PriEval-Protect : Privacy Evaluation & Protection Framework

Introduction

In the digital age of healthcare, privacy concerns have become increasingly critical due to the widespread use of patient data in AI-driven applications and clinical decision-making systems. Hospitals, research institutions, and health-tech companies must implement robust data protection mechanisms to comply with stringent regulations such as the GDPR and HIPAA, while also upholding ethical standards and patient trust [1] [2] . However, existing approaches often address privacy in a fragmented manner—focusing either on compliance scoring, risk estimation, or anonymization techniques in isolation. This siloed perspective leads to limited scalability and adaptability across diverse healthcare data environments.

To address this gap, **PriEval-Protect** introduces a comprehensive and unified framework that integrates Privacy Scoring and Privacy Solutions to systematically assess and enhance data privacy. The Privacy Scoring phase quantifies privacy risk using a multi-factor model that incorporates regulatory context (e.g., HIPAA, GDPR compliance) [2] [1] , data distribution structure (centralized vs. decentralized) [3] [4] , encryption types (symmetric, asymmetric, homomorphic) [5], and privacy metrics such as uncertainty, information gain/loss, i indistinguishability, adversary success rate, time, and error [6]. The scoring is based on a customizable weighted formula that adjusts according to the dataset characteristics and applicable regulatory standards.

Following risk assessment, the Privacy Solutions phase recommends tailored privacy-preserving techniques. This includes prioritizing Federated Learning (FL) to support decentralized model training without exposing raw data [4], and Differential Privacy (DP) to introduce mathematically controlled noise that prevents individual data points from being identifiable [4] —all while maintaining high analytical accuracy.

While prior research has explored privacy metric taxonomies [6], GDPR/ HIPAA compliance scoring models [2] [1], and risk formulas for open data scenarios [7], they lack an integrated, automated decision-making system that considers all relevant privacy dimensions and translates scores into actionable protection strategies.

By unifying risk quantification and privacy-enhancing solutions, PriEval-Protect offers a scalable, adaptable, and practical solution for healthcare stakeholders—enabling them to meet compliance goals, reduce privacy risks, and preserve data utility.

State of Art

I. Privacy Scoring :

Informed by insights from existing literature and practical frameworks, we propose a unified workflow for evaluating privacy in e-health datasets. Our goal is to compute a **composite privacy score** by integrating diverse but complementary factors.

The core factors considered in this model include:

1) Regulatory Context :

a. HIPAA: [1], [8]

The authors outlined HIPAA's privacy and security requirements to safeguard electronic health data, ultimately identifying seven core principles that characterize HIPAA compliance. [8]

The authors conducted an in-depth analysis of compliance factors influencing HIPAA adoption. Their research identified key determinants such as: [1]

- **Security Awareness SA** – The extent to which employees understand and follow security policies.
- **Management Support MS**– The role of executive leadership in promoting HIPAA adherence.
- **Security Culture SC**– The ingrained values and attitudes toward compliance within an institution.
- **Security Behavior SB**– How security training influences daily data handling practices.
- **Security Effectiveness SE**– The ability of HIPAA measures to prevent unauthorized access and breaches.

These factors were integrated into a weighted compliance model, where Security Awareness, Management Support, and Security Culture contribute to Security Behavior and Security Effectiveness. The final compliance score reflects how well an institution adheres to HIPAA mandates.

Integration in our workflow: [1]

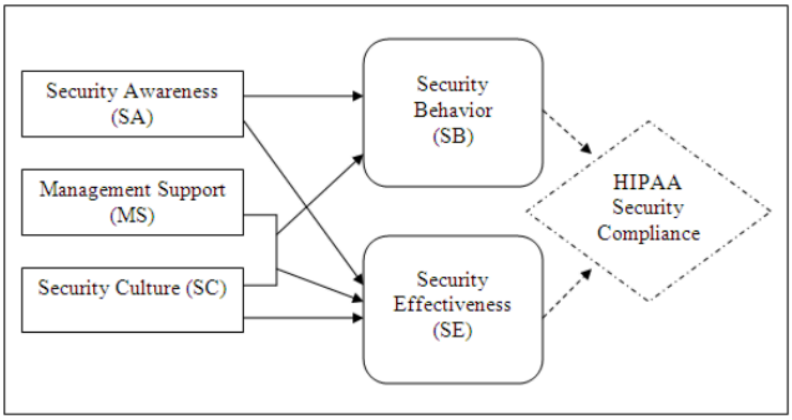

figure 1. The empirically-validated conceptual model of the relevant factors and their effects on HIPAA security compliance in AMCs. [1]

