Privacy, Data Protection, Risk, and Compliance in the Age of Generative AI Systems: A Systematic Mapping Study

Richardson B. da S. Andrade[1], **Geraldo Pereira Rocha Filho**[2], **Gilmar dos Santos Marques**[1], **Edna Dias Canedo**[1]

[1]University of Brasília (UnB), Department of Computer Science
Brasília, DF, Brazil

[2]State University of Southwest Bahia (UESB)
Vitória da Conquista, BA, Brazil

jcrbsa@gmail.com,gilmar.marx@gmail.com

geraldo.rocha@uesb.edu.br,ednacanedo@unb.br

Abstract. ***Research Context****: Generative Artificial Intelligence (GenAI), particularly Large Language Models (LLMs), is advancing rapidly, raising concerns about privacy, data protection, risk management, and regulatory compliance. Despite transformative potential, adoption remains immature and constrained by ethical, technical, and legal limits.* ***Scientific and/or Practical Problem****: Organizations, developers, and end users face risks of privacy violations, re-identification, model inversion, and opacity. Current frameworks such as GDPR and Brazil's LGPD struggle to address GenAI's complexity, leaving gaps between technical safeguards and legal obligations.* ***Proposed Solution and/or Analysis****: We performed a systematic mapping focused on privacy-preserving mechanisms, risk management frameworks, and compliance models applicable to GenAI. We synthesize state-of-the-art approaches, their strengths and limitations, and assess how they enable trustworthy adoption.* ***Related IS Theory****: The study is grounded in information systems governance, privacy by design, and responsible AI, using sociotechnical lenses integrating technological, organizational, and regulatory perspectives.* ***Research Method****: Following a protocol, we searched four major digital libraries (ACM DL, IEEE Xplore, ScienceDirect, SpringerLink), applied inclusion and exclusion criteria, and conducted quality assessment. From 1,138 initial studies, 44 were analyzed in depth, and 15 met all thresholds.* ***Summary of Results****: We identify four principal categories of privacy techniques (differential privacy, federated learning, cryptographic approaches, synthetic data), five risk management frameworks (e.g., NIST AI RMF, MITRE ATLAS/AI Security), and compliance instruments (DPIA, conformity assessment, FRIA). Comparative analyses reveal trade-offs among robustness, scalability, and regulatory alignment.* ***Contributions and Impact to the IS area****: This study consolidates mechanisms for addressing privacy, risk, and compliance in GenAI, highlights gaps between technical safeguards and legal requirements, and distills design implications for trustworthy systems. It supports scholars and practitioners in engineering responsible AI and informs IS research agendas, organizational policies, and regulatory strategies for intelligent information systems.*

1. Introduction

Generative Artificial Intelligence (GenAI), especially systems based on Large Language Models (LLMs), has advanced rapidly in recent years and currently enables sophisticated automation in everyday tasks such as document summarization, language translation, question answering, and human–machine conversational interaction [Warudkar and Jalit 2024]. Beyond these tasks, GenAI supports a wide range of applications including content generation, creative processes, personalization, methodological innovation, intelligent chatbots, and automated decision support [Master et al. 2024]. As a result, GenAI systems are quickly permeating various domains such as healthcare, finance, industry, education, and government, largely driven by the integration of LLM-based services offered by major technology corporations. Despite the transformative potential, the adoption and integration of GenAI solutions have not been originally designed with ethical principles, privacy guarantees, data protection rights, risk management, or compliance with regulatory frameworks in mind [Rocha et al. 2023]. This creates a pressing need for maturity models and mechanisms that support responsible and trustworthy adoption.

From the perspective of Information and Communication Technology (ICT) professionals, additional challenges emerge: the need to align technical and legal aspects throughout the software development life cycle, ensuring privacy by design from requirements elicitation to system implementation [de Paula Porto et al. 2025]. At the same time, academic, industrial, governmental, and societal initiatives have pointed to several limitations: lack of transparency in model training, opacity in decision-making processes, risks of privacy breaches in prompts, exposure to adversarial attacks, and weak control mechanisms to prevent misuse or unintended consequences.

To address these challenges, this article reports the results of a systematic mapping study (SMS) that investigates how privacy preservation, risk management, and regulatory compliance have been approached in the context of GenAI. The study analyzed 1,138 papers retrieved from four major digital libraries, of which 44 were selected and 15 met the established quality threshold. The results show that the main technical mechanisms discussed include differential privacy, federated learning, cryptographic techniques, and synthetic data generation; the main risk management frameworks identified include NIST AI RMF and MITRE AI Security; and the main compliance instruments analyzed are DPIA, CA, and FRIA. The findings highlight the trade-offs between robustness, scalability, and regulatory alignment, while also revealing gaps between technical safeguards and legal requirements. Therefore, this work provides a consolidated view of the state of the art and points to future directions for research and practice on GenAI adoption.

2. Research Method

This study follows the Systematic Mapping Study (SMS) process as outlined by Petersen et al. [Petersen et al. 2015] and Kitchenham and Brereton [Kitchenham and Brereton 2013a]. The protocol was carefully planned to ensure rigor, transparency, and replicability, covering the following stages: definition of the research question, search strategy, selection criteria, study selection, data extraction, and synthesis. To guide the SMS, the following broad research question (RQ) was defined: **RQ.1. How are privacy, risk management, and regulatory compliance being addressed in**

the development, deployment, and adoption of Generative AI (GenAI) systems, and what are the implications for trust and responsible use?

This RQ is intentionally broad, as recommended for SMSs, to capture research trends and classify contributions rather than evaluate causal effects. For the purpose of analysis and synthesis, three complementary lenses were considered: L.1: **Privacy and Adoption:** perceptions and concerns of users and developers, and their implications for trust and organizational uptake of GenAI; L.2: **Technical and Risk Mechanisms:** privacy preserving techniques (e.g., Differential Privacy, Federated Learning, cryptography, synthetic data) and risk management frameworks across the AI lifecycle, and L.3: **Regulatory Compliance:** alignment with data protection and AI governance instruments (e.g., GDPR/LGPD, DPIA, Conformity Assessment under the EU AI Act, FRIA) and associated challenges.

The literature search was performed in four major digital libraries widely used in software engineering research due to their breadth and indexing quality: ACM Digital Library, IEEE Xplore, ScienceDirect, and SpringerLink. Database specific adaptations of the search string were applied to respect query syntax, and the full protocol (search, screening, quality assessment, and extraction) was executed to ensure reproducibility. We developed a search string iteratively, following PICO (Population, Intervention, Comparison, Outcome) as suggested by Kitchenham and Charters [Kitchenham and Brereton 2013b]. The final string was: (`"Generative AI" OR "GenAI"`) AND (`"privacy" OR "data protection" OR "ethic*" OR "GDPR"`) AND (`"risk*" OR "mechanism*" OR "compliance"`). Database specific adaptations were applied to respect query syntax limitations. Searches were conducted between March and September 2025 without time restrictions.

