Preserving Confidentiality: An Extensive Review of Privacy Techniques in Data Science

Yuvraj Panchal, Rashid Sheikh, Shiv Shankar Rajput, Narendra Pal Singh Rathore
Department of Computer Science and Engineering,
Acropolis Institute of Technology and Research, India
yuvrajpanchal3176@gmail.com, narendrasingh@acropolis.in,
rashidsheikh@acropolis.in

Abstract: - In the field of facts technology, privacy-keeping techniques are important to guaranteeing the ethical and stable use of information. This widespread assessment looks at gives an in-intensity exam of privateness-preserving methods, which includes federated getting to know, homomorphic encryption, stable multi-birthday celebration computation (SMPC), anonymization strategies, and differential privateness. This paper explains the theoretical underpinnings, actual-world packages, difficulties, and future opportunities of privacy-keeping strategies in statistics technological know-how, with an emphasis on latest traits. The observe explores anonymization techniques such generalization, suppression, k-anonymity, l-variety, and t-closeness after offering an define of privateness dangers in data science. The standards and makes use of federated gaining knowledge of, homomorphic encryption, SMPC, and differential privacy are then tested, emphasizing how those strategies guard sensitive statistics and sell teamwork in evaluation. Examples from the fields of healthcare, banking, telecommunications, social media, and government are tested to reveal how beneficial privacy-keeping strategies are inside the actual global. The look at additionally lists critical problems such as scalability, usability, and resilience to adversarial assaults and shows viable future paths for this subject's studies and boom. We desire to offer pupils, practitioners, and policymakers a radical knowledge of privateness-preserving strategies and their implications for moral and long-term records-pushed choice-making via this enormous survey.

Key-Words: *- Privacy Preservation, Encryption, Federated learning, Anonymization, Privacy Threats, SMPC.*

I. Introduction

This paper intends to pinpoint the strengths, limitations and comparative effectiveness of a range of privacy-preserving techniques by means of systematic examination across different use cases and data types. By critically assessing state-of-the-art solutions, this survey is expected to help stakeholder make data privacy strategy choices in their own domains [2]. The hazards of sensitive information breaches, abuse, and unauthorized access are growing along with the sophistication and ubiquity of data collection technologies. As such, there has never been a greater demand in data science for strong privacy-preserving strategies. Techniques for protecting privacy cover a wide range of approaches and tools that are intended to protect private data while maintaining usefulness for data analysis and use [5]. In the rapidly changing landscape of data science, the use of a huge quantity of data has become paramount to finding knowledge and innovation across various domains. However, this sudden inflow of data brings significant concerns about its privacy and security. As data collection methods grow more sophisticated and widespread, so do also the threats that come with unauthorized access, misuse or breaches of sensitive information. Henceforth, there has never been such a criticality in data science for robust privacy-preserving techniques. The purpose of this review is to provide an extensive understanding of the present position of privacy-preserving methods within data science. To identify common problems, new trends and new research directions, this paper will need to review, sort out and classify different techniques used in methodologies such as; algorithms and frameworks that are designed to deal with these challenges. Its mission is to join the conversation on privacy through data protection and contribute practical knowledge for researchers, practitioners and policy makers [4]. The increasing amounts of data produced in the current digital era provide tremendous potential for advancement and creativity in a wide range of fields.

II. Methodology

1. Literature Review: The approach begins with a comprehensive literature review to identify and collect relevant research papers, articles and literature on privacy protection strategies in data science. This step ensures that the analysis covers a wide range of methodologies and recent developments in the field [3].

2. Selection criteria: A set of criteria is developed to determine the most appropriate documents for analysis and narrowed down. This includes reviewing projects that address privacy concerns in data science, applying state-of-the-art techniques, and making substantial theoretical or practical contributions to knowledge about privacy protection on the main.

3. Classification Framework: A class framework is evolved to categorize the identified privateness-keeping strategies. This framework normally consists of categories consisting of:

- **Anonymization Techniques**: Methods like k-anonymity, l-variety, t-closeness, and differential privacy.
- **Encryption Techniques**: Including homomorphic encryption, searchable encryption, and stable multi-birthday celebration computation (MPC).
- **Data Perturbation**: Approaches consisting of noise addition, statistics swapping, and artificial records technology.
- **Privacy-maintaining Machine Learning**: Techniques like federated getting to know, secure aggregation, and version inversion prevention.

- **Access Control Mechanisms**: Role-based totally get right of entry to manipulate (RBAC), characteristic-primarily based encryption (ABE), and blockchain-primarily based access control.
- **Other Emerging Techniques**: Such as privacy-retaining statistics mining algorithms and privacy-maintaining information sharing frameworks [7].

4. Algorithm Analysis: For each category of privacy-preserving techniques, specific algorithms and methodologies are analyzed in detail. This entails:

- Describing the underlying concepts of every technique.
- Discussing the strengths and weaknesses in phrases of privateness protection effectiveness.
- Providing examples of real-international programs or case research in which these techniques had been carried out efficiently.
- Comparing one-of-a-kind algorithms inside each category primarily based on standards along with computational performance, scalability, and applicability to specific forms of statistics.

5. Evaluation Metrics: The review paper employs specific assessment metrics to assess the overall performance and effectiveness of privateness-preserving strategies. These metrics might also include:

- **Privacy Metrics**: Quantitative measures which include ε in differential privacy or re-identity threat measures in anonymization strategies.
- **Usability**: Measures to assess the impact of privacy preservation techniques on data usage, such as loss of accuracy in machine learning models or loss of information due to disturbing data.
- **Technical Considerations**: Considerations of computational complexity and efficiency, which are important for assessing the feasibility of using these techniques in real-world applications [9].

