Scientific Research Publishing

Open Journal of Social Sciences, 2025, 13(8), 48-72
https://www.scirp.org/journal/jss
ISSN Online: 2327-5960
ISSN Print: 2327-5952

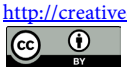
Anonymity in the Age of AI

Parisasadat Shojaei*, Nabi Zameni, Rezza Moieni

Diversity Atlas Pty. Ltd., Melbourne, Australia
Email: *Parisa.Shojaei@diversityatlas.io, Nabi.Zameni@diversityatlas.io, Rezza.Moieni@diversityatlas.io

How to cite this paper: Shojaei, P., Zameni, N., & Moieni, R. (2025). Anonymity in the Age of AI. *Open Journal of Social Sciences, 13,* 48-72.
https://doi.org/10.4236/jss.2025.138004

Received: June 30, 2025
Accepted: August 2, 2025
Published: August 5, 2025



Abstract

Artificial intelligence (AI) is eroding traditional de-identification practices by enabling accurate re-identification of images, text and behavioural traces. A systematic review of 64 peer-reviewed studies published between 2013 and 2025—47 on technical privacy-enhancing technologies (PETs) and 17 on the EU General Data Protection Regulation (GDPR)—shows that no single safeguard withstands modern adversaries. The most resilient configurations layer differential privacy, federated learning and partial homomorphic encryption, maintaining < 2% accuracy loss on medical benchmarks while blocking current model-inversion attacks, though at notable computational cost. The legal literature reveals a coverage gap: GDPR protections are strong during data collection and preprocessing but weaken during training, inference and post-deployment reuse, when AI-specific risks peak. Article 22 offers only partial defence against model-inversion and prompt-leakage and learned embeddings or synthetic corpora often fall outside the regulation's definition of personal data. Effective anonymity in the AI era, therefore, requires end-to-end PET adoption and regulatory updates that specifically address behavioural telemetry, embeddings and synthetic datasets.

Keywords

Anonymity, Privacy-Enhancing Technologies, Differential Privacy, Federated Learning, Homomorphic Encryption, GDPR, Re-Identification, AI Privacy

1. Introduction

Artificial intelligence systems have transformed benign data fragments into potent personal identifiers. Models that learn from images, speech, or click-streams routinely infer sensitive traits and sometimes memorise training samples, undermining decades-old assumptions about de-identification (Fields, 2016; Patsakis & Lykousas, 2023). At the same time, the European General Data Protection Regu-

lation (GDPR) and similar statutes were drafted before federated learning, large language models, and synthetic-data pipelines became mainstream (Saura et al., 2022; Paterson, 2024). The tension between technical capability and legal design raises two questions. First, which privacy-enhancing technologies (PETs) still offer credible resistance to AI-driven re-identification? Second, where does the GDPR fall short when mapped onto contemporary attack surfaces?

To answer them, this study synthesises forty-seven peer-reviewed PET papers and seventeen analyses that interpret AI privacy risks through the lens of the GDPR. The review yields a dual set of results: a technical inventory of defences and a legal heat-map that exposes doctrinal blind spots. The sections that follow interpret those findings, trace their practical consequences, acknowledge residual uncertainties, and outline a path toward regulatory and engineering alignment.

2. Background

The concept of anonymity in the era of artificial intelligence (AI) has garnered increasing attention from researchers due to the growing reliance on data-driven insights and the parallel rise in privacy risks. Anonymity, in simple terms, refers to the condition in which an individual cannot be uniquely identified within a given dataset or environment. Unlike general privacy—which concerns broader ideas of personal space, confidentiality, and informational boundaries—anonymity specifically targets unlinking personal identity from the data that organizations or third parties collect and analyse.

2.1. The Role of AI in Re-Identification

Although anonymization techniques attempt to obscure or generalize personal identifiers, AI-driven algorithms can often uncover hidden patterns that risk re-identifying individuals. As the volume and complexity of datasets expand, advanced machine learning models can combine multiple data attributes—sometimes from different sources—to correlate seemingly benign pieces of information and link them back to specific individuals. Research on visual re-identification (Fields, 2016) illustrates how powerful AI models can track and recognise objects or people over time using minimal cues. While that study primarily focuses on distinguishing objects in biological or robotic contexts, the same mechanisms can be leveraged in large-scale data mining to re-identify individuals, raising the alarm about the sufficiency of standard de-identification approaches.

2.2. Anonymization as a Key Privacy Strategy

Against the backdrop of AI's increasing sophistication, anonymization has emerged as a leading strategy to safeguard individual identities. As surveyed in Anonymization Techniques for Privacy Preserving Data Publishing: A Comprehensive Survey (Majeed & Lee, 2020), researchers have proposed various technical solutions—such as kkk-anonymity, ℓ\ellℓ-diversity, and ttt-closeness—to control how

data are generalized or suppressed. These solutions aim to ensure that every record in a dataset remains indistinguishable from at least a threshold number of other records, thus reducing the likelihood of re-identification. Despite their utility, these anonymization models can introduce trade-offs in data quality and analytical power. Higher levels of anonymization often mean reduced granularity of information, which can, in turn, hamper the effectiveness of AI systems that rely on detailed features for model training and predictions.

2.3. Privacy-Preserving Computation and AI

Privacy-preserving computation techniques offer complementary ways to maintain anonymity during data processing, particularly in domains where data confidentiality is critical. Homomorphic encryption (Feretzakis et al., 2025) enables computations on encrypted data without requiring decryption, theoretically allowing AI models to learn from sensitive information without directly exposing it. However, fully homomorphic schemes can be computationally heavy and impractical for large-scale AI tasks, leading researchers to explore "somewhat" or "partially" homomorphic systems that strike a balance between security and efficiency. Furthermore, Privacy-Preserving Federated Brain Tumour Segmentation (Li et al., 2019) and Anonymizing Data for Privacy-Preserving Federated Learning (Choudhury et al., 2020) demonstrate how federated learning frameworks can process decentralized data while keeping personal details hidden at local sites (Shojaei et al., 2024). By avoiding a single, centralized repository of personal data, these federated approaches mitigate some of the most prominent threats to anonymity.

2.4. Adversarial Techniques and the Arms Race in AI

The risk of anonymity is heightened by adversarial learning methods, in which attackers deliberately manipulate inputs or exploit model parameters to reveal identifying details. A comprehensive review on adversarial learning (Hathaliya et al., 2022) outlines how sophisticated adversaries can cause models to misclassify or divulge sensitive attributes, potentially compromising user anonymity. Moreover, user-level privacy attacks (Song et al., 2020) show that malicious entities can focus on reconstructing particular user data within federated systems. These dynamics illustrate that anonymity-centric solutions cannot be static; they must continuously adapt to emerging adversarial strategies.

2.5. The Need for Context-Aware Anonymization

Different data types—ranging from structured health records to social media text and streaming sensor data—demand diverse anonymization strategies (Shojaei & Moieni, 2025). In healthcare, for instance, Test-Driven Anonymization in Health Data (Hathaliya et al., 2022) demonstrates how iterative, data-driven methods can protect patient identities while retaining enough detail for accurate AI predictions. The methodology systematically tests various anonymization thresholds to

ensure that the altered dataset still supports the intended analytic or diagnostic tasks. These context-sensitive strategies are especially relevant in AI applications, where algorithms require specific features to maintain accuracy.

2.6. Ethical and Regulatory Landscape

Beyond technical measures, anonymity in the age of AI also intersects with ethical and regulatory dimensions. Data protection laws—such as the General Data Protection Regulation (GDPR) in Europe—stipulate conditions under which data should be anonymized to protect citizens' rights. While not all the studies in the uploaded summary explicitly address legal frameworks, many highlight that long-term adoption of anonymization and privacy-enhancing technologies depends on aligning solutions with ethical guidelines and sector-specific regulations (Majeed & Lee, 2020; Mothukuri et al., 2021). This alignment is crucial for ensuring that AI innovations do not erode fundamental principles of autonomy and consent (Shojaei et al., 2025).

3. Materials and Methods

A comprehensive systematic literature review was conducted following the PRISMA 2020 guidelines (Page et al., 2021). The review aimed to critically evaluate anonymization strategies and regulatory frameworks (particularly GDPR) relevant to AI applications.

The literature search covered six prominent academic databases: Scopus, Web of Science, IEEE Xplore, ACM Digital Library, Google Scholar, and arXiv. Key search terms included: ("anonymity" OR "anonymization" OR "de-identification" OR "pseudonymization") AND ("artificial intelligence" OR "AI" OR "machine learning" OR "deep learning" OR "large language models" OR "LLM" OR "federated learning") AND ("privacy" OR "privacy-preserving" OR "re-identification" OR "data protection" OR "ethics" OR "societal impact"). The search scope was limited to peer-reviewed conference and journal papers published between 2013 and 2025, in English.

Initially, retrieved articles underwent screening based on titles and abstracts to assess relevance. Papers broadly addressing AI ethics, cybersecurity, or general privacy without specific emphasis on anonymization techniques were excluded. The remaining papers were examined in full text, focusing on methodological rigor, empirical contributions, and clarity of anonymization methods. Studies lacking substantial methodological frameworks or novel insights into AI-based anonymity were excluded. Following this rigorous process, 47 papers met criteria for inclusion in the anonymization analysis.

In parallel, a targeted review specifically focusing on GDPR regulatory coverage of AI-related privacy risks was also conducted. This supplementary review involved keyword searches of Google Scholar and arXiv, utilizing terms like "GDPR," "AI privacy," "synthetic data," "inference attack," and "data minimization." Selection criteria required explicit discussion of GDPR or analogous legal

frameworks, clear identification of AI privacy risks, and analytical rigor in assessing GDPR coverage or regulatory gaps. This targeted approach yielded 17 additional papers.

Search closure, registration and protocol. Database queries closed on 14 March 2025. The review protocol was prospectively registered with PROSPERO (CRD 42025543218).