We defined objective inclusion (IC) and exclusion (EC) criteria, adapted from established guidelines: **IC1**: Primary studies published in peer-reviewed venues (journals or conferences); **IC2**: Studies addressing the research question; **EC1**: Secondary studies (systematic reviews, mappings, surveys); **EC2**: Non-English publications; **EC3**: Studies not available for download or with fewer than 5 pages; **EC4**: Duplicates or gray literature (blogs, theses, reports) ; **EC5**: Studies are not addressing the research question. The study selection was conducted in several phases: (1) execution of the search string in all selected sources; (2) removal of duplicates using the StArt tool [Fabbri et al. 2016]; (3) title, abstract, and keyword screening; (4) full-text reading for final eligibility; (5) quality assessment using a checklist adapted from Petersen et al. [Petersen et al. 2015]. Disagreements among reviewers were resolved in consensus meetings, as recommended by best practices. Each candidate study was assessed against a Quality Assessment Checklist (QAC) with nine questions covering clarity of objectives, methodological soundness, and relevance to privacy, risk, or compliance in GenAI. Scores were assigned (Yes = 1, Partially = 0.5, No = 0), and a threshold of 7/10 was defined for inclusion.

Data extraction followed a structured form based on Petersen et al. [Petersen et al. 2015], capturing metadata (authors, venue, year), research focus (privacy, risk, compliance), proposed mechanisms, advantages, and limitations. To ensure consistency, one researcher performed the extraction and another validated it. The extracted data were analyzed using both quantitative (publication trends, country distribution, keyword frequencies) and qualitative approaches (thematic coding of

mechanisms and regulatory challenges). Results were synthesized to highlight research gaps and potential directions, consistent with SMS goals of mapping research landscapes.

3. Results

This section reports the results of the systematic mapping. We first present a descriptive overview of the corpus and the screening flow. We then synthesize the evidence on (i) privacy-preserving mechanisms (Section 3.1), (ii) risk-management approaches (Section 3.2), and (iii) regulatory and compliance instruments (Section 3.3). Throughout the section, we cross-reference the comparative matrices in Tables 1, 2, and 3. Figure 1 summarizes the selection pipeline. From 1,138 retrieved records, iterative filtering (year, type, and scope), title, abstract, keyword screening, and a short read of introductions and conclusions reduced the set to 76 candidates. After applying inclusion/exclusion criteria, 44 primary studies remained for full-text analysis. A Quality Assessment Checklist (QAC) retained 15 studies for in-depth synthesis (privacy: 4; risk: 5; compliance: 6), ensuring transparency and reproducibility of the mapping protocol.

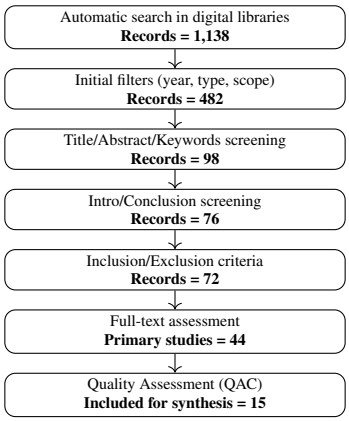

Figure 1. Screening and selection flow adopted in the SMS.

Figure 2 shows a right-skewed distribution: one publication in 2020 and 2022, a rise in 2023 (6 papers), and a marked peak in 2024 (26 papers), with continued activity into 2025 (10 papers up to our search cutoff). This pattern is consistent with the diffusion of LLM-based GenAI and the maturing debate on privacy, risk, and compliance. It also implies that most contributions are recent and, therefore, frameworks and engineering practices remain in active evolution.

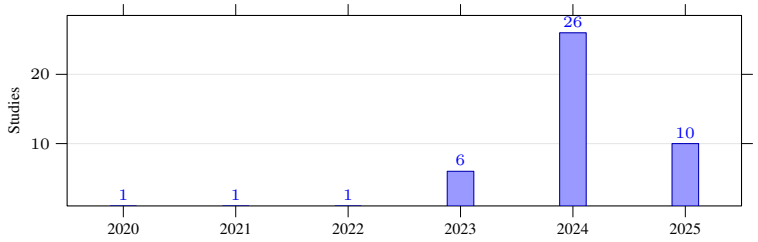

Figure 2. Publications per year in the mapping (2020–2025).

Figure 3 indicates that the United States and India account for the largest share of publications, followed by Germany and China. This concentration suggests regional

influences: EU-based work more often anchors mechanisms in GDPR/DPIA/FRIA and AI Act Conformity Assessment, whereas US-based contributions emphasize governance and operationalization (e.g., NIST AI RMF, MITRE, CSA MRM). We account for these contextual biases in the synthesis. Figure 4 shows that *privacy* is the most frequent theme

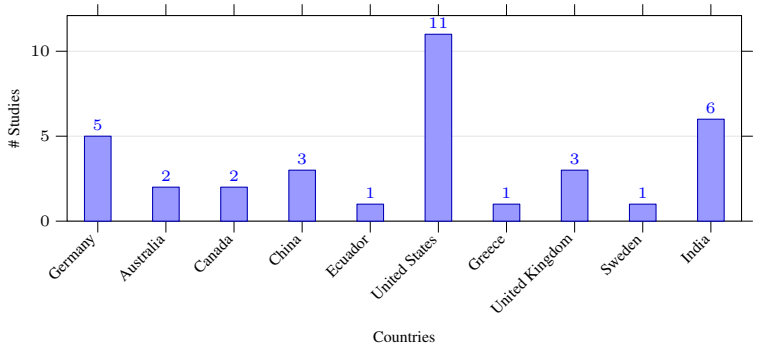

Figure 3. Publications per country in the mapping.

(28 occurrences), followed by *risk* (18 occurrences) and *compliance* (14 occurrences). This mirrors the field's trajectory: immediate concerns with data protection and leakage (training and inference), then structured risk management across the AI lifecycle, and, increasingly, formal regulatory articulation for GenAI.

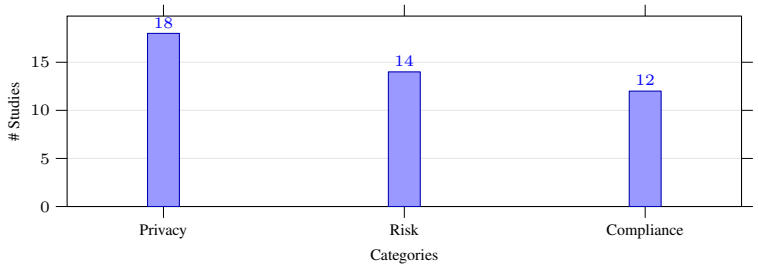

Figure 4. Publications per category in the mapping.

3.1. Privacy-Preserving Mechanisms

We categorize mechanisms by how they act on the training and inference pipeline: (i) *Differential Privacy* (DP), (ii) *Federated Learning* (FL) and decentralized training, (iii) *Cryptography and Secure Multi-Party Computation* (SMPC), and (iv) *Synthetic data generation and anonymization*. Table 1 summarizes coverage across the selected studies.

Differential Privacy (DP) is widely regarded as one of the most robust mechanisms to protect individual data points during both training and inference in machine learning systems. The principle of DP consists in adding calibrated random noise to data, gradients, or outputs in such a way that the presence or absence of any single individual in the dataset remains unnoticeable [Hu et al. 2020, Ouadrhiri and Abdelhadi 2022]. Advanced techniques, such as relevance based adaptive noise and user level DP, have been developed to balance privacy protection with model utility [Wei et al. 2021]. However, these approaches often require careful tuning of privacy budgets (ϵ), as excessive noise addition can lead to significant degradation of model accuracy, particularly in deep learning contexts.

