

Article

A Privacy-Preserving and Attack-Aware AI Approach for High-Risk Healthcare Systems Under the EU AI Act

Konstantinos Kalodanis ^{1,*}, Georgios Feretzakis ², Athanasios Anastasiou ³, Panagiotis Rizomiliotis ¹,
Dimosthenis Anagnostopoulos ¹ and Yiannis Koumpourous ⁴

¹ Department of Informatics & Telematics, Harokopio University of Athens, 176 76 Kallithea, Greece; prizomil@hua.gr (P.R.); dimosthe@hua.gr (D.A.)

² School of Science and Technology, Hellenic Open University, 263 31 Patra, Greece; georgios.feretzakis@ac.eap.gr

³ Biomedical Engineering Laboratory, National Technical University of Athens, 157 72 Zografou, Greece; aanastasiou@biomed.ntua.gr

⁴ Digital Innovation in Public Health Research Lab, University of West Attica, 122 43 Egaleo, Greece; ykoump@uniwa.gr

* Correspondence: kkalodanis@hua.gr

Abstract: Artificial intelligence (AI) has significantly driven advancement in the healthcare field by enabling the integration of highly advanced algorithms to improve diagnostics, patient surveillance, and treatment planning. Nonetheless, dependence on sensitive health data and automated decision-making exposes such systems to escalating risks of privacy breaches and is under rigorous regulatory oversight. In particular, the EU AI Act classifies AI uses pertaining to healthcare as “high-risk”, thus requiring the application of strict provisions related to transparency, safety, and privacy. This paper presents a comprehensive overview of the diverse privacy attacks that can target machine learning (ML)-based healthcare systems, including data-centric and model-centric attacks. We then propose a novel privacy-preserving architecture that integrates federated learning with secure computation protocols to minimally expose data while ensuring strong model performance. We outline an ongoing monitoring mechanism compliant with EU AI Act specifications and GDPR standards to further improve trust and compliance. We further elaborate on an independent adaptive algorithm that automatically tunes the level of cryptographic protection based on contextual factors like risk severity, computational capacity, and regulatory environment. This research aims to serve as a blueprint for designing trustworthy, high-risk AI systems in healthcare under emerging regulations by providing an in-depth review of ML-specific privacy threats and proposing a holistic technical solution.

Keywords: healthcare; high-risk AI systems; EU AI Act; privacy; federated learning; secure computation; privacy attacks; intelligent algorithms



Academic Editor: Domenico Rosaci

Received: 26 February 2025

Revised: 26 March 2025

Accepted: 27 March 2025

Published: 30 March 2025

Citation: Kalodanis, K.; Feretzakis, G.; Anastasiou, A.; Rizomiliotis, P.; Anagnostopoulos, D.; Koumpourous, Y. A Privacy-Preserving and Attack-Aware AI Approach for High-Risk Healthcare Systems Under the EU AI Act. *Electronics* **2025**, *14*, 1385. <https://doi.org/10.3390/electronics14071385>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Artificial intelligence (AI) is now an essential and integral component of the modern healthcare world, making significant contributions to industries by assisting in the implementation of intelligent and sophisticated algorithms. These algorithms have the capacity to enhance and create a large number of clinical and operational processes that range from performing sophisticated analytics that help with diagnostics to enabling automated systems of patient monitoring, as well as delivering personalized and tailored treatment plans for patients. These technologies can process patient data to identify subtle patterns and associations, improving decision-making and patient outcomes. However, the reliance

on sensitive health information and automated decision-making elevates such applications to the category of “high-risk AI systems”, particularly in jurisdictions considering more rigorous legislation like the European Union Artificial Intelligence Act (EU AI Act) [1]. The EU AI Act covers regulations for the assurance that AI technologies, especially those significantly impacting the safety, health, and fundamental rights of the populace, adhere to high levels of standards for transparency, risk handling, data governance, and accountancy [2].

In parallel, privacy and data protection remain pressing concerns. The General Data Protection Regulation (GDPR) sets the foundational framework for processing personal data in the EU. However, integrating GDPR obligations with the EU AI Act requirements introduces new challenges [3]. Healthcare systems using AI generally need massive patient data inputs for maximum model effectiveness, and the balance required for using patient data versus protecting patient privacy is a complicated challenge. In addition, the lack of explainability for AI models can also undermine patient trust and make regulatory conformity complicated when the underlying cause for the decision is not easily explainable [4].

Despite the benefits acknowledged, the need is essential to assess the ethical and societal implications involved in the use of high-risk AI tools in the health sector. In general, the risk for implicit prejudice, algorithmic prejudice, and the likelihood for massive data exposures highlights the need for building robust systems focused on the importance of privacy and openness. In addition, the challenge for progress under the set by ethical and regulatory guidelines highlights the need for the medical sector to bring its practice into line with the guidelines for AI adoption.

The rapid advancement in AI systems, especially in the field of deep learning and transformer architectures, adds more levels of complexity around these concerns. These complex models often require much larger datasets, thus exacerbating concerns around data interchangeability and compatibility by stakeholders operating in the medical sector. One main challenge faced by these stakeholders is the balance required when using vast datasets for improvement in predictive capacity and the need to protect patient data from exposure. These concerns find their solution under the European Union AI Act, where the regulatory lifespan for AI is described through risk classification, detailed documentation, continuous oversight, and after-market surveillance of AI systems [5]. This regulatory structure not only mandates the level of compliance during the launch but also expects continuous commitment towards mitigation against risk, openness, and accountability. Contemporary healthcare is being complemented by AI-supported models focused on optimizing patient care and administrative processes. This is made possible through the increased adoption of electronic health records (EHRs), wearable devices, and mHealth platforms, all of which generate vast quantities of patient-focused data each day. These vast data allow for the specialized use of AI for optimizing patient outcomes through enhanced diagnostic capability, customized treatment plans, and predictive analysis for the ability to recognize health concerns even before their deterioration. However, the vast data coupled with the complexity around the use of machine learning models increases concerns over data security, potential data breaches, and algorithm fairness. Therefore, the regulatory environment provided by the EU AI Act plays an essential part in resolving the concerns by providing not only high efficiency expectations but also fairness, auditability, and data security considerations for AI applications under the medical sector.

Research has proposed various technical solutions—differential privacy, homomorphic encryption, and federated learning—to mitigate tensions between data utility and privacy [6,7]. Yet the question remains: how can these technologies be embedded into a comprehensive system architecture that aligns with both the GDPR and the emerging provisions of the EU AI Act, particularly for high-risk applications? We acknowledge that integrating Privacy Enhancing Technologies (PETs) such as differential privacy, federated

learning, secure multi-party computation, and homomorphic encryption into a unified, robust, and compliant architecture involves inherently complex challenges. Healthcare data, though individually sparse, are indeed fragmented across multiple locations, institutions, and temporal scales. The aggregation, exchange, and storage of these distributed data substantially escalate privacy risks. Recognizing these intrinsic complexities, we emphasize that our proposed solutions aim to provide foundational guidance and initial technical frameworks rather than definitive, exhaustive resolutions to all privacy concerns in healthcare AI. While federated learning can limit direct data exposure by enabling distributed training across institutions, it does not fully address transparency, risk assessment, or suitability demands. Similarly, methods for secure computation—such as homomorphic encryption or secure multiparty computation—offer unique privacy advantages but can introduce performance bottlenecks, complicate system design, and fail to address accountability mechanisms that the legislation requires.

