

AI-Based Approaches for Protecting Personal Data Privacy on Digital Platforms

Priti Jitendra Yadav¹, Aarti Sanjay Gawai²

¹Student, Dept. of Information Technology, D.G Ruparel College of Arts, Science and Commerce, Mumbai, Maharashtra, India

²Assistant Professor, Dept. of Information Technology and Computer Science, D.G Ruparel College of Arts, Science and Commerce, Mumbai, Maharashtra, India

¹pritiyadav20050@gmail.com, ²aartigawai1611@gmail.com

ABSTRACT--Digital platforms collect personal information for communication, shopping, financial transactions, entertainment, and personalized services. The same information can also be exposed through unauthorized access, excessive tracking, profiling, third-party sharing, or accidental disclosure. Conventional controls such as encryption and access management remain necessary, but they do not address every privacy concern when information is collected and processed continuously. This paper reviews how Artificial Intelligence (AI) can support personal-data privacy on digital platforms, with emphasis on machine learning, deep learning, federated learning, and differential privacy. The review considers their roles in identifying sensitive information, detecting unusual access, reducing centralized data exposure, and limiting individual-level disclosure. The reviewed literature also shows that these techniques are often discussed separately, while practical platforms need protection at several stages of the data lifecycle. To address this gap, this paper presents a conceptual framework linking data identification, sensitivity classification, privacy-risk assessment, protection selection, and continuous monitoring. The framework is intended as a practical basis for balancing privacy protection with data utility, scalability, computational cost, and user control. The paper further discusses implementation considerations, limitations, and future research directions for building more adaptive and privacy-aware digital platforms.

Keywords: *Artificial Intelligence; Personal Data Privacy; Digital Platforms; Privacy-Preserving AI; Machine Learning; Federated Learning; Differential Privacy; Sensitive Data Detection.*

1. INTRODUCTION

Digital platforms have become part of routine communication, online shopping, banking, entertainment, education, and many other services. To provide these functions, platforms collect information such as account details, location history, browsing activity, transaction records, preferences, and device

Information, and user interactions. Such information can improve personalization and service quality, but it also creates a persistent privacy concern. Data may be accessed by unauthorized parties, combined with other datasets for profiling, retained longer than necessary, or disclosed without users fully understanding how it will be used.

The scale and speed of digital data processing make privacy management increasingly difficult to perform only through manual review or fixed rules. AI can assist by examining large amounts of information, identifying patterns, classifying sensitive content, and recognizing activity that differs from expected behavior. For example, a machine-learning model can support sensitive-data classification, while an anomaly-detection model can flag unusual access to a user account. Federated learning can reduce the need to move raw training data to a central repository, and differential privacy can reduce the amount of individual information revealed by a model output.

At the same time, AI should not be treated as a privacy solution by itself. AI systems depend on data and may introduce additional privacy risks through model training, information leakage, inappropriate inference, or weak protection of model updates. This creates a two-sided problem: AI can be used to strengthen privacy, but the AI system itself must also be designed and operated in a privacy-aware manner. The central issue is therefore not simply whether AI can protect data, but how different AI-based methods can be combined according to the sensitivity, purpose, and risk associated with the data.

This paper reviews relevant work on AI-assisted privacy protection and develops a conceptual framework for digital platforms. The study focuses on the complete data lifecycle, beginning with identification and classification and continuing through risk assessment, protection, deployment, and monitoring. The objective is to provide

a structured view of how different techniques can complement one another while recognizing the trade-offs between privacy, utility, scalability, and computational cost.

2. LITERATURE REVIEW

Yu et al. [1] examined the development of privacy-protection research in AI and highlighted the growing importance of privacy-preserving approaches as AI systems increasingly depend on large datasets. Their review demonstrates that privacy has become closely connected with the development and use of AI rather than being a separate issue addressed only after data collection.

Morel et al. [2] reviewed AI-driven personalized privacy assistants designed to support users when making privacy-related decisions. Such systems can reduce the difficulty of interpreting complex privacy choices and can help users understand what information is being requested or shared. However, meaningful user control remains important because an automated assistant should support rather than silently replace the user's decision.

Jahns et al. [3] surveyed privacy-preserving deep-learning approaches and discussed methods such as differential privacy and secure computation. These techniques can reduce exposure of sensitive information, but they may require additional processing resources and can influence model performance. The literature therefore suggests that privacy protection should be evaluated together with computational and operational requirements.

Federated learning has received attention because it allows learning to take place across separate devices or data holders without requiring all raw training data to be placed in one central location. Chronis et al. [4] studied federated learning in privacy-preserving recommender systems and showed its relevance to personalization scenarios. Nevertheless, model updates can still contain information related to local training data, so federated learning should not be interpreted as a complete privacy guarantee.

