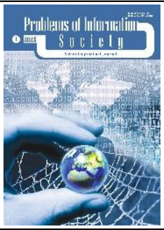

www.jpis.az

Anonymization of personal medical data based on artificial intelligence

Ramiz Shikhaliyev

Institute of Information Technology, B. Vahabzade str., 9A, AZ1141, Baku, Azerbaijan

shikhramiz61@gmail.com

orcid.org/0000-0002-8594-6721

ARTICLE INFO

https://doi.org/10.25045/jpis.v17.i1.04

Article history:

Received 08.09.2025
Received in revised form 15.11.2025
Accepted 20.01.2026

Keywords:

Personal medical data
Patient confidentiality
Data anonymization
Artificial intelligence
De-identification
Differential privacy

ABSTRACT

The widespread digitalization of healthcare has led to the accumulation of substantial volumes of personal medical data (PMD), creating new opportunities for enhancing the quality of medical care, supporting informed clinical decision-making, and advancing scientific research. At the same time, the large-scale accumulation of PMD poses serious risks to cybersecurity and patient privacy. One key mechanism for mitigating these risks is the anonymization of PMD. However, traditional anonymization methods demonstrate significant limitations when processing complex, multidimensional, and unstructured PMD. This article examines approaches to using artificial intelligence (AI) methods for PMD anonymization. It substantiates the need to move from classical statistical, syntactic, and cryptographic models to intelligent and adaptive systems capable of automatically identifying sensitive information, performing context-sensitive transformations, and generating synthetic data. Furthermore, key technical, ethical, and regulatory issues, as well as the risks associated with the use of AI for PMD anonymization, are analyzed.

1. Introduction

The digitalization of healthcare has led to the automation of the collection, storage, and use of personal medical data (PMD). At the same time, the digitalization of healthcare has generated vast volumes of structured and unstructured PMD. These data create a "big data" ecosystem and include electronic health records (EHR), multi-omics profiles, medical images, and Internet of Medical Things (IoMT) data (Gupta and Kumar, 2023). The use of these data enables personalized diagnostics, predictive analytics, and the discovery of new treatments. However, the sensitive nature of PMD makes it a target for patient privacy violations. Unauthorized disclosure or re-identification of these data can lead to discrimination, social stigma, financial losses, and erosion of trust in the healthcare system.

To address these issues, strict legal and ethical frameworks have been established. The European Union's General Data Protection Regulation (GDPR) prescribes principles of data minimization, purpose limitation, and requires anonymization or pseudonymization as key technical and organizational measures (GDPR, 2016). Similarly, the US Health Insurance Portability and Accountability Act (HIPAA) establish national standards for the anonymization of protected health information (HIPAA, 1996).

Traditional data anonymization methods, based primarily on statistical disclosure control and cryptographic methods, form the basis for ensuring data privacy. However, when applied to modern PMDs, their usefulness is significantly reduced, especially for high-dimensional data, which reduces their analytical value. Besides, they are vulnerable to sophisticated linkage attacks using auxiliary datasets and are poorly adapted to unstructured data, such as clinical text records and medical images (Rocheret et al., 2019)

Recent advances in artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), enable the creation of methods for anonymizing PMDs. AI can automate the identification of PMDs and their protection in complex datasets, as well as create adaptive anonymization strategies and the generation of synthetic data to ensure privacy (Yoon et al., 2020; Lison et al., 2021)

The purpose of this article is to comprehensively analyze and systematize the methods of anonymization of PMD with an emphasis on intelligent approaches based on AI, as well as to identify key scientific and practical problems that hinder the widespread implementation of intelligent anonymization of PMD, and to determine the prospects for further research in this area.

The rest of the article is structured as follows: Section 2 provides an overview of related work; Section 3 describes the materials and methods used in this study; Section 4 presents a discussion; Section 5 summarizes the work and outlines directions for future research.

2. Related work

Data protection and anonymization have seen significant advances over the past two decades. Early research focused primarily on statistical and syntactic anonymization models designed to reduce the risk of identity and attribute disclosure in structured datasets.