SE and SB formulas: [1]

SE = .864 + .569*SA + .320*MS x SC

SB = 3.311 + .265*SA +.255*MS x SC

Limits:

- The article misses how to actually extract influencing factors (like awareness or support) from policies and regulations, making the evaluation non-automated and hard to scale.
- Surveys, while useful, are biased and manual—they can't be fully trusted or automated for large-scale compliance checks.

b. GDPR: [2]

The authors proposed and validated the **Benjumea Privacy Scale**, a robust GDPR-based tool to evaluate the quality of privacy policies in mHealth applications. The scale was developed using a two-round eDelphi study with privacy experts, refining a previous checklist of **14 GDPR-derived** items. They added two new items and assigned weights to each based on expert consensus using Likert scale scoring and IQR analysis. The final weighted scale helps measure how clearly and completely privacy policies inform users about data processing, rights, and legal obligations.

Key contributions:

- Validation of 14 core GDPR-based items
- Addition of 2 expert-suggested items
- Weighted scoring (1 or 0.5) to reflect perceived importance
- Final scale provides a transparent, measurable method to assess privacy compliance in mHealth apps

Integration in our workflow:

- We can incorporate the Benjumea scale's items and weights directly into our privacy policy scoring tool to assess textual GDPR compliance.
- The scale's item definitions can guide our feature extraction from policies (e.g., legal basis, data retention, user rights).

GDPR compliance formula: [2]

$$GCS = \frac{\Sigma Xi * Wi}{\Sigma Wi} \times 100$$

With Xi the items from tables below and Wi weights L1=1 or L2=0.5 depending on the factor.

Xi values come from assessing whether a specific GDPR-required element is present in a privacy policy. Values are: 0 missing or non-compliant , 0.5 partially compliant, 1 fully present and compliant.

Note: we can find a detailed description of each item in the article.

Table 4 Round 2 results for Likert-scale items

Item identifier	Median	% Ratings ≥4	IQR
I1	5	100.00%	0
I2	4	63.16%	2
I3	5	100.00%	0
I4	5	100.00%	0
I5	5	100.00%	0
I6	5	94.74%	1
I7	5	100.00%	0.5
I8	5	94.74%	1
I9	4	89.47%	1
I10	5	89.47%	1
I11	5	94.74%	1
I12	4	84.21%	1
I13	4	89.47%	1
I14	5	94.74%	1
T2	5	94.74%	1
T3	4	84.21%	0.5
T4	3	42.11%	2
T5	4	52.63%	1
T6	3	47.37%	1
T7	3	42.11%	1
T8	4	57.89%	2
T9	3	42.11%	2

Abbreviation: IQR, interquartile range.

Table 6 Weight assignment

Items with a weight of L1	I1, I3, I4, I5, I6, I7, I8, I10, I11, I14, T2
Items with a weight of L2	I2, I9, I12, I13, T3

sented in a different way, as shown in ►Table 5. Participants could select one or more answers for this item.

Finally, regarding weight assignment, and based on the results presented in ►Table 4, higher weights were assigned to those items that have been considered "very important" (median 5) at the expense of those considered "important" (median 4). ►Table 6 shows the weight assignment. L1 means a high weight, while L2 means a low weight. We propose values of L1 = 1 and L2 = 0.5.

figure 2. derived GDPR-items and weights to calculate compliance score. [2]

Limits:

- The article does not provide an automated method or algorithm for extracting the presence and compliance of Xi (items) in privacy policies. Instead, it describes a manual expert-driven evaluation process using a checklist-based assessment.
- The study only focuses on GDPR compliance and does not address other global privacy regulations like HIPAA, CCPA, or other sector-specific laws to estimate a general privacy score.