Recent critical analyses point out that many implementations of DP in machine learning relax its formal guarantees to preserve utility. For instance, the adoption of (ϵ, δ)-DP and other relaxations has become widespread, allowing smaller amounts of noise but weakening the theoretical protection offered. In practice, this means that DP-based ML often amounts to "noise addition with a privacy label," which falls closer to traditional statistical disclosure control methods and lacks meaningful ex ante guarantees. Indeed, large ϵ values commonly used in real applications (sometimes well above 1, or even as high as 8–14) essentially nullify the theoretical protection of DP, rendering the privacy guarantees ineffective [Blanco-Justicia et al. 2023].

Moreover, [Blanco-Justicia et al. 2023] showed empirically that standard machine learning techniques to reduce overfitting (e.g., dropout, regularization, transfer learning) can achieve a better trade-off between utility, efficiency, and privacy than many DP implementations. Their findings emphasize that DP often introduces substantial computational overhead without necessarily providing stronger protection against attacks such as membership inference. As a result, while DP remains a theoretically strong framework with desirable composability properties, its practical use in ML and especially in generative AI must be carefully reassessed. Claims of privacy protection under DP should always be accompanied by empirical evaluations, as the actual guarantees depend not only on the ϵ values but also on model behavior, dataset sensitivity, and training procedures. In summary, Differential Privacy provides an important conceptual foundation for privacy-preserving machine learning, but its misuse or overly relaxed implementations may offer little more than superficial protection. For generative AI systems, which rely on large-scale sensitive data, Differential Privacy must be applied with strict parameterization and complemented with other safeguards (e.g., risk management frameworks, federated learning, or encryption) to ensure effective privacy protection in practice.

Federated Learning (FL) and Decentralized Approaches has emerged as a promising paradigm for enabling collaborative training of machine learning models without requiring the centralization of sensitive data. Instead of transmitting raw datasets, only model updates or gradients are shared, thereby reducing direct exposure of personal information [Wu et al. 2021, Mothukuri et al. 2021]. Recent studies highlight that FL is particularly effective when combined with complementary techniques such as Differential Privacy, secure aggregation, and anonymization strategies, which further strengthen resilience against privacy breaches [Ma et al. 2021]. Advanced schemes, including personalized FL, vertical FL, and robust aggregation mechanisms, have been proposed to address challenges posed by heterogeneous and non-IID data distributions across clients, as well as to mitigate the impact of adversarial behaviors [Ma et al. 2021]. Nonetheless, despite its advantages, FL remains vulnerable to inference and reconstruction attacks, as malicious actors may still exploit gradient updates to infer sensitive attributes or reconstruct original data. To address these limitations, recent research emphasizes the integration of lightweight cryptographic primitives, adaptive aggregation methods, and trust-aware mechanisms as necessary enhancements for ensuring stronger security guarantees in GenAI contexts [Mothukuri et al. 2021].

Homomorphic Encryption and Secure Multi-party Computation (SMPC) are fundamental mechanisms that enable computations on sensitive data without exposing the information in plaintext. Homomorphic encryption, in particular, allows operations to be

performed directly on encrypted data, ensuring that only the final results are revealed after decryption. SMPC, in turn, makes it possible for multiple parties to collaborate in joint computations while preserving the confidentiality of their private inputs [Ma et al. 2021].

These approaches have proven especially relevant in multi-party and inter-institutional contexts, such as healthcare and finance, where data privacy is critical [Kaissis et al. 2020]. Nevertheless, the high costs of communication and computation remain significant barriers to large-scale adoption. To address this challenge, recent advances have explored the integration of homomorphic encryption into more efficient schemes, such as *xMK-CKKS*, a multi-key variant proposed for federated learning scenarios. This scheme allows each participant to use a distinct encryption key, requiring the collaboration of all parties for the decryption process. As a result, the risk of information leakage is reduced even in scenarios involving collusion between servers and compromised devices, while also mitigating performance limitations observed in previous techniques [Ma et al. 2021]. Despite these advances, important challenges remain, particularly regarding scalability and resilience against sophisticated adversarial attacks. Still, the combination of homomorphic encryption and secure multi-party computation stands out as one of the most promising strategies to ensure privacy in distributed and GenAI systems.

Synthetic Data Generation and Anonymization: Synthetic data generation, particularly through the use of Generative Adversarial Networks (GANs), has emerged as a promising mechanism to preserve privacy while enabling data sharing in sensitive domains. These techniques aim to produce datasets that closely approximate the joint distribution of real data without exposing individual records, thereby mitigating privacy risks [Yoon et al. 2020]. Recent advances have proposed frameworks such as ADS-GAN, designed specifically for medical data, which introduce a quantifiable mathematical definition of "identifiability." This approach measures the probability of re-identification based on the combination of all attributes of an individual patient. By leveraging conditional GANs, ADS-GAN minimizes identifiability risks while maintaining the statistical fidelity of the generated data. Comparative evaluations across multiple independent datasets demonstrate that ADS-GAN not only reduces the likelihood of re-identification but also preserves model performance when trained on synthetic data, outperforming other state-of-the-art anonymization methods.

Beyond GAN-based approaches, anonymization strategies such as data minimization, controlled release, and de-identification continue to play an important role in reducing re-identification risks [Kaissis et al. 2020]. Nevertheless, these traditional techniques often lack the ability to capture complex data distributions, which may limit their utility for advanced AI applications. Thus, combining synthetic data generation with formal definitions of identifiability and robust anonymization mechanisms represents a more comprehensive solution for enabling privacy preserving data sharing in domains such as healthcare, finance, and government.

Comparative Analysis of Privacy and Data Protection Mechanisms: Table 1 summarizes the contributions of each study with respect to these mechanisms. Overall, Differential Privacy and Federated Learning emerged as the most frequently applied approaches, while SMPC appeared less often due to its computational complexity. Synthetic data generation and anonymization techniques were increasingly discussed as comple-

mentary strategies, reflecting the need to balance utility and privacy in GenAI applications.

Table 1. Comparative analysis of privacy mechanisms identified in the selected studies.

Author	DP	FL	SMPC	Synthetic Data	Anonymization
[Wolfe et al. 2024]	✓	×	×	×	×
[Jadon and Kumar 2023]	✓	✓	×	✓	✓
[Zhang and Boulos 2023]	✓	✓	×	×	✓
[Kamaruddin et al. 2024]	×	×	×	×	✓
[Hopster and Maas 2023]	×	×	×	×	×
[Master et al. 2024]	×	×	×	×	✓
[Hassan et al. 2024]	✓	×	×	✓	×
[Lin et al. 2020]	✓	×	×	✓	×
[Lin et al. 2024]	×	×	×	×	✓
[Park and Madisetti 2025]	×	×	×	×	✓
[Maliakel et al. 2024]	×	✓	×	✓	×
[Gupta et al. 2023]	×	×	×	×	✓
[Kumar et al. 2023]	×	×	×	×	✓
[Kumar et al. 2021]	×	✓	×	×	×

3.2. Risk Management Mechanisms

This subsection discusses the main risk management mechanisms identified in the selected studies, with emphasis on frameworks tailored for generative artificial intelligence (GenAI) and cybersecurity. **AI RMF (NIST Artificial Intelligence Risk Management Framework)** developed by NIST, builds upon the foundations of the *Cybersecurity Framework* (CSF) [Barrett 2018] and adapts them to the domain of artificial intelligence. The framework organizes risk management into four core functions: **Govern**, **Map**, **Measure**, and **Manage**. Each function is designed to assist organizations in identifying, assessing, monitoring, and mitigating risks throughout the lifecycle of AI systems, including GenAI applications.

