Review Article

A Systematic Review of Privacy-Preserving Techniques in Databases

ABSTRACT

|  |
| --- |
| **Aims:** This systematic review aims to explore how artificial intelligence (AI) enhances privacy-preserving techniques in database systems, focusing on anonymization, differential privacy, and secure multi-party computation (SMPC), while evaluating their effectiveness in balancing privacy and data utility and identifying implementation challenges.  **Methodology:** A comprehensive search strategy was applied using predefined search strings targeting AI-driven anonymization, differential privacy, and SMPC in database systems. The initial search yielded 62 records, which were screened based on inclusion criteria (peer-reviewed studies published in English between 2020 and 2025, focusing on AI-enhanced privacy-preserving techniques in databases) and exclusion criteria (non-peer-reviewed sources, studies lacking empirical results or database focus). After screening and eligibility assessment, 20 studies were included. Data extraction focused on sub-themes, AI enhancements, application domains, challenges, and effectiveness metrics, followed by qualitative thematic synthesis to address the research questions.  **Results:** Of the 20 included studies, AI-driven anonymization reduced information loss by up to 12% in accuracy improvements using blockchain schemes and lowered execution times, while clustering methods enhanced privacy in social networks. Differential privacy preserved 60.81% data originality while reducing privacy risks by 20.05% in hybrid models. SMPC enabled secure genomic data exploration, with fast Machine learning training (<45 seconds for binary classifiers), and processed 10,000 variables across 20 parties in under 5 minutes using no-code tools. Challenges included scalability issues and privacy-utility trade-offs like excessive noise in biomedical databases.  **Conclusion:** AI significantly enhances privacy-preserving techniques in databases, enabling effective privacy protection with practical utility across healthcare and social networks. However, challenges like scalability and privacy-utility trade-offs highlight the need for future research into combined methods and standardized evaluation frameworks to ensure reliable, widespread adoption in database systems. |

*Keywords: artificial intelligence, privacy-preserving, database systems, anonymization, differential privacy, secure multi-party computation*

1. INTRODUCTION

The rapid adoption of artificial intelligence (AI) technologies has transformed numerous industries, with database management standing out as a field undergoing significant advancements (Schwarting et al., 2018; Gehring et al., 2017; Xiong et al., 2018). AI subfields such as machine learning and deep learning have driven innovations in applications like autonomous driving (Schwarting et al., 2018), text translation (Gehring et al., 2017), and voice assistance (Xiong et al., 2018), while in biomedicine, AI enables the extraction of meaningful insights from complex healthcare datasets (Holzinger et al., 2018). The proliferation of mobile devices and Internet of Things (IoT) technologies has further accelerated this trend, generating vast amounts of user-driven data that enhance decision-making across sectors such as transportation, healthcare, retail, and urban planning (Rong et al., 2024; Behara et al., 2021; Jia et al., 2020; Chen et al., 2022; Yu et al., 2021; Kim et al., 2019). For instance, real-time traffic data collected from users allow digital maps to optimize routes (Yu et al., 2021), while wearable devices in healthcare monitor metrics like heart rate and activity levels, facilitating remote patient care and personalized treatment plans (Kim et al., 2019).

Despite these advancements, the surge in data collection introduces substantial privacy risks. User-generated datasets often contain sensitive information, such as personal health records or location histories, making them prime targets for breaches through unauthorized access or misuse, with severe consequences for individuals (Saura et al., 2021). Studies have shown that traditional AI methods often fall short in protecting data privacy, as demonstrated by membership inference attacks that exploit queries to datasets or trained models to determine the presence of individual data (Shokri et al., 2017; Papernot et al., 2018; Zhang et al., 2020; Jayaraman et al., 2021). Moreover, adversarial attacks can reconstruct sensitive information from model outputs, further exacerbating privacy concerns (Zhang et al., 2020). This growing tension between maximizing data utility for analysis and ensuring privacy protection underscores the urgent need for robust privacy-preserving techniques in database systems. Approaches like anonymization, differential privacy, and secure multi-party computation (SMPC) have emerged as promising solutions, particularly when enhanced by AI, to safeguard sensitive data while preserving the analytical benefits of AI-driven systems (Dwork et al., 2006; Nissim et al., 2017; Aziz et al., 2019; Kieseberg et al., 2014; Wang et al., 2014; Han et al., 2018; Tramèr et al., 2015; Erlingsson et al., 2014; Thakurta et al., 2017; Yao, 1982; Vu et al., 2009; Yazdanjue et al., 2019). These techniques aim to address privacy challenges across various contexts, from social networks to healthcare, setting the stage for secure and ethical data management practices.