HIPAA	GDPR
• The article does not provide an automated method or algorithm for extracting the presence and compliance of Xi	• The study only focuses on GDPR compliance and does not address other global privacy regulations like HIPAA, CCPA, or other sector-specific laws to estimate a general privacy score.

2) Data distribution : [3], [4]

From multiple literature sources we can conclude that data distribution significantly affects privacy, security, and compliance in healthcare systems. Different studies highlight its role in privacy risk management.

The authors of the first article [3] show that centralized architectures offer simplicity and integration benefits but pose higher privacy risks due to single points of failure, whereas decentralized systems enhance resilience and privacy by distributing data across multiple nodes.

The second article [4] emphasizes that federated learning leverages decentralized data distribution to preserve patient privacy, with privacy-enhancing techniques like differential privacy and homomorphic encryption mitigating risks during model training.

Integration in our workflow: [3], [4]

- Centralized vs. decentralized classification helps in scoring data exposure and integration complexity.
- Privacy trade-offs: Helps calibrate weights in our privacy scoring model (e.g., decentralized systems are resilient but harder to integrate).

Limits:

- Both articles highlight the critical role of data distribution (centralized vs. decentralized) in influencing privacy, security, and compliance, but fail to provide a concrete method or formula to quantify its impact in a privacy score — leaving a gap that we aim to address through structured weighting using **MCDM techniques** like AHP to assign weights based on expert judgment. [9]

3) Encryption Type : [5]

encyrption	type	presence	degree	effect on privacy
	Symmetric	1	80%	70%
	Asymmetric	0	60%	100%

Table. encryption types effects on privacy

The authors of [5] proposed and implemented a hybrid encryption algorithm that combines an optimized version of AES (called P-AES) (symmetric) with RSA (asymmetric) to improve the security and efficiency of medical data storage in cloud databases. The work is motivated by the challenge of protecting sensitive patient data while enabling high-performance access and storage in hospital systems.

They modified the AES algorithm into a parallel version (P-AES) to handle large medical records more efficiently. This was then paired with RSA to secure key exchange, forming a hybrid model.

The approach was tested in a hospital information management system, and results demonstrated faster encryption/decryption, good scalability with long data, and resilience against typical cryptographic attacks.

key findings: [5]

- AES and outer-layer database encryption showed high encryption/decryption efficiency.
- The hybrid encryption algorithm (P-AES + RSA) effectively protects medical data in cloud databases and enhances patient privacy.

Integration in our workflow:

We can conclude from literature that hybrid encryption deserves the highest score because it protects both data and key confidentiality.

No encryption should be penalized in the scoring as it introduces high privacy risk.

This also aligns with GDPR [10] and HIPAA [11] expectations on secure storage and transmission.

Limits:

- Limitations of P-AES: It currently supports only text data, not images or multimedia, and is restricted to AES-128 (128-bit keys only). There remains significant potential for improvement.

- While highlighting the importance of encryption, the article did not provide quantitative weights or formal scoring methods for comparing encryption types — leaving a gap that we aim to address through structured weighting using **MCDM techniques** like AHP to assign weights based on expert judgment. [9]

4) Privacy metrics: [6]

The authors conducted a systematic survey of over **80** technical **privacy metrics** to bring structure to a fragmented landscape. The authors reviewed these metrics across multiple privacy domains (databases, communication systems, social networks, etc.) and introduced a **taxonomy** based on:

- What aspect of privacy the metric measures
- The inputs required
- The type of data involved
- Output characteristics (e.g., uncertainty, indistinguishability, information gain/loss)

They also proposed a framework of nine guiding questions to help choose appropriate metrics for different scenarios.

Key findings :

- Privacy is <u>multidimensional</u>— no single metric can capture all aspects and the need for multi-metric evaluation is critical to avoid bias.
- The authors classified metrics into 8 output categories :
  1. Uncertainty
  2. Information gain/loss
  3. Data similarity
  4. Indistinguishability
  5. Adversary's success probability
  6. Error
  7. Time
  8. Accuracy/Precision

<u>Integration in our workflow :</u>

- The taxonomy can guide which metrics to use when calculating privacy scores across regulatory contexts (GDPR, HIPAA).
- Supports the use of multi-criteria decision-making (AHP) by emphasizing <u>non-uniform importance of metrics</u>.