6. Data Processing and Data Processing: The methodology involves the analysis of case studies and useful data processing using privacy protection techniques. This provides insights in terms of real-world implications, challenges faced and lessons learned from applying these techniques to sectors as diverse as healthcare, finance, social media.

7. Classification and analysis: Finally, the research paper integrates the findings from the literature review, classification process, algorithm analysis, research metrics, and case studies. It identifies general trends, emerging challenges and gaps in current research. Based on the insights gained from the study, recommendations for future research directions are made [11].

III. Privacy Threats in Data Science

Although there are numerous different strategies to glean insights from facts in data technological know-how, there are also worries to non-public privateness associated with these procedures [6]. This section appears at a number of privateness dangers that arise in data science, such as inference attacks, membership inference, characteristic disclosure, and identity publicity [4]. Comprehending these risks is important to formulating efficacious measures to protect privacy. Because facts technology uses such a lot of extraordinary methods to research and interpret facts, there are threats to private privateness. The complex interactions of statistics collection, aggregation, and analysis create possibilities for privacy risks including identity revelation, in which human's identities can be deduced from information units that appear like anonymized. Attribute disclosure additionally carries a risk due to the fact that statistics evaluation tactics can also unintentionally divulge personal data about a person. By the use of statistical patterns in data, club inference attacks can ascertain if specific people are protected in a dataset, so infringing their privateness [19]. Because they allow attackers to infer personal statistics about specific persons from seemingly harmless data, inference assaults get worse privacy troubles [24].

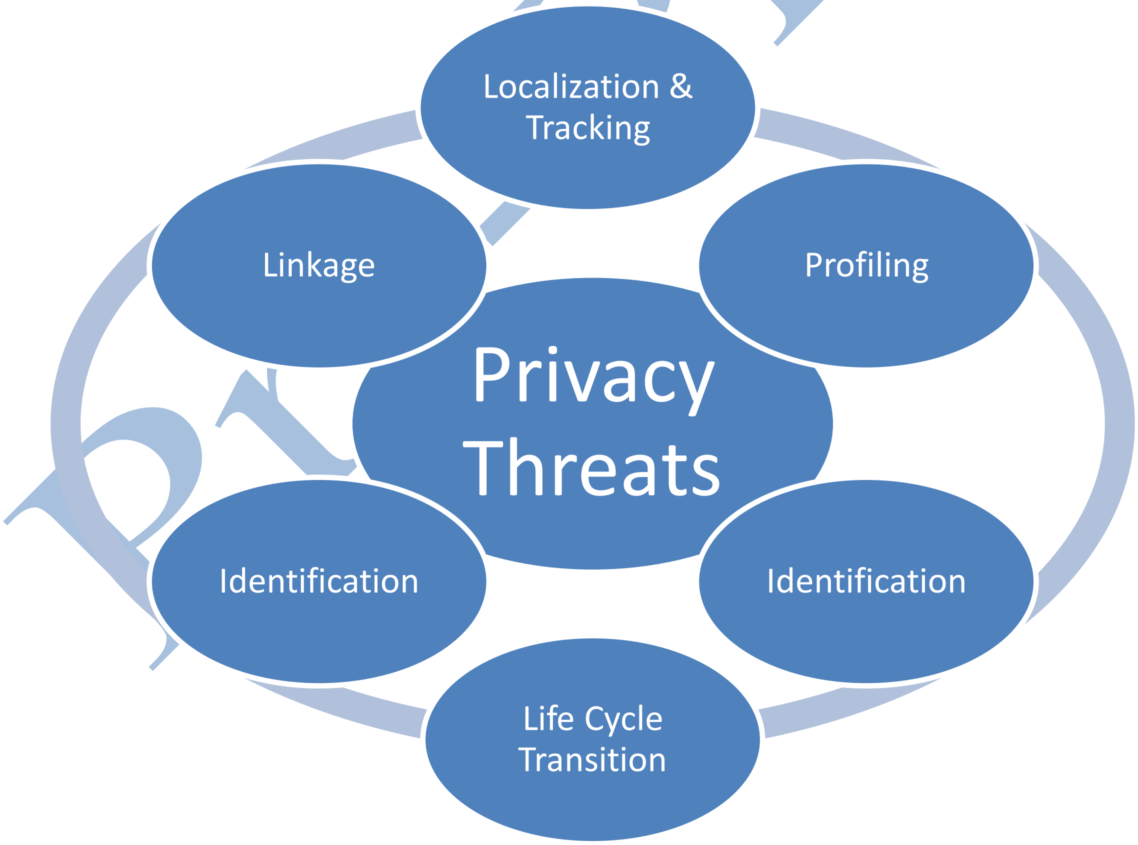

Figure 1 – Privacy Threats in Data Science

IV. Privacy Threats Real-World Scenarios

Medical Data De-anonymization:

- **Scenario**: A scientific research institution collects anonymized patient information to examine disorder styles. However, an outside dataset containing demographic info (like age, gender, and ZIP code) is blended with the anonymized clinical data.
- **Risk**: Demographic information can be linked to medical records to re-identify individuals. These breaches of privacy may lead to the disclosure of serious health conditions or genetic traits, and may lead to discrimination or ill-treatment [14].