More broadly, Privacy Enhancing Technologies (PETs) constitute a diverse range of methods to preserve sensitive information and facilitate privacy law compliances. A study [8] presents a comprehensive theoretical framework describing how PETs such as differential privacy, homomorphic cryptography, federated learning, and secure multi-party computation collectively enhance both privacy and security, particularly healthcare scenarios wherein information exchange is both vital and strictly regulated.

In addressing the many intricate concerns, the work here presents an overall analysis of the various forms of privacy violations potentially encountered by machine learning medical systems, separating violations against data privacy from those against model privacy. Additionally, this study introduces the concept of a privacy-preserving architecture specially designed for high-risk medical applications for artificial intelligence. Our framework integrates federated learning to minimize direct data transfers, incorporates secure computation protocols for additional privacy safeguards, and includes an auditing and explainability subsystem designed to comply with the GDPR and future EU AI Act provisions. In addition, we introduce an adaptive cryptographic algorithm that can dynamically modulate security levels based on real-time context, such as resource availability, the urgency of clinical decisions, and evolving regulatory constraints.

The remainder of this paper is organized as follows: Section 2 reviews related work and highlights the regulatory requirements driving system design. Section 3 provides a detailed overview of privacy attacks in machine learning, distinguishing between attacks on data and model privacy. Section 4 offers a new perspective on how AI adoption in healthcare has progressed from theoretical foundations to practical deployment. Section 5 introduces our proposed framework architecture and describes its main components. Section 6 presents our novel adaptive algorithm that enhances the system's overall security beyond the basic framework. Section 7 discusses the insights, challenges, and implications of aligning this approach with the EU AI Act. Finally, Section 8 concludes this paper by summarizing our contributions and proposing additional directions for future research.

2. Background

The starting point of modern artificial intelligence can be traced back to Rosenblatt's 1958 work [9] in neural computing when he introduced Perceptron as part of one of the very first models of neural networks. The subsequent evolution of artificial neural networks (ANNs) laid down the groundwork that led to today's machine learning and artificial intelligence platforms. Although modern AI research frequently emphasizes deep learning methods due to their effectiveness with large datasets, the issue of healthcare data introduces unique challenges. While cumulatively vast, health-related data are often fragmented

and dispersed geographically across institutions and temporally across different periods, complicating secure aggregation and analysis.

2.1. High-Risk AI Systems in Healthcare

Healthcare systems by their very nature fall under high-risk activities because their impact is immediate and direct on the lives and general wellbeing of the populace, being a fundamental component of the operation of society [10]. The utilization and deployment of AI tools for treatment suggestions and diagnostic categorization can cause the diseases to be misclassified. Misclassification is very dangerous, resulting potentially in improper or tardy medical treatment for the patient when the patient is counting on the right and prompt treatment. These systems fall under the European Union's AI Act under the stringent regulatory structure, where the developer and operator need to demonstrate conformity in risk management, data governance, and dependability in their operation [11]. Although existing standards (e.g., ISO 13485 for medical devices [12]) address some of these needs, the EU AI Act introduces requirements specifically targeted at AI, including traceability, continuous monitoring, and human oversight.

In practice, the "high-risk" designation means that such healthcare AI systems should incorporate continuous validation and monitoring protocols. For instance, all disparities in the effectiveness of the system need to be noted and analyzed for the cause, thus ensuring that all possible negative influences or inaccuracies do not build up over the lifespan. This is beyond the initial development phase and ongoing for the lifespan of AI, where ongoing updates and risk analysis need to maintain compliance.

2.2. Privacy Concerns in AI-Driven Healthcare

Artificial intelligence platforms typically depend on vast databases, thus escalating the chances of infringing patient confidentiality, especially when handling sensitive health data [13]. Confidential data held in the health data of the patient can be breached through improper access, possibly resulting in serious consequences like identity theft and discrimination. AI models call for transparency; AI developers are also under pressure to protect proprietary models, which may conflict with the principle of "right to explanation" or potential auditing under the EU AI Act.

Recent studies emphasize that privacy risks are not solely associated with direct data leakage but can manifest indirectly. For instance, membership inference attacks exploit trained model parameters to determine whether a specific data record was included in the training dataset [14]. Model inversion attacks can potentially reconstruct and reveal some features from the data inputs, thus posing the significant and serious risk of compromising the patient data's confidentiality. These very complex and sophisticated attacks highlight the importance of, and bring sharply into relief, the essential need for the adoption of a multi-layer security defense technique. This technique could ideally combine the best features from federated learning paradigms, cryptography tools, and robust regulatory compliances, all for the sole purpose of providing the maximum level of security and defense against possible invasions.

2.3. State-of-the-Art Privacy Approaches

Federated learning reduces data centralization by enabling decentralized model training across multiple institutions [15]. Yet, without additional privacy safeguards like secure multiparty computation or differential privacy, model parameters shared during aggregation may inadvertently expose sensitive information. Homomorphic encryption allows for computations on encrypted data, ensuring that no raw data are disclosed, but the computational overhead can be significant, particularly when dealing with large neural networks [16]. Secure multiparty computation (SMPC) relies on distributed cryptographic

protocols to perform collaborative computations, but it typically requires substantial communication overhead and intricate protocol design [17].

While these methods offer significant advantages, they address different aspects of the privacy problem. Federated learning primarily addresses data minimization, differential privacy mitigates the risk of re-identification, and secure computation ensures that even intermediate computation steps remain confidential. Consequently, practitioners often need to combine these methods to achieve an end-to-end solution that meets the strict “high-risk” classification standards.

Additionally, the sections of explainability and interpretability remain the subject of widespread scientific study. Cutting-edge AI tools, especially deep neural networks, often act like “black boxes”, posing the challenge of understanding the processes through which specific outcomes are achieved. In the context of the EU AI Act, the ability to provide understandable descriptions is one fundamental prerequisite for high-risk classification systems [18]. For this reason, good explainability tools, like saliency maps and local interpretable model agnostic explanations (LIMEs), are increasingly seen as integral parts of the overall regulatory environment for AI.

In modern scenarios, many widespread applications of AI in the health sector—from AI-assisted radiology to patient flow forecasting—have highlighted the beneficial inputs machine learning can provide towards optimizing the efficiency of healthcare. In addition, the applications also highlighted the dangers associated with data centralization, which could potentially act as points for failures. Scholars have posited the idea that the combination of federated learning and robust cryptography can enable the collective building of models in a non-centralized environment, thus limiting the dependence upon local data repositories, yet these solutions frequently focus on either differential privacy or homomorphic encryption in isolation rather than a concerted strategy that addresses the entire compliance spectrum demanded by regulations like the EU AI Act.

3. Privacy Attacks on ML

In this paper, we introduce a complete list of the various forms of attacks against machine learning (ML)-based models. We categorize the various attacks into two dimensions: data (i.e., data privacy) and model (i.e., model privacy). The following several subsections explain each type independently.