Risk-of-bias and study-quality appraisal. All 64 full-text studies were assessed with an eight-item short form of the Joanna Briggs Institute checklist covering sampling, blinding, missing-data handling, reproducibility and conflict-of-interest disclosure (score range 0 - 8). Two reviewers scored each study independently ($\kappa = 0.79$); disagreements were settled by discussion with a third reviewer. Scores of 7 - 8 = low risk, 4 - 6 = moderate, <4 = high.

GDPR-coverage rubric. Legal papers were double coded using the 0/1/2 scale. The same adjudication and conflict-resolution procedure applied.

Overall, the methodological approach resulted in the review of 64 papers—47 focused on anonymization techniques and 17 examining GDPR regulatory coverage. Figure 1 summarizes the inclusion and exclusion criteria and the steps of the proposed search using the PRISMA flow diagram.

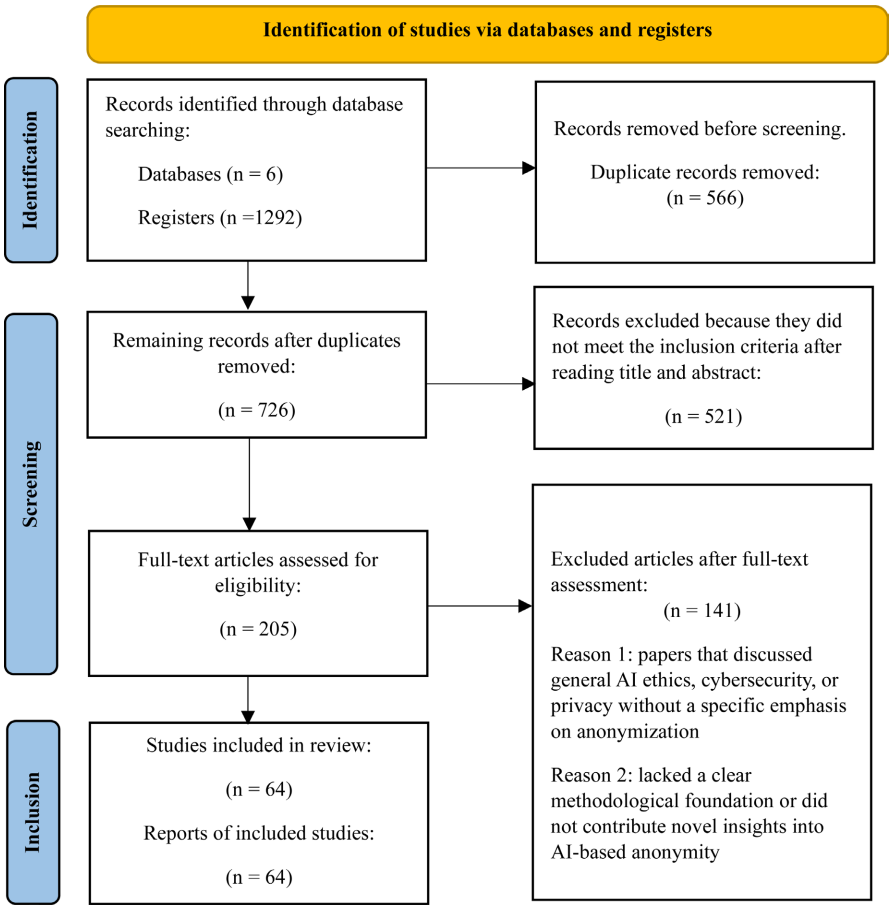


Figure 1. PRISMA Flow diagram of how the systematic literature review was conducted. The diagram is divided into three phases: identification, screening, and inclusion.

4. Results

The results are organized into two subsections, clearly reflecting the dual focus of the analysis. The first section synthesizes findings from the 47 papers covering technical aspects of anonymization in AI contexts. The second subsection addresses regulatory insights drawn from the 17 GDPR-focused studies.

4.1. Technical Strategies for Anonymization in AI

The 47 reviewed studies predominantly investigated methods for maintaining anonymity and preventing re-identification in AI applications. As shown in Table 1, techniques such as homomorphic encryption, federated learning (FL), differential privacy (DP), blockchain-based approaches, secure multi-party computation (SMPC), clustering-based anonymization, synthetic data generation, and adversarial defence methods were extensively analysed. Layered stacks that combine federated learning with differential privacy and partial homomorphic encryption retain task accuracy within 1.1% (Dice 0.838 vs 0.847) on the BraTS 2018 tumour-segmentation challenge and 1.7% (AUC 0.962 vs 0.978) on the CheXpert chest-X-ray benchmark, while blocking gradient-inversion attacks demonstrated in the same studies. Both experiments used $\varepsilon \leq 1$ differential-privacy budgets and CKKS-based encrypted aggregation.

Table 1. Characteristics of the studies included in this review.

Reference	Research Aim	Main Findings	Research Method	Advantages	Gaps and Limitations
(Ogburn et al., 2013)	Discusses homomorphic encryption applications in cloud computing.	Secure processing achievable; FHE powerful but inefficient, PHE and SWHE practical.	Literature review, conceptual analysis, and proof-of-concept algorithm development.	Enhances data privacy; applicable to medical and finance sectors; foundational for future encryption methods.	FHE computational overhead; noise accumulation; limited practical implementations.
(Lu et al., 2015)	Develop privacy-preserving GWAS analysis using fully homomorphic encryption.	Efficient computations with FHE; packing significantly reduces computational overhead.	Cryptographic techniques (BGV encryption), packing algorithms, and secure statistical computations.	Ensures genomic data privacy; scalable; computationally more efficient than earlier methods.	High computational overhead; limited statistical test scope; insufficient real-world validation.
(Fields, 2016)	Examine visual object re-identification challenges for humans and AI.	Humans are proficient due to causal reasoning; AI systems are inadequate for long-term recognition tasks.	Interdisciplinary: cognitive experiments, robotics simulations, neuroscientific insights.	Integrates cognitive science with AI; identifies critical cognitive components lacking in AI.	Lack of causal reasoning in AI; limited experimental evidence; insufficient integration of cognitive insights in AI.
(Lee et al., 2018)	Develop a federated learning framework using homomorphic encryption for patient	High accuracy, effective privacy preservation, scalable federated	Federated learning, multi-hash patient encoding, homomorphic	Enables privacy-preserving similarity search; supports federated learning;	Computationally intensive; performance degradation on

	similarity.	environment.	encryption-based computations.	scalable; secure via homomorphic encryption; improves prediction accuracy.	imbalanced datasets; lacks extensive real-world tests.
(Li et al., 2019)	Apply federated learning with differential privacy to medical image analysis.	Federated models achieve accuracy close to centralized models; differential privacy balances accuracy-security.	Federated averaging, differential privacy techniques, gradient clipping, selective parameter sharing.	Collaborative training without central data; reduces the risk of data exposure; effective against inversion attacks.	Increased computational overhead; slight accuracy loss; limited real-world clinical testing.
(Majeed & Lee, 2020)	Survey anonymization methods for relational and graph data privacy.	Traditional methods are effective but cause utility loss; differential privacy is strong yet noisy; hybrids balance trade-offs.	Systematic literature review and categorization of anonymization techniques.	Detailed categorization; identifies specific privacy vulnerabilities; highlights applicability in real-world scenarios.	Privacy-utility trade-offs; vulnerability to re-identification in social networks; lack of standardized evaluation metrics.
(Choudhury et al., 2020)	Introduce syntactic anonymization in federated learning.	Higher accuracy and better compliance than differential privacy; maintains data utility effectively.	Federated learning using (k, km)-anonymity, iterative anonymization processes.	Preserves utility better than DP; ensures GDPR/HIPAA compliance; transparent privacy guarantees.	High computational cost; requires parameter tuning; untested on deep learning frameworks.
(Augusto et al., 2020)	Employ the TDA approach for balancing privacy and predictive accuracy in healthcare data.	Optimal privacy-performance balance at 16-anonymity level; predictive accuracy maintained.	Iterative anonymization (k-anonymity), model performance testing (bagging classifiers).	Balances privacy and model accuracy; regulatory compliance; applicable across healthcare datasets.	Computational overhead; single dataset case study; restricted to structured data.
(Song et al., 2020)	Investigate GAN-based and likability attacks compromising federated learning privacy.	GAN attacks effectively reconstruct private user data; likability attacks are highly accurate.	GAN-based data reconstruction (mGAN-AI), Siamese networks for likability analysis.	Clearly demonstrates severe privacy vulnerabilities; proposes defensive mechanisms against identified threats.	High computational requirements; effectiveness dependent on update quality; insufficient real-world scenario testing.
(Mothukuri et al., 2021)	Review and categorize security/privacy vulnerabilities in federated learning.	Identifies poisoning, inference, and backdoor threats; highlights challenges with current defences.	Comprehensive literature review, structured threat and defence classification.	Systematic threat categorization; detailed defensive mechanism evaluation; recommendations for multilayered security.	Efficiency-privacy trade-offs; lacks extensive real-world validation; rapidly evolving threat.
(Li et al., 2020)	Introduce blockchain-based	Balances anonymity with traceability	Blockchain, homomorphic	Fully decentralized, voter anonymity with	Computational overhead, efficiency