The AI RMF adopts a broad and neutral definition of risk, encompassing not only potential harms but also opportunities, which broadens its applicability across diverse organizational contexts. This philosophy aligns with the CSF approach, which structures cybersecurity risk management into high-level functions (Identify, Protect, Detect, Respond, and Recover), thereby ensuring flexibility and sector-specific adaptation [Barrett 2018]. Nevertheless, practical implementation of the AI RMF in GenAI contexts faces several challenges. These include the need for subjective interpretations during deployment, significant requirements for specialized technical and human resources, and limited technical detail for critical scenarios such as prompt injection attacks or sensitive data leakage [Shah and Bajpai 2025, Schmitz et al. 2024]. As such, while the AI RMF offers a robust and adaptable conceptual foundation, its effectiveness in practice depends on being complemented by technical safeguards and domain specific frameworks.

Fraunhofer IAIS AI Assessment Catalog developed in Germany, provides a structured framework for evaluating AI systems across multiple dimensions, including fairness, autonomy, transparency, reliability, security, and data protection. These dimensions are further subdivided into risk areas that cluster similar mitigation measures and affected stakeholders or objectives. The catalog relies on quantifiable scales and measurable indicators (e.g., degree of data anonymization) to operationalize its evaluation criteria [Schmitz et al. 2024].

One of the framework's main strengths lies in its ability to explicitly link risks to specific system properties, such as model architecture versus embeddings, allowing for

a more fine-grained and context-aware analysis. Additionally, the catalog supports case-specific risk evaluations, enabling tailored assessments rather than generic checklists. Importantly, it offers concrete evaluative criteria for mitigation strategies, providing actionable insights for practitioners and policymakers. Nevertheless, the framework presents notable limitations. Its high degree of complexity poses significant challenges for large-scale adoption, particularly for organizations lacking specialized expertise. Moreover, the numerous cross-dependencies among dimensions make aggregation of risk scores difficult, reducing the clarity of overall system assessments. Finally, while the catalog provides detailed criteria at the dimension level, it offers limited guidance on synthesizing results into comprehensive, high-level risk profiles. Despite these challenges, the Fraunhofer IAIS AI Assessment Catalog remains a valuable tool for bridging abstract ethical principles and concrete risk mitigation in AI governance.

Criticality Pyramid (German Data Ethics Commission – DEK): The Criticality Pyramid, developed by the German DEK, provides a conceptual framework for classifying AI systems into five levels of criticality, ranging from minimal to unacceptable risk, based on their potential for harm and the scale of impact. This model is particularly valuable for policymakers and regulators as it communicates risk in a clear, intuitive manner, supporting transparent decision-making about acceptable uses of AI technologies [Baloukas et al. 2024]. One of its main strengths lies in its simplicity, offering a high-level categorization that is accessible to both experts and non-experts, thus serving as a useful tool for establishing political priorities and legal boundaries. However, despite its communicative power, the framework lacks operational granularity for practitioners and system engineers. It does not provide concrete methods for implementing technical safeguards or mitigating risks in practice, making it difficult to apply directly in system design or compliance processes. As such, while the Criticality Pyramid is effective as a strategic policy instrument, it must be complemented by more technical and context-specific frameworks to guide implementation and enforcement in real-world GenAI systems.

AI Model Risk Management (MRM) proposed by the Cloud Security Alliance (CSA), emphasizes governance of AI models with a strong focus on validation, monitoring, and documentation throughout the model lifecycle. Unlike broader regulatory frameworks, MRM is pragmatic and engineering oriented, making it particularly suitable for development teams that require actionable controls in practice. Recent work has highlighted its relevance in contexts where Generative AI is integrated into the Software Development Life Cycle (SDLC), stressing the need for responsible adoption guided by fairness, bias mitigation, privacy, transparency, and accountability [Shah and Bajpai 2025]. MRM contributes by offering concrete practices, such as auditing, assessments, and benchmarking, to safeguard model integrity and trust. However, despite its strengths, its scope remains limited when addressing higher level governance challenges, such as ethical oversight, regulatory alignment, and cross organizational accountability, which are crucial for long-term responsible AI adoption.

The MITRE AI Security Framework is a cybersecurity-oriented reference model that provides actionable controls to safeguard AI systems against adversarial threats, including malicious prompt injections, model manipulation, and data leakage. Unlike broader governance oriented frameworks such as the NIST AI RMF, the MITRE approach is highly pragmatic and engineering-focused, offering technical safeguards that

can be directly applied by software development and MLOps teams. It emphasizes proactive defense strategies such as adversarial testing, monitoring pipelines, and resilience measures to ensure AI robustness across deployment environments. The framework's main strength lies in its practical utility, serving as a complement to higher-level ethical or governance frameworks. However, its limitations are equally clear: the MITRE framework prioritizes technical security risks while offering limited guidance on addressing broader ethical, legal, and societal concerns such as fairness, transparency, or accountability [Shah and Bajpai 2025].

Comparative Analysis of Risk Management Mechanisms in GenAI Systems: As shown in Table 2, the NIST AI RMF appears most frequently due to its broad applicability and adaptability, although its implementation remains challenging in practice. The Fraunhofer catalog and the Criticality Pyramid are also referenced, especially in studies emphasizing contextual risk classification and transparency. Meanwhile, CSA MRM and MITRE provide more pragmatic, engineering oriented guidance, focusing on model governance and technical safeguards. However, their scope is narrower, and they tend to underrepresent broader ethical and societal risks. Overall, the mapping reveals that no single framework provides comprehensive coverage, highlighting the importance of integrating complementary mechanisms in order to achieve both technical robustness and ethical accountability in GenAI systems.

Table 2. Comparative analysis of risk management mechanisms identified in the selected studies.

Author	NIST AI RMF	Fraunhofer	Pyramid	MRM	MITRE
[Lee et al. 2024]	–	–	–	–	–
[Domínguez Hernández et al. 2024]	✓	–	–	–	–
[Nidhisree et al. 2024]	–	–	–	–	–
[Humphreys et al. 2024]	–	–	–	–	–
[Hacker et al. 2023]	–	–	–	–	–
[Rauh et al. 2025]	✓	–	–	–	–
[Djeffal 2025]	–	–	–	–	–
[Schmitz et al. 2024]	✓	✓	✓	–	–
[Beltran et al. 2024]	–	–	–	–	–
[Baloukas et al. 2024]	–	–	–	–	–
[Shah and Bajpai 2025]	✓	–	–	✓	✓

3.3. Laws and Regulations

This subsection discusses the main legal and regulatory mechanisms identified in the selected studies. It is important to note that our review excluded recent frameworks that are not yet globally recognized or consolidated, focusing instead on those with more mature adoption in practice and governance debates. **Conformity Assessment (CA)** is a cornerstone of the EU AI Act, applying primarily to high-risk AI systems. Its purpose is to verify ex-ante compliance with a set of legal, technical, and ethical requirements, including data governance, documentation, transparency, human oversight, robustness, and cybersecurity [Thelisson and Verma 2024, Kim et al. 2025]. Successful assessments culminate in a declaration of conformity and CE marking, enabling the system's access to the European market.