3.1. Data Privacy

In healthcare, protecting data privacy is a top priority. Data owners wish to secure the data before training or deploying ML models in many scenarios. For instance, consider a medical research project or a specialized model trained on electronic health records. If the party that owns the data differs from the party that processes it, the private data might still be transmitted over a secure channel yet remain unencrypted on the computing server’s end. This exposes the data to both insider and outsider threats. Consequently, the data are thus compromised, leaking features, membership data, or the very beliefs held by the dataset. These forms follow the list of the fundamental data-attacking forms for privacy: re-identification (de-anonymization), reconstruction, and property inference.

3.1.1. Re-Identification/De-Anonymization

Re-identification attacks attempt to reverse an existing de-identification process. In this scenario, the data are shared under the assumption that they have been anonymized. However, an adversary uses auxiliary information or statistical methods to link data back to specific individuals.

Recent advancements in linkage attacks have shown that even partial datasets can be cross-referenced with publicly available records to pinpoint the identity of a patient, sometimes matching as few as three non-unique attributes [19]. In healthcare, this risk grows significantly when multiple hospitals participate in collaborative initiatives, sharing partially de-identified patient information. Although anonymization techniques like *k*-anonymity, *l*-diversity, and *t*-closeness remain prominent, studies have demonstrated that these techniques can be circumvented, especially in the presence of high-dimensional data [20]. For instance, an attacker might use unique combinations of features—such as ZIP codes, birth dates, or rare medical conditions—to match records to specific individuals.

Moreover, re-identification attacks are no longer limited to traditional tabular data. Emerging research indicates that textual and imaging data are equally vulnerable, with sophisticated natural language processing (NLP) and computer vision models used to re-link masked data back to their source [21]. These developments underscore the need to enforce strict control over data-sharing protocols and to implement robust encryption whenever feasible.

3.1.2. Reconstruction Attacks

Reconstruction attacks involve rebuilding substantial portions of the dataset. Even where identifier fields will not necessarily be disclosed, the attacker can use patterns, partial disclosures, or even the models' outputs for the sensitive portion reconstruction for each subject [22]. These data fragments can pose serious concerns over privacy, perhaps even leaking latent health diseases or private information.

One hallmark of reconstruction attacks is their ability to capitalize on aggregate statistics released as part of standard data-sharing practices [23]. For instance, when healthcare institutions publish summary statistics or partial model gradients—often in the context of federated learning—an attacker can combine these fragments with auxiliary data to approximate entire patient profiles. Recent work has shown that reconstruction can even occur in near-real-time, utilizing advanced generative models to fill in missing details in patient profiles [24]. This has led researchers to emphasize differential privacy mechanisms that rigorously limit the amount of information any model output can reveal about individual data entries.

An additional consideration is the data reconstruction potential in imaging exams, which is very relevant for radiology and pathology. An attacking party could theoretically reconstruct partial medical images showing tumors or other sensitive data and thus make conclusions about the patient's health. To mitigate the foregoing dangers, some institutions incorporate both perturbations and encryptions into their image-sharing processes. However, finding the balance between data distortion, the minimum required for the preservation of models' accuracies, and robust security remains an open research question [25].

3.1.3. Property Inference Attacks

Property inference attacks allow adversaries to deduce characteristics of the training dataset that were not explicitly targeted during the learning process. For instance, if a deep learning model was trained to perform face recognition, the dataset might not explicitly record how many subjects wear glasses. Yet an attacker could still glean such information by probing model outputs across multiple inputs.

A key driver behind property inference attacks is the fact that neural networks often encode latent patterns beyond their primary classification or regression tasks [26]. In a healthcare setting, even benign features such as demographic distributions or lifestyle factors (e.g., smoking and alcohol consumption) could be exploited if an attacker can

systematically query the model. In 2022, a new approach called gradient-based property extraction was proposed, showing that if adversaries obtain access to intermediate gradient updates in a federated environment, they can infer population-level statistics within a few training iterations [27]. This finding raises crucial questions about how gradient-sharing protocols should be designed, especially in collaborative healthcare research, where multiple institutions synchronize updates for a global model.

Moreover, property inference attacks can worsen fairness concerns, as they could expose implicit correlations used by the model that relate to sensitive demographic or socioeconomic attributes [28]. This possibility necessitates the integration of both privacy and fairness auditing tools within the learning pipeline, ensuring that the model remains compliant with ethical and regulatory expectations.

3.2. Model Privacy

In the context of machine learning, the idea of privacy goes beyond the data alone and covers the models themselves, including architectures and learned parameters. For example, machine-learning-as-a-service hosted by the cloud often incurs fees for the use of models through API; thus, the stolen or reverse-engineered model can threaten the provider's revenue and reveal confidential insights.

3.2.1. Model Extraction Attack

Also known as a black-box attack, a model extraction attack occurs when an adversary repeatedly queries an ML model to replicate its behavior. By capturing input-output pairs, the attacker trains a substitute model that closely approximates the original.

Recent studies show that model extraction can be performed even with limited queries, using adaptive sampling strategies that identify the model's decision boundaries more efficiently [29]. In healthcare, this is particularly concerning when AI solutions are integrated into telemedicine platforms or diagnostic APIs where usage-based fees apply. Aside from not needing to pay for the licensing cost, the adversaries may also be able to obtain valuable patterns that were obtained by mistake during the training of the models.

3.2.2. Membership Inference Attack

Membership inference attacks determine whether a specific record was part of the training dataset. In healthcare, knowing that someone's data were used for training can disclose they were once admitted to a particular hospital or received specific treatments.

Recent research in 2022 has shown that membership inference can succeed with high accuracy under realistic settings, especially in over-parameterized or poorly regularized models [30]. As an additional point of interest, the membership is made even more vulnerable when the federated learning scenario is taken into consideration, as it allows for access to partial global models. This highlights the importance of implementing differential privacy or cryptography for each and every update to the training iteration database.

3.2.3. Model Inversion Attack

Model inversion aims to recover or estimate the original data from the models' resulting outputs. This is essentially a serious violation of privacy, given the potential exposure of confidential and sensitive data, including medical imaging or genomic data.

In 2022, an enhanced form of inversion was demonstrated, which uses a two-step process: first, approximate class-level features from the model, and second, refine these features via generative adversarial networks (GANs) to match real-world data distributions [31]. Within the medical field, attackers can potentially generate similar medical images associated with a given pathology, thus acquiring sensitive patient medical data from the dataset.

3.2.4. Shadow Model Attack

A shadow model attack employs a secondary model—trained on data the attacker controls—to mimic the target model. Once the shadow model is sufficiently accurate, the attacker can probe it more extensively, circumventing rate limits or detection on the target system.

Since 2022, more sophisticated shadow model frameworks have appeared, leveraging transfer learning to drastically reduce the data requirements for building an effective mimic model [32]. In healthcare scenarios, even minimal gleaned insights—such as threshold values for positive diagnoses—can compromise both patient privacy and the proprietary edge of specialized AI solutions.

3.2.5. Adversarial ML Attacks

Adversarial ML attacks involve feeding imperceptibly modified inputs into a trained model, thereby causing misclassifications [33]. In healthcare, such manipulations could mask signs of disease or generate false positives in diagnostic systems.

