured Data Formats: This problem's main obstacle is developing mechanisms that protect privacy for both structured and unstructured data. Unstructured data, like text documents or multimedia files, lacks a predetermined framework, whereas structured data, such that found in databases, is arranged in tables. Creating techniques and algorithms that can successfully safeguard sensitive data in a variety of data types is the first step in creating tangible solutions.

Robust and Scalable Methods for Heterogeneous Datasets on a Large Scale: Privacy-preserving strategies need to be robust enough to manage heterogeneous information and scalable to handle huge datasets as the amount and diversity of data continue to increase. This entails creating algorithms and systems that, without sacrificing efficacy or efficiency, can securely and efficiently manage data from a variety of sources, formats, and scales.

Maintaining Data in Its Original Form for Analytics: This situation presents a problem in maintaining data in its original, native form while enabling relevant data analytics. By using this method, substantial data modifications are not necessary, enabling analysis to be done directly on the raw data. The objective is to achieve a balance between privacy and usefulness so that important insights can be obtained without disclosing private information.

Creating Methods Beyond Anonymization: One popular privacy tactic is anonymization, which entails deleting personally identifiable data. However, new methods must be developed in response to growing privacy issues such identity exposure, discrimination, and spying. This challenge compels academics to investigate novel approaches that surpass conventional anonymization, offering improved defense against changing privacy threats.

Optimizing Data Utility While Preserving Privacy: A basic problem is striking a balance between data utility and privacy. Data analysis should not be unduly restricted by privacy-preserving measures. Finding the ideal balance guarantees that businesses can protect individuals' privacy while gaining insightful knowledge from the data. This calls for the thorough evaluation of trade-offs and the creation of methods that maximize privacy as well as benefit.

5.CONCLUSION

The field of AI-driven big data cybersecurity for cloud environments is evolving, with new problems as well as opportunities to explore when it comes to privacy-preserving solutions. The comprehensive analysis of current approaches highlights the widespread use of data anonymization, with standout methods including data distribution, randomization, K anonymity, L diversity, T closeness, and multidimensional sensitivity-based anonymization (MDSBA). Although these techniques make a substantial contribution to the privacy of structured data, the literature also identifies important gaps, particularly when it comes to resolving the privacy issues raised by the fact that over 80% of modern data is unstructured. In order to address the

issues raised, practical solutions that protect privacy in both structured and unstructured data must be developed. Scalability, resilience, and maintaining data in its original format for analytical purposes should all be prioritized. Moreover, the necessity of creative methods beyond anonymization underscores the dynamic nature of the threat landscape and the requirement for flexible privacy preservation tactics. To sum up, the investigation of privacy-preserving methods in AI-driven big data cybersecurity for the cloud indicates the need for all-encompassing and flexible solutions that fill in the gaps and guarantee the efficient protection of sensitive data in the constantly changing field of cloud-based cybersecurity

References

1. Almasoud, A. S., Abdelmaboud, A., Alsubaei, F. S., Hamza, M. A., Yaseen, I., Abaker, M., ... &Rizwanullah, M. (2022). Deep Learning with Image Classification Based Secure CPS for Healthcare Sector. Computers, Materials & Continua, 72(2).
2. Atlam, H. F., Azad, M. A., Altamimi, M., &Fadhel, N. (2022). Role of Blockchain and AI in Security and Privacy of 6G. In AI and Blockchain Technology in 6G Wireless Network (pp. 93-115). Singapore: Springer Nature Singapore.
3. Attkan, A., & Ranga, V. (2022). Cyber-physical security for IoT networks: a comprehensive review on traditional, blockchain and artificial intelligence based key-security. Complex & Intelligent Systems, 8(4), 3559-3591.
4. Bellagarda, J. S., & Abu-Mahfouz, A. M. (2022). An updated survey on the convergence of distributed ledger technology and artificial intelligence: Current state, major challenges and future direction. IEEE Access, 10, 50774-50793.
5. Bugshan, N., Khalil, I., Rahman, M. S., Atiquzzaman, M., Yi, X., &Badsha, S. (2022). Toward trustworthy and privacy-preserving federated deep learning service framework for industrial internet of things. IEEE Transactions on Industrial Informatics, 19(2), 1535-1547.
6. Chakraborty, C. (Ed.). (2022). Digital Health Transformation with Blockchain and Artificial Intelligence. CRC Press.
7. Chawla, G., & Rizvi, S. W. A. (2022, November). Healthcare Data Security in Cloud Environment. In International Conference on Intelligent Vision and Computing (pp. 245-253). Cham: Springer Nature Switzerland.