One of the most fundamental approaches is k-anonymity, proposed by Sweeney (2002), which guarantees that each record is indistinguishable from at least k-1 other records with respect to a set of quasi-identifiers. However, subsequent research has shown that k-anonymity is vulnerable to attacks related to homogeneity and background knowledge, especially in medical datasets with asymmetric diagnosis distributions. To address these issues, l-diversity was proposed. (Machanavajjhala et al., 2006), which ensures the diversity of sensitive attributes within equivalence classes. However, further analysis revealed the limitations of l-diversity in the presence of semantically similar sensitive attributes. This limitation motivated to the development of the t-closeness model, which imposes a bound on the distance between the distribution of sensitive attributes in anonymized groups and their global distribution in the dataset (Li et al., 2007). Although t-closeness improves robustness to attribute disclosure, its use is often accompanied by significant information loss and increased computational complexity, which limits the scalability of the method in large-scale healthcare systems.

In parallel with the development of syntactic models, differential privacy (DP) was proposed as a mathematically rigorous data protection paradigm that provides formal guarantees against re-identification (Dwork, 2006). DP-based approaches add calibrated noise to query results or training processes, thereby ensuring that the presence or absence of an individual has a negligible effect on the output. DP is actively researched in the fields of healthcare analytics and ML, including the development of privacy-preserving models (Abadi et al., 2016). However, a number of studies emphasize the fundamental trade-off between the level of privacy and the clinical utility of data, as excessive noise can lead to reduced diagnostic and prognostic accuracy.

Cryptographic methods represent another significant area of research in the field of PMD protection. In particular, approaches such as homomorphic encryption allow computations to be performed directly on encrypted data, which theoretically ensures secure analysis without revealing sensitive information (Acar et al., 2018). However, despite their high level of security, cryptographic methods often prove impractical for real-world medical applications due to high computational costs and limited compatibility with modern ML methods. Therefore, more lightweight cryptographic and hybrid approaches have been proposed to achieve a balance between privacy and usability in the analysis of structured medical data (Aminifar et al., 2019).

Table 1 provides a summary of the key characteristics, advantages, and limitations of traditional data anonymization models applied to data anonymization.

The qualitative assessment “Resilience to background knowledge” in Table 1 is derived from well-established theoretical and empirical results reported in the literature. Specifically, k-anonymity is classified as having low resilience to background knowledge because it protects only against identity disclosure, but remains vulnerable to homogeneity and background knowledge attacks, particularly in medical datasets with skewed distributions of sensitive attributes (Sweeney, 2002). The l-diversity model improves upon k-anonymity by enforcing diversity of sensitive attributes within equivalence classes; however, its resilience is considered medium, as it does not adequately address semantic similarity between sensitive values and remains vulnerable to skewness and similarity attacks (Machanavajjhala et al., 2006).

The t-closeness model demonstrates higher resilience to background knowledge attacks because it constrains the distance between the distribution of sensitive attributes in anonymized groups and the overall dataset distribution, thereby limiting an adversary's ability to exploit auxiliary information (Li et al., 2007). Nevertheless, this increased protection is achieved at the cost of higher information loss and computational complexity.

Differential privacy is categorized as providing absolute resilience to background knowledge attacks, since its semantic privacy guarantee is independent of an adversary's prior knowledge and formally bounds the information gain that can be obtained from the data release Dwork, 2006; Abadi et al., 2016). Similarly, cryptographic approaches such as homomorphic encryption are considered to offer absolute resilience, as data remain encrypted during processing and no plaintext information is revealed; however, their practical applicability is limited by significant computational overhead and scalability constraints (Acar et al., 2018).

With the rapid growth of unstructured PMD, particularly clinical text descriptions, research has shifted to ML-based de-identification methods. Natural language processing (NLP) approaches, including recurrent neural networks and transformer-based models, have demonstrated high accuracy in identifying PMD in clinical texts (Dernoncourt et al., 2017). Compared to rule-based systems, ML methods are more robust to linguistic variability and contextual ambiguity, while reducing the labor costs of manual model maintenance and tuning (Di Cerbo and Trabelsi, 2018). The combination of these advantages makes ML-based de-identification a practical solution for large-scale clinical text processing.