Recent breakthroughs in adversarial attack generation rely on gradient-based methods that systematically identify “weak spots” in a model’s decision boundary [34]. Notably, these can be adapted to clinical imaging tasks, where minute pixel perturbations may be indistinguishable from the human eye but drastically change AI-driven diagnostic outcomes. Furthermore, with the rise of multi-modal healthcare data (e.g., combining text, image, and sensor streams), adversarial approaches have become increasingly flexible, targeting multiple data channels concurrently [35].

3.2.6. Membership Memorization Attack (MMA)

This study [36] introduced a distinct perspective on how models “memorize” data. In an MMA, adversaries aim to determine whether a specific record was part of the training process by exploiting the model’s tendency to store fine-grained details from the training set.

In 2022, researchers discovered that large language models and large-scale vision transformers could inadvertently memorize unique training samples if they were repeated enough times in the dataset or if the training included specific “trigger phrases” [37]. In healthcare, large clinical language models used for textual patient histories or pathology reports may inadvertently reveal confidential patient identifiers if they are not carefully sanitized before training. As these models grow larger, the memorization phenomenon—and thus privacy risk—can intensify.

3.2.7. Model-Reuse Attacks

Model-reuse attacks deliberately embed hostile logic—sometimes called “backdoors”—into reusable model components or feature extractors. When integrated into a larger ML pipeline, a malicious actor can trigger specific outcomes by providing a “trigger” input. More recent work from 2022 demonstrates that advanced backdoor injection can remain latent through multiple fine-tuning rounds, only to be activated at a specific inference step [38]. In a healthcare context, this can be calamitous: the diagnostic paradigm annotating silently after some “trigger” pattern is seen during the scan, underestimating the sickness severity or overestimating the sickness severity. For this, secure model provenance checks, layered cryptography, and strong anomaly detection are needed to resist model-reuse attacks.

Figure 1 provides a high-level taxonomy of privacy attacks across the machine learning lifecycle, grouped into Data Privacy Attacks (re-identification/de-anonymization, reconstruction, property inference) and Model Privacy Attacks (model extraction, membership

inference, model inversion, shadow-model attacks, adversarial ML, membership memorization, and model-reuse attacks).

Privacy Attacks on Machine Learning in Healthcare

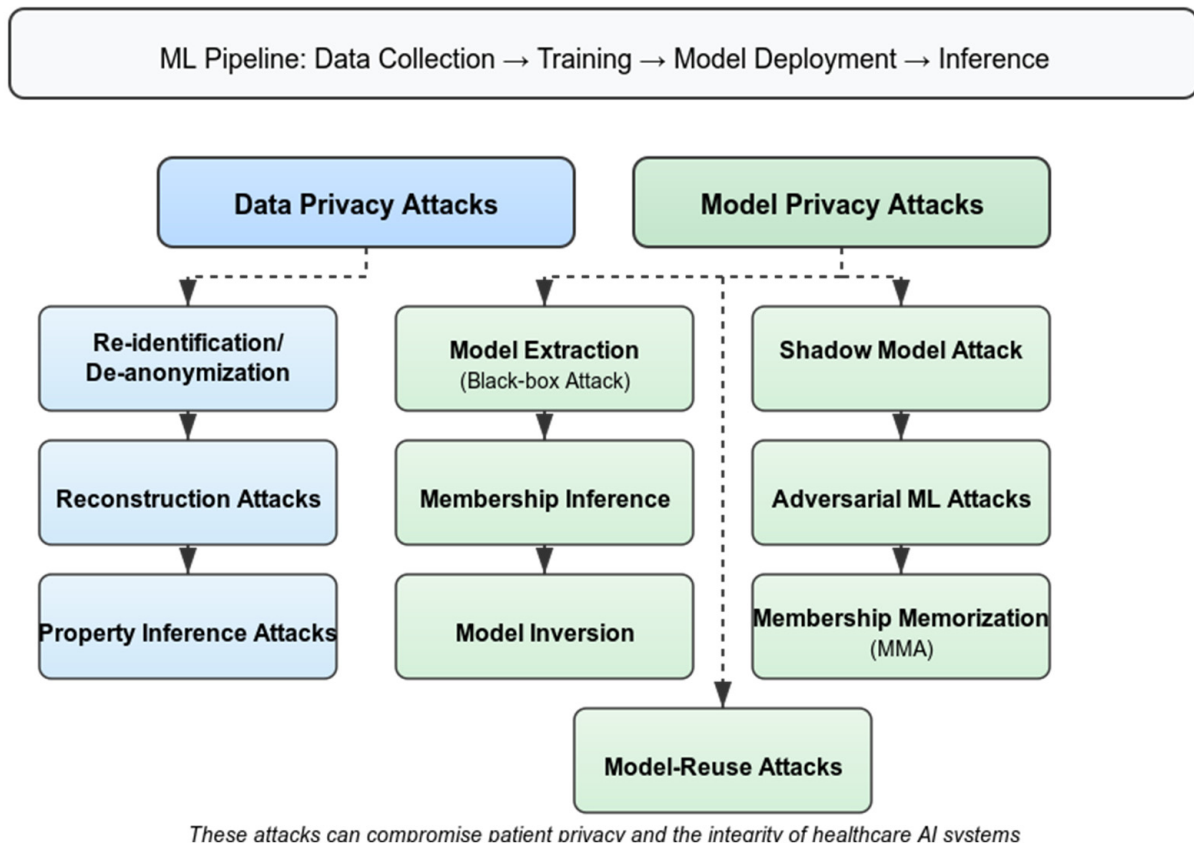


Figure 1. Overview of Data Privacy and Model Privacy Attacks in the ML pipeline.

Overall, these privacy attacks highlight the complex interplay between data security, model integrity, and patient confidentiality in modern AI-driven healthcare systems. By examining both data-centric and model-centric attacks, stakeholders can design more resilient architectures—incorporating encryption, secure multi-party computation, differential privacy, or other advanced mitigations—that align with both technological best practices and regulatory mandates in the EU AI Act.

4. AI Adoption in Healthcare: From Theory to Practice

The growing influence of AI in the healthcare domain extends far beyond theoretical proofs of concept and controlled laboratory experiments. Over the past few years, substantial progress has been made in translating ML-driven methods into clinically actionable tools. This change is being driven by the convergence of a number of factors, such as the development of increasingly sophisticated computing infrastructure, the accumulation of large-scale medical databases, the evolution of regulatory frameworks, and the ever-increasing need for care that is centered on the patient. The transition from conceptual models to actual implementation is fraught with challenges that require an extensive analysis of clinical validation, ethical and regulatory considerations, and a strong information technology infrastructure.

4.1. Clinical Validation and Real-World Data

Clinical validation of AI solutions is crucial to determine their reliability, applicability in diverse settings, and safety for use in real-world patient care. Although the initial efficacy of a model can be gauged with small-scale tests or simulated datasets, these are inadequate to represent the complex realities involved with actual medical data. Current research has highlighted the necessity of collective approaches from multiple institutions to collect extensive patient records and thus minimize the biases that can be introduced through the use of homogeneous datasets [39].

However, the quest for diversity in datasets is hindered by several privacy issues and logistical challenges. Diverse data schemas are normally seen in healthcare facilities, leading to the challenges of uniform data curation. In addition, the chances of privacy violations increase with every new dataset aggregation, particularly when such data are from different entities with no shared security system. Federated learning has been proposed as a possible solution, where model training is possible with collaborative efforts without raw data aggregation. Federated environments require additional investment in cryptographic computations, secure aggregation protocols, and data normalization.