Location Data in Mobility Studies:

- **Scenario**: A transportation business enterprise collects GPS statistics from customers' smartphones to analyze site visitor styles and improve service efficiency.
- **Risk**: Even if the facts is anonymized, unique styles of motion can display private data. For example, frequent visits to positive locations (like medical facilities, locations of worship, or political occasions) can disclose touchy factors of a character's existence, undermining their privateness [17].

Social Media and Behavioral Analytics:

- **Scenario**: Social media structures track customers' interactions, likes, and stocks to personalize content and target advertisements.
- **Risk**: Aggregating behavioral information across structures can create distinctive user profiles. Algorithms can predict political affairs, sexual orientation, or economic status based totally on reputedly risk-free interactions. This information can be exploited by means of malicious actors for targeted harassment or manipulation [7].

V. Privacy-Preserving Approaches

Implementing privacy-retaining techniques in facts technological know-how faces several challenges that should be addressed to make sure effectiveness and usefulness. Here, we're going to delve into these demanding situations scalability, usability, and resilience to assaults and recommend capability solutions or future studies instructions for each:

1. Scalability

- ✓ **Challenge**: Many privateness-keeping strategies, including differential privateness or homomorphic encryption, may be computationally intensive and won't scale nicely with big datasets or actual-time statistics streams. These scalability problems may hinder their useful application in situations that require smaller statistical analyzes and responses [32].

Future Directions/ Potential Solutions:

- Optimized Algorithms: Many green algorithms and protocols designed for scalability ensure compromised external privacy. Includes courses on flexible cryptographic operations, parallel computing strategies, and optimized statistical systems.
- **Hardware Acceleration**: Investigate the use of specialized hardware or committed secure hardware modules to boost up privateness-preserving computations, making them more possible for real-time applications.
- **Distributed Computing**: Explore dispensed methods where computations are dispensed throughout more than one node or processors, leveraging parallelism to deal with massive-scale statistics efficiently whilst preserving privateness [33].

2. Usability

- **Challenge**: Privacy-maintaining techniques regularly require specialized knowledge in cryptography and facts privateness, which can be a barrier for adoption with the aid of non-experts which includes facts scientists, analysts, or utility builders. Complex protocols and integration with present systems can also hinder usability [22].

Future Directions/ Potential Solutions:

- **User-Friendly Tools and Libraries**: Easy-to-use software libraries, APIs, or structures were designed that difficult to understand the mechanisms for keeping complicated privacy. Tools along with IBM's "Fully Homomorphic Encryption Toolkit" and Microsoft's "Simple Encrypted Arithmetic Library (SEAL)" are steps on this course.
- **Integration with Standard Tools**: Integrate privateness-protective features into broadly used facts analytics and device studying frameworks (e.g., TensorFlow, PyTorch) to simplify adoption and reduce the getting to know curve for operators [12].
- **Education and Training**: Training packages and educational materials may be supplied to increase attention and expertise of privacy protection techniques among facts specialists, permitting them to successfully enforce these strategies in exercise.

3. Resilience to Attacks

- **Challenge**: Privacy protection mechanisms need to be sturdy to a number of attacks, along with statistical simulation attacks, aspect-channel assaults, and hostile assaults aimed toward exploiting or records leakage vulnerabilities via context around.

Future Directions/ Potential Solutions:

- **Formal Security Analysis**: Conduct rigorous security audits and formal proofs to make sure the privateness protection machine is resistant to regarded attacks. This includes cryptographic primitives with established safety residences and defenses which are continuously updated towards rising threats.
- **Adversarial Testing**: Use adversary checking out protocols to assess the robustness of privateness protection systems in opposition to state-of-the-art assaults, including distinct privacy strategies that may withstand hypothetical assaults
- **Privacy-Preserving AI**: Explore the alignment of privateness protection techniques with artificial intelligence (AI) and machine mastering (ML) for transformative protection which could come across and mitigate privacy breaches in actual time [17].

VI. Techniques of anonymization

In facts technology, anonymization is a crucial tactic for maintaining privacy. A tremendous analysis of anonymization techniques, including generalization, suppression, k-anonymity, l-variety, and t-closeness, is given on this segment. It clarifies the underlying ideas of each method as well as its blessings, drawbacks, and practical makes use of in a range of fields. In data technology, anonymization is a vital tactic for maintaining privacy [20]. A full-size analysis of anonymization strategies, consisting of generalization, suppression, k-anonymity, l-variety, and t-closeness, is given in this phase. It clarifies the underlying ideas of each approach as well as its blessings, drawbacks, and sensible makes use of in a number of fields [3]. Anonymization Strategies make sure that no person can be diagnosed from the facts, shielding humans privateness. In order to save you identification, generalization includes substituting greater popular values for precise ones. Datasets containing identifiable information are selectively removed using suppression [23]. K-anonymity ensures that, in terms of quasi-identifier traits, any report may be distinguished from at least k-1 other records. By guaranteeing that touchy features have at the least l exclusive values within each equivalency elegance, L-range improves k-anonymity [18].