Table 1. Comparison of traditional data anonymization models

Model/Criterion	**Guarantee type**	**Resilience to background knowledge**	**Impact on data usefulness**	**Computational complexity**	**Key drawback**	**Key references**
k-anonymity	Syntactic	Low	High (generalization)	Medium	Vulnerability to homogeneity attacks	(Sweeney, 2002)
l-diversity	Syntactic	Medium	High	Medium	Ignoring semantic closeness of attributes	(Machanavajjhala et al., 2006)
t-closeness	Syntactic	High	Very high	High (EMD calculation)	Complexity of parameterization and calculations	(Li et al., 2007)
Differential privacy	Semantic	Absolute	Controlled reduction (noise)	Low/Medium	Privacy-precision tradeoff, difficulty of setting up ε	(Dwork, 2006; Abadi et al., 2016)
Homomorphic encryption	Cryptographic	Absolute	Zero (for encrypted transactions)	Very high	Impractical for complex calculations on big data	(Acar et al., 2018)

In recent years, AI-based anonymization systems have expanded beyond detection tasks to include adaptive data transformation and synthesis mechanisms. Generative models, particularly generative adversarial networks, have been successfully applied to create synthetic medical data that preserves key statistical characteristics of the original datasets while reducing the risk of re-identification (Jordon et al., 2022). Synthetic data generation is considered a promising approach to ensuring privacy at the design stage, particularly in the context of data sharing and reproducible research in healthcare (Chen et al., 2021).

Furthermore, distributed learning paradigms such as federated learning (FL) have attracted considerable attention as a means of mitigating the privacy risks associated with centralized storage of PMDs. By training models locally and transmitting only aggregated parameters, FL significantly limits direct access to PMDs and can be further strengthened by integrating DP mechanisms (Li et al., 2019).

Despite these advances, recent research highlights the remaining challenges associated with privacy attacks on ML models, including membership inference and model inversion attacks (Shokri et al., 2017). Moreover, increasing regulatory attention to transparency and accountability issues has highlighted the limitations of black-box AI models, stimulating the development of explainable AI approaches for privacy-focused systems (Goodman and Flaxman, 2017).

Table 2 provides a comparative overview of intelligent data anonymization methods based on artificial intelligence, including their applicability areas, advantages, and limitations.

Overall, the literature reviewed in this section demonstrates significant progress in the field of PMD anonymization, but also reveals the fragmentation of statistical, cryptographic, and AI-based approaches. This circumstance highlights the need to develop integrated intelligent anonymization systems that combine adaptive AI mechanisms with formal privacy guarantees and regulatory compliance.

Table 2. Comparison of intelligent anonymization methods based on AI

Method/Aspect	Principle of operation	Key benefits	Main challenges	Application example	Key references
Named entity recognition based on Deep Learning	Semantic understanding and text markup	High accuracy on unstructured data, robust to variations	Dependence on the quality and volume of labeled training data	De-identification of clinical records in EHR	(Lison et al., 2021; Dernoncourt et al., 2017; Di Cerbo and Trabelsi, 2018)
Adaptive anonymization	Dynamic strategy selection based on context	Optimizing the privacy-utility balance for a specific task	Complexity of determining the reward function; computational costs	Personalized anonymization of data streams from IoT devices	(Lison et al., 2021; Fredrikson et al., 2014)
Generative models (GANs/VAEs)	Creating synthetic data that mimics the distribution of real data	Fundamentally eliminate the risk of re-identification	Assessing the adequacy and privacy of synthetic data; GAN collapse mode	Creating open datasets for medical AI competitions	(Yoon et al., 2020; Jordon et al., 2022; Chen et al., 2021; Yale et al., 2020)
DP-SGD	Introducing noise into the model optimization process	Formal mathematical guarantees of privacy	Decreased final accuracy of the model; difficulty in tuning hyperparameters (ε, δ)	Training a diagnostic model on data from multiple hospitals with privacy guarantees	(Abadi et al., 2016; Phong and Phuong, 2023)
Federated learning	Distributed training without sharing raw data	Data remains with the owner; regulatory barriers are reduced	Communication overhead; data heterogeneity between nodes	Joint development of a model for predicting diabetes complications by a consortium of clinics	(Li et al., 2019; Shokri et al., 2017)

3. Material and methods

3.1. Problem Definition

The key challenge addressed in this study is the increasing inadequacy of traditional data anonymization models when dealing with modern, complex, and heterogeneous PMDs. Modern healthcare information systems generate a wide variety of data, including multivariate genomic and multi-omics data, continuous real-time information streams from IoMT devices, EHRs, unstructured clinical records, and medical images. These types of data are characterized by high interdependence of features, temporal correlations, and semantic richness, which increases both the analytical value and the risk of disclosing sensitive information.