A second challenge to clinical validation is matching the exact parameters of algorithmic performance measures to the broad clinical outcomes. An algorithm highly successful in disease progression or onset prediction in a controlled setting may not be able to cope with confounding factors or clinical protocol variability between sites. In actual application, broad validation includes not only sensitivity and specificity but also the effect of the model on patient care pathways, treatment cost, and overall quality of life. This broad validation process is increasingly being adopted by researchers and regulatory agencies, and therefore, increasingly, there is a need for open reporting of therapeutic benefit and possible adverse effects [40].

In addition, the evaluation of real-world applications needs to consider the ongoing evolution of medical knowledge and techniques. The emergence of new drugs or diagnostic methods can reduce the efficacy of an artificial intelligence model based on past data, requiring constant retraining and model updates. In a federated setting, organizations need to securely and reliably handle cryptographic keys, model versions, and audit logs during the upgrade process. The complex interplay between changing data environments and the need for effective privacy-preserving mechanisms highlights the challenges of real-world deployments.

Furthermore, real-world validation must consider the ongoing advancement of medical knowledge and procedures. The introduction of new medicines or screening techniques can diminish the relevance of an AI model built on historical data, requiring continuous retraining and adjustment. In a federated context, institutions must consistently and securely maintain cryptographic keys, model versions, and auditing logs during upgrades. The interaction between evolving data environments and the necessity for effective privacy-preserving solutions underscores the intricacy of practical implementation.

4.2. Ethical and Regulatory Hurdles

As artificial intelligence transitions from the realm of abstractions to fundamental decision-making components of healthcare protocols, the risk of ethical practice and regulatory compliance grows. In the European setting, the upcoming EU AI Act puts into the spotlight the need to develop open standards for AI systems with a direct impact on human health or wellbeing. This is supported by far-reaching regulations such as the GDPR, which address the issue of protecting personal data. Developers thus face a sophisticated compliance environment in which any transgression—regarding data use, algorithmic bias, or user transparency—can have draconian legal and ethical repercussions [41].

A major ethical concern involves the potential exacerbation of health disparities. Even if a model is shown to be validated across datasets, certain subgroups—e.g., ethnic minorities or those with uncommon diseases—may be underrepresented. The underrepresentation of these subgroups may lead to reduced predictive performance of the model, ultimately resulting in suboptimal or even harmful results. The resolution of these disparities requires thorough fairness analysis during both development and implementation. This process encompasses meticulous collection of demographic data, application of bias mitigation strategies such as reweighting and adversarial debiasing, and ongoing performance evaluation across varied patient demographics [42].

In addition, transparency and interpretability have become essential drivers for building trust between patients and healthcare professionals. While black-box models exhibit high accuracy, their lack of transparency could be a source of suspicion if their internal workings are not revealed. Healthcare professionals need explanations that connect predictions to meaningful clinical scenarios. Researchers are exploring explainable AI (XAI) methods—such as attention mechanisms, saliency maps, or instance-level contributions—to reduce the gap between the output of a model and human comprehension. However, the use of secure encryption techniques (e.g., homomorphic encryption or secure multiparty computation) can make some aspects of interpretability invisible, and novel solutions that balance security with transparency need to be developed.

Finally, the accountability principle is of significant importance. In high-risk environments, determining the accountable party for errors—whether it be healthcare provider, the AI vendor, or the data custodian—poses a significant challenge. The European Union's AI Act and supporting guidelines emphasize the importance of well-defined liability frameworks, and developers are required to outline the model development process, testing procedures, and the subsequent modifications. These legal frameworks, along with well-defined ethical guidelines, are intended to make the environment safer but also add more burdens to developers and healthcare organizations to remain in continuous compliance and to document extensively.

4.3. Infrastructure and Interoperability

The successful deployment of artificial intelligence in the healthcare industry depends on a strong, secure, and flexible infrastructure. The majority of healthcare institutions still struggle with issues surrounding legacy systems that lack native encryption, standardized interfaces, or adequate computational power needed for the execution of advanced algorithms in real time. Plugging AI solutions into such an infrastructure requires an integrated approach to system upgrades, such as hardware accelerators (e.g., GPUs or TPUs. Graphics Processing Units, computer chips that can handle big datasets efficiently, are the major drivers for the significant ramp-up in computational power required for the training of deep models. Google has produced Tensor Processing Units that are custom-designed integrated circuits designed specifically for machine learning purposes, providing better efficiency and performance during the training process and the inference process involved in the operation of deep neural networks. GPUs and TPUs enable faster computations and improved scalability of AI algorithms (essential for real-time decision-making and predictive analytics in healthcare applications) and software components (e.g., container orchestration platforms such as Kubernetes 1.32.2) that facilitate deployment and scalability.

Interoperability, however, is the enabler of mass adoption of AI. A number of different healthcare professionals work with different electronic health record systems and data models, including Health Level Seven (HL7), Fast Healthcare Interoperability Resources (FHIR v5.0.0), or Digital Imaging and Communications in Medicine (DICOM for imaging) [43]. HL7 is a set of international standards for the sharing, integration, exchange, and retrieval of electronic health data, and the primary focus remains on clinical and administrative data. FHIR, on the other hand, is a new standard that has been created by HL7 International on top of current web technologies in order to make exchanging data faster and more efficient by overcoming some weaknesses in earlier versions of HL7. Conversely, DICOM is the standard for medical imaging data storage, modification, and transmission to guarantee fit with imaging systems and devices. Data types and formats frequently exchanged within these interoperable healthcare systems include JavaScript Object Notation (JSON), Extensible Markup Language (XML), YAML Ain't Markup Language (YAML), Comma-Separated Values (CSVs), plain text files, and binary image formats (e.g., JPEG, PNG, TIFF). These diverse data format facilitate standardized communication, structured storage, and flexible integration across heterogeneous healthcare systems. Integration of this heterogeneous formatting will help to enable continuous data exchange. Standardized APIs such as Representational State Transfer (REST) APIs that enable efficient and stateless communication using JSON or XML data formats, or Simple Object Access Protocol (SOAP) APIs that rely on XML messaging for the sake of structured information exchange, can be utilized to normalize a variety of sources. REST APIs are widely adopted due to their simplicity, scalability, and suitability for web-based applications. On the other hand, SOAP APIs offer built-in standards regarding the security and compliance of transactions, making them suitable for applications that require strict formal agreements.

The importance of security is high, especially in the context of application programming interfaces (APIs). Authentication and authorization mechanisms, such as JSON Web Tokens (JWTs), OAuth 2.0, and OpenID Connect 1.0 (OIDC), have been very popular. JWT provides a compact and URL-safe way of expressing claims between parties involved, thus for allowing secure and efficient identity authentication and access control in distributed systems. OAuth 2.0 provides secure delegated access without the need for direct sharing of credentials, while OIDC builds upon OAuth 2.0 by adding an identity layer, thus allowing for user authentication. Additionally, API gateways, in combination with the Transport Layer Security (TLS) protocol, secure communications, thus protecting sensitive health data from unauthorized access and cyber attacks. The task becomes increasingly difficult when cross-border data sharing is involved, since different jurisdictions have different privacy legislation. Such cross-border complexities bring to the forefront the need for multi-level compliance protocols and adaptive cryptographic solutions that can adapt to local legislation without sacrificing collaborative potential.