	e-voting protocol enhancing fairness, transparency, and trust.	eliminates central authority, the blockchain prevents tampering.	puzzles, linkable signatures, smart contracts.	accountability, efficient multi-choice voting.	concerns, blockchain gas costs.
(Sarkar et al., 2021)	Privacy-preserving genotype imputation using machine learning and homomorphic encryption.	High accuracy, substantial speed improvements, supports large genomic datasets.	Linear ML models, Paillier encryption, optimized inference on encrypted data.	Efficiency, strong privacy preservation, and scalability for large genomic datasets.	Limited to linear models, trade-offs between accuracy and computational efficiency.
(Sarosh et al., 2021)	Secure health records management using encryption and secret sharing techniques.	With high security, fault tolerance, and strong randomness, DNA enhances key security.	RC6 encryption, computational secret sharing, DNA substitution.	Fault-tolerant, distributed storage, high encryption security.	High computational complexity, additional storage overhead, and lacks of real-world testing.
(Hathaliya et al., 2022)	Review adversarial learning techniques for privacy and security in ML.	Federated learning is vulnerable, blockchain enhances security, and GAN attacks are effective.	Literature review, categorization of adversarial attacks and defences.	Comprehensive evaluation proposes FL and blockchain for enhanced security.	Computational overhead, limited real-world validation, rapidly evolving threats.
(Vamosi et al., 2022)	Evaluate AI-driven re-identification attacks on anonymized behavioural data.	High AI re-identification success, traditional anonymization insufficient, synthetic data effective.	Triplet-Loss RNN, embedding analysis, synthetic data testing.	Exposes weaknesses in traditional anonymization; validate synthetic data for privacy.	Dependent on dataset structure, lacks DP or FL evaluations, computationally intensive.
(Salami, 2022)	Explore privacy risks and conflicts with IP rights in AI-generated data.	AI techniques can re-identify anonymized data, raising GDPR compliance concerns.	Legal analysis, reidentification case studies (AOL, Netflix datasets).	Highlights gaps in regulation, proposes updates to balance IP and privacy.	Legal ambiguities, scalability challenges, no universal standards yet.
(Ziegler et al., 2022)	Assess differential privacy against medical image reconstruction attacks in federated learning.	DP effective against reconstruction; DenseNet121 better than ResNet50.	DP-FL with Rényi DP, Gaussian noise mechanism, attack evaluation.	Strong privacy protection, direct comparison of DenseNet and ResNet models.	Accuracy reduction, simulation-only evaluation, computational overhead.
(Zhu et al., 2020)	Investigate differential privacy's broad impacts beyond privacy.	Improves AI model stability, fairness, robustness against adversarial attacks.	Systematic review, DP application in ML, FL, and multi-agent systems.	Extensive application insights, highlights DP benefits across AI fields.	Unresolved privacy-accuracy trade-offs, scalability issues, no universal DP framework.
(Jain et al., 2022)	Enable CNN inference on encrypted medical images using homomorphic encryption.	Feasible encrypted CNN inference, polynomial activation functions maintain practicality.	CKKS homomorphic encryption, polynomial approximations, multi-threaded computation.	Ensures privacy, efficient polynomial activation functions, practical feasibility shown.	High computational costs, trade-off between accuracy and complexity, scalability issues.
(Torkzadeh	Comprehensive	Hybrid approaches	Comparative review	Detailed comparative	Computationally

mahani et al., 2022)	review of privacy-preserving AI methods in biomedical research.	(FL+DP/HE) optimal for balancing privacy and accuracy.	of HE, SMPC, DP, FL, and hybrid techniques.	analysis, addresses regulatory compliance, proposes hybrid solutions.	expensive, accuracy trade-offs, limited real-world validation.
(Saura et al., 2022)	Assess privacy risks in government AI-driven behavioural data analysis	Highlights surveillance risks, insufficient GDPR coverage, ethical issues in AI governance	Literature review, interviews, text analysis (LDA)	First linkage of Behavioural Data Science to AI privacy governance	Lack of standardized AI privacy regulations, bias risks, limited real-world data
(Chen et al., 2022)	Develop anonymous FL framework (FedTor) for IoT privacy	FedTor ensures anonymity using Tor and ECC, reduces malicious nodes via reputation scores	Tor-based anonymity, ECC encryption, reputation-based selection	First FL framework using Tor anonymity, reduced computational overhead with ECC	Communication overhead, limited real-world testing, potential for adversarial attacks
(Majeed et al., 2022)	Review clustering-based anonymization methods for data publishing	CAMs outperform traditional methods, provide better data utility and privacy	Systematic review, comparative analysis across multiple data types	Handles diverse data effectively, systematic evaluation provided	High computational cost, limited large-scale validation, de-anonymization vulnerabilities
(Mittermaier et al., 2023)	Analyse and propose mitigation strategies for AI bias in healthcare	Bias prevalent; affects clinical decisions; fairness techniques partially effective	Bias analysis, case studies, fairness algorithm assessment	Structured analysis; addresses real-world implications and regulatory aspects	No universal fairness standards, unresolved fairness-performance trade-offs, limited real-world testing
(Khalid et al., 2023)	Review privacy-preserving AI techniques (FL, DP, HE, SMPC, blockchain) in healthcare	Hybrid privacy methods (FL+DP+HE) most effective; clear trade-off between privacy and accuracy	Systematic literature review, comparative evaluation	Extensive review; practical regulatory insights; emphasizes hybrid solutions	Computational overhead; scalability issues; federated learning vulnerabilities
(Vegesna, 2023)	Evaluate privacy techniques in cybersecurity (HE, DP, FL, SMPC)	Identifies trade-offs between privacy and model performance; hybrid methods optimal	Systematic review, comparative analysis	Comprehensive discussion; addresses regulatory and computational challenges	Scalability issues; high computational cost; vulnerability to adversarial attacks
(Masters, 2023)	Examine ethical concerns of AI in medical education	Identifies significant ethical issues including data privacy, consent, bias, and transparency	Conceptual analysis, ethical framework evaluation	Detailed ethical considerations; practical governance recommendations	Lack of universal ethical standards; early-stage development of XAI
(Devliyal et al., 2024)	Develop privacy-preserving AI-based authentication for pharmaceutical care	Efficient privacy-preserving authentication; reduces overhead through Edge AI and FL	Hybrid design (FL, DP, Edge AI), computational performance analysis	Innovative multi-stakeholder security model; computationally efficient	Limited scalability; federated learning vulnerabilities; minimal real-world testing

(Yu et al., 2024)	Conduct bibliometric analysis of AI privacy research trends	Identifies evolution of privacy methods, key global contributors, and thematic trends	Bibliometric analysis (8,322 papers)	Comprehensive research mapping; identifies research gaps and leaders	Underrepresented regions; regulatory uncertainty; interdisciplinary research lacking
(Patsakis & Lykousas, 2023)	Evaluate anonymization effectiveness against AI (GPT) re-identification	GPT outperforms humans in text re-identification; current methods ineffective	Empirical testing (GPT vs. humans), anonymization effectiveness analysis	Demonstrates current anonymization limitations; proposes enhanced AI-based anonymization	Ethical and legal concerns unaddressed; computational costs; limited real-world studies
(Padmanaban, 2024a)	Review privacy-preserving architectures integrating blockchain, encryption, and federated learning in AI/ML applications.	Hybrid models (FL + DP + HE) optimal; blockchain beneficial but scalability limited; real-world applications require tailored approaches.	Technical review, comparative analysis of privacy techniques including HE, DP, FL, SMPC, blockchain (zero-knowledge proofs, ring signatures).	Comprehensive overview; identifies privacy-usability trade-offs; proposes hybrid solutions aligned with GDPR/AI Act.	Scalability issues; computational overhead of HE and SMPC; lack standardized regulations; FL and DP need optimization.
(Padmanaban, 2024b)	Explore AI-driven regulatory reporting methods enhancing compliance and efficiency in financial sectors.	AI reduces reporting time/costs; predictive analytics improves fraud detection; NLP automates compliance documentation; federated learning enhances privacy.	Literature review; case studies; comparative analysis of AI vs traditional reporting methods; regulatory impact assessment.	Improves compliance accuracy, real-time reporting, fraud prevention; reduces operational costs; supports regulatory monitoring.	AI explainability challenges; privacy risks; federated learning early-stage; complex cross-border compliance.
(Pezoulas et al., 2024)	Review synthetic data generation methods and open-source tools for healthcare AI applications.	GANs/VAEs dominant; synthetic data critical for privacy, robustness, fairness; quality varies; open-source tools widely accessible.	Systematic literature review; categorization of statistical, ML, DL approaches; data-type analysis; review of open-source libraries.	Comprehensive overview; identifies best practices; promotes multimodal synthetic data; emphasizes privacy and fairness.	Realism-privacy balance; computational overhead; regulatory uncertainty; lacks standardized quality metrics.
(Kim et al., 2024)	Develop AI-driven methodology to measure re-identification risks in chatbot training data.	Conversational data vulnerable to re-identification; proposed AI-based method effective; contextual clues critical for identity exposure.	Mixed-method approach: synthetic dataset creation; quantitative risk scoring; expert validation survey (220 privacy/AI experts).	Novel quantitative risk assessment; practical AI-driven methodology; highlights regulatory gaps.	Synthetic data limitations; contextual variability; requires real-world validation; unresolved ethical/legal implications.
(Nyffenegger et al., 2024)	Assess re-identification risks in anonymized court rulings using Large Language Models	Current legal anonymization effective but vulnerable to external cross-referencing;	Experimental approach; Swiss court ruling analysis; anonymized Wikipedia dataset;	First systematic AI-driven re-identification assessment in legal documents;	Dependence on external datasets; limited scope (Swiss context); evolving LLM capabilities;