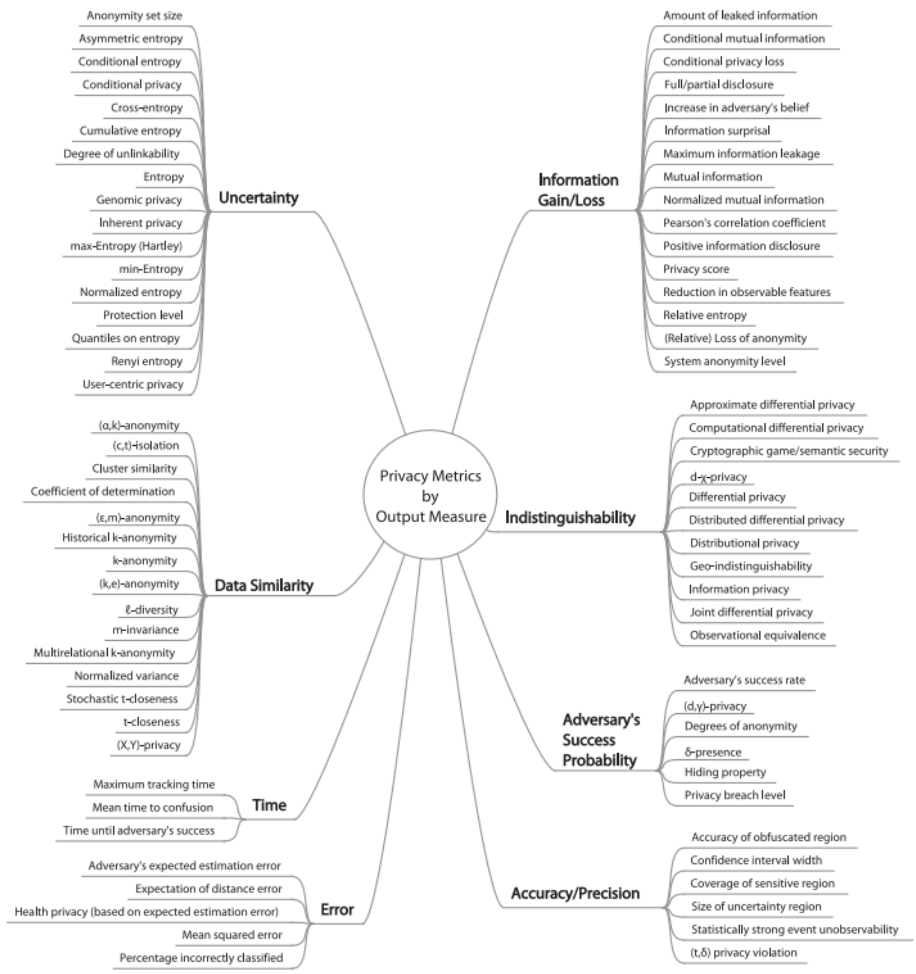

figure 3. Taxonomy of privacy metrics, classified by output [6]

Limits :

- Does not assign numerical weights or practical prioritizations (we will complement it with AHP model). [9]
- Many metrics are domain-specific—may require adaptation to work well in medical/e-health data but we notice the presence of genome security which might be the closest to the healthcare domain .
- Limited tool support or validated implementations: lack of mentioned libraries or specific implementations.

⇒ To operationalize our framework, we will leverage tools like PyCANON [12] for computing technical metrics such as *k-anonymity, l-diversity, and t-closeness*. However, accurate calculation of these metrics depends on the correct identification of quasi-identifiers (QIs) and sensitive attributes (SAs).

For this purpose, we will use Large Language Models (LLMs) to extract and classify attributes from datasets and schema documentation.

❖ LLMs use justification: [13]

The author conducted a comprehensive study exploring automated quasi-identifier (QID) recognition using multiple techniques, including deep learning, causal inference, and heuristic optimization. The work introduced several architectures and approaches, aiming to overcome the limitations of traditional privacy protection methods, particularly for identifying QIDs in tabular datasets. They worked with diverse datasets, both medical (heart+disease , diabetes) and non medical (adult, nursery and car).