Another dimension of infrastructure preparedness is fault tolerance and resilience. Healthcare systems should not be subject to extended service outages due to encryption activities or network overload. With additional artificial intelligence technologies being deployed—e.g., automated alerts in intensive care units—healthcare organizations need to ensure that any breakdown of a cryptographic framework, model-serving interface, or network link does not compromise patient safety. Distributed storage, redundancy of key elements, and fallback procedures (e.g., fallback to older protocols in the case of an AI system failure) can mitigate these threats; however, these measures add an additional layer of design complexity.

The move from pilot projects to enterprise-wide, large-scale, artificial intelligence solution deployment requires technical ability and readiness within the organization and strong financial justification. Clinicians, information technology professionals, and administrative

staff need to work together to determine which AI technologies bring significant improvements in patient care and financial performance. For instance, the rollout of an AI-powered triage system in emergency rooms has the potential to reduce wait times significantly and reduce human errors; however, the development of real-time federated analytics for multi-site data may require significant operational costs and involve significant training. It is a challenging endeavor to find the right balance between these elements, which more often than not requires incremental rollouts and ongoing return on investment analysis.

The above considerations highlight the fact that the integration of artificial intelligence in the healthcare field is a multifaceted effort that requires harmonization of clinical, ethical, regulatory, and infrastructural factors. The shift from proof-of-concept validation platforms to widespread clinical use involves ongoing interaction among researchers, policymakers, clinicians, and technology vendors. While recent evidence points toward the potential of AI in radiomics, genomics, and long-term patient monitoring [44], the successful deployment of these technologies will be supported by a strategic plan that addresses not just efficacy but also patient safety, data confidentiality, and system reliability.

5. Proposed Privacy-Preserving Framework

We now turn to our Privacy-Preserving Federated Learning (PPFL) architecture, which aims to satisfy four key principles: (1) Data Minimization; (2) Security and Confidentiality; (3) Transparency and Auditability; and (4) Regulatory Compliance with GDPR and the EU AI Act. The framework comprises three main components:

- Federated Learning Coordinator (FLC);
- Secure Computation Layer (SCL);
- Compliance and Explainability Module (CEM).

Federated Learning, which has received wide recognition largely owing to the novel federated averaging algorithm that was pioneered by Google, enables a collaborative model training strategy that encompasses several distributed institutions collaborating with one another. This high-level technique is accomplished without direct information-sharing or pooling of raw information being demanded, something that is a big positive in terms of privacy protection. Having raw patient information localized and contained within their facilities, this technique effectively pre-empts potential privacy risks that would otherwise be dealt with by data-sharing mechanisms. The locally trained models, that is to say those trained on-site at each institution, are pooled centrally. The exact aggregation procedure guarantees that privacy and security of sensitive information are strictly maintained at all times in cooperation. In addition, there is a thorough overview presented on federated machine learning in the context of the healthcare domain specifically, encompassing a detailed examination of its clinical applications, as well as the complex technical architecture underlying it [45].

5.1. Federated Learning Coordinator

The FLC is responsible for initiating training rounds, distributing the global model to local nodes at each healthcare institution, aggregating local weight updates, and disseminating the updated global model. A unique aspect of our coordinator is the dual-layer aggregation mechanism; before the final aggregation step, each local update is encrypted and partially aggregated using SMPC protocols, reducing the risk of single-point data leaks.

One of the critical novelties of the FLC lies in its capability to orchestrate cross-institutional training schedules based on dynamic data availability. This ensures that hospitals with lower patient volumes or delayed data uploads can still participate without stalling the entire training process. Recent work has highlighted how scheduling algorithms can optimize both performance and communication overhead in federated learning,

particularly in environments with heterogeneous network conditions [46]. By integrating such scheduling heuristics into our FLC, we aim to maintain a balanced computational workload across all participants, preventing bottlenecks and enhancing scalability.

In addition, the FLC has a risk assessment function that makes use of minimal metadata (e.g., dataset size and sensitivity classification) to determine the required level of encryption. If the metadata indicate the existence of highly sensitive types of information—e.g., genomic data or sets of orphan diseases—the FLC automatically enforces stronger cryptographic practices. This adaptive approach is in line with the Adaptive Crypto Orchestrator (ACO) presented in Section 6, thus ensuring that the FLC model distribution aligns with contemporaneous risk assessments.

5.2. Secure Computation Layer

To address the privacy and security concerns that remain in standard federated learning, the Secure Computation Layer integrates Homomorphic Encryption, SMPC, and Differential Privacy. These components ensure that model updates and intermediate computations remain confidential, even in untrusted or partially malicious environments.

In designing the SCL, we emphasize modularity so that individual healthcare institutions can choose different encryption tiers based on their computational resources. For instance, a resource-rich institution might opt for fully homomorphic encryption (FHE) to maintain maximum confidentiality, while smaller clinics with limited computing infrastructure may choose partially homomorphic encryption schemes with reduced overhead. Such a tiered approach aligns with emerging best practices in hybrid healthcare consortia, where not all nodes possess identical IT capabilities [47].

Additionally, the SCL manages a privacy budget for differential privacy mechanisms, dynamically adjusting the noise injection levels to preserve model utility. If the training process detects that the global model's accuracy falls below a predefined threshold, the SCL can temporarily relax noise parameters, maintaining a minimal guarantee of privacy while re-calibrating for acceptable performance. Conversely, if membership inference risks rise (as flagged by the CEM's risk modules), the SCL escalates noise levels for subsequent training epochs.

5.3. Compliance and Explainability Module

The CEM acts as a post hoc auditing system that verifies whether certain practices align with EU AI Act stipulations, including risk management, transparency logs, and an explainability interface. Clinicians and healthcare administrators can use a user-facing dashboard to understand high-level reasoning behind AI-driven decisions.

Our design for the CEM adopts a multi-level logging approach; every model update and cryptographic key exchange is recorded in an immutable ledger, accessible only to authorized auditors. Each hospital encrypts their data locally using institution-specific cryptographic keys before transmission. This encryption prevents other hospitals from accessing or analyzing the raw data, ensuring secure aggregation only at the Federated Learning Coordinator (FLC). The FLC, therefore, receives data in an encrypted format, performing aggregation without decrypting individual hospital data, thereby preserving privacy. Such granular yet secure traceability has become increasingly important in high-risk AI systems, helping stakeholders demonstrate ongoing compliance with the EU AI Act's risk management obligations.

In parallel, the CEM's explainability interface supports both retrospective and real-time analyses of AI decisions. Clinicians can request post hoc explanations of diagnostic outputs, while administrators can monitor aggregated trends for potential biases across demographic subgroups. To address the tension between encryption and explainability,

the CEM leverages partially decrypted intermediate representations—or saliency maps derived from homomorphically encrypted gradients—to provide clinically relevant insights without exposing raw patient data. The double-layer approach ensures the integrity of the system’s cryptographic protections, even as it generates interpretive results that are used to inform clinical decisions.