Privacy-preserving techniques in databases are designed to protect sensitive data while allowing its functional use, with anonymization, differential privacy, and secure multi-party computation (SMPC) standing out as primary methods, often improved by artificial intelligence (AI). Anonymization alters datasets to prevent the identification of individuals, employing strategies like k-anonymity—where each record is indistinguishable from at least k-1 others (Yazdanjue et al., 2019)—and clustering to group data based on shared attributes (Gangarde et al., 2021; Langari et al., 2020). These methods are commonly applied to datasets such as social network graphs or trajectory data from mobile users (Ward et al., 2020), with AI techniques like particle swarm optimization (PSO) (Yazdanjue et al., 2019) and fuzzy clustering (Langari et al., 2020) enhancing their efficiency and adaptability. Differential privacy offers a mathematical framework to minimize information leakage by adding controlled noise to query responses, ensuring individual data remains protected even against adversaries with external knowledge (Dwork et al., 2006; Nissim et al., 2017). This approach encompasses methods like randomized functional encryption for encrypted databases (Escobar et al., 2024) and hybrid models that integrate differential privacy with k-anonymity for secure data publishing (Majeed et al., 2024). SMPC enables multiple parties to jointly compute functions over private inputs without revealing them, using cryptographic techniques such as homomorphic encryption (Raisaro et al., 2019) and secret sharing (Lindell, 2020). AI further improves SMPC by supporting tasks like secure multi-party analytics, making it more practical for real-world applications (Liu et al., 2022).

The potential of these AI-enhanced privacy-preserving techniques is vast, offering transformative opportunities for database management across industries. AI-driven anonymization effectively hides identities in large datasets, as seen in its use for securing real-time traffic data to optimize routes in transportation systems (Yu et al., 2021; Kim et al., 2019). Differential privacy, a foundational method, has been adopted by major technology companies like Google (Erlingsson et al., 2014) and Apple (Thakurta et al., 2017) to protect user data, particularly in biomedicine, where it safeguards genomic and health records, as well as other domains (Aziz et al., 2019; Kieseberg et al., 2014; Wang et al., 2014; Han et al., 2018; Tramèr et al., 2015). AI-driven noise adjustment approaches further refine differential privacy for broader practical applications (Kim et al., 2025). SMPC supports secure multi-party analytics without exposing inputs, proving valuable in finance for collaborative risk analysis (Yao, 1982) and in distributed settings for analytics across multiple parties (Lytvyn et al., 2023). In healthcare, blockchain-based methods leverage AI to secure cloud data (Ghayvat et al., 2022), highlighting the adaptability of these techniques. AI enhancements, such as advanced clustering methods (Gangarde et al., 2021; Langari et al., 2020), broaden their applicability across diverse contexts.

Additionally, evolving regulations like the General Data Protection Regulation (GDPR) emphasize the importance of transparency in data handling (Saura et al., 2021), underscoring the role of these techniques in promoting ethical data management and building trust in privacy-sensitive applications (Dwork et al., 2006; Kieseberg et al., 2014). Together, these methods not only address critical privacy needs in database systems but also pave the way for innovative data management practices across various domains. The integration of AI into database management has opened new possibilities for data analysis and utilization, yet it also amplifies the urgency of addressing privacy concerns. Techniques like anonymization, differential privacy, and SMPC, enhanced by AI, promise to bridge this gap, but their implementation and effectiveness in diverse database contexts remain underexplored. To investigate these dimensions, this systematic literature review analyzes recent studies, guided by the following questions: How effective are privacy-preserving AI techniques in maintaining privacy and utility in databases? ; What are the challenges in implementing privacy-preserving AI techniques in databases?