However, while CA establishes preventive safeguards prior to commercialization, several challenges remain. First, the process demands significant financial and technical capacity, which creates barriers for small and medium-sized enterprises (SMEs) [Wörsdörfer 2024]. Second, scholars highlight the absence of detailed procedural guidelines for conducting assessments, raising risks of inconsistent application across jurisdictions and sectors [Thelisson and Verma 2024]. Moreover, critiques from Metcalf et al.

[Metcalf 2025] warn that CA may be vulnerable to regulatory capture, as powerful industry actors could shape standards and enforcement in ways that privilege incumbents while undermining fundamental rights protections. Thus, while CA is pivotal for the EU's risk-based governance model, its effectiveness depends on independent oversight, transparency, and continuous refinement to avoid becoming a purely bureaucratic exercise.

Data Protection Impact Assessment (DPIA) established under the GDPR, is a complementary mechanism to the AI Act that applies when the processing of personal data entails high risks to individuals' rights and freedoms. DPIA requires organizations to evaluate proportionality, necessity, and mitigation measures before data processing, thereby embedding privacy-by-design principles into AI development [Thelisson and Verma 2024]. Over the years, the DPIA has been consolidated in regulatory practice across Europe, with national data protection authorities issuing detailed guidance. Despite its maturity, the DPIA faces notable limitations. Its scope is restricted to personal data, leaving out broader risks such as systemic bias, lack of transparency, or societal harms that are common in generative AI systems. Moreover, as Zhang and Meng [Zhang and Meng 2025] argue in the context of Legal Judgment Prediction, traditional impact assessments often fail to capture the opacity and technical complexity of modern AI models, which can undermine both accountability and trust. Consequently, while the DPIA remains a crucial instrument for protecting privacy, it must be complemented by additional frameworks such as CA and proposed Fundamental Rights Impact Assessments (FRIA) to ensure a holistic governance approach for generative AI.

Fundamental Rights Impact Assessment (FRIA) was introduced in the EU AI Act as a complementary governance mechanism. Unlike Conformity Assessment (CA) and Data Protection Impact Assessment (DPIA), FRIA broadens the scope by requiring users of high-risk AI systems to systematically evaluate potential impacts on human rights, marginalized groups, democracy, and the environment [Thelisson and Verma 2024, Briggs and Cross 2024]. This mechanism extends responsibility beyond providers, involving both public and private entities that deploy such systems. FRIA integrates ethical and social considerations into decision-making processes, while also fostering transparency and accountability in sensitive domains. However, practical challenges remain: its implementation may overlap or conflict with other regulatory assessments (such as CA and DPIA), and there is still a lack of clear methodologies and standardized mitigation criteria. Additionally, FRIA may increase bureaucratic burden, particularly for small and medium-sized enterprises, potentially hindering its large-scale adoption [Thelisson and Verma 2024, Kim et al. 2025].

Comparative Analysis of Compliance Mechanisms for GenAI Systems. Table 3 presents a comparative analysis of the main compliance mechanisms identified in the selected studies. Conformity Assessment (CA) appears as the core element of the AI Act, DPIA remains essential for GDPR compliance, and FRIA emerges as an innovative mechanism that extends evaluation beyond data protection to encompass broader fundamental rights. Overall, DPIA is the most frequently discussed mechanism, while CA and FRIA appear more selectively, reflecting both the regulatory focus on risk classification and the growing recognition of human rights in GenAI governance.

Table 3. Comparative analysis of compliance mechanisms for GenAI systems.

Author	CA	DPIA	FRIA
[Khowaja et al. 2024]	–	✓	–
[Vigna 2022]	✓	✓	–
[Wörsdörfer 2024]	–	✓	–
[Briggs and Cross 2024]	–	✓	✓
[Thelisson and Verma 2024]	✓	✓	✓
[Zhang et al. 2026]	–	✓	✓
[Kim et al. 2025]	✓	✓	–
[Meza et al. 2025]	–	–	–
[Sovrano et al. 2025]	✓	–	–
[Teo 2024]	–	✓	✓
[Metcalf 2025]	–	–	–

RQ.1 Summary: The mapping shows that privacy-preserving mechanisms for GenAI are mostly concentrated on *Differential Privacy* and *Federated Learning*, although both face practical limitations in scalability and robustness. Users remain concerned about privacy leaks, re-identification, and transparency, which directly affect trust and adoption. Risk management relies mainly on high-level frameworks such as the NIST AI RMF, complemented by more technical but narrower approaches like MITRE AI Security and CSA MRM. Regulatory compliance is fragmented: DPIA (GDPR) is widely adopted, CA (AI Act) introduces preventive safeguards but faces feasibility issues, and FRIA emerges as a promising yet immature mechanism to integrate broader fundamental rights. Overall, the findings highlight persistent gaps between technical safeguards and legal requirements, raising challenges for the responsible adoption of GenAI.

4. Discussion of Results

This section critically reflects on the findings of the mapping study, structured into three main perspectives: (i) privacy, trust, and adoption (4.1), (ii) risk management approaches (4.2), and (iii) regulatory compliance mechanisms (4.3). Each subsection contrasts theoretical proposals with the practical limitations observed across the analyzed studies, aiming to consolidate implications for the design and governance of GenAI systems.

4.1. User and Developer Perceptions: Privacy, Trust, and Adoption

User Privacy Perceptions. Users remain primarily concerned with the collection, storage, and potential misuse of personal data by GenAI systems. Key risks include unauthorized extraction, re-identification, profiling through inferred attributes, and leakage of sensitive information via model outputs or attacks such as model inversion and membership inference [Golda et al. 2024, Liu et al. 2025, Diro et al. 2025]. Concerns are amplified by the proliferation of deepfakes, which compromise not only privacy but also public trust in digital ecosystems [Gupta and Rathore 2024, Al-Kfairy et al. 2024, Wang et al. 2023]. Interestingly, user trust appears paradoxical: end-users often value systems with human-like characteristics, user-friendly interfaces, and richer functionalities over those with robust privacy-preserving mechanisms. This finding suggests a potential "privacy usability trade-off." As [Djeffal 2025] notes, reflexive prompt engineering can mitigate these tensions by embedding ethical and legal awareness directly into user interactions.

Developer Privacy Perceptions. Developers face the dual challenge of complying with complex regulations (e.g., GDPR) while implementing technical safeguards such as Differential Privacy, Federated Learning, and encryption [Feretzakis et al. 2024]. However,

empirical studies show that in practice, developers often neglect privacy mechanisms when using GenAI tools for software testing and coding tasks, inadvertently exposing sensitive data or proprietary information [Chen et al. 2025]. This gap is partly explained by the difficulty of interpreting legal frameworks that employ vague or principle-based language [de Paula Porto et al. 2025]. Conversely, developers with stronger technical-legal literacy leverage GenAI to improve productivity, debugging, and automated testing, but their practices vary significantly. [Petrovska et al. 2024] emphasize the need to embed responsible GenAI practices into software engineering education, given the growing expectation from industry partners that graduates will use these tools responsibly.

Impact on Trust and Adoption Privacy risks directly undermine user trust and consequently slow organizational adoption of GenAI. Studies confirm that privacy concerns are particularly salient in high-stakes sectors such as healthcare and finance [Zhang et al. 2026]. At the same time, adoption is not uniformly inhibited: when perceived benefits are substantial or organizational reputation is strong, users may tolerate higher risks [Kumar et al. 2021]. Transparency, explainability, and effective privacy controls act as trust mediators, softening the negative impact of privacy concerns and enabling conditional adoption.