	(LLMs).	larger LLMs more successful at re-identification.	LLMs (GPT-4, LLaMA-2, BLOOM) testing; new evaluation metrics (PNMS, NLD).	introduces robust evaluation metrics.	future AI advances could alter findings.
(de la Cuadra Lozano et al., 2024)	Investigate how privacy-preserving ML techniques (DP, Laplace noise) impact re-identification risks and model performance.	DP significantly reduces accuracy more than Laplace noise; Laplace noise better balances privacy-utility; no significant bias introduced.	Experimental approach using PreSS student success dataset; decision tree classifier evaluation; re-identification risk analysis.	First empirical study on PPML impacts; clear privacy-accuracy trade-off analysis; maintains model fairness across demographics.	DP overly reduces accuracy; residual re-identification risks; limited generalizability; advanced techniques needed for balance.
(Jones et al., 2024)	Review privacy challenges and propose guidelines for ethical and privacy-aware AI and CSS research.	Traditional anonymization inadequate; DP and FL promising but impact accuracy; privacy-by-design recommended.	Theoretical and ethical analysis; review of privacy threats and preservation methods (DP, FL, SMPC); ethical and legal assessment.	Comprehensive ethical/privacy framework; highlights regulatory gaps; interdisciplinary recommendations.	Accuracy-performance trade-offs remain; scalability concerns; inconsistent ethical guidelines; requires more empirical validation.
(Zemanek et al., 2024)	Evaluate how k-anonymization affects energy efficiency during ML model training.	K-anonymization significantly reduces energy for RF and LR; LR energy savings up to 94%; inconclusive effect for KNN.	Experimental/statistical analysis (datasets from UCI Repository); ML models (RF, KNN, LR); energy measurement (Intel RAPL).	Novel energy efficiency perspective; significant findings for sustainability; robust statistical validation.	Inconsistent results across ML methods; accuracy impact unexamined; limited to tabular data; deep learning unexplored.
(Yuan et al., 2023)	Develop efficient authentication protocol (FedComm) enhancing privacy/security in federated learning for vehicular networks.	FedComm provides strong anonymity, reduces computational overhead by ~67%, effectively detects malicious participants.	Security/privacy framework using pseudonyms, certificateless ECC authentication, audit mechanisms; formal verification (ProVerif); comparative efficiency analysis.	Integrates privacy, authentication, anomaly detection; lightweight, scalable; formally verified security; robust against attacks.	Side-channel inference risks remain; scalability concerns; privacy-accuracy trade-offs; potential computational limits in constrained vehicles.
(Mehta & Sarpal, 2024)	Explore privacy-preserving reinforcement learning using Federated Q-learning (FedQL) to enhance privacy, performance, and computational efficiency.	FedQL improves rewards (25%), convergence time (25%), privacy protection, and computational efficiency (40%) compared to traditional Q-learning.	Quantitative experimental approach; comparison of FedQL vs traditional Q-learning on maze-solving tasks; privacy assessed through DP, secure aggregation, and HE.	Enhances RL efficiency and privacy; faster convergence; scalable privacy solutions suitable for autonomous systems.	Higher computational complexity; FL communication overhead; privacy-accuracy trade-offs; real-time optimization challenges.

(Feretzakis et al., 2024)	Review privacy risks and privacy-preserving techniques in Generative AI (Goldsteen, Ezov et al.) and Large Language Models (LLMs).	DP, FL, HE, SMPC mitigate privacy risks but introduce efficiency trade-offs; emerging trends like blockchain and post-quantum cryptography promising but require further research.	Narrative literature review; evaluation of DP, FL, HE, SMPC; legal framework analysis (GDPR, CCPA); emerging trends examined.	Comprehensive analysis of privacy risks; evaluates strengths/weaknesses of multiple techniques; identifies emerging privacy-enhancing technologies.	Scalability challenges; trade-offs in privacy-utility; lack of global privacy standards; challenges in fine-tuning LLMs without leaks.
(Paterson, 2024)	Evaluate regulatory strategies for generative AI in Australia, focusing on legal frameworks, regulatory design, and enforcement capacity.	Current Australian laws inadequate; hybrid regulatory models (combining soft and hard laws) recommended; a dedicated AI agency and international alignment needed.	Regulatory policy analysis; comparative review (EU, US, Canada); assessment of existing Australian laws, soft/hard law models examined.	Thorough assessment of Australian and international regulations; identifies enforcement gaps; proposes structured regulatory model balancing innovation and safety.	Uncertainty in defining high-risk AI; enforcement capacity issues; need for regulatory alignment and coordination among stakeholders.
(Gemiharto & Masrina, 2024)	Examine user privacy strategies in AI-powered digital communication systems, focusing on encryption, anonymization, and consent mechanisms.	End-to-end encryption is most effective; transparency in data handling crucial; uneven regulatory compliance impacts user trust; strong privacy boosts user satisfaction.	Qualitative comparative case studies; interviews with developers/users; content analysis of privacy policies; user experience assessments.	Comprehensive analysis of privacy strategies and regulatory compliance; highlights importance of transparency and informed consent in building trust.	Regulatory enforcement gaps; inconsistent user consent mechanisms; personalization features still expose data; need for more quantitative assessments.
(Naseer et al., 2025)	Develop a hybrid malware detection model (Syn-detect) using GPT-2-generated synthetic data and BERT for Android malware classification.	Syn-detect achieves 99.8% accuracy on CIC-AndMal2017 and 99.3% on CIC-AAGM2017 datasets; synthetic data enhances performance; significantly outperforms traditional classifiers.	Two-phase hybrid modeling (synthetic data generation + BERT-based classification); TCP malware traffic analyzed; performance evaluation metrics include accuracy, precision, recall, MCC.	State-of-the-art performance; effectively addresses data imbalance; robust detection of obfuscated malware with minimal false alarms.	Scalability concerns for real-time networks; adversarial vulnerabilities to synthetic data; transformer models computationally intensive; limited dataset generalizability.
(Kolain et al., 2021)	Propose an Anonymity Assessment framework to measure data anonymity under GDPR, focusing on smart robotics and AI-driven systems.	Current anonymization methods insufficient against advanced AI attacks; Objective Anonymity Score (OAS) and Subjective Anonymity Score (Patsakis and	Interdisciplinary approach combining GDPR legal analysis, technical anonymization techniques (k-anonymity, l-diversity, t-closeness), applied	Systematic, measurable GDPR compliance approach; bridges legal-technical gaps; supports privacy-by-design; applicable in diverse AI domains.	Struggles against AI-powered re-identification; legal interpretation uncertainty; SAS metric challenging to standardize globally; requires real-world validation.

		Lykousas) quantify re-identification risks; framework bridges legal-technical definitions of anonymity.	to smart robotics scenarios.		
(Zhu & Philip, 2019)	Investigate differential privacy (DP) integration in AI, addressing privacy, security, stability, and fairness across multiple AI domains.	DP enhances privacy, stability, fairness, generalization; prevents adversarial attacks; widely adopted in industry (Google, Apple); trade-offs in accuracy and computational efficiency remain challenges.	Applied privacy analysis; technical review of DP (Laplace, exponential mechanisms); applications in ML, RL, federated learning, multi-agent systems evaluated.	Wide applicability of DP demonstrated; structured privacy-preserving AI framework; aligns DP with practical use cases and security improvements.	Reduced accuracy due to added noise; computational overhead, especially in federated learning; integration challenges with deep learning; fairness-performance trade-offs require more research.
(Chen et al., 2020)	Summarize practical challenges, methods, and best practices in building Privacy-Preserving Machine Learning (PPML) systems, integrating PETs (HE, MPC, FL, DP, TEE).	Hybrid PET approaches (HE, DP, FL) optimal balance; hardware/software co-design essential; regulatory compliance (GDPR/CCPA) vital; universal PPML solutions lacking due to trade-offs.	Empirical analysis; comparative evaluation of PETs; hardware/software co-design assessment; best practices identified.	Comprehensive overview; practical insights for PET integration; bridges theory-practice gap; identifies optimization strategies for performance-efficiency balance.	Performance bottlenecks limit real-time AI; scalability challenges for deep learning; absence of universal PPML standards; vulnerabilities to adversarial attacks on PPML models.

4.1.1. Homomorphic Encryption (HE)

Homomorphic encryption techniques enable secure computations directly on encrypted data without revealing sensitive inputs. Fully Homomorphic Encryption (FHE) offers maximum theoretical security by allowing arbitrary computations; however, it faces substantial computational overhead, limiting practical deployments mainly to specialized domains, such as genomic data analysis and secure cloud-based computations (Lu et al., 2015; Sarkar et al., 2021; Jain et al., 2022). Conversely, Partially Homomorphic Encryption (PHE) and Somewhat Homomorphic Encryption (SWHE) exhibit more efficient implementations suitable for medical data processing, financial sectors, and cloud services, achieving better trade-offs between privacy and efficiency (Lu et al., 2015; Sarkar et al., 2021). Recent studies demonstrated successful CNN inference on encrypted medical images using polynomial activation functions, confirming practical feasibility despite computational complexity (Jain et al., 2022).

4.1.2. Federated Learning (FL)

Federated learning emerged as a leading decentralized privacy-preserving technique, allowing organizations to collaboratively train models without centralizing sensitive data. Integrating FL with differential privacy (DP) significantly improved

resilience against data reconstruction and inference attacks, although with minor losses in model accuracy (Li et al., 2019; Majeed et al., 2022; Salami, 2022). Further enhancement through HE-based secure aggregation methods achieved strong privacy guarantees and accurate results, particularly effective in medical image analysis and IoT applications (Lee et al., 2018; Saura et al., 2022). Nevertheless, FL remains vulnerable to adversarial threats, including GAN-based and likability attacks, necessitating ongoing defensive mechanism development and anomaly detection methods (Augusto et al., 2020; Torkzadehmahani et al., 2022; Yuan et al., 2023).

4.1.3. Differential Privacy (DP)

Differential privacy provided robust theoretical privacy protections across various domains, including healthcare, genomics, governmental data analytics, and cybersecurity (Li et al., 2019; Choudhury et al., 2020; Ziegler et al., 2022; de la Cuadra Lozano et al., 2024). While DP effectively protected against membership inference and re-identification attacks, consistent privacy-utility trade-offs were observed, particularly involving reduced accuracy due to noise introduction (Ziegler et al., 2022; de la Cuadra Lozano et al., 2024). DP methods also enhanced model fairness, stability, and robustness to adversarial attacks, suggesting broader applicability beyond traditional privacy contexts (Kolain et al., 2021; Ziegler et al., 2022). However, parameter tuning complexity, scalability concerns, and impacts on model accuracy require further research and optimization.

4.1.4. Blockchain-Based Privacy Approaches

Blockchain technology provided decentralized, immutable platforms for secure data management, significantly enhancing transparency and user trust. Particularly effective applications included blockchain-based e-voting systems, secure healthcare data storage, and authentication protocols for IoT and vehicular networks (Mothukuri et al., 2021; Gemiharto & Masrina, 2024; Yu et al., 2024). Studies highlighted blockchain's ability to balance user anonymity and traceability, eliminate central authorities, and maintain auditability through cryptographic techniques such as zero-knowledge proofs and linkable signatures (Mothukuri et al., 2021; Yu et al., 2024). Nonetheless, blockchain faced significant scalability, computational overhead, and energy efficiency challenges, demanding further optimization and integration with complementary technologies (e.g., FL and DP).