Traditional data anonymization approaches fail to provide a satisfactory balance between robust privacy protection and data utility, particularly in modern healthcare environments characterized by high-dimensional and unstructured PMD. In particular, they are poorly applicable to high-dimensional and unstructured PMDs, where over-generalization or suppression leads to significant information loss, reducing the validity of subsequent clinical analytics and ML applications (Rocher et al., 2019; Fredrikson et al., 2014). In addition, traditional models typically rely on limited knowledge of potential attacks and assume a static threat environment, making them vulnerable to sophisticated re-identification attacks using auxiliary information from open sources, including social media, government databases, and commercial data brokers.

In parallel with technical challenges, the regulatory environment governing personal data is becoming increasingly strict, heterogeneous, and fragmented. The GDPR and HIPAA impose overlapping and sometimes conflicting requirements regarding consent, purpose limitation, cross-border data transfers, and de-identification standards. Compliance with these regulations requires not only the use of appropriate anonymization methods but also the ability to document, justify, and demonstrate privacy risk assessments and decisions made to transform personal data.

Taken together, technical and regulatory requirements highlight the need to develop more intelligent, automated, and context-aware approaches to PMD anonymization that can adapt to their characteristics, conditions of use, and evolving threats. Such approaches must ensure scalable processing, maintain analytical and clinical utility, and demonstrate compliance with regulatory requirements. Consequently, the task of PMD anonymization must fundamentally encompass both technical challenges of maintaining utility, legal issues of regulatory compliance, and ethical aspects of protecting patient privacy.

3.2. Problem solution

An intelligent PMD anonymization conceptual framework is proposed, which consists of three stages: (1) intelligent detection and classification, (2) adaptive privacy risk assessment and anonymization strategy selection, and (3) implementation via a secure and explainable pipeline (fig. 1).

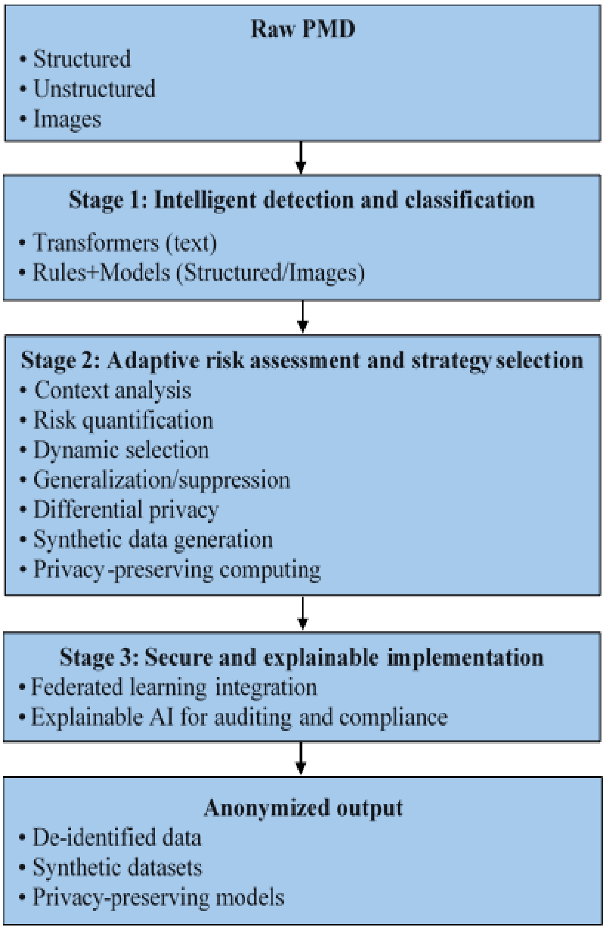
Raw PMD
• Structured
• Unstructured
• Images
Stage 1: Intelligent detection and classification
• Transformers (text)
• Rules+Models (Structured/Images)
Stage 2: Adaptive risk assessment and strategy selection
• Context analysis
• Risk quantification
• Dynamic selection
• Generalization/suppression
• Differential privacy
• Synthetic data generation
• Privacy-preserving computing
Stage 3: Secure and explainable implementation
• Federated learning integration
• Explainable AI for auditing and compliance
Anonymized output
• De-identified data
• Synthetic datasets
• Privacy-preserving models

Fig.1. PMD anonymization conceptual framework

3.1.1. *Intelligent detection and classification*

This stage of the proposed system involves implementing a hybrid detection and classification system designed to identify sensitive and personally identifiable information in various PMD formats. A hybrid system combining ML models and rule-based verification mechanisms enables high detection accuracy while simultaneously ensuring regulatory compliance.