Differential privacy provides another layer of protection by limiting the influence of an individual record on a statistical result or model output. This approach is useful when aggregate information is needed but individual-level disclosure should be restricted. Recent work [5] demonstrates the continuing interest in integrating

differential privacy into recommendation systems while preserving acceptable service utility.

Taken together, the reviewed studies show that the techniques address different stages of the privacy problem. AI-based assistants can improve user awareness, machine learning can support classification and risk analysis, federated learning can reduce centralized exposure of raw data, and differential privacy can restrict individual-level disclosure [1]–[5]. The broader research gap is that a digital platform needs a coordinated approach rather than a single isolated privacy mechanism. This paper addresses that gap through an integrated conceptual framework.

3. RESEARCH GAP AND OBJECTIVES

The literature provides useful techniques for protecting personal information, but several practical gaps remain. First, privacy controls are often considered at the level of a particular technique rather than across the complete data lifecycle. Second, the same protection level is not necessarily suitable for every type of information. Third, privacy protection may reduce service utility, increase processing requirements, or affect model accuracy. Finally, AI systems themselves may become a source of privacy exposure if training data, model outputs, or exchanged updates are not adequately protected. The main objective of this research is to organize existing AI-based privacy techniques into a single conceptual process for digital platforms. The specific objectives are: (i) to identify major categories of personal information and their privacy implications; (ii) to examine how AI can support sensitivity classification and privacy-risk assessment; (iii) to study the complementary roles of federated learning and differential privacy; (iv) to propose a continuous monitoring and response stage; and (v) to identify practical trade-offs, limitations, and future research opportunities.

4. METHODOLOGY

This study follows a literature-review-based conceptual methodology. Instead, existing academic and technical literature is examined to identify recurring themes related to AI-assisted privacy protection on digital platforms. The selected literature is considered according to the problem addressed, AI technique used, privacy mechanism, expected benefit, limitation, and relevance to the digital-platform data lifecycle.

4.1. Research Approach

A structured review approach is used to organize the literature into four broad areas: data identification and classification, privacy-risk assessment, privacy-preserving processing, and continuous monitoring. This organization makes it possible to compare techniques that are normally discussed independently and to consider how they may operate together.

4.2. Data Identification and Classification

The first stage is to identify the information handled by a platform and determine its sensitivity. AI-based classification can assist in recognizing identity information, location history, behavioral data, financial information, and other sensitive categories. Classification provides a basis for applying stronger controls to higher-risk information rather than treating all data in exactly the same way.

4.3. Privacy-Risk Assessment

After classification, the platform can assess the level of privacy risk associated with a particular data item or processing activity. Factors may include the sensitivity of the information, purpose of processing, access pattern, exposure level, and consequences of unauthorized disclosure. Machine-learning techniques can support this stage by identifying patterns that differ from expected behavior and by prioritizing activities that require additional review.

4.4. AI-Based Privacy Protection

The framework then selects a protection mechanism according to the assessed risk. Federated learning is suitable when the platform needs collaborative model training while reducing the movement of raw information. Differential privacy can be introduced when model outputs or aggregate results could reveal information about individuals. Anomaly detection can be used to identify suspicious access, while automated sensitive-data detection can support classification and data-handling decisions. These techniques are complementary and may be combined when the application's sensitivity requires multiple layers of protection.

4.5. Continuous Monitoring and Response

Privacy protection cannot be considered complete when the initial controls are applied. User behavior, platform functions, data flows, and attack patterns can change over time. AI-based monitoring can therefore examine

ongoing activity and flag deviations from expected patterns. Depending on the risk, the platform may restrict access, request additional verification, apply stronger protection, or notify a responsible user or administrator. This stage converts privacy from a one-time configuration into an ongoing operational process.

5. CONCEPTUAL FRAMEWORK

The proposed framework connects the stages of the data lifecycle in a sequence that can be adapted to different digital platforms. The process begins with personal-data identification and sensitivity classification. The classified information is then evaluated for privacy risk, after which an appropriate AI-based protection mechanism is selected. The final stage continuously observes activity and provides a response when the privacy condition changes.