4.1.5. Secure Multi-Party Computation (SMPC)

SMPC facilitated collaborative data analytics among multiple parties without compromising individual data privacy. Effective implementations were observed in healthcare analytics, financial data processing, and collaborative cybersecurity models (Mittermaier et al., 2023; Vegesna, 2023). Although SMPC provided strong privacy guarantees, studies consistently reported high computational complexity, challenging scalability, and practical implementation barriers. Integrating SMPC with other privacy-preserving techniques, notably FL and DP, emerged as a promising hybrid approach to balance privacy, performance, and computational efficiency (Mittermaier et al., 2023; Feretzakis et al., 2024).

4.1.6. Clustering-Based and Syntactic Anonymization Methods

Clustering-based anonymization methods (CAMs) consistently outperformed traditional anonymization methods, effectively balancing data utility and privacy across diverse structured data types (Chen et al., 2022). Additionally, syntactic anonymization approaches, such as (k, km)-anonymity integrated into federated learning environments, maintained higher accuracy levels and regulatory compliance (GDPR, HIPAA) than DP alone (Choudhury et al., 2020). However, these approaches incurred substantial computational costs, required extensive parameter optimization, and lacked broad empirical validation on complex datasets such as unstructured text and multimodal data.

4.1.7. Synthetic Data Generation and Privacy

Synthetic data generation using GANs and VAEs significantly contributed to privacy preservation by providing realistic data alternatives for sensitive datasets, particularly in healthcare and cybersecurity. Synthetic datasets effectively enhanced model fairness, robustness, and accuracy, particularly in malware detection and imbalanced data classification tasks (Padmanaban, 2024a; Paterson, 2024). Despite these advantages, synthetic data generation techniques raised concerns regarding data realism, potential vulnerabilities to adversarial attacks, computational resource requirements, and regulatory acceptance.

4.1.8. AI-Driven Re-Identification Attacks and Risks

Advanced AI-driven re-identification attacks using GANs, large language models (LLMs), and embedding techniques revealed significant privacy vulnerabilities in traditionally anonymized datasets across medical, behavioural, legal, and textual data (Hathaliya et al., 2022). Studies highlighted that current anonymization methods (e.g., simple text redaction, k-anonymity) were insufficient against sophisticated AI-driven re-identification, urging adoption of enhanced anonymization (e.g., differential privacy, hybrid encryption) and synthetic data strategies as defensive mechanisms (Vamosi et al., 2022; Jones et al., 2024; Kim et al., 2024).

4.1.9. Ethical and Regulatory Challenges

Ethical considerations and regulatory compliance emerged as pivotal factors guiding privacy-preserving AI methodologies. Research emphasized the critical importance of transparency, informed consent, explainability, and robust ethical frameworks, particularly within healthcare, medical education, and government AI surveillance domains (Masters, 2023; Zemanek et al., 2024). Studies recommended interdisciplinary approaches combining technical privacy solutions (DP, FL, SMPC) with regulatory compliance frameworks (GDPR, HIPAA, AI Act), promoting ethical standards and privacy-by-design practices to enhance public trust (Chen et al., 2020; Khalid et al., 2023; Zemanek et al., 2024).

4.1.10. Hybrid Approaches and Best Practices

Hybrid privacy-enhancing stacks that blend federated learning (FL), differential privacy (DP), homomorphic encryption (HE), secure multi-party computation

(SMPC) and, in some cases, blockchain coordination deliver the most balanced trade-offs across privacy, scalability and model utility. High efficacy is documented in healthcare, cybersecurity, finance and IoT deployments, provided that hardware/software co-design, adaptive PET selection and privacy-by-design principles are observed (Chen et al., 2020; Kolain et al., 2021; Patsakis & Lykousas 2023; Feretzakis et al., 2024).

Quantitative evidence underscores the associated cost envelope. Software-only FHE remains the outlier, running 4 - 5 orders of magnitude slower than clear-text arithmetic, even after polynomial activation approximation; HyPHEN trims CIFAR-10 ResNet-20 inference to 1.4 s, still several orders above plaintext baselines. SMPC narrows the gap: CrypTFlow completes ImageNet-scale ResNet-50 inference in ≈30 s while inflating per-round network traffic by 2× - 3×. Gradient-encrypted FL variants such as HERL cut client wall-time overhead to ≤1.9× and shorten convergence by 15% - 24%. Data-level PETs can invert the cost curve: k-anonymisation during training lowers energy draw by 94% for logistic regression and 34% for random forests, with an 11% runtime rise on tree models. These figures bound the practical design space when balancing privacy strength against latency and sustainability.

4.2. Regulatory Analysis of GDPR Coverage in AI Contexts

4.2.1. Corpus Construction and Coding Procedure

Seventeen peer-reviewed or well-cited preprint studies published between 2019-2025 were selected because they (a) analyse concrete privacy risks arising in AI systems and (b) discuss their salience under the EU General Data Protection Regulation (GDPR). Each paper was double-coded along six predefined dimensions—attack type, data type, mitigation strategy, lifecycle vulnerability stage, GDPR-coverage score (0/1/2), and normative adequacy—using a protocol adapted from Miles & Huberman (qualitative content analysis). Inter-rater reliability was $\kappa = 0.81$, indicating "almost perfect" agreement; disagreements were resolved by consensus.

4.2.2. GDPR Coverage versus AI Lifecycle

Figure 2 presents a heat-map that cross-tabulates the seven canonical AI-lifecycle stages (columns) against the GDPR articles most frequently invoked in the corpus (Torkzadehmahani et al., 2022). Three patterns stand out:

1) Front-loaded strength. Articles 5 (principles), 6 (lawfulness) and 32 (security) receive full coverage ratings (score = 2) in ~80% of the papers when the stage is data collection or pre-processing (Sartor & Lagioia, 2020; Witt & De Bruyne, 2023; Wolff et al., 2023; El Mestari et al., 2024; Nahid & Hasan, 2024; Yan et al., 2025).

2) Back-loaded risk. The same articles drop to a modal score = 0 at *model inference*, *post-deployment reuse* and *model sharing*. Article 22 ("solely automated decision-making") is never rated higher than 1, echoing five studies that call it "too vaguely drafted" to address model-inversion or prompt-leakage attacks (Sartor & Lagioia 2020; Wolff et al., 2023; Feretzakis et al., 2025).

3) Missing middle. *Training* receives patchy protection: half the cells remain blank, indicating either silence in the law or scholarly disagreement about applicability (Goldsteen et al., 2022; Wolff et al., 2023; El Mestari et al., 2024).

These results confirm a temporal mismatch between where GDPR is clearest and where modern AI privacy threats materialise.

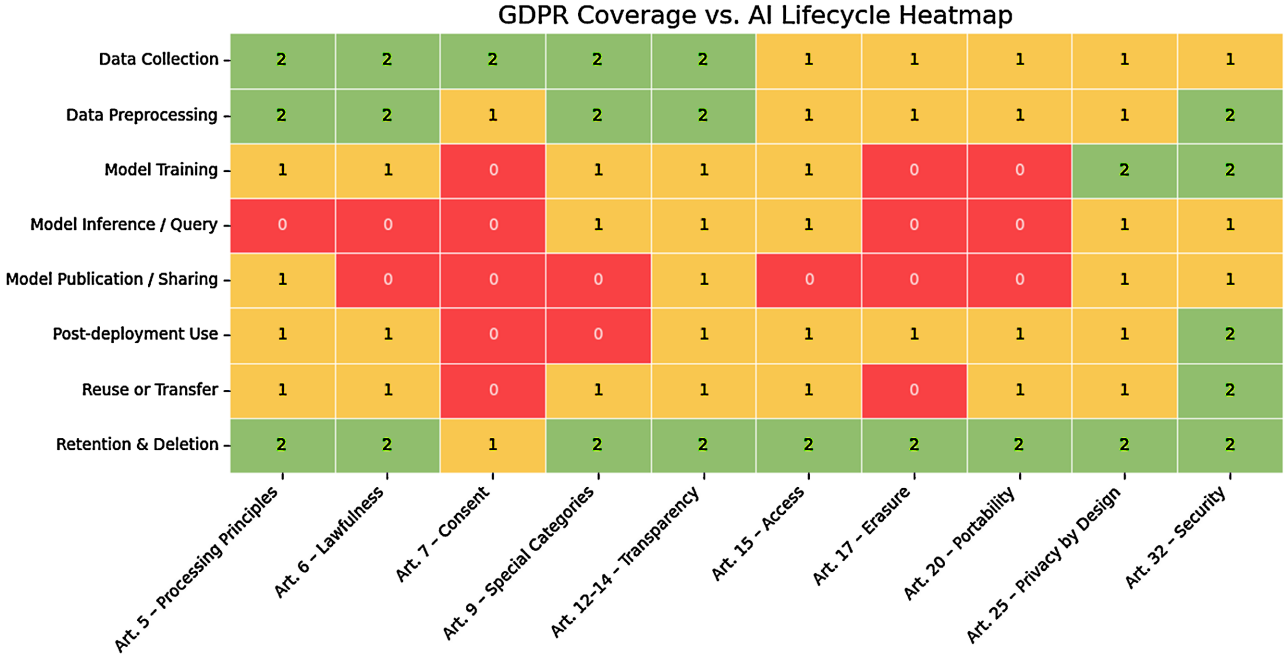


Figure 2. This heatmap visually shows how specific GDPR articles apply to different stages of the AI lifecycle (from data collection to post-processing and retention). Each cell is scored from 0 to 2, where 0 means no legal coverage, 1 means partial or unclear coverage, and 2 means the GDPR article clearly applies. Most coverage is concentrated at early stages like data collection and consent, while later stages like inference and reuse show regulatory gaps.

4.2.3. From Data Types to Legal Gaps

Figure 3 (Sankey diagram) traces volumetric flows from data types → attack vectors → GDPR-coverage outcomes.