	Direction	Description
1	**QIDLearningLibrary**	Creation of *QIDLearningLib*, a Python library with metrics purposed for QID detection aiming for open-source publication.
2	**LLM Experiment**	Using a Language Model (LLM) to identify QIDs in the Adult dataset, testing its prediction accuracy under different contexts.
3	**Mutant Autoencoder**	Design of a deep learning model for QID recognition, analyzing feature weights, tested on synthetic and real data.
4	**Custom Loss DNN**	A deep neural network with a specialized loss function designed for QID metrics, replacing standard metrics like MSE.
5	**Evolutionary Algorithm**	Use of evolutionary algorithms to optimize QID recognition by evolving the best attribute configurations through simulated generations.
6	**Causal VAE | GAN with Isolation Forest Ensemble**	Development of a robust, explainable model (Causal VAE_GAN with Isolation Forests Ensemble), adaptable to any dataset with efficient runtime.

Table 1. Summary of Research Directions for QID Recognition [13]

Key findings :

- **Mutant Autoencoder** : Designed to detect QIDs via weight convergence, *but* was unstable and partially abandoned.

- **CustomLossDNN** : A deep neural network with a custom loss function combining distinction, separation, and k-anonymity. Conceptually strong *but* limited by complex backpropagation and computational inefficiency.
- **Evolutionary Algorithm** : Explored QID combinations using fitness functions. Results were good, *but* the method was costly (70+ hours), lacked causal modeling, and had low explainability.
- **CVAE_GAN_IFE** : The most advanced approach combining causal discovery, VAE-GAN reconstruction, and isolation forest ensemble for QID detection. Outperformed ground truth in most datasets *but* was limited to QIDs (not SAs), computationally intensive, and unaware of regulatory requirements.
- **LLM Experiment** : A Large Language Model from Perplexity.ai [14] was tested for classifying QIDs, SAs, and NSAs using different levels of context. Performance improved with better prompt design, showing the potential of prompt engineering in enhancing attribute classification.

⇒ We chose to integrate LLMs because:

- Traditional approaches are often limited to quasi-identifier detection only, lack support for identifying sensitive attributes, and require high computational resources or manual effort.
- LLMs offer a scalable, lightweight, and context-aware solution that can classify QIDs, SAs, and NSAs directly from data and documentation—making them ideal for our goal of automating privacy risk assessment across diverse healthcare datasets.

Detailed key findings for the LLM approach :

- The LLM Perplexity.ai [14] was tested on the dataset, using known ground-truth labels for QIDs, SAs, and NSAs.
- It was evaluated under three levels of context:
  1. Attribute list only
  2. Reference to the Adult dataset with attribute list
  3. Full dataset upload with attribute list
- Performance improved with increased context, as measured by the *Jaccard index*, confirming the value of contextual input.
- The LLM was able to correctly predict QIDs, SAs, and NSAs, showing strong potential for automation in early-stage attribute classification.
- This experiment also revealed that prompt engineering is a key lever for improving LLM performance in privacy-related tasks.

Context	$J(QID_{real}, QID_{predicted})$	$J(SA_{real}, SA_{predicted})$	$J(NSA_{real}, NSA_{predicted})$
1 - None	0.333	0.333	0.111
2 - Reference	0.454	0.333	0.111
3 - Whole Dataset	0.545	0.333	0.143

Table 2. Jaccard Indexes for QIDs, SAs, and NSAs [13]

Integration in our workflow :

- LLMs will be used in the **preprocessing phase** of our framework to automatically classify QIDs, SAs, and NSAs. This classification will enable precise metric computation via tools like PyCANON (k-anonymity, l-diversity, t-closeness).

- Using prompt engineering techniques, we can adapt the LLMs to various data types and context, improving flexibility and reducing manual workload.

Limits :

- Current LLMs **rely heavily on well-crafted prompts**; poor prompt design may yield suboptimal results.