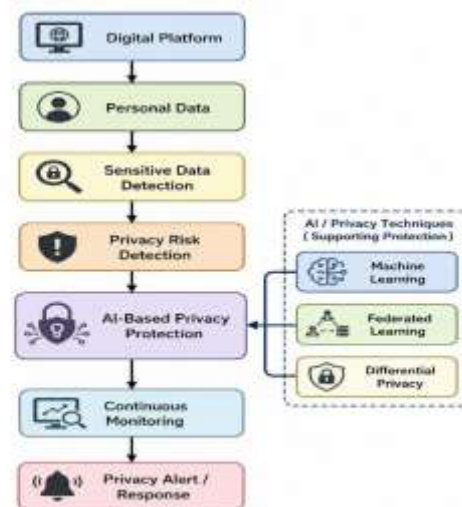


Figure 1. AI-Based Personal Data Privacy Protection Framework

Fig. 1. Source: Authors' own framework developed from the reviewed literature [1]–[15].

The framework is intentionally layered. A low-risk dataset may require classification and conventional access controls, whereas highly sensitive information may require additional privacy-preserving learning and output protection. This layered design also supports adaptation: if monitoring identifies a new risk, the platform can reassess the data and modify the protection level instead of relying on a fixed rule.

6. RESULTS AND DISCUSSION

The reviewed literature indicates that AI can make privacy management more responsive to changes in platform activity. Automated classification can reduce the manual effort required to identify sensitive information, while anomaly detection can draw attention

to unusual access or processing behavior [1], [2]. The main contribution of the framework is therefore not a new algorithm but an organized process that connects these functions across the data lifecycle.

Federated learning is particularly relevant to services that require personalization while attempting to reduce centralized storage of raw user information. Keeping training data closer to its source can reduce the amount of sensitive material held in one repository [4]. However, the exchanged model updates can still create privacy concerns. For sensitive applications, federated learning may therefore need to be combined with additional mechanisms such as differential privacy.

The privacy-utility trade-off is another important finding. Stronger privacy controls can increase processing requirements, reduce model accuracy, or limit the usefulness of a service [3], [5]. Applying identical protection to every dataset may therefore be inefficient. A risk-based approach is more practical because it allows the platform to apply stronger protection where the expected privacy impact is higher while avoiding unnecessary restrictions on lower-risk processing.

AI Approach	Privacy Role	Limitation
Machine Learning	Risk/access detection	Data-dependent
Deep Learning	Complex anomaly analysis	Less interpretable
Federated Learning	Distributed training	Update leakage risk
Differential Privacy	Output protection	Utility trade-off
Sensitive Data Detection	Data classification	Possible misclassification

TABLE I: Comparison of AI-based approaches for personal-data privacy protection.

This table shows why using a layered system is better than picking just one technique for every situation. Machine learning helps with sorting and risk analysis, deep learning handles complex patterns, federated learning reduces the need for central storage of raw data, and differential privacy limits how much individual data is revealed. Together, these methods should be used based on how sensitive the data is and what the platform needs to run smoothly.

7. IMPLEMENTATION CONSIDERATIONS

To use an AI-based privacy framework, you need to understand how data flows through a digital platform. The first step is to create a data inventory that records the main sources of data, where it is stored, how it is processed, and who might access it. This list helps determine which data needs stronger privacy measures. Data classification can then be used to separate information according to its level of sensitivity. Automated classification techniques can assist in identifying personal or confidential information and assigning appropriate protection levels. Privacy rules can subsequently be connected with these categories so that access and processing decisions follow consistent and documented procedures. This also makes it easier to review privacy-related decisions when required. Another important consideration is the location at which AI models perform their processing. Not every task needs to be transferred to a centralized server. When possible, sensitive information can be processed close to its original source. Local or edge-based processing may reduce unnecessary data movement and therefore limit exposure. For applications that require collaborative model training, federated learning can provide an alternative by allowing participating systems to train a shared model without directly exchanging their original datasets. However, model updates must still be protected against unauthorized access and possible information leakage.

Differential privacy can also be considered when analytical results, statistical information, or model outputs may reveal details about individuals. Rather than relying on a single privacy mechanism, differential privacy should be incorporated as one component of a broader protection strategy. The effectiveness of these measures also depends on continuous monitoring of platform activity. Monitoring can involve reviewing access records, system logs, data transfers, authentication events, and model-related operations. Unexpected changes in these activities may indicate misuse, unauthorized access, or other privacy-related risks. Appropriate responses can include restricting access, requesting additional authentication, modifying protection settings, or sending alerts to responsible personnel. Human oversight is particularly important when an automated privacy decision may significantly affect a user. Automated systems can assist with routine decisions, but important or uncertain cases should remain subject to review by an authorized person. This provides

an additional layer of accountability and reduces the possibility of relying completely on an incorrect automated decision.