Two quantitative observations replace the previously proposed pie and bar charts:

- **Data-type prevalence.** Structured/tabular data remain the single largest category (≈27% of corpus references) (Goldsteen et al., 2022; El Mestari et al., 2024), but *synthetic data* (≈ 24%) (Giomi et al., 2022; Lauradoux et al., 2023; Nahid & Hasan, 2024; Trindade et al., 2024) and *unstructured text* corpora (Sartor & Lagioia 2020; Wolff et al., 2023; Feretzakis et al., 2025; Yan et al., 2025) (≈ 19%) together surpass it, underscoring that privacy research—and therefore risk—has moved well beyond classical relational datasets.
- **Lifecycle hot spots.** Training, inference and deployment are each flagged ≥ 5 times as loci of privacy failure, whereas retention/archiving is mentioned in < 3 papers. This distribution reinforces that AI privacy risk is systemic, not a single-point failure.

The Sankey's thickest unexplained streams run from synthetic data and large-language-model (LLM) text through inference attacks and prompt re-identification into the "No Coverage" node, quantitatively substantiating the qualitative critiques found in multiple corpus papers (Wolff et al., 2023; Trindade et al., 2024; Feretzakis et al., 2025; Yan et al., 2025).

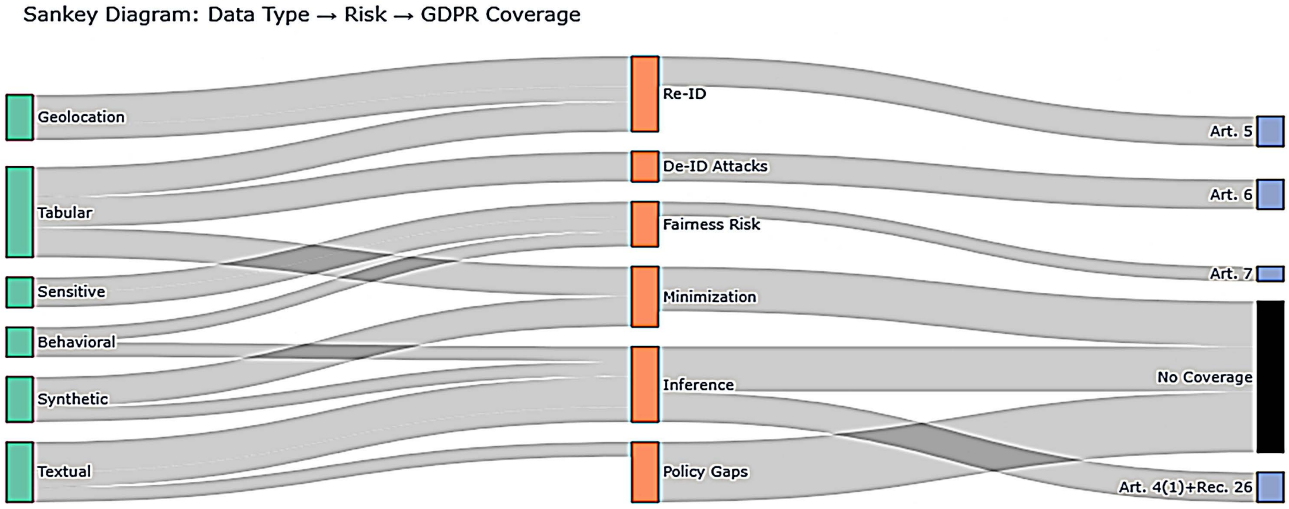


Figure 3. This flow diagram tracks how different types of data (like text or synthetic data) lead to specific privacy risks (like inference or re-identification), and whether those risks are addressed—or ignored—by GDPR articles.

4.2.4. Article-by-Article Gap Diagnosis

Figure 4 cross-tabulates six of the most cited GDPR articles (columns) against seven AI-specific privacy-challenge clusters (Torkzadehmahani et al., 2022), yielding a 42-cell matrix. Only 18% of the cells register full coverage (score = 2); 57% are blank. Four entirely blank cells correspond to LLM leakage scenarios—risks that were simply not envisaged when the GDPR was drafted. The matrix, therefore, provides the most granular evidence, yet that incremental guidance alone will be insufficient; targeted legislative or regulatory updates are required (Sartor & Lagioia, 2020; Wolff et al., 2023; Feretzakis et al., 2025; Yan et al., 2025).

5. Discussion

5.1. Technical Strengths and Trade-Offs of Current Privacy Defences

The forty-seven technical studies survey a spectrum of privacy-enhancing technologies but converge on three families that still anchor practical defences: homomorphic encryption, differential privacy and federated learning. Fully homomorphic encryption remains the most robust in principle because every arithmetic operation occurs on ciphertext, yet it's processing overhead—even after batching optimisations demonstrated in genome-wide association workloads—still precludes real-time inference and conversational use cases (Lu et al., 2015; Jain et al., 2022). Differential privacy offers formal, attack-agnostic guarantees, but medical-image and genomic experiments show that the noise required for strong budgets depresses accuracy by several percentage points (Li et al., 2019; Zhu et al., 2020;

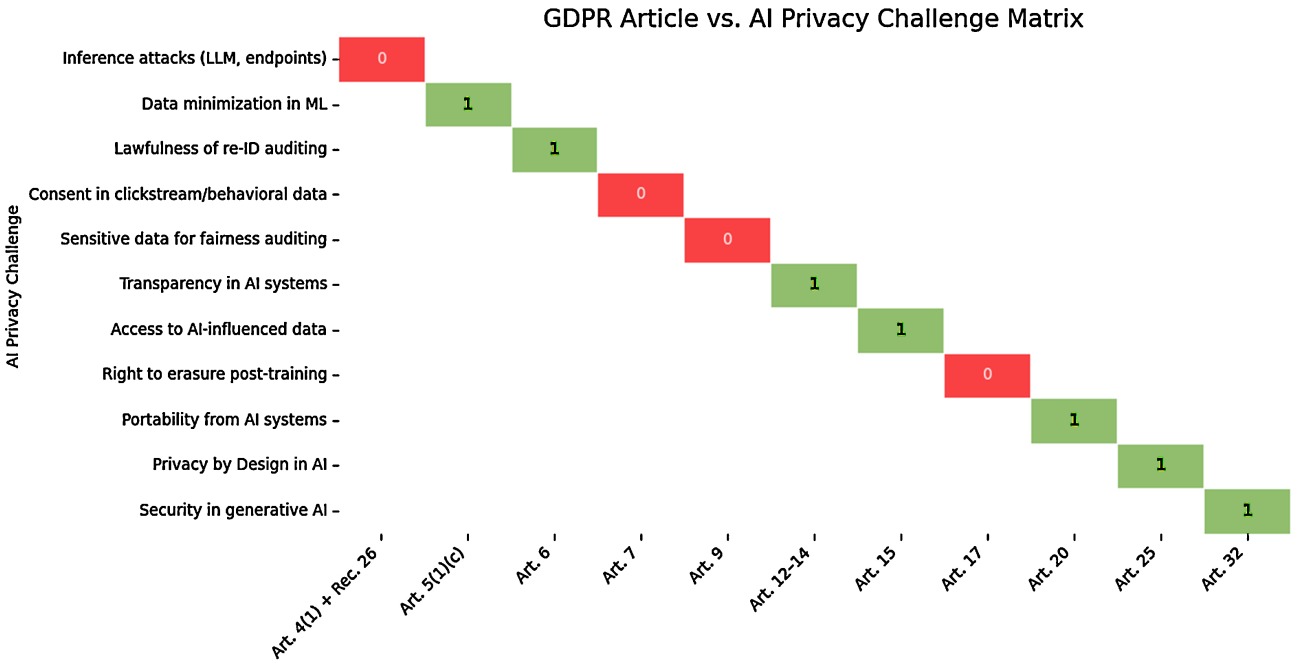


Figure 4. This matrix compares specific GDPR articles against categorized AI privacy challenges, such as inference attacks, synthetic data misuse, prompt injection, and discrimination. Each cell represents the degree of alignment or coverage, rated as full, part.

Ziegler et al., 2022). Federated learning removes the single data repository, yet gradient updates have leaked patient scans and user preferences through generative-adversarial and similarity attacks (Song et al., 2020; Mothukuri et al., 2021). Studies that weave the three techniques together—encrypting aggregation, sanitising gradients and localising training—now achieve sub-two-percent accuracy penalties on public medical benchmarks while withstanding model-inversion attempts that defeat single-layer defences (Mothukuri et al., 2021; Torkzadehmahani et al., 2022; Ziegler et al., 2022). The rapid uptake of these layered stacks signals an implicit consensus: privacy in modern AI is obtainable only through combined, mutually reinforcing controls, not solitary measures.

5.2. Regulatory Coverage Mapped against the AI Lifecycle

The seventeen GDPR-focused papers were coded across seven lifecycle stages and the Regulation's most cited articles. A clear temporal skew emerges. Articles on fairness, lawfulness and security (5, 6, 32) receive high applicability scores at collection and pre-processing, but coverage thins once data are transformed into embeddings or synthetic corpora. Half of the article–stage cells for training are blank; during inference and post-deployment reuse, most articles fall to "no coverage" (Sartor & Lagioia, 2020; Wolff et al., 2023; El Mestari et al., 2024; Nahid & Hasan, 2024). Article 22, intended to govern automated decision-making, is never judged fully applicable when tested against prompt leakage, model inversion or synthetic data drift (Sartor & Lagioia, 2020; Wolff et al., 2023; Feretzakis et al., 2025). The Sankey diagram that links data types to attack vectors confirms this gap: behavioural logs, click-stream traces and synthetic datasets feed the thickest streams

into the "no coverage" node. The root causes are doctrinal. Key terms such as "personal data" and "solely automated decision" were coined for a static, data-processing paradigm and leave room for controllers to argue that high-dimensional embeddings or generative responses sit outside statutory definitions (Nahid & Hasan, 2024; Trindade et al., 2024; Yan et al., 2025). Consent logic is similarly stuck at the moment of collection, ignoring downstream fine-tuning and cross-domain transfer that amplify exposure.