To process unstructured PMD (e.g., clinical text), a modern NLP model based on transformers, pre-trained and fine-tuned on extensive biomedical and clinical corpora such as BioBERT and GatorTron, is applied. Through contextualized representation learning, such models significantly outperform traditional approaches to named entity recognition (NER) based on rules or statistical methods (Lison et al., 2021). In particular, they are capable of detecting not only explicit identifiers (e.g., patient names, addresses, or identification numbers), but also indirect identifiers and quasi-identifiers, the sensitivity of which is determined by semantic context, feature co-occurrence, or specific medical knowledge. Also, such models identify clinically significant and rare concepts, including unique diagnoses or combinations of treatment interventions, which may increase the risk of re-identification even in the absence of direct identifiers (Lison et al., 2021).

For structured data and medical imaging, the system integrates additional detection mechanisms to ensure comprehensive identifier detection. Model-based detectors, including convolutional neural networks, are used to identify sensitive information embedded in images, such as text overlays commonly found in radiology images (Korytkowski et al., 2023). Additionally, structured attributes are analyzed using checklists derived from established regulatory standards, such as the HIPAA Safe Harbor de-identification criteria. These checklists provide a deterministic level of verification, ensuring that all legally relevant identifiers are consistently detected and labeled, regardless of model uncertainty or data sparseness (Benitez and Malin, 2010).

3.1.2. *Adaptive privacy risk assessment and anonymization strategy selection*

In this stage, the adaptive privacy strategy engine evaluates the context of the identified data, its intended use case, and applicable regulatory thresholds. It quantifies the risk of re-identification using metrics such as ε in DP or empirical risk models (Dwork, 2006). Based on this assessment, the system dynamically selects an anonymization strategy from a set of tools.

For low-dimensional and well-structured datasets where a minor loss of analytical utility is acceptable, classic anonymization operations such as generalization and suppression are applied. These methods reduce data granularity by replacing specific attribute values with broader categories or selectively removing high-risk attributes. Such transformations are particularly effective for structured demographic and administrative variables, as they preserve semantic consistency while significantly reducing the risk of re-identification. The choice of generalization hierarchies and suppression thresholds is determined by the assessed privacy risk and the intended analytical purpose, ensuring compliance with regulatory requirements and maintaining the suitability of the data for analysis.

In cases of statistical analysis, summary reporting, or training ML models, where mathematically provable privacy guarantees are required, the proposed framework uses DP mechanisms. By injecting calibrated noise into queries, gradients, or model parameters, the DP ensures that the inclusion or exclusion of any individual record has a limited impact on the results, formally controlled by a privacy parameter ε. The framework provides optimal noise calibration strategies adapted to the PMD distribution, allowing for explicitly balancing privacy protection and analytical accuracy (Yang et al., 2020). These mechanisms are particularly effective for population-level studies and for privacy-preserving model training pipelines.

For tasks requiring high data efficiency, such as exploratory research, algorithm development, or system validation, the platform supports the generation of synthetic data as an advanced alternative to traditional anonymization. Using modern generative models, including tabular diffusion probability models (TabDDPMs) (Kotelnikov et al.,2023) and generative adversarial networks (GANs) (Bae et al.,2020), augmented with DP guarantees, the platform creates synthetic datasets that accurately approximate the statistical properties of the original data. These models preserve complex multivariate dependencies, temporal patterns, and clinically relevant correlations, minimizing the risk of disclosure and preventing the recovery of PMD. As a result, synthetic data can act as a secure and efficient replacement for confidential PMD for research and development (Yale et al., 2020).

In settings where direct data exchange is impossible due to legal, ethical, or organizational constraints, this architecture relies on confidential computing techniques. These include lightweight cryptographic operations, secure aggregation protocols, and FL approaches that enable collaborative analysis of distributed data sources without centralized storage. In such federated environments, the source data remains local to each participant, and exchange occurs only through secure model updates or aggregated statistics, minimizing the risk of leaks and ensuring regulatory compliance. This approach is particularly effective for cross-institutional healthcare research and distributed model training.