The reliability of the AI model should also be evaluated before deployment. Important measures include privacy protection capability, false-positive and false-negative rates, processing time, computational requirements, scalability, and overall service quality. A model with very high prediction accuracy is not necessarily the best choice if achieving that accuracy requires excessive computation or creates additional privacy risks. In practical applications, the selected model should therefore provide a reasonable balance between accuracy, privacy, efficiency, and usability.

8. CHALLENGES AND LIMITATIONS

8.1. Data Quality and Availability

The effectiveness of AI-driven privacy protection systems depends on the quality and availability of data. Real-world datasets may contain missing values, outdated records, unequal representation, and classification errors, which can affect prediction accuracy and privacy risk assessment. Furthermore, access to suitable datasets containing personal or confidential information is often restricted due to privacy concerns, making it difficult for researchers to obtain representative data and evaluate proposed frameworks under realistic conditions.

8.2. Privacy–Utility Trade-off

Balancing privacy protection and system effectiveness is a major challenge, as increasing privacy restrictions may reduce the accuracy or usefulness of analytical results. Techniques such as differential privacy can introduce noise into data and require additional computational resources, potentially increasing processing time and reducing efficiency. Therefore, a single privacy configuration may not be suitable for every digital platform, and the appropriate level of protection depends on data sensitivity, potential disclosure risks, processing requirements, and the platform's intended functionality.

8.3. Model and Training Risks

Protecting stored databases alone does not guarantee the privacy of an AI-based system, as sensitive information may also be exposed through training data, model parameters, updates, or generated outputs. Even when raw records are not directly shared, models may reveal

information through learned patterns and predictions. Therefore, privacy risks must be considered throughout the entire AI lifecycle, including data preparation, model training, deployment, output generation, and updates. Additional safeguards may be required to reduce the possibility of sensitive information being inferred from trained models.

8.4. Users and Administrators

User trust and administrator acceptance are important challenges for AI-driven privacy systems, particularly when automated decisions are difficult to understand or explain. A lack of transparency regarding privacy-related decisions may reduce confidence in the system, especially when such decisions affect access to information or essential services. A practical framework should therefore provide clear explanations for significant automated actions and allow authorized users to review important decisions, seek assistance, and override automated outcomes when necessary. This approach can improve accountability while maintaining the efficiency of automation.

8.5. Changing Platform Behavior

Digital platforms continuously evolve through new data sources, services, software updates, changing user behavior, and emerging security threats. Consequently, privacy controls that are effective under particular conditions may become insufficient as platform requirements and associated risks change. Regular assessments are therefore necessary to review privacy policies, classification rules, model performance, accessibility measures, and monitoring procedures. Continuous monitoring and periodic updates can help ensure that the proposed framework remains relevant and effective as digital environments evolve.

8.6. Scope of the Present Study

This study presents a conceptual approach rather than a framework validated through extensive experimental implementation. Therefore, the discussion should not be interpreted as evidence of measured performance under real-world workloads. The study brings together multiple AI-based and privacy-preserving techniques within a common conceptual structure, but experimental evaluation is required to establish their actual effectiveness, computational requirements, scalability, and privacy benefits under practical conditions.

9. FUTURE SCOPE

Future research can evaluate the proposed framework using representative workloads from different digital platforms and application domains to investigate its practical feasibility and measurable benefits. Such evaluations should consider privacy protection levels, classification accuracy, computational cost, scalability, processing time, and user experience across different workloads and system sizes. Further research could also explore adaptive privacy management, in which protection levels are adjusted according to data sensitivity, processing objectives, user behavior, access patterns, and estimated privacy risks. Stronger privacy controls could be applied in high-risk situations, while less restrictive measures could be used for lower-risk applications to reduce unnecessary computational overhead. Another promising direction is the integration of federated learning and differential privacy, with experiments examining their combined effectiveness and the additional computational and communication requirements involved. Establishing common evaluation standards would also make it easier to compare different privacy-preserving configurations. In addition, AI-driven privacy assistants could help users understand privacy settings, identify sensitive information, and make informed decisions about data sharing. Such assistants should provide clear explanations, enforce appropriate safeguards, and maintain meaningful human oversight of important privacy decisions, ensuring that AI supports rather than limits user control. Large language models may also offer opportunities to interpret privacy requirements, review data practices, assist analysts, and support preliminary privacy assessments. However, their reliability, potential for unintended data exposure, and risks of unauthorized access must be carefully evaluated before their use in sensitive environments. Finally, future studies should examine technical, organizational, and regulatory requirements together, since technical controls alone may not provide adequate protection without effective organizational procedures, transparent policies, appropriate governance, and human oversight. Combining these elements could provide a stronger foundation for protecting personal information on AI-based digital platforms.

