International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 3, May-Jun-2026, ISSN (Online): 2395-566X

Privacy-Preserving Technologies for Secure Data Management in Digital Systems: A Review of Encryption, Anonymization, Differential Privacy, Federated Learning and Secure Computation

Gawade Aditi Dattatray[1] Prof. C. A. Kolse[2] Prof. M. C. Kolse[3] Prof. S. K. Said[4]
Department Of Artificial Intelligence And Data Science Jaihind College Of Engineering Pune, Maharashtra, India

Abstract—Digital platforms, cloud services, online transactions and data-driven applications collect and process large volumes of personal and confidential information every day, increasing the risk of data breaches, unauthorized access, identity theft and misuse. Protecting privacy while still allowing data to be used for legitimate purposes has therefore become a core requirement of modern systems. This paper reviews the principal privacy-preserving technologies and how they can be combined for secure data management. It covers data encryption at rest and in transit, anonymization models such as k-anonymity and l-diversity together with the risk of re-identification, differential privacy, which adds calibrated noise to protect individuals, federated learning, which trains models without centralizing raw data, homomorphic encryption, which allows computation on encrypted data, secure multi-party computation and trusted execution environments. Attacks such as membership inference that motivate these defences are discussed. A layered architecture for privacy-aware data management is proposed, with a research gap analysis, objectives and a comparison of techniques by protection level, performance cost and maturity. Regulatory drivers, including India's Digital Personal Data Protection Act 2023 and the GDPR, are considered. The review concludes that no single technique is sufficient and that layered combinations matched to the threat model provide the best balance of privacy, utility and performance.

Index Terms—Privacy-preserving technology, data security, differential privacy, federated learning, homomorphic encryption, secure multi-party computation, anonymization, data protection

I. INTRODUCTION

1.1 Motivation

Every online purchase, health consultation, banking transaction and social-media interaction leaves a data trail. Organizations use this data to improve services, but large stores of personal information are attractive targets and can be misused even without a breach. Laws such as the European GDPR [1] and India's Digital Personal Data Protection Act 2023 [2] now require organizations to process personal data lawfully, minimize its collection and protect it with appropriate safeguards.Traditional security focuses on keeping outsiders away from data. Privacy-preserving technologies (PPTs) go further: they aim to let data be used for analysis or machine learning while limiting what can be learned about any individual, even by the organization performing the analysis.This paper reviews the main PPTs, explains their guarantees and costs, proposes an architecture that combines them, and identifies research gaps.

1.2 Objectives Of The Review

The objectives of this review are:

- **O1:** To study the principles, guarantees and limitations of major privacy-preserving technologies.
- **O2:** To review attacks that motivate the use of these technologies.
- **O3:** To propose a layered architecture combining PPTs for secure data management.
- **O4:** To compare techniques by protection level, computational cost and maturity.
- **O5:** To discuss implementation challenges and regulatory alignment.

II. OVERVIEW OF KEY CONCEPTS

2.1 Privacy Threats

Privacy can be violated by direct breaches, by linking anonymized data with external sources, or by inferring information from statistics or trained models. Narayanan and Shmatikov [3] re-identified users in an anonymized movie-rating dataset by matching it with public reviews, and Shokri *et al.* [4] showed that an attacker can determine whether a specific record was in a model's training set. These attacks show that removing names is not enough.

2.2 Families Of Techniques

Fig. 1 classifies PPTs into four families: protecting data at rest and in transit, transforming data before release, limiting what outputs reveal, and computing on data without exposing it.

© 2026 IJSRET

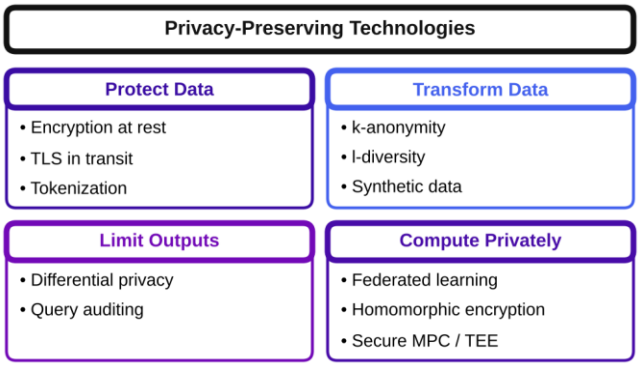


Fig. 1. Families of privacy-preserving technologies.