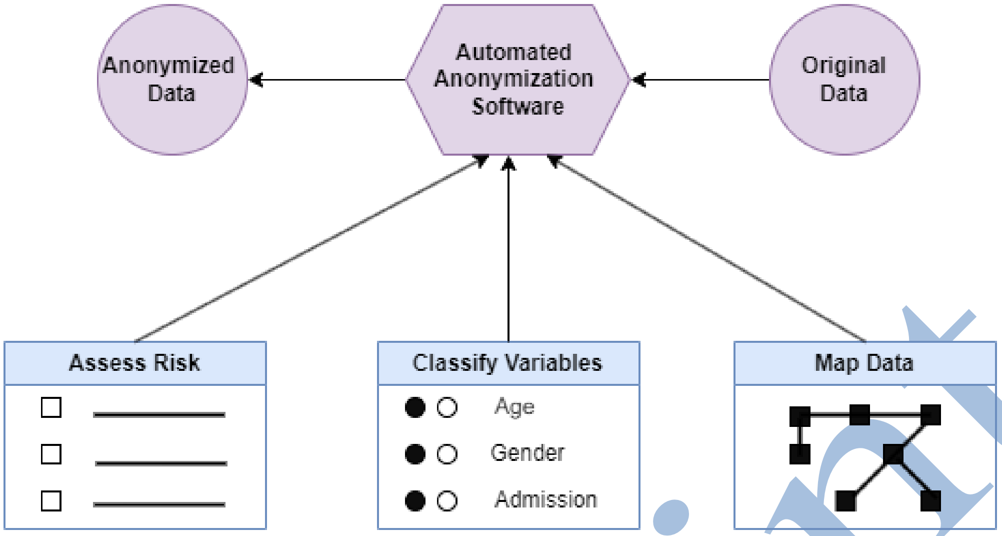

Figure 2 – Techniques of anonymization

VII. Differential privacy

Privacy exceptions have become the de facto measure for privacy protection in data analytics. This section explores methods including Laplace, Gaussian, and exponential mechanisms while digging into the theoretical underpinnings and real-world applications of differential privacy It looks at current developments in data science and explores how to use differential privacy has been used in a variety of contexts. When it comes to protecting privacy in data analytics, differential privacy is the gold standard [7]. This section explores methods including Laplace, Gaussian, and exponential methods that dig into the theoretical principle of differential privacy and mechanisms in order for the presence or absence of one's data to have no consequences the outcome of a calculation is minimally affected, Differential Privacy protects individual privacy Provides differential privacy, the Laplace method Adds random noise to the query results obtained from the Laplace distribution. An exponential process like this maintains privacy by probabilistically selecting outputs based on their utility. The Gaussian filter provides a strong privacy guarantee, which suppresses the query results with Gaussian noise [2][17].

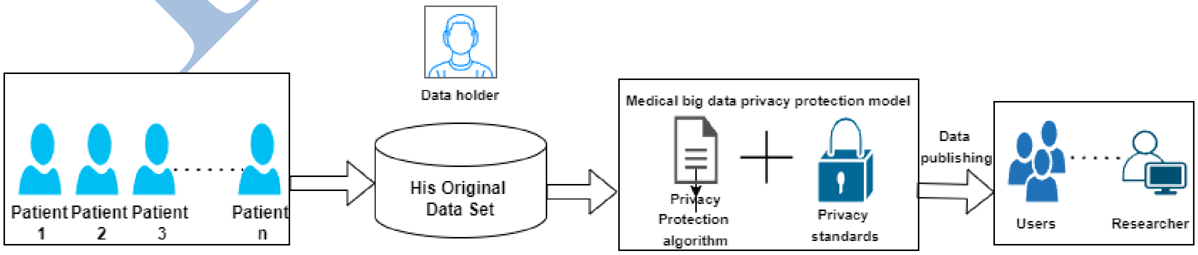

Figure 3 - Medical big data privacy protection model.

VIII. Homomorphic Encryption

Hidden information audits are possible thanks to uniform, privacy-preserving encryption. This section provides an in-depth review of isomorphic encryption, dividing it into three categories: fully isomorphic, partially isomorphic, and partially isomorphic Focuses on secure data processing analysis environments, exploring the challenges, development and implementation of isomorphic encryption use data science. This section examines isomorphic encryption in detail, dividing it into three categories: fully homogeneous encryption (FHE), somewhat homogeneous encryption (SHE), and partially homogeneous encryption (PHE) supports combination or multiple PHEs over encrypted data, while SHE supports combination all only multiplication Allows restricted numbers. The strongest variant, FHE, allows encrypted data to be accounted for arbitrarily while maintaining anonymity in the audit process. By preventing the data from ever having to be decrypted, homologous encryption reduces the chances of unwanted access. While FHE still requires more computers, new developments have increased its use in practical situations [1][25].