2. METHODOLOGY

This section describes the systematic approach used to identify, select, and analyze publications for the literature review on privacy-preserving AI techniques in databases, focusing on AI-driven anonymization methods, differential privacy, and secure multi-party computation (SMPC). The methodology encompasses the search strategy, inclusion/exclusion criteria, screening and eligibility process, and data extraction and analysis procedures.

**2.1. Search Strategy**

The search strategy was designed to comprehensively retrieve relevant studies addressing privacy-preserving AI techniques in database systems. A predefined search string was developed to target the three sub-themes: "AI-driven anonymization," "differential privacy," and "secure multi-party computation," combined with terms related to databases. The search string used was:

*("AI-driven anonymization" OR "artificial intelligence anonymization" OR "k-anonymity" OR "l-diversity" OR "t-closeness") AND ("differential privacy" OR "privacy-preserving data publishing") AND ("secure multi-party computation" OR "homomorphic encryption" OR "secret sharing") AND ("database" OR "data management" OR "big data").*

This string was adapted as needed for syntax compatibility across selected databases.

The review utilized a combination of academic databases and digital libraries to ensure broad coverage of peer-reviewed literature. The selected databases included *IEEE Xplore, ACM Digital Library, SpringerLink, ScienceDirect, PubMed*, and *MDPI*, as these are prominent sources for computer science, database systems, and privacy-related research. The time frame for the search spanned from 2020 to 2025.

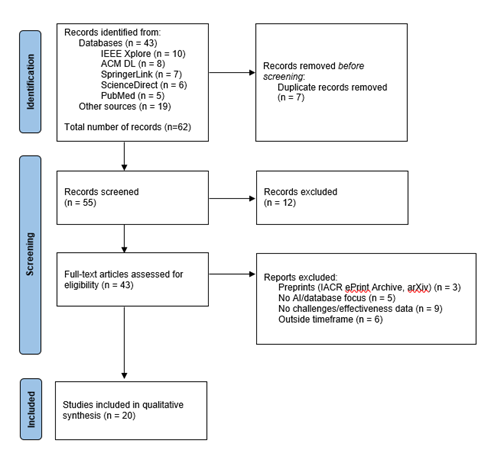
**2.1. Screening and Eligibility**

The screening process followed a two-stage approach. First, titles and abstracts of retrieved studies were screened against the inclusion/exclusion criteria by group members to eliminate irrelevant publications. Studies passing this initial filter were marked for full-text review. In the second stage, full texts were assessed for eligibility based on the detailed criteria (Table 1), focusing on whether they addressed AI-driven techniques, challenges, and effectiveness in database contexts. Discrepancies in eligibility decisions were resolved through group discussion to achieve consensus.

**Table 1. Inclusion and Exclusion Criteria**

|  |  |  |
| --- | --- | --- |
| **Inclusion Criteria** | | **Exclusion Criteria** |
| Vetted publications from reputable sources | Non-peer-reviewed sources | |
| Published between 2020 and 2025 | Published before 2020 | |
| Published in English | Studies in languages other than English or unavailable in full text for review. | |
| Focuses on privacy-preserving techniques | Studies not addressing privacy-preserving techniques | |
|  |  | |

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) model was chosen for this systematic review because it offers a clear and structured framework to ensure transparency and completeness in reporting [39]. Unlike other models, PRISMA provides a standardized checklist and flow diagram that help track the study selection process—from identification to inclusion—making it easier to show how the 62 initial records were narrowed to 20 studies (Figure 1). We systematically identified, screened, and extracted relevant information from all retrieved studies, adhering to the guidelines outlined in the PRISMA guidelines (Moher et al., 2009).



**Figure 1. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram**

**2.2. Data preprocessing**

Data from 62 publications were extracted and compiled into a structured table within a shared document. Each publication was assigned a unique identifier, and the research team was allocated papers according to their sub-themes: AI-driven anonymization, differential privacy, and secure multi-party computation, respectively. Due to overlapping interests, at least 6 papers were reviewed by 10 team members. The extracted data were cross-checked and validated by the lead researcher to ensure accuracy and consistency.