4.2. Risk Management for GenAI Systems

Engineering-oriented approaches. [Shah and Bajpai 2025] advocate embedding technical and organizational controls across the AI lifecycle, including human-in-the-loop oversight, open-source gate-keeping, continuous audits, and adoption of structured frameworks such as MITRE AI Security and CSA Model Risk Management. While these approaches provide actionable guidelines, they fall short in specifying technical implementations such as prompt filtering, output sanitization, or runtime monitoring key for preventing inference-time data leaks.

Operational-technical approaches. [Baloukas et al. 2024] propose an integrated framework combining differential anonymization with quantifiable quality metrics and automatic generation of legal licenses tied to data risk levels. This bridges the gap between anonymization and regulatory obligations, particularly for training data sharing. However, its scope is limited to controlled environments and does not extend to emergent risks in inference stages, such as prompt injection or training data extraction. In sum, while the literature offers strong conceptual and sector-specific frameworks (e.g., NIST AI RMF, Fraunhofer catalogs), practical challenges persist in operationalizing risk controls for GenAI, especially in adversarial contexts.

4.3. Regulatory Compliance for GenAI Systems

GDPR Compliance. Studies such as [Khowaja et al. 2024] highlight GDPR violations in current GenAI practices, including personal data collection without explicit consent and inadequate purpose limitation. Although these works correctly diagnose risks, most discussions remain principle-based, without mapping how requirements such as data minimization or erasure translate into technical safeguards for GenAI architectures (e.g., fine-tuning or in-context prompting).

Opacity, Re-identification, and Hallucinations. [Zhang and Meng 2025] stress that LLM opacity undermines GDPR rights such as explanation and erasure, as models may regenerate sensitive data even after deletion requests. [Kim et al. 2025] reinforce that transparency gaps in training datasets hinder effective DPIAs under the AI Act, particularly for high-risk systems. Yet, few studies offer scalable technical mitigations against re-identification, with proposals like Federated Learning and Differential Privacy rarely validated in real-world GenAI deployments.

Intersection of AI Act and GDPR. [Teo 2024] illustrate how the EU AI Act requires Conformity Assessments (CA) for high-risk systems, which overlap with GDPR provisions on data quality and documentation. Their *careAI* tool proposes integrating DPIA and CA to avoid duplication. However, conflicts remain: while the AI Act allows for provider-led self-assessment in many cases, GDPR mandates independent oversight for high-risk data processing. Studies such as [Wörsdörfer 2024] and [Metcalf 2025] further warn that regulatory capture and resource asymmetries may undermine effective enforcement, especially for SMEs.

Broader ethical and social dimensions. [Briggs and Cross 2024] caution that without mechanisms such as Fundamental Rights Impact Assessments (FRIA), GenAI risks undermining established human rights at scale, particularly for marginalized groups and democratic processes. FRIA offers a means of integrating social and ethical risks, but suffers from methodological ambiguity and potential overlap with existing assessments [Thelisson and Verma 2024]. Overall, the literature underscores a regulatory landscape still in flux: while GDPR provides strong legal foundations, GenAI challenges its enforceability; the AI Act introduces complementary governance mechanisms but raises concerns about feasibility and coherence; and FRIA emerges as a necessary but underdeveloped safeguard for broader rights.

5. Threats to Validity

As with any Systematic Mapping Study (SMS), the findings of this work are subject to limitations that may affect their validity and generalizability. **Construct validity** concerns whether the study design and protocol adequately capture the intended research questions. Our search string was iteratively refined using the PICO strategy and applied to four major digital libraries (ACM DL, IEEE Xplore, ScienceDirect, and SpringerLink). While this ensured broad coverage, the choice of keywords (e.g., "Generative AI" instead of narrower terms such as "LLMs") may have biased retrieval toward general discussions, potentially omitting domain specific studies. In addition, gray literature and industry reports were excluded, which may underrepresent emerging practices in the fast-evolving GenAI landscape. **Internal validity** relates to the rigor of study selection and data extraction. To mitigate bias, inclusion and exclusion criteria were predefined and applied in multiple stages (screening by title/abstract/keywords, full text reading, and quality assessment). A Quality Assessment Checklist (QAC) with a scoring threshold was used to ensure consistency. Nevertheless, selection decisions may still have been influenced by subjective judgments, particularly in borderline cases. Mitigation included consensus meetings between reviewers.

Conclusion validity addresses whether the results and interpretations are sound and supported by the data. Quantitative analyses (e.g., publication trends, country distribution) relied on descriptive statistics, while qualitative synthesis depended on thematic coding. Although triangulation between reviewers was applied, the limited number of included high-quality studies (15) reduces statistical power, which must be considered when generalizing findings. Comparative tables were designed to increase transparency but may oversimplify nuanced contributions. **External validity** refers to the generalizability of results beyond the analyzed sample. The mapping focused exclusively on peer reviewed literature published in English, which may limit applicability to non-English-speaking contexts or industry practice. Furthermore, given the rapid evolution of GenAI, many relevant technical mechanisms or regulatory responses may not yet be represented in the academic literature. Thus, findings should be interpreted as an overview of the state of research rather than a definitive assessment of real world adoption. Overall, threats were mitigated by following established SMS guidelines, using multiple digital libraries, adopting transparent selection and quality assessment criteria.

6. Conclusion

This study aimed to provide a comprehensive overview of the state of the art regarding privacy and data protection, risk management, and regulatory compliance in the context of Generative Artificial Intelligence (GenAI) systems. To this end, we conducted a systematic mapping study following a transparent and replicable protocol, which resulted in the selection and in-depth analysis of fifteen high quality studies out of 1,138 initially retrieved records. The results identified four major categories of privacy preserving mechanisms (differential privacy, federated learning, cryptography and secure multi-party computation, and synthetic data generation/anonymization), five risk management frameworks (including NIST AI RMF, MITRE AI Security Framework, and CSA Model Risk Management), and three compliance instruments (DPIA, CA, and FRIA). The comparative analysis revealed important trade-offs between technical robustness, scalability, and regulatory alignment, highlighting both the potential and limitations of each approach.

As a contribution, this study consolidates and systematizes recent scientific production, offering a reference for researchers and practitioners interested in designing and adopting GenAI systems in a more responsible manner. Academically, the findings extend the understanding of how technical and regulatory mechanisms have been investigated, while also identifying relevant gaps, such as the lack of empirical evaluation of privacy techniques in real-world GenAI deployments. In practice, the results can inform organizations and policymakers in selecting mitigation and compliance mechanisms that are better suited to their contexts. Finally, the mapping also underscores the need for future research that more effectively integrates technical and legal solutions, for example by developing explainability metrics applicable to LLMs, methodologies for continuous auditing, and governance models that reconcile organizational efficiency with the protection of fundamental rights. In this way, the study contributes to advancing the debate on responsible GenAI adoption and paves the way for new investigations in a rapidly evolving field.

Artifact Availability

The supporting data for this work is available at `https://zenodo.org/uploads/17254610`.

Acknowledgment

This study was financed in part by the Conselho Nacional de Desenvolvimento Científico e Tecnológico CNPq (Grant Nº 300883/2025-0).

References

Al-Kfairy, M., Mustafa, D. G., Kshetri, N., Insiew, M., and Alfandi, O. (2024). Ethical challenges and solutions of generative ai: An interdisciplinary perspective. *Informatics*, 11:58.