3.1.3. Secure and explainable implementation

In this stage, the anonymization strategy selected by the adaptive decision module is implemented through secure modular pipelines that explicitly integrate both privacy mechanisms and explainability components. This approach ensures not only the technical efficiency of the applied transformations but also their transparency, verifiability, and compliance with regulatory and ethical requirements.

ML-based analytical tasks, the platform supports deployment within the feature learning paradigm, enabling collaborative model training across multiple healthcare institutions without centralizing sensitive PMDs. In such a federated environment, anonymization and protection of PMDs are performed locally at each participating node, such as a hospital or clinical repository. To limit potential information leakage at the record level, methods such as differentially confidential stochastic gradient descent (DP-SGD) (Phong and Phuong, 2023) are applied. Only confidential, encrypted, or securely aggregated model parameters are transmitted to the central coordination server, enabling global aggregation without direct access to the original data. This architecture significantly reduces the attack surface, minimizes regulatory risks during data transfer, and supports the principle of data minimization while maintaining high analytical performance of the model.

ML-based anonymization processes, explainable AI (XAI) methods (Arrieta et al., 2020) are integrated into the platform's structure as the primary audit mechanism. Specifically, local interpretable model-independent explanations (LIME) and SHapley additive explanations (SHAP) are used to generate interpretable and human-readable explanations for the PMD anonymization decisions applied. These methods are employed to quantify the contribution of individual attributes, attribute interactions, and data segments to model outputs, thereby enabling the identification of direct identifiers, quasi-identifiers, and latent re-identification vectors that pose an elevated disclosure risk. The resulting explanatory profiles inform the selection and parametrization of targeted anonymization operations, such as generalization, suppression, aggregation, or noise injection, while explicitly justifying the scope and intensity of each transformation. Moreover, LIME- and SHAP-based explanations facilitate a formal assessment of the privacy-utility trade-off by measuring the extent to which anonymization attenuates the influence of sensitive attributes on model behavior while preserving the analytical utility and predictive validity of clinically relevant features. The explanations and associated transformation metadata are systematically recorded as auditable artifacts, documenting sensitivity assessments, anonymization rationales, and post-transformation impacts on both disclosure risk and model performance. These audit trails provide verifiable evidence of compliance with data protection regulations, including the GDPR and HIPAA, while also enhancing institutional transparency, methodological reproducibility, and stakeholder trust in privacy-preserving PMD analytics.

4. Discussion

Despite its obvious prospects, AI-based systems are associated with a number of significant difficulties that need to be overcome.

Thus, the privacy–utility trade-off in PMD anonymization is not a fixed compromise but a context-dependent optimization problem, where stronger privacy guarantees inevitably reduce data granularity, while preserving high analytical and clinical utility requires adaptive, risk-aware anonymization strategies.

The key tradeoff between privacy and utility of PMD remains, but has become more nuanced and multifaceted. Accurately assessing the retention of clinically relevant findings, causal relationships, and model fairness necessitates the development of domain-specific utility metrics that go beyond standard statistical accuracy.

Moreover, the AI models themselves used for anonymization, such as generative models or named entity recognition models, can become targets for attacks. Attacks aimed at forcing a generative model to memorize and reproduce training examples, as well as attacks on data membership inference, pose a serious threat to both anonymization and FL (Shokri et al., 2017; Hayes et al., 2019).

Effective protection requires a comprehensive, multi-layered approach, including protection of PMD, rigorous model auditing, and, where necessary, formal verification of critical components. In this regard, the implementation of XAI methods is crucial: it not only enhances trust in the system but also ensures compliance with regulatory and ethical requirements. Regulators and ethics committees must understand the rationale behind automated anonymization decisions, especially when these decisions impact the analytical and clinical value of the data used for research.

Compliance with regulatory requirements for the processing of personal data is a dynamic and ever-evolving challenge. The distinction between anonymization and pseudonymization, enshrined in the GDPR, creates legal uncertainty, meaning that perfect, irreversible anonymization is often viewed as a theoretical ideal rather than a practically achievable goal. Therefore, AI systems must be designed in accordance with the principles of "privacy by default" and "accountability by default," ensuring documentation of all transformations and providing a quantifiable assessment of the residual risk.