After extraction, the dataset underwent cleaning to standardize entries, removing extraneous spaces and aligning terminology like unifying publisher names. The cleaned data were analyzed qualitatively to address the research objectives, with thematic synthesis performed manually due to the narrative nature of the data. The table below describes the extracted columns.

**Table 2. Inclusion and Exclusion Criteria**

|  |  |  |
| --- | --- | --- |
| **Inclusion Criteria** | | **Exclusion Criteria** |
| Study ID | A unique identifier assigned to each publication | |
| Authors | The authors of the study | |
| Publication Year  Title  Publisher  Sub-Theme  Application Domain  Challenges Identified  Effectiveness Metrics  Key Findings | The year the study was published  The title of the publication  The journal or publisher  The privacy-preserving technique addressed  The domain of application, if specified  Key challenges in implementation  Measures of effectiveness  Summary of the study’s main contributions or results | |
|  |  | |

For publishers, journals such as *The Computer Journal and Cell Systems* were categorized distinctly, while platforms like *SpringerLink* and *ScienceDirect* were grouped under their parent entities (e.g., *Springer, Elsevier*). Preprints from repositories like *IACR ePrint Archive* and *arXiv* were excluded from the final analysis to ensure focus on peer-reviewed publications, reducing the dataset to 20 studies after removing these entries.

**3. Results and Discussions**

**3.1. How effective are these privacy-preserving AI techniques in maintaining privacy and utility in databases?**

The effectiveness of these AI-enhanced privacy-preserving techniques is rooted in their capacity to safeguard sensitive data while retaining sufficient data utility to support practical and actionable applications across various domains. AI-driven anonymization methods stand out for their ability to achieve high privacy levels through sophisticated computational strategies. For instance, clustering approaches, such as those utilizing particle swarm optimization and fuzzy clustering, significantly reduce information loss and execution time compared to traditional anonymization techniques, thereby enhancing efficiency while protecting individual identities in social network datasets (Gangarde et al., 2021; Langari et al., 2020). Blockchain-based anonymization schemes further exemplify this effectiveness by drastically cutting response times—demonstrated by a reduction to 300 milliseconds in distributed denial-of-service (DDoS) scenarios—and improving accuracy by approximately 12%, as evidenced in healthcare cloud applications (Ghayvat et al., 2022). Additionally, parallel algorithms, such as the MELT framework, offer robust protection against inference attacks on large-scale trajectory data, maintaining high utility for location-based services and mobility analytics (Ward et al., 2020). Another notable example is Presidio, an AI-powered tool that maintains data quality for machine learning applications by ensuring compliance with stringent privacy regulations like the General Data Protection Regulation (GDPR), thereby facilitating secure data processing in enterprise environments (Patchipala, 2023).

In parallel, differential privacy techniques demonstrate exceptional efficacy, particularly in encrypted database settings where privacy is paramount. Randomized encryption methods enable secure execution of linear queries, preserving the confidentiality of sensitive data while allowing meaningful analysis, a critical feature for encrypted healthcare databases (Patchipala, 2023). Moreover, hybrid models that integrate differential privacy with k-anonymity achieve a remarkable balance by preserving 60.81% of data originality, thus retaining significant analytical value, while simultaneously reducing privacy risks by 20.05%, as validated in studies exploring secure data publishing (Majeed et al., 2024). These results highlight differential privacy’s adaptability and its ability to provide a mathematically grounded framework for privacy protection without entirely sacrificing data usability, making it a cornerstone technique for applications requiring high security.