Baloukas, C., Papadopoulos, L., Demestichas, K., Weissenfeld, A., Schlarb, S., Aramburu, M., Redó, D., García, J., Gaines, S., Marquenie, T., Eren, E., and Erdogan Peter, I. (2024). A Risk Assessment and Legal Compliance Framework for Supporting Personal Data Sharing with Privacy Preservation for Scientific Research. In *Proceedings of the 19th International Conference on Availability, Reliability and Security*, ARES '24, New York, NY, USA. ACM. event-place: Vienna, Austria.

Barrett, M. P. (2018). Framework for improving critical infrastructure cybersecurity version 1.1. Technical Report NIST.CSWP.04162018, NIST Cybersecurity Framework.

Beltran, M. A., Ruiz Mondragon, M. I., and Han, S. H. (2024). Comparative Analysis of Generative AI Risks in the Public Sector. In *Proceedings of the 25th Annual International Conference on Digital Government Research*, dg.o '24, pages 610–617, New York, NY, USA. Association for Computing Machinery. event-place: Taipei, Taiwan.

Blanco-Justicia, A., Sánchez, D., Domingo-Ferrer, J., and Muralidhar, K. (2023). A critical review on the use (and misuse) of differential privacy in machine learning. *ACM Comput. Surv.*, 55(8):160:1–160:16.

Briggs, M. and Cross, M. (2024). Generative AI: Threatening Established Human Rights Instruments at Scale. In *2024 4th International Conference on Applied Artificial Intelligence (ICAPAI)*, pages 1–8.

Chen, K., Zhou, X., Lin, Y., Feng, S., Shen, L., and Wu, P. (2025). A survey on privacy risks and protection in large language models. *J. King Saud Univ. Comput. Inf. Sci.*, 37(7).

de Paula Porto, D., Prado, R. D. C. V., dos Santos Marques, G., Serrano, A. L. M., de Mendonça, F. L., and Canedo, E. D. (2025). Ethical requirements in the age of artificial intelligence: A systematic literature review. *Simpósio Brasileiro de Sistemas de Informação (SBSI)*, pages 663–672.

Diro, A., Kaisar, S., Saini, A., Fatima, S., Pham, H. C., and Erba, F. (2025). Workplace security and privacy implications in the genai age: A survey. *J. Inf. Secur. Appl.*, 89:103960.

Djeffal, C. (2025). Reflexive Prompt Engineering: A Framework for Responsible Prompt Engineering and AI Interaction Design. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '25, pages 1757–1768, New York, NY, USA. Association for Computing Machinery.

Domínguez Hernández, A., Krishna, S., Perini, A. M., Katell, M., Bennett, S., Borda, A., Hashem, Y., Hadjiloizou, S., Mahomed, S., Jayadeva, S., Aitken, M., and Leslie,

D. (2024). Mapping the individual, social and biospheric impacts of Foundation Models. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '24, pages 776–796, New York, NY, USA. ACM. Rio de Janeiro,Brazil.

Fabbri, S., Silva, C., Hernandes, E., Octaviano, F., Di Thommazo, A., and Belgamo, A. (2016). Improvements in the start tool to better support the systematic review process. In *Proceedings of the 20th international conference on evaluation and assessment in software engineering*, pages 1–5.

Feretzakis, G., Anastasiou, A., Pitoglou, S., Paxinou, E., Gkoulalas-Divanis, A., Kalodanis, K., Tsapelas, I., Kalles, D., and Verykios, V. (2024). Securing a generative ai-powered healthcare chatbot. *Studies in health technology and informatics*, 321:195–199.

Golda, A., Mekonen, K., Pandey, A., Singh, A., Hassija, V., Chamola, V., and Sikdar, B. (2024). Privacy and security concerns in generative ai: A comprehensive survey. *IEEE Access*, 12:48126–48144.

Gupta, M., Akiri, C., Aryal, K., Parker, E., and Praharaj, L. (2023). From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy. *IEEE Access*, 11:80218–80245.

Gupta, R. and Rathore, B. (2024). Exploring the generative ai adoption in service industry: A mixed-method analysis. *Journal of Retailing and Consumer Services*, pages –.

Hacker, P., Engel, A., and Mauer, M. (2023). Regulating ChatGPT and other Large Generative AI Models. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, pages 1112–1123, New York, NY, USA. Association for Computing Machinery. event-place: Chicago, IL, USA.

Hassan, U., Zhu, J., Chen, D., and Cheung, S.-C. S. (2024). DPGEM: Differentially Private Generative Model with Exponential Mechanism. In *2024 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6. ISSN: 2157-4774.

Hopster, J. K. G. and Maas, M. M. (2023). The technology triad: disruptive AI, regulatory gaps and value change. *AI and Ethics*, 4(4):1051–1069. Publisher: Springer Science and Business Media LLC.

Hu, R., Guo, Y., Li, H., Pei, Q., and Gong, Y. (2020). Personalized federated learning with differential privacy. *IEEE Internet of Things Journal*, 7:9530–9539.

Humphreys, D., Koay, A., Desmond, D., and Mealy, E. (2024). AI hype as a cyber security risk: the moral responsibility of implementing generative AI in business. *AI and Ethics*, 4(3):791–804. Publisher: Springer Science and Business Media LLC.

Jadon, A. and Kumar, S. (2023). Leveraging Generative AI Models for Synthetic Data Generation in Healthcare: Balancing Research and Privacy. In *2023 International Conference on Smart Applications, Communications and Networking (SmartNets)*, pages 1–4.

Kaissis, G., Makowski, M., Rückert, D., and Braren, R. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*, 2:305 – 311.

Kamaruddin, S., Uphaday, N. K., Selamat, H. S., Mohd Saufi, N. N., Wan Rosli, W. R., and Mohamad, A. M. (2024). The Legal Paradigm of Generative AI in Malaysia and India: Problems and Prospects. In *2024 International Conference on Artificial Intelligence and Emerging Technology (Global AI Summit)*, pages 413–417.

Khowaja, S. A., Khuwaja, P., Dev, K., Wang, W., and Nkenyereye, L. (2024). ChatGPT Needs SPADE (Sustainability, PrivAcy, Digital divide, and Ethics) Evaluation: A Review. *Cognitive Computation*, 16(5):2528–2550. Publisher: Springer Science and Business Media LLC.

Kim, B.-J., Jeong, S., Cho, B.-K., and Chung, J.-B. (2025). AI Governance in the Context of the EU AI Act. *IEEE Access*, 13:144126–144142.

Kitchenham, B. and Brereton, P. (2013a). A systematic review of systematic review process research in software engineering. *Information and Software Technology*, 55(12):2049–2075.

Kitchenham, B. A. and Brereton, P. (2013b). A systematic review of systematic review process research in software engineering. *Inf. Softw. Technol.*, 55(12):2049–2075.

Kumar, K., Kuldeep, and Bhushan, B. (2023). Augmenting Cybersecurity and Fraud Detection Using Artificial Intelligence Advancements. In *2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, pages 1207–1212.

Kumar, M., Sharma, S., Singh, J., and Dwivedi, Y. (2021). 'okay google, what about my privacy?': User's privacy perceptions and acceptance of voice based digital assistants. *Comput. Hum. Behav.*, 120:106763.

Lee, H.-P. H., Yang, Y.-J., Von Davier, T. S., Forlizzi, J., and Das, S. (2024). Deepfakes, Phrenology, Surveillance, and More! A Taxonomy of AI Privacy Risks. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA. Association for Computing Machinery. event-place: Honolulu, HI, USA.

