A Survey on Privacy-Enhancing Techniques in the Era of Artificial Intelligence

Elias Dritsas[1], Maria Trigka[1], and Phivos Mylonas[2]

[1] Department of Computer Engineering and Informatics,
University of Patras, Patras, Greece
{dritsase,trigka}@ceid.upatras.gr
[2] Department of Informatics and Computer Engineering,
University of West Attica, Athens, Greece
mylonasf@uniwa.gr

Abstract. In the era of Big Data and Artificial Intelligence (AI), the unprecedented scale and complexity of data collection, processing, and analysis pose significant privacy challenges. This paper presents the first findings of a survey in progress providing a comprehensive and focused overview of privacy-enhancing technologies (PETs) designed to mitigate these challenges and ensure the protection of sensitive information. We explore cryptographic and non-cryptographic techniques including differential privacy (DP), homomorphic encryption (HE), secure multi-party computation (SMPC), and federated learning (FL). Each of these techniques is examined in terms of its basic principles and advantages. Also, some key challenges for implementing PETs are briefly discussed. Finally, we conclude the survey by denoting our future research directions in the field.

Keywords: Data Privacy · Privacy-Preserving AI · Differential Privacy · Homomorphic Encryption, Secure Multi-Party Computation · Federated Learning · Artificial Intelligence.

1 Introduction

In an era characterized by data explosion and AI's rapid advancement, the proliferation of data-driven technologies has led to increasing concerns about data privacy. Safeguarding individual privacy has emerged as a paramount concern [2]. The integration of AI into various domains — from cloud computing [12], healthcare [10], [7], and social networks [11] to smart cities [15]— promises unprecedented benefits but also poses significant risks to personal privacy.

AI systems often require vast data to train models, which can inadvertently expose personal information. This exposure can lead to unauthorized data access, identity theft, and other privacy breaches. Privacy-preserving AI seeks to address these risks by developing techniques that protect sensitive information while still enabling the powerful capabilities of AI. The challenge lies in striking a balance between leveraging the full potential of AI and ensuring robust privacy protections.

Several critical factors have motivated the adoption of privacy-preserving in AI workflow. Firstly, AI systems often process large amounts of personal and sensitive data. So, privacy-preserving techniques ensure that individual privacy is maintained, safeguarding sensitive information from misuse or unauthorized access. Secondly, users are more likely to trust and engage with AI applications that guarantee the privacy and security of their personal information. Trust is crucial for the widespread adoption of AI technologies. Thirdly, privacy-preserving techniques reduce the risk of data breaches and cyberattacks. By ensuring that data remains private and secure even when processed, organizations can mitigate the potential damage caused by such incidents. Finally, privacy-preserving AI (PPAI) enables safe data sharing and collaboration between organizations without exposing sensitive information. PPAI employs techniques such as DP, FL, HE, and SMPC to achieve these goals. By integrating these methods, AI can be both powerful and respectful of privacy, ensuring a balance between innovation and the protection of individuals.

This paper capitalises on the relevant literature and presents the first results of a survey analysis in progress; especially, Section 2 focuses on the major privacy-preserving techniques and applications, describing a taxonomy of the main existing cryptographic and non-cryptographic techniques. Additionally, in Section 3, the paper examines and lists the key challenges of implementing privacy-preserving research techniques. Lastly, Section 4 concludes the paper and sets future research directions.

2 Techniques for Privacy-Enhancing

AI techniques like machine learning and deep learning can be used to analyze patterns and behaviours in a way that preserves user privacy. This approach is often referred to as privacy-preserving or privacy-enhancing AI.

In *Encrypted Data Analysis*, AI algorithms can be trained on encrypted data using techniques like *HE* or *SMC*. This allows the AI model to perform computations on the encrypted data without ever decrypting it, preserving the privacy of the underlying information [14].

HE is a form of encryption that enables computations to be carried out directly on encrypted data, producing an encrypted result that, when decrypted, matches the result of operations performed on the plaintext data. There are several types of HE schemes, including partially HE (PHE) and fully HE (FHE) [13]. With PHE, it's possible to perform only one type of mathematical operation (either addition or multiplication) on encrypted data. FHE schemes allow for both addition and multiplication operations to be performed on encrypted data. This means that complex computations can be carried out on encrypted data without ever needing to decrypt it [12].

SMPC [15] is a cryptographic technique that allows multiple parties to jointly compute a function over their private inputs while keeping those inputs confidential. The primary goal of MPC is to enable collaboration and computation over sensitive data without revealing that data to any of the involved parties.

Each party privately encrypts their input using cryptographic techniques such as secret sharing. This process ensures that no single party has access to the complete input data. The parties then collaborate to perform the desired computation on the encrypted inputs. This is typically achieved through a series of cryptographic protocols that allow the parties to perform operations on their shares of the inputs while ensuring that they never learn anything about the other parties' inputs. Once the computation is complete, the parties combine their shares of the output to obtain the final result. Importantly, no party learns anything about the other parties' inputs beyond what can be inferred from the output.