III. REVIEW OF EXISTING WORK

Anonymization. Sweeney [5] formalized k-anonymity, which requires every record to be indistinguishable from at least k−1 others on quasi-identifiers such as age, gender and postal code. Machanavajjhala *et al.* [6] showed that k-anonymity can still leak sensitive attributes when all records in a group share the same value, and proposed l-diversity. The de-anonymization attack of Narayanan and Shmatikov [3] demonstrated that sparse, high-dimensional data are especially hard to anonymize.

Differential privacy. Dwork and Roth [7] presented differential privacy as a mathematical guarantee that the output of an analysis changes very little whether or not any one individual's data is included, achieved by adding carefully calibrated noise. Abadi *et al.* [8] extended this to deep learning with differentially private stochastic gradient descent, clipping each example's gradient and adding noise during training.

Federated learning. McMahan *et al.* [9] introduced federated averaging, in which devices train locally and share only model updates. Bonawitz *et al.* [10] designed a secure aggregation protocol so that the server learns only the sum of updates, not any individual update. Kairouz *et al.* [11] surveyed open problems including communication cost, heterogeneous data and attacks on updates.

Computing on encrypted data. Paillier [12] proposed an additively homomorphic public-key scheme, and Gentry [13] constructed the first fully homomorphic encryption scheme, allowing arbitrary computation on ciphertexts. CKKS [14] made approximate arithmetic on encrypted real numbers practical for machine learning. Yao [15] introduced secure two-party computation, the foundation of secure multi-party computation, and trusted execution environments such as Intel SGX [16] offer hardware-isolated computation as a faster alternative with different trust assumptions. Table I summarizes the reviewed works.

TABLE I Summary of Literature Survey

Ref.	Author(s) / Year	Technique / Approach	Key Contribution	Limitation / Gap
[5]	Sweeney, 2002	k-anonymity model	Formal anonymity via indistinguishable groups	Vulnerable to homogeneity attacks
[6]	Machanavajjhala et al., 2007	l-diversity	Protects sensitive attribute diversity	Still weak against background knowledge
[3]	Narayanan and Shmatikov, 2008	De-anonymization of sparse data	Re-identified users via auxiliary data	Attack paper; no defence
[7]	Dwork and Roth, 2014	Differential privacy foundations	Provable, composable privacy guarantee	Accuracy loss at strong privacy
[8]	Abadi et al., 2016	DP stochastic gradient descent	Private deep learning with accounting	Lower model accuracy
[9]	McMahan et al., 2017	Federated averaging	Training without centralizing data	Updates can leak information
[10]	Bonawitz et al., 2017	Secure aggregation protocol	Server sees only aggregate updates	Communication overhead
[13]	Gentry, 2009	Fully homomorphic encryption	Arbitrary computation on ciphertexts	Very high computational cost
[4]	Shokri et al., 2017	Membership inference attack	Showed models leak training membership	Motivates DP; no full fix

© 2026 IJSRET

IV. IDENTIFIED RESEARCH GAPS

The following gaps are identified:

- Homomorphic encryption and secure computation remain much slower than plaintext computation for large workloads.
- Choosing the privacy budget in differential privacy is difficult, and its effect on accuracy is poorly understood by practitioners.
- Federated learning is vulnerable to inference and poisoning attacks unless combined with other techniques.
- Integrated frameworks that combine several PPTs within one data-management system are rare.
- Mapping technical guarantees to legal requirements such as the DPDP Act is not well established.

V. SYSTEM ARCHITECTURE

The architecture in Fig. 2 applies privacy protection at every stage of the data lifecycle, from collection to sharing of results.