8. Fadi, O., Karim, Z., & Mohammed, B. (2022). A survey on Blockchain and Artificial intelligence technologies for enhancing security and privacy in smart environments. IEEE Access, 10, 93168-93186.
9. Firouzi, F., Jiang, S., Chakrabarty, K., Farahani, B., Daneshmand, M., Song, J., &Mankodiya, K. (2022). Fusion of IoT, AI, edge–fog–cloud, and blockchain: Challenges, solutions, and a case study in healthcare and medicine. IEEE Internet of Things Journal, 10(5), 3686-3705.
10. Gajmal, Y. M., &Udayakumar, R. (2022). Privacy and utility-assisted data protection strategy for secure data sharing and retrieval in cloud system. Information Security Journal: A Global Perspective, 31(4), 451-465.
11. Hewage, C. T., Khattak, S. K., Ahmad, A., Mallikarachchi, T., Ukwandu, E., &Bentotahewa, V. (2022). Multimedia Privacy and Security Landscape in the Wake of AI/ML. Social Media Analytics, Strategies and Governance, 203-228.
12. Jagatheesaperumal, S. K., Rahouti, M., Ahmad, K., Al-Fuqaha, A., & Guizani, M. (2022). The duo of artificial intelligence and big data for industry 4.0: Applications, techniques, challenges, and future research directions. IEEE Internet of Things Journal, 9(15), 12861-12885.
13. Kataria, D., Walid, A., Daneshmand, M., Dutta, A., Enright, M. A., Gu, R., ... &Darema, F. (2022, October). Artificial Intelligence And Machine Learning. In 2022 IEEE Future Networks World Forum (FNWF) (pp. 1-70). IEEE.
14. Kersten, A., & Robinson, I. A. (2022). Data Protection or Data Utility?.
15. Müftüoğlu, Z., Kızrak, M. A., &Yıldırım, T. (2022). Privacy-Preserving Mechanisms with Explainability in Assistive AI Technologies. Advances in Assistive Technologies: Selected Papers in Honour of Professor Nikolaos G. Bourbakis–Vol. 3, 287-309.
16. Sandeepa, C., Siniarski, B., Kourtellis, N., Wang, S., & Liyanage, M. (2022). A survey on privacy for B5G/6G: New privacy challenges, and research directions. Journal of Industrial Information Integration, 100405.
17. Sharma, A., Sharma, V., Jaiswal, M., Wang, H. C., Jayakody, D. N. K., Basnayaka, C. M. W., & Muthanna, A. (2022). Recent trends in AI-based intelligent sensing. *Electronics*, *11*(10), 1661.

18. Sharma, A., Sharma, V., Jaiswal, M., Wang, H. C., Jayakody, D. N. K., Basnayaka, C. M. W., & Muthanna, A. (2022). Recent Trends in AI-Based Intelligent Sensing. Electronics 2022, 11, 1661.
19. Tabassum, I., Bazai, S. U., Zaland, Z., Marjan, S., Khan, M. Z., & Ghafoor, M. I. (2022, December). Cyber Security's Silver Bullet-A Systematic Literature Review of AI-Powered Security. In 2022 3rd International Informatics and Software Engineering Conference (IISEC) (pp. 1-7). IEEE.
20. Tatineni, S. (2022). INTEGRATING AI, BLOCKCHAIN AND CLOUD TECHNOLOGIES FOR DATA MANAGEMENT IN HEALTHCARE. Journal of Computer Engineering and Technology (JCET), 5(01).