Under HIPAA's "expert determination" approach, an AI-based risk assessment model can act as an expert, but this requires a high level of verification, transparency, and trust from the medical and legal communities. Furthermore, cross-border data exchange also creates complexity, requiring dynamic adjustments to anonymization strategies depending on the recipient's jurisdiction.

Successful implementation of such systems requires close collaboration between data scientists, medical experts, ethicists, and legal experts. Ethical considerations extend beyond privacy protection to include bias analysis. If training data for AI anonymization contains systemic biases, this could lead to inconsistent protection or distorted information across different demographic groups, exacerbating healthcare inequities. Patient-centered approaches, potentially based on dynamic consent models and blockchain technologies (Nugent et al., 2016) can provide users with more detailed control over the anonymization process and the use of their data, ensuring that technical practices comply with ethical principles of autonomy.

5. Conclusion

This article examines the transition from traditional to intelligent approaches to anonymizing personal medical data. Classical statistical and syntactic models, while fundamental, are poorly suited to the complexity of modern medical data. Cryptographic methods and DP provide stronger formal guarantees, but often at the expense of high utility or practicality. Intelligent AI-based anonymization constitutes a necessary evolutionary step, providing context-sensitive, adaptive, and utility-preserving data protection. The proposed framework combines advances in NLP, generative AI, DP, and advanced artificial intelligence into a unified, comprehensive approach.

Key challenges include the vulnerability of AI components to sophisticated attacks, the "black-box" problem, the difficulty of standardizing risk assessments, and a constantly evolving, fragmented regulatory framework. Nevertheless, AI should be regarded as a critical component of a multi-layered security and privacy architecture that ensures protection at all levels of personal data processing.

Future research will be interdisciplinary and should focus on several key areas: 1) Developing hybrid AI models that combine the formal rigor of differential privacy with the adaptability of DL; 2) Creating standardized, clinically validated utility metrics and benchmark datasets to evaluate anonymization methods; 3) Developing privacy-augmented AI to make automated decisions transparent and auditable; 4) Exploring decentralized and patient-centric paradigms, such as combining functional programming with data warehouses. and 5) Promoting international harmonization of regulations and certification of AI-based anonymization tools. By addressing these challenges, intelligent anonymization can become a robust pillar of a secure, ethical, and collaborative healthcare data ecosystem.

Acknowledgments

This work was supported by the Azerbaijan Science Foundation – Grant No. AEF-MCG-2023-1(43)-13/04/1-M-04

References

Abadi, M., Chu, A., Goodfellow, I., McMahan, H.B., Mironov, I., Talwar, K., Zhang, L. (2016). Deep learning with differential privacy. Proceedings of the ACM Conference on Computer and Communications Security, 308–318.

Acar, A., Aksu, H., Uluagac, A.S., Conti, M. (2018). A survey on homomorphic encryption schemes: Theory and implementation. ACM Computing Surveys, 51(4), 1–35.

Aminifar, A., Lamo, Y., Pun, K.I., Rabbi, F. (2019). A practical methodology for anonymization of structured health data. Proceedings of the Scandinavian Conference on Health Informatics, 1–7.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., and Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82-115.

Bae H, Jung D, Choi HS, Yoon S. (2020). AnomiGAN: Generative Adversarial Networks for Anonymizing Private Medical Data. Pac Symp Biocomput., 25, 563-574.

Benitez, K., and Malin, B. (2010). Evaluating re-identification risks with respect to the HIPAA privacy rule. Journal of the American Medical Informatics Association, 17(2), 169-177.

Chen, R.J., Lu, M.Y., Chen, T.Y., Williamson, D.F., Mahmood, F. (2021). Synthetic data in machine learning for medicine and healthcare. Nature Biomedical Engineering, 5(6), 493–497.

Dernoncourt, F., Lee, J.Y., Uzuner, O., Szolovits, P. (2017). De-identification of patient notes with recurrent neural networks. Journal of the American Medical Informatics Association, 24(3), 596–606. (

Di Cerbo, F., and Trabelsi, S. (2018). Towards personal data identification and anonymization using machine learning techniques. ADBIS Short Papers and Workshops, 1–11.

Dwork, C. (2006). Differential privacy. In Automata, Languages and Programming. 1–12. Springer.

European Parliament and Council. (2016). Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Official Journal of the European Union, L119, 1–88.