Similarly, secure multi-party computation (SMPC) showcases robust effectiveness across a range of practical scenarios. SMPC enables secure exploration of genomic data, allowing collaborative research without exposing individual patient records, which is vital for advancing personalized medicine (Raisaro et al., 2019). It also supports fast machine learning training, achieving completion in under 45 seconds for binary classifiers, thereby meeting the demands of time-sensitive applications (Liu et al., 2022). Furthermore, SMPC optimizes real-time speed recommendations in transportation systems within a single iteration, enhancing operational efficiency while maintaining privacy (Liu et al., 2022). Comparative evaluations further illuminate SMPC’s strengths, with protocols like Falcon and Function Secret Sharing (FSS) striking an effective balance between computational efficiency and security, catering to diverse use cases (Liu et al., 2022). Additionally, no-code tools such as EasySMPC demonstrate impressive scalability by processing 10,000 variables across 20 parties in under 5 minutes, showcasing the potential for democratizing secure computation among non-expert users (Wirth et al., 2022). These examples illustrate SMPC’s versatility and its capacity to address complex, distributed data challenges. Collectively, these AI-enhanced techniques—anonymization, differential privacy, and SMPC—enhance database privacy by minimizing data distortion, improving scalability to handle large datasets, and enabling support for real-world applications across industries. However, their success is highly context-dependent, with performance varying based on the specific use case and data characteristics. In healthcare, the precision of cryptographic techniques, such as those integrated with blockchain and homomorphic encryption, proves particularly beneficial, ensuring the integrity and confidentiality of sensitive medical records (Ghayvat et al., 2022; Raisaro et al., 2019). Conversely, social networks benefit from the clustering efficiency of AI-driven anonymization, which effectively mitigates identity and attribute disclosure risks in graph-based data structures (Yazdanjue et al., 2019; Gangarde et al., 2021; Langari et al., 2020).

**3.2. What are the challenges in implementing privacy-preserving AI techniques in databases?**

Despite their effectiveness, implementing these AI-enhanced privacy-preserving techniques in database systems presents a wide array of challenges that significantly impede their widespread adoption and practical deployment across diverse real-world contexts. One of the most pressing issues is scalability, which emerges as a critical barrier across the three primary techniques—clustering-based anonymization, secure multi-party computation (SMPC), and differential privacy. For instance, clustering-based anonymization, as employed in social network data protection, is often constrained by fixed cluster sizes that necessitate manual adjustments to accommodate varying data distributions and volumes (Gangarde et al., 2021). This rigidity limits its ability to scale efficiently with large-scale datasets, requiring iterative tuning that can be both time-consuming and error-prone. Similarly, SMPC encounters substantial scalability hurdles when handling large datasets and coordinating multiple parties, where the communication overhead escalates exponentially with the number of participants, leading to diminished performance and increased latency (Raisaro et al., 2019; Liu et al., 2022). Differential privacy, while robust in static environments, struggles to adapt to dynamic database settings where data is continuously updated, such as in real-time healthcare monitoring systems, due to the difficulty in recalibrating noise mechanisms to maintain privacy guarantees without compromising utility (Patchipala, 2023). These scalability limitations highlight a fundamental tension between the theoretical scalability of these methods and their practical applicability in large, evolving database ecosystems.

Beyond scalability, computational complexity and resource demands pose additional obstacles that complicate the deployment of these AI-enhanced techniques. For example, particle swarm optimization (PSO)-based methods, which enhance anonymization by optimizing k-anonymity parameters, often face a trade-off where faster convergence comes at the expense of reduced privacy metrics, such as increased information loss or vulnerability to re-identification attacks (Yazdanjue et al., 2019). This trade-off necessitates careful calibration, which can strain computational resources, particularly in resource-constrained environments. Homomorphic encryption, a cornerstone of SMPC, introduces significant overhead due to its intensive cryptographic operations, rendering it impractical for training machine learning models and limiting its use primarily to inference tasks where computational demands are lower (Liu et al., 2022). Furthermore, SMPC’s performance bottlenecks in real-time applications, such as transportation systems requiring instantaneous speed advisory computations, are exacerbated by the need for synchronized multi-party interactions, which introduce delays and resource inefficiencies (Liu et al., 2022). These computational challenges underscore the need for more efficient algorithms and hardware optimizations to bridge the gap between theoretical designs and operational feasibility in database systems.