10. CONCLUSION

As digital platforms become more prevalent, the need to safeguard personal information during its collection, processing, storage and usage is increasing. While

traditional security controls are still necessary, they don't address every privacy-related issue in large-scale, constantly evolving data environments. Additional capabilities of AI can be utilized to identify sensitive data, recognize unusual behavior or risks, and assist with privacy-protection processing.

The papers explored various roles that machine learning, deep learning, and federated learning play in digital-platform privacy lifecycles, as well as their respective roles in different domains. The literature reviewed presented a conceptual framework that links data identification, sensitivity classification, privacy-risk assessment, protection selection, and continuous monitoring. This was done using the following methods: The framework stresses the importance of treating privacy as an ongoing process, rather than a single control applied at the end of data processing. Efforts on utility and computational cost can be influenced by greater privacy, while AI models may face additional privacy risks. The proposed structure does not assert that a single approach is adequate for all circumstances. It provides a methodical framework for choosing and combining methods that are appropriate for the risk, data type, and desired outcome. This approach can be advanced through future experimental validation, adaptive controls, privacy-preserving learning, and transparent AI-based privacy assistance.

REFERENCES

- [1] S. Yu, F. Carroll, and B. L. In IEEE Access, vol, Bentley provided insights into privacy protection research in AI. 12, 41704–41726, 2024, doi: 10.1109/ACCESS.2024.3378126.
- [2] V. Morel, L. H. Iwaya, and S. In IEEE Access, Fischer-Hübner reviewed the literature on AI-Driven Privacy Assistants. 13, 160982–161002, 2025, doi: 10.1109/ACCESS.2025.3609188.
- [3] E. Jahns, M. Stojkov, and M. A. Kinsy, "Privacy-Preserving Deep Learning: A Survey on Theoretical Foundation, Software Framework and Hardware Accelerators" IEEE Access, vol. 13, 67821–67855, 2025, doi: 10.1109/ACCESS.2025.3561721.
- [4] C. Chronis, I. Varlamis, Y. Himeur, A. N. Sayed, T. M. Al-Hasan, A. Nhlabatsi, F. Bensaali, and G. In IEEE Open Journal of the Computer Society, Dimitrakopoulos authored a paper on the application of Federated Learning to privacy-protecting reconstruction systems. 5, 227–247, 2024, doi: 10.1109/OJCS.2024.3396344.

[5] M. Amiri, G. Quirchmayr, and E. "Using a Local Distance-Based Methodology to Introduce Differential Privacy for Recommender Systems," Weippl wrote in IEEE Access. vol. 44, Issue 1, Issue 2. 13, 161824–161843, 2025, doi: 10.1109/ACCESS.2025.3605563.

[6] C. Dwork and A. Roth, "The Algorithmic Foundations of Differential Privacy," Foundations and Trends in Theoretical Computer Science, vol. 9, no. 3–4, pp. 211–407, 2014, doi: 10.1561/04000000042.

[7] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep Learning with Differential Privacy," in Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, pp. 308–318, 2016.

[8] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, vol. 54, pp. 1273–1282, 2017.

[9] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth, "Practical Secure Aggregation for Privacy-Preserving Machine Learning," in Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, pp. 1175–1191, 2017, doi: 10.1145/3133956.3133982.

[10] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership Inference Attacks Against Machine Learning Models," in 2017 IEEE Symposium on Security and Privacy, pp. 3–18, 2017, doi: 10.1109/SP.2017.41.

[11] L. Zhu, Z. Liu, and S. Han, "Deep Leakage from Gradients," in Advances in Neural Information Processing Systems, vol. 32, 2019.

[12] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," ACM Computing Surveys, vol. 41, no. 3, pp. 1–58, 2009, doi: 10.1145/1541880.1541882.

[13] M. Al-Rubaie and J. M. Chang, "Privacy-Preserving Machine Learning: Threats and Solutions," IEEE Security & Privacy, vol. 17, no. 2, pp. 49–58, 2019, doi: 10.1109/MSEC.2018.2888775.

[14] R. Xu, N. Baracaldo, and J. Joshi, "Privacy-Preserving Machine Learning: Methods, Challenges and Directions," arXiv preprint arXiv: 2108.04417, 2021.

[15] "Differential Privacy in Deep Learning: A Literature Survey," Neurocomputing, vol. 589, Art. no. 127663, 2024, doi: 10.1016/j.neucom.2024.127663.

BIOGRAPHIES



Ms. Priti Jitendra Yadav.
M.Sc. Information Technology.
Student, IT Dept. D. G. Ruparel
College.



Ms. Aarti Sanjay Gawai
M.Sc. Information Technology
Asst. Prof., IT/CS Dept.
D. G. Ruparel College.