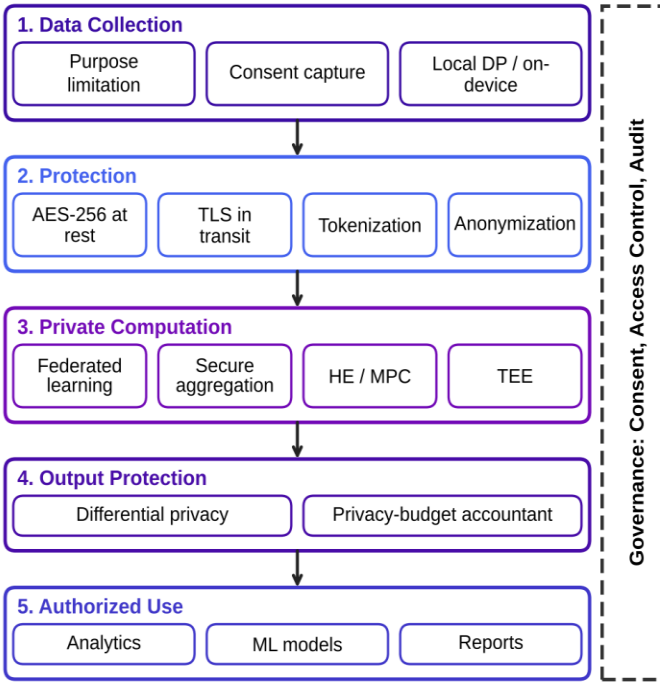


Fig. 2. Proposed layered architecture for privacy-aware data management.

5.1 Data Collection Layer

Only data needed for a stated purpose is collected, with recorded consent. Where possible, local differential privacy or on-device processing is applied before data leaves the user's device.

5.2 Protection Layer

Data is encrypted in transit with TLS and at rest with AES-256 [17]. Direct identifiers are replaced by tokens, and datasets intended for release are generalized to meet k-anonymity and l-diversity requirements.

5.3 Privacy-Preserving Computation Layer

Machine learning across institutions uses federated learning with secure aggregation. Sensitive joint computations use secure multi-party computation or homomorphic encryption, and high-performance tasks may run inside trusted execution environments.

D. Output Protection Layer

Statistics and models released to analysts are protected with differential privacy, and a privacy-budget accountant tracks cumulative loss across queries.

5.4 Governance Layer

Access control, consent management, retention policies and audit logs enforce legal and organizational rules and allow compliance to be demonstrated.

VI. IMPLEMENTATION METHODOLOGY

The working of the system is shown in Fig. 3. A mechanism M satisfies ε-differential privacy if, for any two datasets D and D′ that differ in one record and any set of outputs S,

$$Pr[M(D) \in S] \leq e^{\varepsilon} \cdot Pr[M(D') \in S] \quad (1)$$

A numeric query f is made private with the Laplace mechanism by adding noise scaled to its sensitivity Δf:

$$M(D) = f(D) + Lap(\Delta f / \varepsilon) \quad (2)$$

Smaller values of ε give stronger privacy but noisier results. In federated learning, the global model is updated as a weighted average of client models:

$$w = \Sigma_k (n_k / n) \cdot w_k \quad (3)$$

© 2026 IJSRET

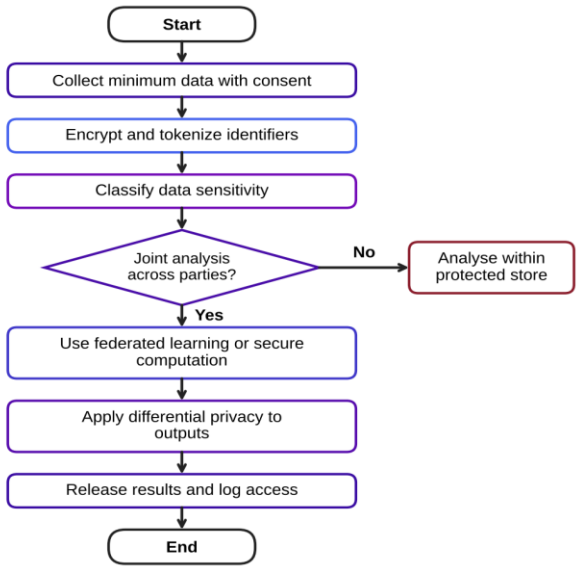


Fig. 3. Privacy-preserving data processing workflow.

VII. PERFORMANCE COMPARISON

Table II compares the techniques. Encryption is mature and fast but protects data only while it is not being used; anonymization is simple but weak against linkage; differential privacy gives provable guarantees at some cost in accuracy; and cryptographic computation offers strong protection at high computational cost.