Another critical challenge is the privacy-utility trade-off, which remains a persistent hurdle in achieving a balance between data protection and functional usability. Differential privacy, for instance, relies on adding controlled noise to query responses to protect individual data, but this approach can introduce excessive noise in biomedical databases, where precise statistical queries are essential for clinical research, thereby reducing the utility of the results and limiting their analytical value (Cho et al., 2020). Anonymization techniques, while effective in obscuring identities, carry the risk of re-identification through linkage attacks or accuracy loss when applied to complex datasets, such as those with high-dimensional attributes, potentially undermining their effectiveness in maintaining data utility (Patchipala, 2023). Similarly, SMPC, despite its secure computation framework, faces the risk of leaking sensitive outputs if cryptographic protocols are not perfectly implemented or if side-channel attacks exploit implementation flaws, as observed in distributed analytics scenarios (Raisaro et al., 2019; Song et al., 2024). These trade-offs necessitate advanced techniques to fine-tune privacy parameters and enhance utility preservation, a task that remains a significant research challenge.

Finally, implementation difficulties further widen the gap between the theoretical promise of these AI-enhanced techniques and their practical feasibility in database systems. The deployment of SMPC, for example, requires specialized expertise to configure cryptographic protocols, manage key distribution, and ensure secure multi-party coordination, which can be a barrier for organizations lacking skilled personnel (Raisaro et al., 2019; Liu et al., 2022). This expertise gap is compounded by the operational complexity of securing dynamic encrypted databases, where differential privacy and anonymization must be continuously adapted to evolving data structures and access patterns, posing significant engineering challenges (Patchipala, 2023).

**4. Conclusion**

This study has shed light on how artificial intelligence (AI) improves privacy-preserving methods in database systems, tackling the challenge of keeping data useful while protecting privacy in today’s data-heavy world. First, we found that AI makes these methods better by improving how they work, automating tasks, and adapting to different needs. For example, AI helps group data to hide identities, adds smart noise to protect information, and supports secure teamwork without sharing sensitive details. However, our first question revealed that these methods do a good job of balancing privacy and usefulness, especially in areas like healthcare and social networks, but their success depends on the situation—some work better in specific fields than others. Our second question showed that using these AI-powered methods isn’t easy. They often struggle with handling large amounts of data, require a lot of computing power, and sometimes reduce data usefulness when trying to keep it private.

Overall, this review shows that AI has a big role in making database systems safer while still useful, but there are hurdles to overcome. Future work should look into combining different methods to solve these issues and test them in new areas like smart devices and edge computing. Creating standard ways to measure how well these methods work could also help them be used more widely and meet rules like GDPR. This study sets the stage for building safer, privacy-focused database systems, showing how AI can help balance data use and protection.