- The use of external LLM platforms raises **data confidentiality concerns**, particularly in sensitive domains like healthcare. To address this, we can either self-host the LLMs in secure environments or, when using external APIs, strictly limit input to schema-level data (e.g., attribute names only). Techniques like few-shot prompt engineering can be used to guide the model effectively without exposing real data, ensuring both utility and privacy.

❖ Diagram : Privacy Framework workflow

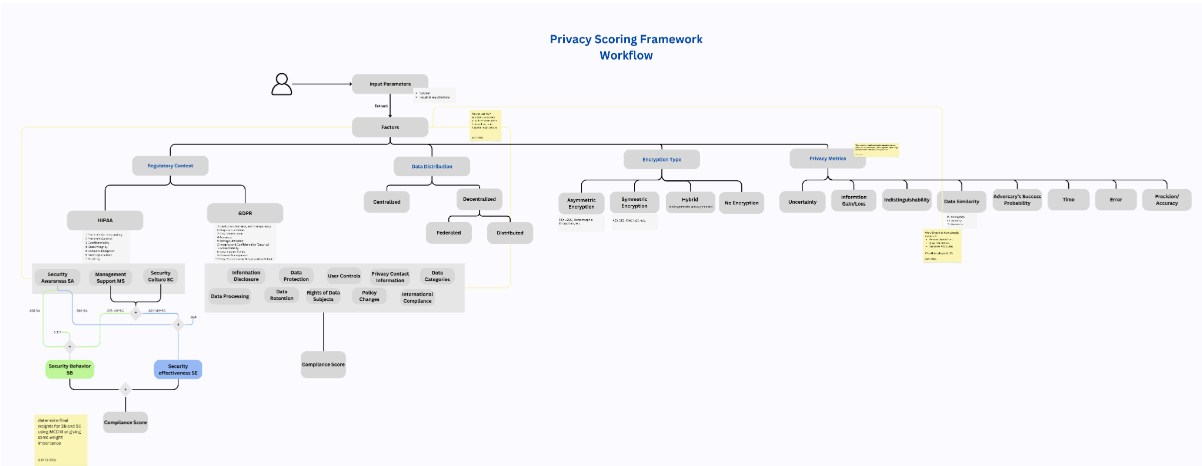

figure 4. PriEval-Protect Privacy Scoring Workflow

-> Please navigate to this link for a clearer view.

Description:

The Privacy Scoring Framework Workflow diagram visually encapsulates the *multi-phase* logic of the *PriEval-Protect* framework, which systematically assesses privacy risks in e-health datasets. At its core, the diagram integrates four foundational components—**regulatory context**, **data distribution**, **encryption type**, and **privacy metrics**—to calculate a composite privacy score.

The *regulatory context* branch incorporates both GDPR and HIPAA, using validated scoring models like the Benjumea Privacy Scale and the HIPAA compliance formulas based on Security Behavior (SB) and Security Effectiveness (SE). These scores are influenced by institutional factors such as awareness, management support, and security culture.

The *data distribution* component distinguishes between centralized and decentralized architectures (federated /distributed), which affect exposure risk and compliance feasibility.

Encryption types are categorized into symmetric, asymmetric, hybrid, or none, with hybrid earning the highest weight due to its resilience and efficiency.

The *privacy metrics* section includes technical indicators like k-anonymity, l-diversity, t-closeness, differential privacy, indistinguishability, and adversary success probability—essential for quantifying re-identification risk and information leakage.

Multi-Criteria Decision Making (MCDM) techniques like AHP are used to assign weights across all factors, ensuring flexibility based on context.

Natural Language Processing (NLP) and Large Language Models (LLMs) to automate the extraction of key information from unstructured inputs such as hospital privacy policies, regulatory documents, and dataset metadata.

This diagram not only guides the risk quantification process but also lays the foundation for recommending tailored *Privacy Solutions*.