TABLE II Comparison of Privacy-Preserving Techniques

Technique	Protection	Overhead	Maturity
Encryption (AES/TLS)	High (storage)	Low	Very high
Anonymization	Low–Medium	Low	High
Differential privacy	High (provable)	Low–Medium	Medium–High
Federated learning	Medium	Medium	Medium
Homomorphic encryption	Very high	Very high	Medium
Secure MPC	Very high	High	Medium

VIII. OPEN ISSUES AND FUTURE DIRECTIONS

Table III summarizes key challenges. Future research directions include hardware acceleration of homomorphic encryption, standardized privacy-budget guidelines, synthetic data generation with formal guarantees, and toolkits that let developers apply PPTs without deep cryptographic expertise. Post-quantum cryptography will also be needed to keep encrypted archives secure in the long term.

TABLE III Challenges and Mitigation

Challenge	Mitigation
Computational cost	Hardware acceleration; hybrid HE + TEE
Privacy–utility trade-off	Careful ε selection; larger datasets
Attacks on federated learning	Secure aggregation, DP, robust aggregation
Implementation complexity	Open-source libraries; reusable frameworks
Legal compliance	Map controls to DPDP Act / GDPR duties

IX. CONCLUSION

Privacy-preserving technologies allow organizations to gain value from data while respecting the rights of individuals. The reviewed literature shows that each technique addresses a different risk: encryption protects stored and transmitted data, anonymization and differential privacy limit what released data reveals, and federated learning, homomorphic encryption and secure computation enable analysis without exposing raw data. Because each has limitations, a layered combination guided by the threat model and legal requirements offers the most reliable path to confidential, secure and trustworthy data management.

9.1 Acknowledgment

The authors express their sincere gratitude to the Department of Artificial Intelligence and Data Science, Jaihind College of Engineering, Pune, for providing the facilities, guidance and encouragement needed to carry out this work. The authors also thank the faculty members and colleagues whose valuable suggestions helped improve the quality of this review.

REFERENCES

1. European Parliament and Council, “Regulation (EU) 2016/679 (General Data Protection Regulation),” Off. J. Eur. Union, L 119, 2016.
2. Government of India, The Digital Personal Data Protection Act, 2023 (Act No. 22 of 2023). New Delhi: Ministry of Law and Justice, 2023.
3. A. Narayanan and V. Shmatikov, “Robust de-anonymization of large sparse datasets,” in Proc. IEEE Symp. Security Privacy, 2008, pp. 111–125.

© 2026 IJSRET

4. R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership inference attacks against machine learning models," in Proc. IEEE Symp. Security Privacy, 2017, pp. 3–18.
5. L. Sweeney, "k-anonymity: A model for protecting privacy," Int. J. Uncertain. Fuzziness Knowl.-Based Syst., vol. 10, no. 5, pp. 557–570, 2002.
6. A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "ℓ-diversity: Privacy beyond k-anonymity," ACM Trans. Knowl. Discov. Data, vol. 1, no. 1, Art. no. 3, 2007.
7. C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," Found. Trends Theor. Comput. Sci., vol. 9, no. 3–4, pp. 211–407, 2014.
8. M. Abadi et al., "Deep learning with differential privacy," in Proc. ACM CCS, 2016, pp. 308–318.
9. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. AISTATS, 2017, pp. 1273–1282.
10. K. Bonawitz et al., "Practical secure aggregation for privacy-preserving machine learning," in Proc. ACM CCS, 2017, pp. 1175–1191.
11. P. Kairouz et al., "Advances and open problems in federated learning," Found. Trends Mach. Learn., vol. 14, no. 1–2, pp. 1–210, 2021.
12. P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in Proc. EUROCRYPT, 1999, pp. 223–238.
13. C. Gentry, "Fully homomorphic encryption using ideal lattices," in Proc. 41st ACM Symp. Theory Comput. (STOC), 2009, pp. 169–178.
14. J. H. Cheon, A. Kim, M. Kim, and Y. Song, "Homomorphic encryption for arithmetic of approximate numbers," in Proc. ASIACRYPT, 2017, pp. 409–437.
15. A. C. Yao, "Protocols for secure computations," in Proc. 23rd IEEE Symp. Found. Comput. Sci. (FOCS), 1982, pp. 160–164.
16. V. Costan and S. Devadas, "Intel SGX explained," IACR Cryptology ePrint Archive, Rep. 2016/086, 2016.
17. National Institute of Standards and Technology, "Advanced Encryption Standard (AES)," FIPS PUB 197, 2001 (updated 2023).

© 2026 IJSRET