In short, this all-encompassing framework emphasizes the critical interaction between federated coordination, secure computation, and continuous compliance monitoring. By weaving risk assessment and explainability into every component, the PPFL architecture aims to establish a trustworthy foundation for AI deployment in high-stakes healthcare scenarios.

Figure 2 shows our revised Privacy-Preserving Federated Learning (PPFL) architecture, emphasizing hospital-level encryption with institution-specific keys. By performing local encryption prior to transmitting data, each hospital ensures that no identifiable patient information ever leaves its premises. The FLC then orchestrates and aggregates encrypted updates in conjunction with the Secure Computation Layer (SCL), while the Compliance and Explainability Module (CEM) handles system-wide monitoring, auditing, and interpretability.

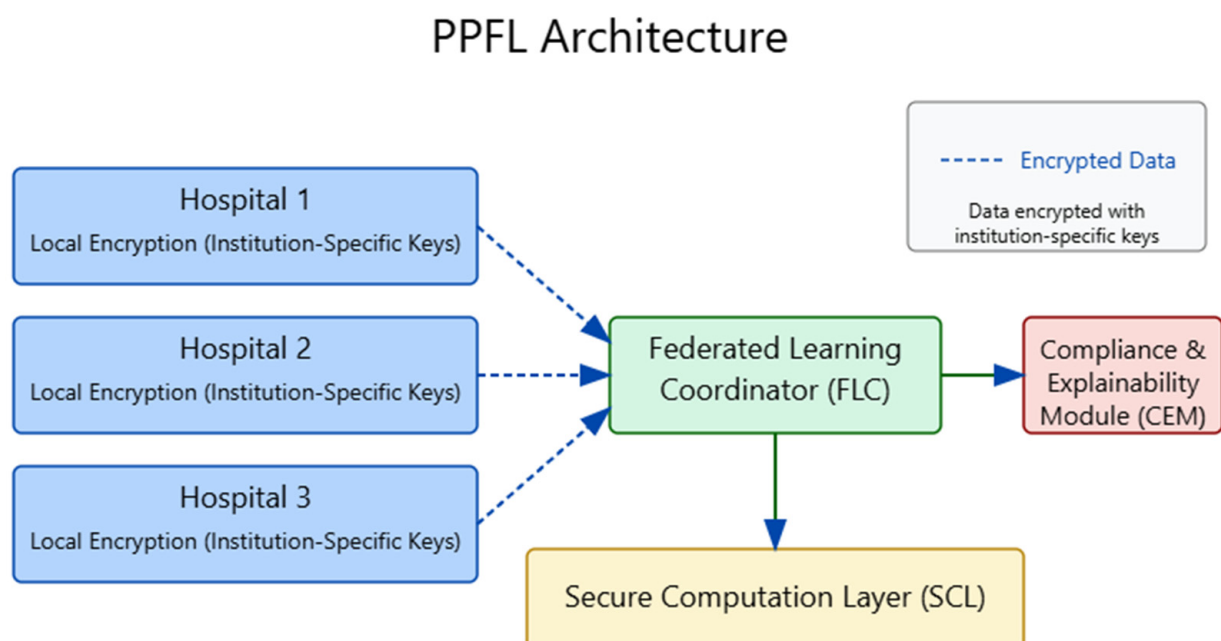


Figure 2. Updated PPFL architecture with institution-specific encryption. This figure highlights how each hospital encrypts patient data locally—using institution-specific cryptographic keys—before sending any information to the Federated Learning Coordinator (FLC). The FLC receives only ciphertext, thus preventing any raw data leaks. The Secure Computation Layer (SCL) applies advanced privacy techniques, while the Compliance and Explainability Module (CEM) provides auditing and interpretability to meet EU AI Act requirements.

In order to illustrate how an individual patient’s data record progresses throughout our proposed system, Figure 3 presents a UML-style activity diagram. Starting from data collection and preprocessing at the hospital, the diagram shows encryption, federated model training, secure aggregation, and the final auditing stages. Notably, each step highlights where risk assessments and cryptographic safeguards are enforced, ensuring privacy compliance before the data are used for training or reporting.

Activity Diagram: Patient Data Flow in Privacy-Preserving Healthcare AI

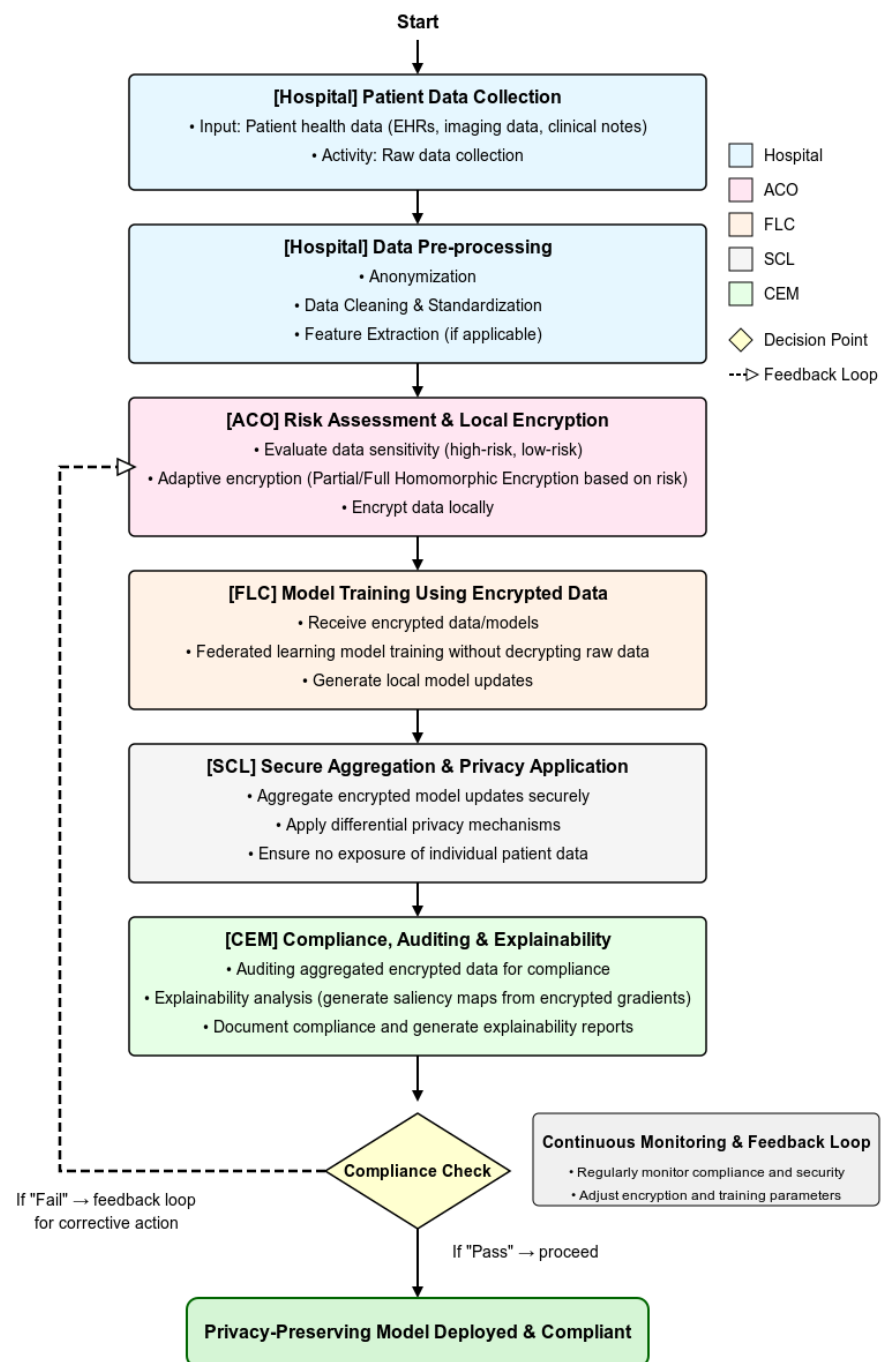


Figure 3. UML activity diagram for patient data flow in privacy-preserving healthcare AI. Each step shows how a single patient record is collected, pre-processed, encrypted, and subsequently used for federated model training. The final compliance and auditing phase ensures that privacy regulations are continuously met. Feedback loops allow for risk re-assessment and adaptive cryptographic settings.