II. Privacy Solutions : (upcoming work, more research will be conducted once phase 1 is done): [4]

References:

https://estudogeral.uc.pt/handle/10316/118106
[1] J. W. Brady, “Securing Health Care: Assessing Factors That Affect HIPAA Security Compliance in Academic Medical Centers,” in *2011 44th Hawaii International Conference on System Sciences*, Jan. 2011, pp. 1–10. doi: 10.1109/HICSS.2011.368.
[2] “A Proposal for a Robust Validated Weighted General Data Protection Regulation-Based Scale to Assess the Quality of Privacy Policies of Mobile Health Applications: An eDelphi Study - PubMed.” Accessed: Mar. 21, 2025. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/37591261/
[3] M. Abughazalah, W. Alsaggaf, S. Saifuddin, and S. Sarhan, “Centralized vs. Decentralized Cloud Computing in Healthcare,” *Appl. Sci.*, vol. 14, no. 17, Art. no. 17, Jan. 2024, doi: 10.3390/app14177765.
[4] X. Gu, F. Sabrina, Z. Fan, and S. Sohail, “A Review of Privacy Enhancement Methods for Federated Learning in Healthcare Systems,” *Int. J. Environ. Res. Public. Health*, vol. 20, no. 15, Art. no. 15, Jan. 2023, doi: 10.3390/ijerph20156539.
[5] F. Zhang, Y. Chen, W. Meng, and Q. Wu, “Hybrid Encryption Algorithms for Medical Data Storage Security in Cloud Database,” Feb. 28, 2019, *Social Science Research Network, Rochester, NY*: 3388492. Accessed: Mar. 21, 2025. [Online]. Available: https://papers.ssrn.com/abstract=3388492
[6] I. Wagner and D. Eckhoff, “Technical Privacy Metrics: A Systematic Survey,” *ACM Comput Surv*, vol. 51, no. 3, p. 57:1-57:38, Jun. 2018, doi: 10.1145/3168389.
[7] A. Ali-Eldin, A. Zuiderwijk, and M. Janssen, “A Privacy Risk Assessment Model for Open Data,” in *Business Modeling and Software Design*, B. Shishkov, Ed., Cham: Springer International Publishing, 2018, pp. 186–201. doi: 10.1007/978-3-319-78428-1_10.
[8] H. S. Gardiyawasam Pussewalage and V. A. Oleshchuk, “Privacy preserving mechanisms for enforcing security and privacy requirements in E-health solutions,” *Int.*

J. Inf. Manag., vol. 36, no. 6, Part B, pp. 1161–1173, Dec. 2016, doi: 10.1016/j.ijinfomgt.2016.07.006.
[9] "Multi-Criteria Decision Making (MCDM) Methods and Concepts." Accessed: Mar. 21, 2025. [Online]. Available: https://www.mdpi.com/2673-8392/3/1/6
[10] "Art. 32 GDPR – Security of processing," General Data Protection Regulation (GDPR). Accessed: Mar. 22, 2025. [Online]. Available: https://gdpr-info.eu/art-32-gdpr/
[11] S. Alder, "HIPAA Encryption Requirements - 2025 Update," The HIPAA Journal. Accessed: Mar. 22, 2025. [Online]. Available: https://www.hipaajournal.com/hipaa-encryption-requirements/
[12] J. Sáinz-Pardo Díaz and Á. López García, "A Python library to check the level of anonymity of a dataset," *Sci. Data*, vol. 9, no. 1, p. 785, Dec. 2022, doi: 10.1038/s41597-022-01894-2.
[13] S. A. Simões, "Automated Data Privacy Protection using Deep Learning and Causality Techniques," in *Automated Data Privacy Protection using Deep Learning and Causality Techniques*, Sep. 2024. Accessed: Mar. 23, 2025. [Online]. Available: https://estudogeral.uc.pt/handle/10316/118106
[14] "Perplexity," Perplexity AI. Accessed: Mar. 23, 2025. [Online]. Available: https://www.perplexity.ai

Proposer un nom pour votre approche

Introduction

// là vous allez parler du contexte + problématique + objectifs

(une page max) avec à chaque fois vous intégrez des références à des travaux existants qui ont traité le même contexte et la même problématique

Etat de l'art

// cette partie sera divisée en deux sous parties :

- la première sera consacrée aux travaux de recherche existants pour la classification des attributs (LLM et autre) finir par un tableau qui montre les limites des approches trouvées
- la seconde traite les travaux étudiés pour faire l'organigramme

NB: N'oubliez pas d'ajouter les références des articles lus à chaque fois