Moreover, AI models can be trained using non-cryptographic techniques such as *DP* [17], achieved by adding statistical noise or randomness to the computation process (e.g., training data) ensuring that individual contributions to the data cannot be reliably distinguished. This noise is added in a controlled manner to balance privacy guarantees with the usefulness of the output. This ensures that even if the model learns from sensitive information, it cannot memorize or reveal specific details about individual users.

In *FL*, AI models are trained across multiple decentralized devices or servers, with each device holding its data. The model is trained locally on each device, and only the model updates (not the raw data) are aggregated centrally. This approach allows for collaborative model training without sharing raw data, preserving user privacy [8]. HE is increasingly used in FL to enhance privacy and security. [16] provides an overview of privacy-preserving enhancing techniques specifically in FL, such as SMC and HE, and discusses their impact on model accuracy and training efficiency. HE allows computation on encrypted data, which is crucial in FL as it involves multiple parties collaborating without sharing raw data.

Besides, AI algorithms can be specifically designed to operate on encrypted or anonymised data - *Privacy-Preserving Algorithms* - while still extracting meaningful insights. For example, techniques like SMC enable multiple parties to jointly compute a function over their inputs without revealing them. On the other hand, AI models can be used to generate synthetic data - *Privacy-Preserving Data Generation*- that mimics the statistical properties of real data while ensuring individual privacy. This synthetic data can then be used to train AI models or share with third parties without disclosing sensitive information.

By leveraging these PPAI techniques, organizations can benefit from the insights derived from user data while respecting user privacy and complying with privacy regulations. This approach enables a wide range of applications, including personalized recommendations, predictive analytics, and risk assessment, while minimizing the risk of privacy breaches. Table 1 summarizes some recent papers on privacy-preserving AI applications, highlighting their domain, techniques used and key insights.

Table 1. Summary of recent papers on applications of PPAI.

Domain	Techniques Used	Applications	Key Insights
Healthcare [3]	DP	Biomedical prediction models	Ensures privacy of patient data while maintaining model performance
Finance [4]	FL, MPC	Credit scoring assessment	Enhances credit scoring ML models without sharing sensitive data
Smart Cities [9], [1]	FL	Transportation systems traffic flow traffic monitoring mobility-aware systems	Demonstrates effective traffic control while protecting privacy
e-commerce, Social media [6]	FL	Personalized recommendation systems (PRS)	Enhance privacy in PRS
IoT [5]	DP	Real-time data analysis in IoT networks	Ensures privacy in real-time IoT data analysis applications
IoT Healthcare	HE, FL	medical data	Builds a novel HE-based FL prototype system that preserves user privacy

3 Challenges for Implementing PPAI

Despite the advancements, implementing privacy-preserving AI presents several challenges. These include computational overhead, scalability issues, and the complexity of integrating PETs into existing AI workflows. Ongoing research aims to develop more efficient algorithms, standardize protocols, and enhance the practical deployment of privacy-preserving techniques.

The future of privacy-preserving AI holds promise, with continuous innovations aimed at reconciling the need for data-driven insights with the imperative of protecting individual privacy. As stakeholders from academia, industry, and government collaborate, the development of robust, scalable, and user-friendly privacy-enhancing solutions will be crucial in ensuring that AI advancements benefit society without compromising privacy. Developing more efficient and scalable privacy-preserving solutions in AI involves addressing several key areas:

- Efficiency and Scalability:
  - Computational Overhead: Techniques like homomorphic encryption (especially FHE) and SMPC are often computationally intensive due to the advanced mathematical constructs involved and require more resources than conventional encryption methods making them impractical for large-scale and real-time AI applications. FHE schemes tend to be more computationally intensive and less practical than PHE schemes

  - Communication Costs: Federated learning and SMPC can incur high communication costs, particularly when dealing with large models and datasets. Reducing these costs is crucial for scalability.
- Accuracy and Utility Trade-offs:
  - Privacy vs. Utility: Ensuring strong privacy guarantees while maintaining high model accuracy and utility remains a significant challenge. Balancing these trade-offs is essential for practical adoption.
  - Noise Calibration: Techniques such as DP require careful calibration of noise to balance privacy and utility, which is not always straightforward.
- Interoperability and Standardization:
  - Lack of Standards: There is a lack of standardized protocols and frameworks for implementing privacy-preserving techniques across different platforms and applications.
  - Interoperability Issues: Ensuring that privacy-preserving solutions can seamlessly integrate with existing AI frameworks and tools is a challenge.
- Usability and Adoption
  - Complexity. Implementing privacy-preserving techniques can be complex and requires specialized knowledge, which can hinder widespread adoption.
  - User Awareness. There is a need to increase awareness and understanding of privacy-preserving techniques among developers and end-users.
- Security Assumptions
  - Robustness Against Attacks: Ensuring that privacy-preserving methods are robust against various types of attacks, such as model inversion and membership inference attacks, remains an ongoing challenge.
  - Adversarial Settings: Research is needed to develop methods that can operate securely in adversarial settings where participants may not be fully trusted.