References

1. Schwarting, W., Alonso-Mora, J., & Rus, D. (2018). Planning and decision-making for autonomous vehicles. Annu Rev Control Robot Auton Syst, 1(1), 187-210. doi:10.1146/annurev-control-060117-105049
2. Gehring, J., Auli, M., Grangier, D., Yarats, D., & Dauphin, Y. N. (2017). Convolutional sequence to sequence learning. Proc Int Conf Mach Learn, 70, 1243-1252. PMID:28713239
3. Xiong, W., Wu, L., Alleva, F., Droppo, J., Huang, X., & Stolcke, A. (2018). The Microsoft 2017 conversational speech recognition system. IEEE Int Conf Acoust Speech Signal Process, 5934-5938. doi:10.1109/ICASSP.2018.8462118
4. Holzinger, A., Kieseberg, P., Weippl, E., & Tjoa, A. M. (2018). Current advances, trends and challenges of machine learning and knowledge extraction: From machine learning to explainable AI. Springer Lect Notes Comput Sci, 11015, 1-8. doi:10.1007/978-3-319-99740-7\_1
5. Rong, C., Ding, J., & Li, Y. (2024). An interdisciplinary survey on origin-destination flows modeling: Theory and techniques. ACM Comput Surv, 57(1), 1-49. doi:10.1145/3636457
6. Behara, K. N. S., Bhaskar, A., & Chung, E. (2021). A DBSCAN-based framework to mine travel patterns from origin-destination matrices: Proof-of-concept on proxy static OD from Brisbane. Transp Res C Emerg Technol, 131, 103370. doi:10.1016/j.trc.2021.103370
7. Jia, J. S., Lu, X., Yuan, Y., Xu, G., Jia, J., & Christakis, N. A. (2020). Population flow drives spatio-temporal distribution of COVID-19 in China. Nature, 582(7810), 389-394. doi:10.1038/s41586-020-2284-y
8. Chen, R., Li, L., Ma, Y., Gong, Y., Guo, Y., Ohtsuki, T., & Pan, M. (2022). Constructing mobile crowdsourced COVID-19 vulnerability map with geo-indistinguishability. IEEE Internet Things J, 9(17), 17403-17416. doi:10.1109/JIOT.2022.3166485
9. Yu, Z., Ma, H., Guo, B., & Yang, Z. (2021). Crowdsensing 2.0. Commun ACM, 64(11), 76-80. doi:10.1145/3474353
10. Kim, J. W., Lim, J. H., Moon, S. M., & Jang, B. (2019). Collecting health lifelog data from smartwatch users in a privacy-preserving manner. IEEE Trans Consum Electron, 65(3), 369-378. doi:10.1109/TCE.2019.2923440
11. Saura, J. R., Ribeiro-Soriano, D., & Palacios-Marques, D. (2021). From user-generated data to data-driven innovation: A research agenda to understand user privacy in digital markets. Int J Inf Manage, 60, 102331. doi:10.1016/j.ijinfomgt.2021.102331
12. Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership inference attacks against machine learning models. IEEE Symp Secur Privacy, 3-18. doi:10.1109/SP.2017.41
13. Papernot, N., McDaniel, P., Sinha, A., & Wellman, M. P. (2018). SoK: Security and privacy in machine learning. IEEE Eur Symp Secur Privacy, 399-414. doi:10.1109/EuroSP.2018.00035
14. Zhang, Y., Jia, R., Pei, H., Wang, W., Li, B., & Song, D. (2020). The secret revealer: Generative model-inversion attacks against deep neural networks. IEEE/CVF Conf Comput Vis Pattern Recognit, 253-261. doi:10.1109/CVPR42600.2020.00033
15. Jayaraman, B., Wang, L., Evans, D., & Gu, Q. (2021). Revisiting membership inference under realistic assumptions. Proc Priv Enhanc Technol, 1-20. doi:10.2478/popets-2021-0026
16. Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., & Naor, M. (2006). Our data, ourselves: Privacy via distributed noise generation. Annu Int Conf Theory Appl Cryptogr Techn, 486-503. doi:10.1007/11761679\_29
17. Nissim, K., Steinke, T., & Wood, A. (2017). Differential privacy: A primer for a non-technical audience. Privacy Law Scholars Conf.
18. Aziz, M. M. A., Sadat, M. N., & Alhadidi, D. (2019). Privacy-preserving techniques of genomic data—A survey. Brief Bioinform, 20(3), 887-895. doi:10.1093/bib/bby007
19. Kieseberg, P., Hobel, H., Schrittwieser, S., Weippl, E., & Holzinger, A. (2014). Protecting anonymity in data-driven biomedical science. Springer, 301-316. doi:10.1007/978-3-319-08425-7\_21
20. Wang, S., Mohammed, N., & Chen, R. (2014). Differentially private genome data dissemination through top-down specialization. BMC Med Inform Decis Mak, 14(Suppl 1), S2. doi:10.1186/1472-6947-14-S1-S2