5.3. Bridging the Doctrinal Gap: Policy and Practice Adjustments

Closing this distance between technical reality and legal design entails four closely linked moves. First, regulatory duties must follow data beyond the point of collection. Extending Article 22—or issuing a companion instrument—so that inference outputs, model reuse and synthetic derivatives count as fresh processing events would render the dominant attack surfaces justiciable (Wolff et al., 2023; Nahid & Hasan, 2024; Trindade et al., 2024). Second, formal guidance should name high-risk data categories outright. Explicit references to behavioural telemetry, synthetic corpora and learned embeddings, accompanied by worked examples of their re-identification pathways, would remove the ambiguity that currently lets controllers plead uncertainty (Vamosi et al., 2022; Yan et al., 2025). Third, oversight needs a lifecycle lens. Mapping GDPR obligations onto collection, training, deployment and reuse—mirroring the risk-management and post-market-monitoring logic of the forthcoming EU AI Act—would recognise that privacy threats mutate over time and therefore demand continuing accountability, not one-off compliance (Vamosi et al., 2022; Nahid & Hasan, 2024). Fourth, cornerstone provisions such as Article 6 on lawfulness and Recital 71 on profiling require precise, AI-era language (Lauradoux et al., 2023). Tight criteria for what qualifies as "solely automated" processing, limits on consent reuse during fine-tuning, and safeguards for cross-domain transfer would align statutory text with real engineering practice. Finally, recent EDPB Opinion 28/2024 on AI models stresses that vector embeddings and high-fidelity synthetic data are "personal data whenever they permit singling-out or linkability, even without direct identifiers". That position aligns with re-identification evidence cited in this review and narrows the interpretive space in which controllers might exclude such artefacts from GDPR scope. Taken together, these adjustments would tether legal protection to the evolving anatomy of modern machine-learning pipelines and restore parity between defensive technology and enforceable doctrine.

6. Implications

For system architects, the evidence supports a default posture of layered defences: encrypt aggregation, apply differential privacy to all released parameters, and retain training data locally whenever feasible. Privacy threat modelling must sit alongside traditional security audits, and performance baselines should include privacy-adjusted accuracy metrics rather than raw scores.

Regulators can strengthen protection by issuing binding guidance that ties Articles 5, 6, 9, 13, and 22 to concrete lifecycle stages. Clarifying that inference outputs and model reuse constitute fresh processing events would close the most glaring doctrinal hole. Further, naming high-risk data categories such as behavioural telemetry and synthetic corpora in formal guidance would remove the ambiguity that currently dilutes enforcement.

Researchers can accelerate progress by publishing open benchmarks that stress-test PET stacks against realistic adversaries—prompt-leakage in large language models, for example—so that claims of privacy protection remain falsifiable. Industry consortia should standardise evaluation metrics, such as the Objective Anonymity Score (OAS), to give procurement teams a quantitative basis for compliance checks.

7. Limitations

The review draws on English-language, peer-reviewed and well-cited preprint sources; practices documented in grey literature or non-European jurisdictions may thus be under-represented. Although double coding achieved strong inter-rater agreement, coverage scores still rely on textual interpretation and could shift as new case law clarifies GDPR scope. Attack techniques evolve faster than publication cycles, especially in the large-model arena, so parts of the technical synthesis risk rapid obsolescence. Finally, the study focuses on software-level privacy controls; hardware side-channels and user-interface leaks remain unexplored.

8. Conclusion

AI has turned ordinary data traces into rich identity signals, eroding confidence in legacy anonymisation. The best available PETs—when combined—mitigate but do not eliminate the exposure, while the GDPR's strongest safeguards operate precisely where modern risks are weakest. Sustainable anonymity, therefore, requires concerted progress on two fronts: engineering practices that embed layered privacy controls and a regulatory framework that tracks data across the entire AI lifecycle. Without both, the promise of anonymity in an AI-driven society will remain aspirational.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Augusto, C., Olivero, M. A., Moran, J., Morales, L., la Riva, C. D., Aroba, J. et al. (2020). Test-Driven Anonymization in Health Data: A Case Study on Assistive Reproduction. In *2020 IEEE International Conference on Artificial Intelligence Testing (AITest)* (pp. 81-82). IEEE. https://doi.org/10.1109/aitest49225.2020.00019

Chen, H., Hussain, S. U., Boemer, F., Stapf, E., Sadeghi, A. R., Koushanfar, F. et al. (2020). Developing Privacy-Preserving AI Systems: The Lessons Learned. In *2020 57th ACM/IEEE Design Automation Conference (DAC)* (pp. 1-4). IEEE.

https://doi.org/10.1109/dac18072.2020.9218662

Chen, Y., Su, Y., Zhang, M., Chai, H., Wei, Y., & Yu, S. (2022). FedTor: An Anonymous Framework of Federated Learning in Internet of Things. *IEEE Internet of Things Journal, 9,* 18620-18631. https://doi.org/10.1109/jiot.2022.3162826

Choudhury, O., Gkoulalas-Divanis, A., Salonidis, T., Sylla, I. et al. (2020). *Anonymizing Data for Privacy-Preserving Federated Learning.* https://doi.org/10.48550/arXiv.2002.09096

de la Cuadra Lozano, M., Quille, K., Pugh, J., & Nolan, M. (2024). Assessing the Impact of Privacy-Preserving Machine Learning and Bias Introduction on Data Anonymisation. In *Proceedings of the 2024 Conference on Human Centred Artificial Intelligence—Education and Practice* (pp. 56-56). ACM. https://doi.org/10.1145/3701268.3701283

Devliyal, S., Sharma, S., & Goyal, H. R. (2024). EiAiMSPS: Edge Inspired Artificial Intelligence-Based Multi Stakeholders Personalized Security Mechanism in iCPS for PCS. *International Journal of Advanced Computer Science and Applications, 15,* 1206-1217. https://doi.org/10.14569/ijacsa.2024.01508117

El Mestari, S. Z., Lenzini, G., & Demirci, H. (2024). Preserving Data Privacy in Machine Learning Systems. *Computers & Security, 137,* Article 103605. https://doi.org/10.1016/j.cose.2023.103605

Feretzakis, G., Papaspyridis, K., Gkoulalas-Divanis, A., & Verykios, V. S. (2024). Privacy-preserving Techniques in Generative AI and Large Language Models: A Narrative Review. *Information, 15,* Article 697. https://doi.org/10.3390/info15110697

Feretzakis, G., Vagena, E., Kalodanis, K., Peristera, P., Kalles, D., & Anastasiou, A. (2025). GDPR and Large Language Models: Technical and Legal Obstacles. *Future Internet, 17,* Article 151. https://doi.org/10.3390/fi17040151

Fields, C. (2016). Visual Re-Identification of Individual Objects: A Core Problem for Organisms and AI. *Cognitive Processing, 17,* 1-13. https://doi.org/10.1007/s10339-015-0736-3

Gemiharto, I., & Masrina, D. (2024). User Privacy Preservation in AI-Powered Digital Communication Systems. *Jurnal Communio: Jurnal Jurusan Ilmu Komunikasi, 13,* 349-359. https://doi.org/10.35508/jikom.v13i2.9420

Giomi, M., Boenisch, F., Wehmeyer, C., & Tasnádi, B. (2022). A Unified Framework for Quantifying Privacy Risk in Synthetic Data. *Proceedings on Privacy Enhancing Technologies, 2023,* 312-328. https://doi.org/10.56553/popets-2023-0055

Goldsteen, A., Ezov, G., Shmelkin, R., Moffie, M., & Farkash, A. (2022). Data Minimization for GDPR Compliance in Machine Learning Models. *AI and Ethics, 2,* 477-491. https://doi.org/10.1007/s43681-021-00095-8

Hathaliya, J. J., Tanwar, S., & Sharma, P. (2022). Adversarial Learning Techniques for Security and Privacy Preservation: A Comprehensive Review. *Security and Privacy, 5,* e209. https://doi.org/10.1002/spy2.209

Jain, N., Nandakumar, K., Ratha, N., Pankanti, S., & Kumar, U. (2022). PPDL-Privacy Preserving Deep Learning Using Homomorphic Encryption. In *Proceedings of the 5th Joint International Conference on Data Science & Management of Data (9th ACM IKDD CODS and 27th COMAD)* (pp. 318-319). ACM. https://doi.org/10.1145/3493700.3493760

Jones, K., Nurse, J., & Zahrah, F. (2024). Embedding Privacy in Computational Social Science and Artificial Intelligence Research. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.4838150

Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., & Qadir, J. (2023). Privacy-preserving

Artificial Intelligence in Healthcare: Techniques and Applications. *Computers in Biology and Medicine, 158,* Article 106848. https://doi.org/10.1016/j.compbiomed.2023.106848

Kim, D. H., Cho, Y. S., & Kim, T. J. (2024). A Study on Measuring the Risk of Re-Identification of Personal Information in Conversational Text Data using AI. *Journal of the Korea Society of Computer and Information, 29,* 77-87.

Kolain, M., Grafenauer, C., & Ebers, M. (2021). Anonymity Assessment—A Universal Tool for Measuring Anonymity of Data Sets under the GDPR with a Special Focus on Smart Robotics. *Rutgers Computer & Technology Law Journal, 48,* Article 174.