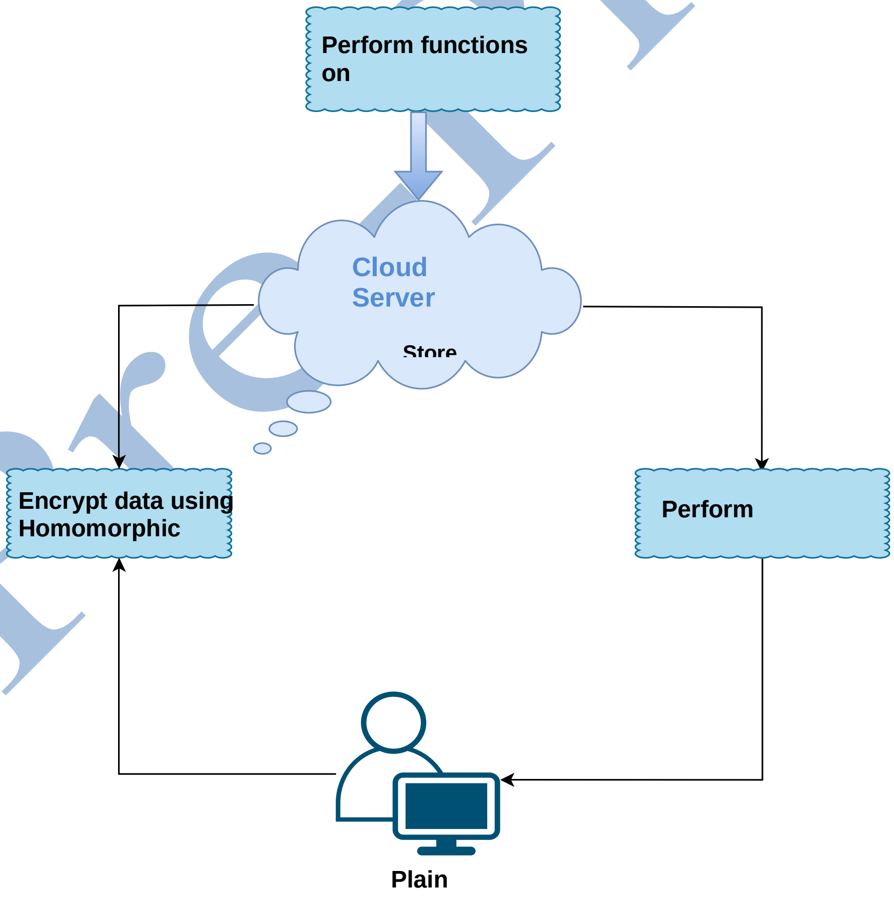

Figure 4 - Homomorphic Encryption

IX. Secure Multi-party computation (SMPC)

Collaborative sensitive computing is facilitated by the use of secure, multi-stakeholder computing, which protects individual contributors from disclosure [8]. This section examines the basic concepts of SMPC, as well as protocols such as Shamir's secret sharing, Yao's entangled circuits, and secure work assessment They are provided by secure mass-processing computing (SMPC), which protects individual privacy it is possible This section examines the basic ideas of SMPC, and protocols such as Shamir's secret sharing, Yao's entangled circuits, and secure work evaluation using these protocols to enable more participants to value the work accounted for simultaneously on their inputs in private. SMPC applications can be found in sectors such as banking, telecommunications and healthcare. It can perform risk assessment, fraud detection and collaborative analysis while protecting privacy [15][26]. Despite its capabilities, SMPC has limitations due to computational complexity and communication costs. Ongoing research seeks to overcome these obstacles to enhance the scalability and efficiency of SMPC systems, thus expanding their relevance in practical settings [28].

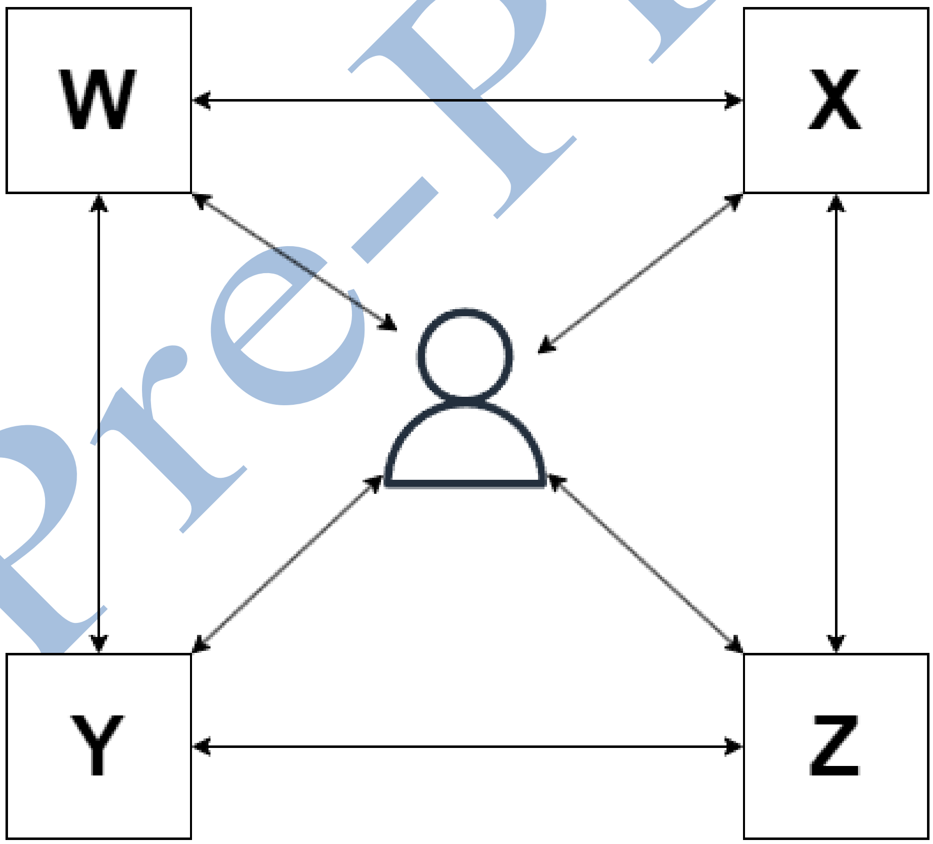

Figure 5 - Multiparty sharing data among each other with any third party using a specified protocol.