4 Conclusions and Future Directions

Privacy-enhancing AI is a critical area of research that addresses the need to protect individual privacy in the age of big data and AI. Techniques like differential privacy, homomorphic encryption, secure multi-party computation and federated learning offer promising solutions to the privacy challenges posed by modern AI systems. While existing privacy-enhancing technologies provide strong privacy guarantees, they often come with significant computational/communication overhead. Our future research will focus on existing techniques for improving the efficiency and scalability of these technologies to make them more practical for large-scale, real-world applications. Federated learning primitives and algorithms are another area of active research we will emphasize; targeting hybrid techniques as well, namely those that combine federated learning with homomorphic encryption, differential privacy, and secure multiparty computation, to name a few, to provide robust privacy guarantees. Moreover, we plan to investigate existing libraries and tools that can easily incorporate these technologies for privacy-enhancing AI into existing workflows and frameworks.

To sum up, by continuing to innovate and address the challenges in the field, researchers and practitioners can ensure that the benefits of AI are realized without compromising individual privacy. The advancements will pave the way for more secure, trustworthy, and privacy-respecting AI systems.

References

1. Al-Huthaifi, R., Li, T., Huang, W., Gu, J., Li, C.: Federated learning in smart cities: Privacy and security survey. Information Sciences **632**, 833–857 (2023)
2. Curzon, J., Kosa, T.A., Akalu, R., El-Khatib, K.: Privacy and artificial intelligence. IEEE Transactions on Artificial Intelligence **2**(2), 96–108 (2021)
3. Eicher, J., Bild, R., Spengler, H., Kuhn, K.A., Prasser, F.: A comprehensive tool for creating and evaluating privacy-preserving biomedical prediction models. BMC medical informatics and decision making **20**, 1–14 (2020)
4. He, H., Wang, Z., Jain, H., Jiang, C., Yang, S.: A privacy-preserving decentralized credit scoring method based on multi-party information. Decision Support Systems **166**, 113910 (2023)
5. Husnoo, M.A., Anwar, A., Chakrabortty, R.K., Doss, R., Ryan, M.J.: Differential privacy for iot-enabled critical infrastructure: A comprehensive survey. IEEE Access **9**, 153276–153304 (2021)
6. Javeed, D., Saeed, M.S., Kumar, P., Jolfaei, A., Islam, S., Islam, A.N.: Federated learning-based personalized recommendation systems: An overview on security and privacy challenges. IEEE Transactions on Consumer Electronics (2023)
7. Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., Qadir, J.: Privacy-preserving artificial intelligence in healthcare: Techniques and applications. Computers in Biology and Medicine p. 106848 (2023)
8. Li, Z., Sharma, V., Mohanty, S.P.: Preserving data privacy via federated learning: Challenges and solutions. IEEE Consumer Electronics Magazine **9**(3), 8–16 (2020)
9. Liu, Y., James, J., Kang, J., Niyato, D., Zhang, S.: Privacy-preserving traffic flow prediction: A federated learning approach. IEEE Internet of Things Journal **7**(8), 7751–7763 (2020)
10. Louassef, B.R., Chikouche, N.: Privacy preservation in healthcare systems. In: 2021 International Conference on Artificial Intelligence for Cyber Security Systems and Privacy (AI-CSP). pp. 1–6. IEEE (2021)
11. Majeed, A., Khan, S., Hwang, S.O.: A comprehensive analysis of privacy-preserving solutions developed for online social networks. Electronics **11**(13), 1931 (2022)
12. Park, J., Kim, D.S., Lim, H.: Privacy-preserving reinforcement learning using homomorphic encryption in cloud computing infrastructures. IEEE Access **8**, 203564–203579 (2020)
13. Pulido-Gaytan, L.B., Tchernykh, A., Cortés-Mendoza, J.M., Babenko, M., Radchenko, G.: A survey on privacy-preserving machine learning with fully homomorphic encryption. In: Latin American High Performance Computing Conference. pp. 115–129. Springer (2020)
14. Qin, H., He, D., Feng, Q., Khan, M.K., Luo, M., Choo, K.K.R.: Cryptographic primitives in privacy-preserving machine learning: A survey. IEEE Transactions on Knowledge and Data Engineering (2023)
15. Qu, Y., Nosouhi, M.R., Cui, L., Yu, S.: Privacy preservation in smart cities. In: Smart cities cybersecurity and privacy, pp. 75–88. Elsevier (2019)

16. Rafi, T.H., Noor, F.A., Hussain, T., Chae, D.K.: Fairness and privacy preserving in federated learning: A survey. Information Fusion **105**, 102198 (2024)
17. Zhu, T., Ye, D., Wang, W., Zhou, W., Philip, S.Y.: More than privacy: Applying differential privacy in key areas of artificial intelligence. IEEE Transactions on Knowledge and Data Engineering **34**(6), 2824–2843 (2020)