Lin, Y., Bao, L.-Y., Li, Z.-M., Si, S.-Z., and Chu, C.-H. (2020). Differential privacy protection over deep learning: An investigation of its impacted factors. *Computers & Security*, 99:102061.

Lin, Y., Gao, Z., Du, H., Niyato, D., Kang, J., and Liu, X. (2024). Incentive and Dynamic Client Selection for Federated Unlearning. In *Proceedings of the ACM Web Conference 2024*, WWW '24, pages 2936–2944, New York, NY, USA. Association for Computing Machinery. event-place: Singapore, Singapore.

Liu, Y., Huang, J., Li, Y., Wang, D., and Xiao, B. (2025). Generative ai model privacy: a survey. *Artif. Intell. Rev.*, 58:33.

Ma, J., Naas, S.-A., Sigg, S., and Lyu, X. (2021). Privacy-preserving federated learning based on multi-key homomorphic encryption. *International Journal of Intelligent Systems*, 37:5880 – 5901.

Maliakel, P. J., Ilager, S., and Brandic, I. (2024). FLIGAN: Enhancing Federated Learning with Incomplete Data using GAN. In *Proceedings of the 7th International Workshop on Edge Systems, Analytics and Networking*, EdgeSys '24, pages 1–6, New York, NY, USA. Association for Computing Machinery. event-place: Athens, Greece.

Master, S., Chirputkar, A., and Ashok, P. (2024). Unleashing Creativity: The Business Potential of Generative AI. In *2024 2nd World Conference on Communication & Computing (WCONF)*, pages 1–6.

Metcalf, T. (2025). AI safety and regulatory capture. *AI & SOCIETY*. Publisher: Springer Science and Business Media LLC.

Meza, J., Saltos, M. F. L., Campoverde, E. U. V., Cejas, M. C. N., and Castro, V. C. A. (2025). AI Regulation for Ecuador. In *2025 Eleventh International Conference on eDemocracy & eGovernment (ICEDEG)*, pages 311–316. ISSN: 2573-1998.

Mothukuri, V., Parizi, R., Pouriyeh, S., ping Huang, Y., Dehghantanha, A., and Srivastava, G. (2021). A survey on security and privacy of federated learning. *Future Gener. Comput. Syst.*, 115:619–640.

Nidhisree, C., Paul, A., Venunadh, A., and Bhowmick, R. S. (2024). Generative AI Under Scrutiny: Assessing the Risks and Challenges in Diverse Domains. In *2024 IEEE 6th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA)*, pages 243–248.

Ouadrhiri, A. E. and Abdelhadi, A. M. (2022). Differential privacy for deep and federated learning: A survey. *IEEE Access*, 10:22359–22380.

Park, J. H. and Madisetti, V. K. (2025). CAPRI: A Context-Aware Privacy Framework for Multi-Agent Generative AI Applications. *IEEE Access*, 13:43168–43177.

Petersen, K., Vakkalanka, S., and Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Inf. Softw. Technol.*, 64:1–18.

Petrovska, O., Clift, L., Moller, F., and Pearsall, R. (2024). Incorporating Generative AI into Software Development Education. In *Proceedings of the 8th Conference on Computing Education Practice*, CEP '24, pages 37–40, New York, NY, USA. Association for Computing Machinery. event-place: Durham, United Kingdom.

Rauh, M., Marchal, N., Manzini, A., Hendricks, L. A., Comanescu, R., Akbulut, C., Stepleton, T., Mateos-Garcia, J., Bergman, S., Kay, J., Griffin, C., Bariach, B., Gabriel, I., Rieser, V., Isaac, W., and Weidinger, L. (2025). Gaps in the Safety Evaluation of Generative AI. In *Proceedings of the 2024 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '24, pages 1200–1217, San Jose, California, USA. AAAI Press.

Rocha, L. D., Silva, G. R. S., and Dias Canedo, E. (2023). Privacy compliance in software development: A guide to implementing the lgpd principles. In *Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing*, pages 1352–1361.

Schmitz, A., Mock, M., Görge, R., Cremers, A. B., and Poretschkin, M. (2024). A global scale comparison of risk aggregation in AI assessment frameworks. *AI and Ethics*, 5(2):1407–1432. Publisher: Springer Science and Business Media LLC.

Shah, J. A. and Bajpai, G. (2025). Responsible Generative AI for Software Development Life Cycle. In *2025 IEEE World AI IoT Congress (AIIoT)*, pages 0056–0061.

Sovrano, F., Hine, E., Anzolut, S., and Bacchelli, A. (2025). Simplifying software compliance: AI technologies in drafting technical documentation for the AI Act. *Empirical Software Engineering*, 30(4). Publisher: Springer Science and Business Media LLC.

Teo, S. A. (2024). Artificial intelligence and its 'slow violence' to human rights. *AI and Ethics*, 5(3):2265–2280. Publisher: Springer Science and Business Media LLC.

Thelisson, E. and Verma, H. (2024). Conformity assessment under the EU AI act general approach. *AI and Ethics*, 4(1):113–121. Publisher: Springer Science and Business Media LLC.

Vigna, F. (2022). Co-regulation Approach for Governing Big Data: Thoughts on Data Protection Law. In *Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance*, ICEGOV '22, pages 59–63, New York, NY, USA. Association for Computing Machinery. event-place: Guimarães, Portugal.

Wang, Y., Pan, Y., Yan, M., Su, Z., and Luan, T. (2023). A survey on chatgpt: Ai–generated contents, challenges, and solutions. *IEEE Open Journal of the Computer Society*, 4:280–302.

Warudkar, S. and Jalit, R. (2024). Unlocking the Potential of Generative AI in Large Language Models. In *2024 Parul International Conference on Engineering and Technology (PICET)*, pages 1–5.

Wei, K., Li, J., Ding, M., Ma, C., Su, H., and Poor, H. (2021). User-level privacy-preserving federated learning: Analysis and performance optimization. *IEEE Transactions on Mobile Computing*, 21:3388–3401.

Wolfe, R., Slaughter, I., Han, B., Wen, B., Yang, Y., Rosenblatt, L., Herman, B., Brown, E., Qu, Z., Weber, N., and Howe, B. (2024). Laboratory-Scale AI: Open-Weight Models are Competitive with ChatGPT Even in Low-Resource Settings. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '24, pages 1199–1210, New York, NY, USA. ACM. Rio de Janeiro, Brazil.

Wu, C., Wu, F., Lyu, L., Huang, Y., and Xie, X. (2021). Communication-efficient federated learning via knowledge distillation. *Nature Communications*, 13.

Wörsdörfer, M. (2024). Biden's Executive Order on AI: strengths, weaknesses, and possible reform steps. *AI and Ethics*, 5(2):1669–1683. Publisher: Springer Science and Business Media LLC.

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

Zhang, C. and Meng, Y. (2025). Bridging the divide: technical research and application on legal judgment prediction. *Artificial Intelligence and Law*. Publisher: Springer Science and Business Media LLC.

Zhang, P. and Boulos, M. (2023). Generative ai in medicine and healthcare: Promises, opportunities and challenges. *Future Internet*, 15:286.

Zhang, Y., Tian, J., and Deng, F. (2026). In generative ai we trust: an exploratory study on dimensionality and structure of user trust in chatgpt. *Interacting with Computers*, 38:58–76.