X. Federated learning

Pattern education of decentralized statistics resources is made possible through federated learning, which modifies traditional machine getting to know fashions. This phase describes the basics of federated mastering, as well as techniques for model aggregation, federated studying differential privateness, and stable aggregation protocols. It looks at how federated learning may be applied to industries together with aspect computing, IoT, and healthcare. The transformation technique There is one which lets in you to educate fashions on decentralized facts resources [10]. This segment describes the fundamentals of merged gaining knowledge of, in addition to methods for merging fashions, merged getting to know gap privacy, and stable joining procedures. Integrated gaining knowledge of lets in joint model training whilst protecting records privacy by way of storing facts throughout gadgets and only sharing version updates [6]. However, obstacles consisting of transmission prices and statistics resources pose severe limitations to vast adoption. Federal Learning will remodel collaborative system studying to address these issues by means of providing progressive solutions in network design, pattern series techniques, and privateness rules [9][27].

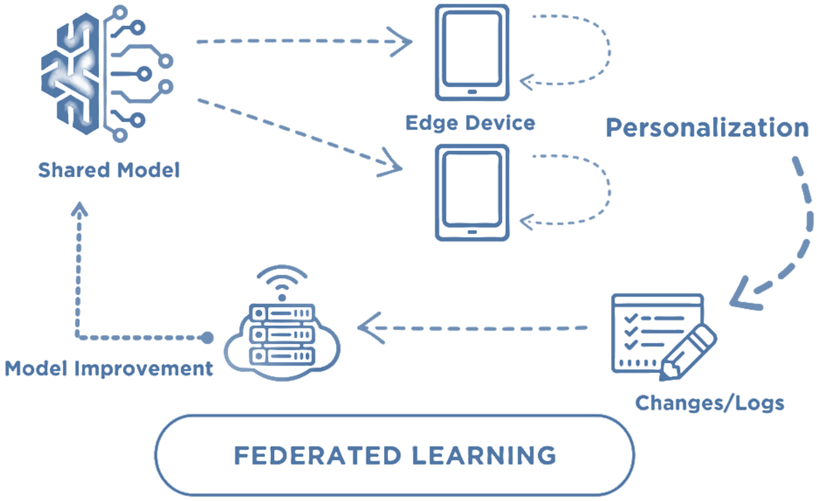

Figure 6 – Federated learning [27]

XI. Comparison Between Different Privacy-Preserving Techniques

Among privacy protection techniques, many methods have emerged to protect touch data by allowing valuable information analysis and device identification Here, we compare and evaluate 3 main methods: integrated learning, uniform encryption, and secure multi-party computation (SMPC), their Let us highlight the strengths and weaknesses [27].

Federated Learning

- ✓ **Definition**: Integrated learning allows instructional machine learning models to be applied across decentralized devices (such as smartphones), which will soon replace statistics. Instead, there is a collection of version updates that hold accounts of the gadgets [27].

Strengths:

- **Privacy Preservation**: The raw records remain in one's gadget, reducing the risk of data propagation throughout model training.
- **Scalability**: Suitable for large data sets distributed across multiple devices by using parallel accounting.
- **Data Diversity**: It encourages multiple sources of data, improving the robustness and generality of the model.

Limitations:

- **Communication Overhead**: There should be frequent conversations between servers and critical devices, which can be bandwidth-extensive [29].
- **Heterogeneity**: Changes in instrument capacity and record distribution can complicate sample collection and assembly.
- **Security Concerns**: It is easy for attacks to occur throughout the model aggregation phase, leaking about affecting statistics from potentially efficient models.

Homomorphic Encryption

- ✓ **Definition**: Homomorphic encryption pre-loads computers into the encrypted account to complete it all at once, allowing the record to be analyzed and privacy managed [22].

Strengths:

- **Data Confidentiality**: The data remains encrypted for the duration of the calculation, making it confidential.
- **Versatility**: It supports many mathematical functions, allowing it to be used in a wide range of mathematics.
- **Security Guarantees**: A strong mathematical foundation provides theoretical assurance about leaks of records.

Limitations:

- **Computational Overhead**: Encryption and decryption schemes can be computationally steeply-priced, affecting the performance of big information sets and complicated programs.
- **Limited Operations**: Some capabilities (e.g., programming, complex queries) are hard to carry out efficiently under current uniform structures.
- **Key Management**: Strong middle practices are needed to save you master compromise and hold safety [29].

Secure Multi-Party Computation (SMPC)

- ✓ **Definition**: SMPC allows more than one events to collectively estimate work on their non-public inputs, at the same time as maintaining that records personal [23].

Strengths:

- **Privacy Preservation**: Each party's input stays personal at some stage in the calculation.
- **Flexibility**: It supports arbitrary operations and calculations and isn't confined to unique kinds of operations.
- **Fault Tolerance**: Flexible in contracting a subset of teams, relying on the selected protocol.

Limitations:

- **Complexity**: Designing and imposing SMPC protocols may be complex and requires know-how in cryptography.
- **Communication Overhead**: It's approximately the essential interactions among elements, which affect scalability and performance [31].
- **Trust Assumptions**: It commonly assumes a semi-realistic adversary model, wherein believe amongst all parties concerned is required to comply with the plan efficaciously [33].