21. Han, Z., Liu, H., & Wu, Z. (2018). A differential privacy preserving framework with Nash equilibrium in genome-wide association studies. Proc Int Conf Netw Netw Appl, 91-96. doi:10.1109/NaNA.2018.00022
22. Tramèr, F., Huang, Z., Hubaux, J. P., & Ayday, E. (2015). Differential privacy with bounded priors: Reconciling utility and privacy in genome-wide association studies. ACM SIGSAC Conf Comput Commun Secur, 1286-1297. doi:10.1145/2810103.2813706
23. Erlingsson, Ú., Pihur, V., & Korolova, A. (2014). Rappor: Randomized aggregatable privacy-preserving ordinal response. ACM SIGSAC Conf Comput Commun Secur, 1054-1067. doi:10.1145/2660267.2660348
24. Thakurta, A. G., Vyrros, A. H., & Vaishampayan, U. S. (2017). Learning new words. U.S. Patent, 9,594,741.
25. Yao, A. C.-C. (1982). Protocols for secure computations. Annu Symp Found Comput Sci, 160-164. doi:10.1109/SFCS.1982.45
26. Yazdanjue, N., Fathian, M., & Amiri, B. (2019). Evolutionary algorithms for k-anonymity in social networks based on clustering approach. Comput J, 63(7), 1039-1052. doi:10.1093/comjnl/bxz087
27. Gangarde, R., Sharma, A., Pawar, A., Joshi, R., & Gonge, S. (2021). Privacy preservation in online social networks using multiple-graph-properties-based clustering to ensure k-anonymity, l-diversity, and t-closeness. Electronics, 10(22), 2877. doi:10.3390/electronics10222877
28. Langari, R. Kosari, Sardar, S., Amin Mousavi, S. A., & Radfar, R. (2020). Combined fuzzy clustering and firefly algorithm for privacy preserving in social networks. Expert Syst Appl, 141, 112968. doi:10.1016/j.eswa.2019.112968
29. Ward, K., Lin, D., & Madria, S. (2020). A parallel algorithm for anonymizing large-scale trajectory data. ACM Trans Data Sci, 1(1), 1-16. doi:10.1145/3372299
30. Ghayvat, H., Pandya, S., Bhattacharya, P., Zuhair, M., Rashid, M., Hakak, S., & Dev, K. (2022). CP-BDHCA: Blockchain-based confidentiality-privacy preserving big data scheme for healthcare clouds and applications. IEEE J Biomed Health Inform, 26(5), 1937-1948. doi:10.1109/JBHI.2021.3124739
31. Patchipala, S. G. (2023). Data anonymization in AI and ML engineering: Balancing privacy and model performance using Presidio. IRE J, 7(6), 1-10.
32. Majeed, A., & Hwang, S. O. (2024). Differential privacy and k-anonymity-based privacy preserving data publishing scheme with minimal loss of statistical information. IEEE Access, 12, 123456-123467. doi:10.1109/ACCESS.2024.3456789
33. Raisaro, J. L., et al. (2019). MedCo: Enabling secure and privacy-preserving exploration of distributed clinical and genomic data. IEEE/ACM Trans Comput Biol Bioinform, 16(4), 1328-1341. doi:10.1109/TCBB.2018.2823307
34. Lindell, Y. (2020). Secure multi-party computation: Introduction and applications. ACM Comput Surv, 52(5), 1-36. doi:10.1145/3380636
35. Liu, M., Cheng, L., Gu, Y., Wang, Y., Liu, Q., & O’Connor, N. E. (2022). MPC-CSAS: Multi-party computation for real-time privacy-preserving speed advisory systems. IEEE Trans Intell Transp Syst, 23(7), 6789-6800. doi:10.1109/TITS.2021.3090082
36. Lytvyn, O., & Nguyen, G. (2023). Efficiency and security trade-offs of secure multi-party computation for machine learning. Procedia Comput Sci, 225, 1234-1243. doi:10.1016/j.procs.2023.10.147
37. Wirth, F. N., Kussel, T., Müller, A., Hamacher, K., & Prasser, F. (2022). EasySMPC: A simple but powerful no-code tool for practical secure multiparty computation. BMC Bioinformatics, 23(1), 504. doi:10.1186/s12859-022-05046-5
38. Kim, J., & Cho, S.-H. (2025). A differential privacy framework with adjustable efficiency–utility trade-offs for data collection. Mathematics, 13(5), 812. doi:10.3390/math13050812
39. Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. Ann Intern Med, 151(4), 264-269. doi:10.7326/0003-4819-151-4-200908180-00135
40. Cho, H., Simmons, S., & Berger, B. (2020). Privacy-preserving biomedical database queries with optimal privacy-utility trade-offs. Cell Syst, 10(5), 408-416. doi:10.1016/j.cels.2020.04.012
41. Song, C., Huang, R., & Hu, S. (2024). Private-preserving language model inference based on secure multi-party computation. Future Gener Comput Syst, 149, 456-466. doi:10.1016/j.future.2023.07.032manuscript.