Fredrikson M, Lantz E, Jha S, Lin S, Page D, Ristenpart T. (2014). Privacy in Pharmacogenetics: An End-to-End Case Study of Personalized Warfarin Dosing. Proc USENIX Secur Symp., 17-32.

Goodman, B., and Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a “right to explanation”. AI Magazine, 38(3), 50–57. (Scopus, Q3/ Web of Science, Q4, in SCIE)

Gupta, N.S., and Kumar, P. (2023). Perspective of artificial intelligence in healthcare data management: A journey towards precision medicine. Computers in Biology and Medicine, 162, 107051.

Hayes, J., Melis, L., Danezis, G., and De Cristofaro, E. (2019). LOGAN: Evaluating privacy leakage of generative models using generative adversarial networks. Proceedings on Privacy Enhancing Technologies, 2019(1), 133-152.

Jordon, J., Szpruch, L., Houssiau, F., Bottarelli, M., Cherubin, G., Maple, C., Weller, A. (2022). Synthetic data – what, why and how? arXiv preprint.

Korytkowski, M., Nowak, J., Scherer, R. (2023). Detecting Sensitive Data with GANs and Fully Convolutional Networks. In: Nguyen, N.T., et al. Intelligent Information and Database Systems. ACIIDS 2023. Lecture Notes in Computer Science, 13995.

Kotelnikov, A., Baranchuk, D., Rubachev, I., Babenko, A. (2023). TabDDPM: Modeling tabular data with diffusion models. Proceedings of the International Conference on Machine Learning, 2023, 17564–17575.

Li, N., Li, T., Venkatasubramanian, S. (2007). t-closeness: Privacy beyond k-anonymity and l-diversity. Proceedings of the International Conference on Data Engineering, 106–115.

Li, W., Milletari, F., Xu, D., Rieke, N., Hancox, J., Zhu, W., Kainz, B. (2019). Privacy-preserving federated brain tumor segmentation. Proceedings of the Workshop on Machine Learning in Medical Imaging, 133–141.

Lison, P., Pilán, I., Sánchez, D., Batet, M., Øvrelid, L. (2021). Anonymization models for text data: State of the art, challenges and future directions. Proceedings of the Annual Meeting of the Association for Computational Linguistics, 4188–4203.

Machanavajjhala, A., Gehrke, J., Kifer, D., Venkitasubramaniam, M. (2006). l-diversity: Privacy beyond k-anonymity. Proceedings of the International Conference on Data Engineering, 24.

Nugent, T., Upton, D., and Cimpoesu, M. (2016). Improving data transparency in clinical trials using blockchain smart contracts. F1000Research, 5, 2541.

Pakhale K. (2023). Comprehensive Overview of Named Entity Recognition: Models, Domain-Specific Applications and Challenges. arXiv:2309.14084

Phong L. T., and Phuong T. T., (2023). Differentially private stochastic gradient descent via compression and memorization. Journal of Systems Architecture, 135, 102819.

Rocher, L., Hendrickx, J.M., de Montjoye, Y.-A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. Nature Communications, 10(1), 3069.

Shokri, R., Stronati, M., Song, C., Shmatikov, V. (2017). Membership inference attacks against machine learning models. IEEE Symposium on Security and Privacy, 3–18.

Sweeney, L. (2002). k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(5), 557–570.

U.S. Department of Health and Human Services. (1996). The Health Insurance Portability and Accountability Act of 1996 (HIPAA). Public Law, 104–191.

Yale, A., Dash, S., Dutta, R., Guyon, I., Pavao, A., and Bennett, K. P. (2020). Generation and evaluation of privacy-preserving synthetic health data. Neurocomputing, 416, 244-255.

Yang, M., Lyu, L., Zhao, J., Zhu, T., and Lam, K. Y. (2020). Local differential privacy and its applications: A comprehensive survey. arXiv preprint arXiv:2008.03686.

Yoon, J., Drumright, L.N., van der Schaar, M. (2020). Anonymization through data synthesis using generative adversarial networks (ADS-GAN). IEEE Journal of Biomedical and Health Informatics, 24(8), 2378–2388.

How to cite: Ramiz Shikhaliyev (2026). Anonymization of personal medical data based on artificial intelligence. Problems of Information Society, 1, 35–43. https://doi.org/10.25045/jpis.v17.i1.04