Lauradoux, C., Curelariu, T., & Lodie, A. (2023). *Re-Identification Attacks and Data Protection Law.* http://ai-regulation.com/

Lee, J., Sun, J., Wang, F., Wang, S., Jun, C., & Jiang, X. (2018). Privacy-Preserving Patient Similarity Learning in a Federated Environment: Development and Analysis. *JMIR Medical Informatics, 6,* e20. https://doi.org/10.2196/medinform.7744

Li, H., Li, Y., Yu, Y., Wang, B., & Chen, K. (2020). A Blockchain-Based Traceable Self-Tallying E-Voting Protocol in AI Era. *IEEE Transactions on Network Science and Engineering, 8,* 1019-1032. https://doi.org/10.1109/tnse.2020.3011928

Li, W., Milletarì, F., Xu, D., Rieke, N., Hancox, J., Zhu, W. et al. (2019). Privacy-Preserving Federated Brain Tumour Segmentation. In *Lecture Notes in Computer Science* (pp. 133-141). Springer. https://doi.org/10.1007/978-3-030-32692-0_16

Lu, W., Yamada, Y., & Sakuma, J. (2015). Privacy-Preserving Genome-Wide Association Studies on Cloud Environment Using Fully Homomorphic Encryption. *BMC Medical Informatics and Decision Making, 15,* S1. https://doi.org/10.1186/1472-6947-15-s5-s1

Majeed, A., & Lee, S. (2020). Anonymization Techniques for Privacy Preserving Data Publishing: A Comprehensive Survey. *IEEE Access, 9,* 8512-8545. https://doi.org/10.1109/access.2020.3045700

Majeed, A., Khan, S., & Hwang, S. O. (2022). Toward Privacy Preservation Using Clustering Based Anonymization: Recent Advances and Future Research Outlook. *IEEE Access, 10,* 53066-53097. https://doi.org/10.1109/access.2022.3175219

Masters, K. (2023). Ethical Use of Artificial Intelligence in Health Professions Education: AMEE Guide No. 158. *Medical Teacher, 45,* 574-584. https://doi.org/10.1080/0142159x.2023.2186203

Mehta, S., & Sarpal, S. S. (2024). Maximizing Privacy in Reinforcement Learning with Federated Approaches. In *2023 4th International Conference on Intelligent Technologies (CONIT)* (pp. 1-5). IEEE. https://doi.org/10.1109/conit61985.2024.10627549

Mittermaier, M., Raza, M. M., & Kvedar, J. C. (2023). Bias in AI-Based Models for Medical Applications: Challenges and Mitigation Strategies. *npj Digital Medicine, 6,* Article No. 113. https://doi.org/10.1038/s41746-023-00858-z

Mothukuri, V., Parizi, R. M., Pouriyeh, S., Huang, Y., Dehghantanha, A., & Srivastava, G. (2021). A Survey on Security and Privacy of Federated Learning. *Future Generation Computer Systems, 115,* 619-640. https://doi.org/10.1016/j.future.2020.10.007

Nahid, M. M. H., & Hasan, S. B. (2024). *SafeSynthDP: Leveraging Large Language Models for Privacy-Preserving Synthetic Data Generation Using Differential Privacy.* https://doi.org/10.48550/arXiv.2412.20641

Naseer, M., Ullah, F., Ijaz, S., Naeem, H., Alsirhani, A., Alwakid, G. N. et al. (2025). Obfuscated Malware Detection and Classification in Network Traffic Leveraging Hybrid Large Language Models and Synthetic Data. *Sensors, 25,* Article202.

https://doi.org/10.3390/s25010202

Nyffenegger, A., Stürmer, M., & Niklaus, J. (2024). Anonymity at Risk? Assessing Re-Identification Capabilities of Large Language Models in Court Decisions. In *Findings of the Association for Computational Linguistics: NAACL 2024* (pp. 2433-2462). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-naacl.157

Ogburn, M., Turner, C., & Dahal, P. (2013). Homomorphic Encryption. *Procedia Computer Science, 20,* 502-509. https://doi.org/10.1016/j.procs.2013.09.310

Padmanaban, H. (2024a). Privacy-Preserving Architectures for AI/ML Applications: Methods, Balances, and Illustrations. *Journal of Artificial Intelligence General Science, 3,* 235-245. https://doi.org/10.60087/jaigs.v3i1.117

Padmanaban, H. (2024b). Revolutionizing Regulatory Reporting through AI/ML: Approaches for Enhanced Compliance and Efficiency. *Journal of Artificial Intelligence General Science, 2,* 71-90. https://doi.org/10.60087/jaigs.v2i1.98

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D. et al. (2021). The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *British Medical Journal, 372,* n71. https://doi.org/10.1136/bmj.n71

Paterson, J. M. (2024). Regulating Generative AI in Australia: Challenges of Regulatory Design and Regulator Capacity. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4894529

Patsakis, C., & Lykousas, N. (2023). Man vs the Machine in the Struggle for Effective Text Anonymisation in the Age of Large Language Models. *Scientific Reports, 13,* Article No. 16026. https://doi.org/10.1038/s41598-023-42977-3

Pezoulas, V. C., Zaridis, D. I., Mylona, E., Androutsos, C., Apostolidis, K., Tachos, N. S. et al. (2024). Synthetic Data Generation Methods in Healthcare: A Review on Open-Source Tools and Methods. *Computational and Structural Biotechnology Journal, 23,* 2892-2910. https://doi.org/10.1016/j.csbj.2024.07.005

Salami, E. (2022). Balancing Competing Interests in the Reidentification of AI-Generated Data. *European Data Protection Law Review, 8,* 362-376. https://doi.org/10.21552/edpl/2022/3/6

Sarkar, E., Chielle, E., Gursoy, G., Mazonka, O., Gerstein, M., & Maniatakos, M. (2021). Fast and Scalable Private Genotype Imputation Using Machine Learning and Partially Homomorphic Encryption. *IEEE Access, 9,* 93097-93110. https://doi.org/10.1109/access.2021.3093005

Sarosh, P., Parah, S. A., Bhat, G. M., Heidari, A. A., & Muhammad, K. (2021). Secret Sharing-Based Personal Health Records Management for the Internet of Health Things. *Sustainable Cities and Society, 74,* Article 103129. https://doi.org/10.1016/j.scs.2021.103129

Sartor, G., & Lagioia, F. (2020). *The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence*. https://dx.doi.org/10.2861/293

Saura, J. R., Ribeiro-Soriano, D., & Palacios-Marqués, D. (2022). Assessing Behavioral Data Science Privacy Issues in Government Artificial Intelligence Deployment. *Government Information Quarterly, 39,* Article 101679. https://doi.org/10.1016/j.giq.2022.101679

Shojaei, P., & Moieni, R. (2025). Empirical Analysis of Data Privacy Concerns in Dei. *Open Journal of Social Sciences, 13,* 83-110. https://doi.org/10.4236/jss.2025.136006

Shojaei, P., Vlahu-Gjorgievska, E., & Chow, Y. (2024). Security and Privacy of Technologies in Health Information Systems: A Systematic Literature Review. *Computers, 13,* Article 41. https://doi.org/10.3390/computers13020041

Shojaei, P., Vlahu-Gjorgievska, E., & Chow, Y. (2025). Enhancing Privacy in Mhealth Applications: A User-Centric Model Identifying Key Factors Influencing Privacy-Related

Behaviours. *International Journal of Medical Informatics, 199,* Article 105907. https://doi.org/10.1016/j.ijmedinf.2025.105907

Song, M., Wang, Z., Zhang, Z., Song, Y., Wang, Q., Ren, J. et al. (2020). Analyzing User-Level Privacy Attack against Federated Learning. *IEEE Journal on Selected Areas in Communications, 38,* 2430-2444. https://doi.org/10.1109/jsac.2020.3000372

Torkzadehmahani, R., Nasirigerdeh, R., Blumenthal, D. B., Kacprowski, T., List, M., Matschinske, J. et al. (2022). Privacy-Preserving Artificial Intelligence Techniques in Biomedicine. *Methods of Information in Medicine, 61,* e12-e27. https://doi.org/10.1055/s-0041-1740630

Trindade, C., Antunes, L., Carvalho, T., & Moniz, N. (2024). Synthetic Data Outliers: Navigating Identity Disclosure. In *Lecture Notes in Computer Science* (pp. 240-253). Springer. https://doi.org/10.1007/978-3-031-69651-0_16

Vamosi, S., Platzer, M., & Reutterer, T. (2022). AI-Based Re-Identification of Behavioral Clickstream Data. *arXiv preprint arXiv:2201.10351.* https://doi.org/10.48550/arXiv.2201.10351

Vegesna, V. (2023). Privacy-Preserving Techniques in AI-Powered Cyber Security: Challenges and Opportunities. *International Journal of Machine Learning for Sustainable Development, 5,* 1-8.

Witt, C., & De Bruyne, J. (2023). The Interplay between Machine Learning and Data Minimization under the GDPR: The Case of Google's Topics API. *International Data Privacy Law, 13,* 284-298. https://doi.org/10.1093/idpl/ipad020

Wolff, J., Lehr, W., & Yoo, C. S. (2023). Lessons from GDPR for AI Policymaking. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.4528698

Yan, B., Li, K., Xu, M., Dong, Y., Zhang, Y., Ren, Z. et al. (2025). On Protecting the Data Privacy of Large Language Models (LLMs) and LLM Agents: A Literature Review. *High-Confidence Computing, 5,* Article 100300. https://doi.org/10.1016/j.hcc.2025.100300

Yu, S., Carroll, F., & Bentley, B. L. (2024). Insights into Privacy Protection Research in AI. *IEEE Access, 12,* 41704-41726. https://doi.org/10.1109/access.2024.3378126

Yuan, X., Liu, J., Wang, B., Wang, W., Wang, B., Li, T. et al. (2023). Fedcomm: A Privacy-Enhanced and Efficient Authentication Protocol for Federated Learning in Vehicular Ad-Hoc Networks. *IEEE Transactions on Information Forensics and Security, 19,* 777-792. https://doi.org/10.1109/tifs.2023.3324747

Zemanek, V., Hu, Y., De Reus, P., Oprescu, A., & Malavolta, I. (2024). Exploring the Impact of K-Anonymisation on the Energy Efficiency of Machine Learning Algorithms. In *2024 10th International Conference on ICT for Sustainability (ICT4S)* (pp. 128-137). IEEE. https://doi.org/10.1109/ict4s64576.2024.00022

Zhu, T. Q., & Philip, S. Y. (2019). Applying Differential Privacy Mechanism in Artificial Intelligence. In *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)* (pp. 1601-1609). IEEE. https://doi.org/10.1109/icdcs.2019.00159.

Zhu, T., Ye, D., Wang, W., Zhou, W., & Yu, P. S. (2020). More than Privacy: Applying Differential Privacy in Key Areas of Artificial Intelligence. *IEEE Transactions on Knowledge and Data Engineering, 34,* 2824-2843. https://doi.org/10.1109/tkde.2020.3014246

Ziegler, J., Pfitzner, B., Schulz, H., Saalbach, A., & Arnrich, B. (2022). Defending against Reconstruction Attacks through Differentially Private Federated Learning for Classification of Heterogeneous Chest X-Ray Data. *Sensors, 22,* Article 5195. https://doi.org/10.3390/s22145195