Comparison Summary

- **Privacy Preservation**: Federated Learning and SMPC provide sturdy privacy ensure via keeping uncooked facts or personal enter confidential. Homomorphic Encryption additionally guarantees confidentiality of the statistics all through computation [34].
- **Performance**: Federal research is bendy but require steady verbal exchange. Isomorphic encryption is versatile but computationally intensive. SMPC is flexible but consists of significant communication expenses.
- **Applicability**: Integrated gaining knowledge of is appropriate for disbursed information structures. Isomorphic encryption works in a number of calculations but is restricted by using excessive computational value. SMPC supports desired performance but calls for cautious protocol layout.

XII. Applications of privacy-preserving techniques

A high-level precis of how privacy management policies are actually implemented in many topics is provided in this part. It provides case studies and uses examples from healthcare, banking, telecommunications, social media, and government agencies to highlight the benefits of privacy management strategies in protecting and creating records weaknesses and procedures for privacy protection to ensure compliance with privacy regulatory guidelines A wide variety is used in many areas, and ensures that accounts are used ethically and securely. A greater precis of privacy strategies used effectively in healthcare, banking, telecommunications, social media, and authority industries is provided in this section Privacy strategies ensure that comply with audit, security, and prison recommendations at the same time as the treatment of individual remedy indicators -And if desired, enables static data to be analyzed [12]. These methods help protect financial transactions, detect fraud, and compare risks at the same time if security protection privacy within the financial institution Similarly, non-public telecommunications holding methods hit personal data are used to model consumer behavior and analyze network visitors to protect Furthermore, privacy strategies also preserve behavioral privacy rights through informational choices that facilitate authorities and social media domains [14].

XIII. Challenges and future Directions

In this research paper, we evaluated and compared several key approaches to privacy protection government learning, uniform encryption, and secure multi-party computing (SMPC) focusing on their strengths, restrictions, and roles in protecting sensitive data. While privacy protection practices have come a long way, there are still many hurdles to overcome and room for improvement. Privacy protection techniques, although highly advanced, are not widely used and face many obstacles to being as effective as they could be This section highlights important issues including feasibility, with emphasis on usability, resistance to hostile attacks, and integration of new technologies [27].

Key Findings:

1. **Privacy Preservation**: Either choice gives a robust technique to privacy protection. Federal inspections assure the confidentiality of the accounts by way of storing unripe data in nearby gadgets and collecting the maximum sensitive model updates. Isomorphic encryption allows computation over encrypted information, maintaining confidentiality through the years. SMPC lets in more than one events to together account even while retaining the enter non-public [16].
2. **Performance Considerations**: State studies afford a compromise between privacy and measurability, although they face the challenges of information alternate. Isomorphic encryption ensures strong security but can be afflicted by computational overhead, specially for complicated operations. However, the SMPC gives flexibility via calling for cautious coordination of process and communication necessities between parties [22].
3. The favored technique is based on precise software necessities, together with facts sharing, computational complexity, and privacy options. Integrated information acquisition is appropriate for conditions with distributed report resources, at the identical time uniform

encryption and SMPC prevail in conditions requiring strict privacy or shared computing, respectively [11].

Recommendations for Future Research:

1. **Enhancing Efficiency**: Efficient algorithms and implementations must be developed for each method to reduce computing charges and increase scalability. This consists of guides on network optimization, computational fee discount in isotropic encryption, and software program simplification in SMPC [21].
2. **Integration and Interoperability**: Explore methods to combine privateness and security strategies into current computing engineering frameworks and structures (e.g. TensorFlow, PyTorch). Developing communication requirements and protocols to facilitate smooth adoption across discrete tasks and structures.
3. **Security and Resilience**: Address vulnerabilities in privacy techniques and functionality attack vectors, including resisting birthday celebration-channel assaults with improved homomorphic encryption and making sure robustness towards malicious adversaries in SMPC of the protocol [14]
4. **Usability and Accessibility**: Focus on growing consumer-first-rate gadget, libraries, and course substances to democratize access to privacy-retaining techniques. This includes simplifying techniques to be used and use, as well as training for scientists and conservation practitioners.
5. **Emerging Technologies**: Explore the mechanisms of privateness upkeep and the interface between rising technologies which consist of blockchain and disparate privateness. Find new ways to increase privacy even as allowing high-quality analytics and tool detection liability [9].

XIV. Conclusion

By growing studying approximately the ones tips, we're able to promote powerful and affordable responses that protect privateness. These efforts are wanted not to optimize the safety of man or woman privateness in truth-pushing environments however to comprise validation into information practices in public sectors and to alter statistics inspire responsible behavior. As privacy issues coincide with technological advances, relentless innovation and collaboration may be vital to form the privacy-respecting fate of accounting technology and its predecessors. In the big information and facts science age, privateness-maintaining strategies are critical to permit the responsible, moral, and stable use of facts. A huge assessment of the maximum latest techniques, applications, difficulties, and potentialities for privateness safety has been given in this survey examine. With regard to privacy protection, this survey takes a look at has given a thorough evaluation of the maximum current strategies, programs, problems, and future directions. Research and innovation in privacy-preserving approaches are important to provide moral and lengthylasting information-driven choice-making, mainly as information keeps to multiply and privateness concerns develop. Privacy-retaining strategies enable people, corporations, and groups to fully utilize the transformational strength of data whilst shielding the right to privateness of people with the aid of selling trust, safety, and openness in statistics practices [21][27].

References

[1] Kou, G., Peng, Y., Shi, Y., & Chen, Z. (2007). Privacy-preserving data mining of medical data using data separation-based techniques. *Data science journal*, *6*, S429-S434.

[2] Boulemtafes, A., Derhab, A., & Challal, Y. (2020). A review of privacy-preserving techniques for deep learning. *Neurocomputing*, *384*, 21-45.

[3] Carvalho, T., Moniz, N., Faria, P., & Antunes, L. (2022). Survey on privacy-preserving techniques for data publishing. *arXiv preprint arXiv:2201.08120*.

[4] Singh, A. P., & Parihar, M. D. (2013). A review of privacy preserving data publishing technique. *International Journal of Emerging Research in Management &Technology*, *2*(6), 32-38.

[5] Taric, G. J., & Poovammal, E. (2017). A survey on privacy preserving data mining techniques. *Indian Journal of Science and Technology*.

[6] Nayak, Gayatri, and Swagatika Devi. "A survey on privacy preserving data mining: approaches and techniques." *International Journal of Engineering Science and Technology* 3.3 (2011): 2127-2133.

[7] Rashid, Asmaa Hatem, and Norizan Binti Mohd Yasin. "Privacy preserving data publishing." *International Journal of Physical Sciences* 10.7 (2015): 239-247.

[8] Iezzi, Michela. "Practical privacy-preserving data science with homomorphic encryption: an overview." *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 2020.

[9] Rajesh, N., K. Sujatha, and A. Arul Lawrence. "Survey on privacy preserving data mining techniques using recent algorithms." *International Journal of Computer Applications* 133.7 (2016): 30-33.

[10] Vaghashia, Hina, and Amit Ganatra. "A survey: privacy preservation techniques in data mining." *International Journal of Computer Applications* 119.4 (2015).

[11] Bertino, Elisa, Dan Lin, and Wei Jiang. "A survey of quantification of privacy preserving data mining algorithms." *Privacy-preserving data mining: Models and Algorithms* (2008): 183-205.

[12] Vaghashia, Hina, and Amit Ganatra. "A survey: privacy preservation techniques in data mining." *International Journal of Computer Applications* 119.4 (2015).

[13] Kiran, P., and N. P. Kavya. "A survey on methods, attacks and metric for privacy preserving data publishing." *International Journal of Computer Applications* 53.18 (2012).

[14] https://www.researchgate.net/publication/220670223_Survey_on_Privacy_Preserving_Data_Mining

[15] Aggarwal, Charu C., and Philip S. Yu. "A condensation approach to privacy preserving data mining." *International Conference on Extending Database Technology*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004.

[16] Churi, Prathamesh P., and Ambika V. Pawar. "A systematic review on privacy preserving data publishing techniques." *Journal of Engineering Science & Technology Review* 12.6 (2019).

[17] Aggarwal, Charu C., and Philip S. Yu. *A general survey of privacy-preserving data mining models and algorithms*. Springer US, 2008.

[18] Kiran, Ajmeera, and D. Vasumathi. "A comprehensive survey on privacy preservation algorithms in data mining." *2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*. IEEE, 2017.

[19] Sreedhar, K. C., M. N. Faruk, and B. Venkateswarlu. "A genetic TDS and BUG with pseudo-identifier for privacy preservation over incremental data sets." *Journal of intelligent & fuzzy systems* 32.4 (2017): 2863-2873.

[20] Murugaboopathi, G., and V. Gowthami. "Slicing based efficient privacy preservation technique with multiple sensitive attributes for safe data distribution." *Journal of Intelligent & Fuzzy Systems* 40.2 (2021): 2661-2668.

[21] Agrawal, Dakshi, and Charu C. Aggarwal. "On the design and quantification of privacy preserving data mining algorithms." *Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. 2001.

[22] https://www.researchgate.net/figure/Summary-of-Privacy-Threats_fig3_341345881

[23] https://www.wallarm.com/what/data-anonymization

[24] https://www.nature.com/articles/s41598-022-19544-3/figures/1

[25] https://www.geeksforgeeks.org/what-is-secure-multiparty-computation/

[26] https://www.researchgate.net/figure/Fully-Homomorphic-Encryption-FHE-26_fig1_324692976

[27] https://medium.com/@natasha.gitlin/federated-learning-9659494608d8

[28] Cabrero-Holgueras, José, and Sergio Pastrana. "Sok: Privacy-preserving computation techniques for deep learning." *Proceedings on Privacy Enhancing Technologies* (2021).

[29] Keshk, Marwa, et al. "Privacy-preserving big data analytics for cyber-physical systems." *Wireless Networks* 28.3 (2022): 1241-1249.

[30] Shah, Alpa, and Ravi Gulati. "Privacy preserving data mining: techniques, classification and implications-a survey." *Int. J. Comput. Appl* 137.12 (2016): 40-46.

[31] Ratra, Ritu, and Preeti Gulia. "Privacy preserving data mining: techniques and algorithms." *International Journal of Engineering Trends and Technology* 68.11 (2020): 56-62.

[32] Torkzadehmahani, Reihaneh, et al. "Privacy-preserving artificial intelligence techniques in biomedicine." *Methods of information in medicine* 61.S 01 (2022): e12-e27.

[33] Ram Mohan Rao, P., S. Murali Krishna, and A. P. Siva Kumar. "Privacy preservation techniques in big data analytics: a survey." *Journal of Big Data* 5.1 (2018): 33.

[34] Pramanik, M. Ileas, et al. "Privacy preserving big data analytics: A critical analysis of state-of-the-art." *Wiley*

Interdisciplinary Reviews: Data Mining and Knowledge Discovery 11.1 (2021): e1387.