6. Adaptive Cryptographic Algorithm

While the PPFL architecture satisfies minimum privacy and compliance requirements, we present a standalone, adaptive crypto algorithm that adjusts encryption and privacy levels dynamically in response to real-time operating conditions. This new system is named the Adaptive Crypto Orchestrator (ACO), which can operate in parallel with the PPFL architecture or be integrated into other artificial intelligence systems in the healthcare industry that need robust confidentiality controls.

6.1. Design Principles and Motivation

The core motivation behind the ACO is the heterogeneous nature of healthcare settings, where each institution might face unique computational constraints, network bandwidth limitations, or data sensitivity requirements. In a traditional federated learning setup, encryption parameters are often fixed from the outset, ignoring the changing context of model training and deployment. This rigidity can lead to less-than-optimal performance—e.g., excessive data over-encryption and excessive computational overhead or under-encryption and exposure of the system to more privacy threats [48]. The ACO addresses these problems by tracking operation and risk metrics in real time, enabling on-the-fly reconfiguration of encryption mechanisms. Moreover, the emergent need for real-time or near-real-time analytics in healthcare calls for dynamic approaches that can gracefully degrade the level of cryptographic rigor when latency-sensitive tasks arise. For instance, an AI-driven triage system in an emergency department may need instantaneous predictions, even if that temporarily reduces the depth of encryption. Conversely, if a sudden cybersecurity alert surfaces, the ACO can immediately escalate all encryption routines, mandating fully homomorphic encryption (FHE) or multi-party computation (MPC) with robust differential privacy layers [49].

6.2. Algorithmic Workflow

We illustrate the key steps of ACO in Algorithm 1 below, illustrating how it synchronizes various levels of encryption to respond to real-time requirements and attacks.

The cryptographic methods employed—partial homomorphic encryption (PHE), secure multi-party computation (SMPC), fully homomorphic encryption (FHE), and enhanced SMPC (ESMPC)—differ greatly in terms of implementation complexity and computational expenses.

PHE: Provides rather less processing overhead than other approaches [50]. Usually used in basic arithmetic calculations. Ideal for scenarios with low-risk data where efficiency is prioritized.

SMPC: Involves moderate computational complexity with significant communication overhead due to distributed cryptographic protocols [51]. Suitable for medium-risk scenarios requiring a balance between security and computational efficiency.

FHE: Carries high computational overhead due to intensive encryption and decryption processes [52]. Generally suitable only for high-risk environments where maximum security is essential and computational resources are abundant.

ESMPC: Enhanced version of SMPC with integrated differential privacy, offering robust privacy guarantees at moderate-to-high computational and communication costs [53]. Optimal in high-risk scenarios demanding strong privacy protection but less computational cost than FHE.

Implementations for these cryptographic techniques may leverage established frameworks such as Microsoft SEAL or Palisade for homomorphic encryption and Sharemind or MP-SPDZ for secure multi-party computation.

Analysis of Privacy Protection Against Attacks (Section 3):

PHE: Effectively mitigates reconstruction attacks and basic data exposure but remains vulnerable to inference-based attacks such as membership inference or property inference due to limited operational encryption.

SMPC: Strong protection against reconstruction and membership inference attacks, although some vulnerabilities to model extraction or property inference attacks persist if intermediate computations are observable or compromised.

FHE: Offers the strongest defense across all described attack vectors (re-identification, reconstruction, property inference, model extraction, inversion, shadow models, and ad-

versarial attacks), as computations remain encrypted throughout the process. However, computational overhead is substantial.

ESMPC: Significantly improves defense against inference attacks (membership inference, property inference) compared to basic SMPC, especially when combined with differential privacy. Provides robust protection with balanced computational demands.

Thus, careful consideration of risk scenarios, computational resources, and privacy requirements guides the adaptive selection among these cryptographic protocols within our Adaptive Crypto Orchestrator (ACO).

Algorithm 1: Adaptive Crypto Orchestrator (ACO)

Input:

- System context C , which includes the following:
Resource availability (e.g., CPU, GPU). Refers to computational resources required for encryption and secure computations.
Network bandwidth.
Patient data sensitivity.
Threat-level indicators.

Output:

- Updated model parameters or inference outputs, encrypted according to the chosen cryptographic protocol.

Initialize: Define threshold sets $\{T_{low}, T_{med}, T_{high}\}$, each mapped to a corresponding cryptographic scheme (e.g., partial homomorphic encryption, SMPC, fully homomorphic encryption).

Procedure:

1. Monitor Context: Continuously observe changes in C . For instance, if network latency spikes or an urgent clinical scenario arises, set a flag for “bandwidth-limited” or “time-critical”.
2. Protocol Selection: Based on the flagged risk scenario, choose the appropriate cryptographic mechanism:
 - If the scenario is low-risk, apply partial homomorphic encryption for more efficient processing.
 - If the scenario is medium-risk, employ secure multiparty computation (SMPC) to balance performance and security.
 - If the scenario is high-risk, activate fully homomorphic encryption or enhanced SMPC with differential privacy.
3. Recalibrate: At each training or inference cycle, measure key performance indicators (e.g., accuracy, latency, resource usage) alongside threat-level changes. If conditions differ from the current protocol’s assumptions, switch to the protocol that matches the new context.
4. Log and Audit: Record cryptographic choices, risk evaluations, and any relevant contextual variables in an immutable audit log. This step ensures compliance with regulations such as the EU AI Act and GDPR.
5. Produce Output: Return the updated model parameters or inference results, using the protocol indicated by the most recent risk assessment.

End Procedure

Note: The ACO may also interface directly with a Compliance and Explainability Module (CEM) in a broader federated learning framework (PPFL). Any protocol changes triggered by the ACO automatically update the compliance ledger and any associated explainability workflows.

Figure 4 illustrates how the ACO selects and adjusts cryptographic protocols (e.g., partial homomorphic encryption, secure MPC, fully homomorphic encryption with dif-

ferential privacy) based on the system context (resource availability, network bandwidth, data sensitivity, threat indicators). The diagram emphasizes continuous adaptation, real-time monitoring of context changes, and risk-based protocol selection that culminates in returning the encrypted model or inference under the chosen secure setting.

Adaptive Crypto Orchestrator (ACO) Algorithm Flow

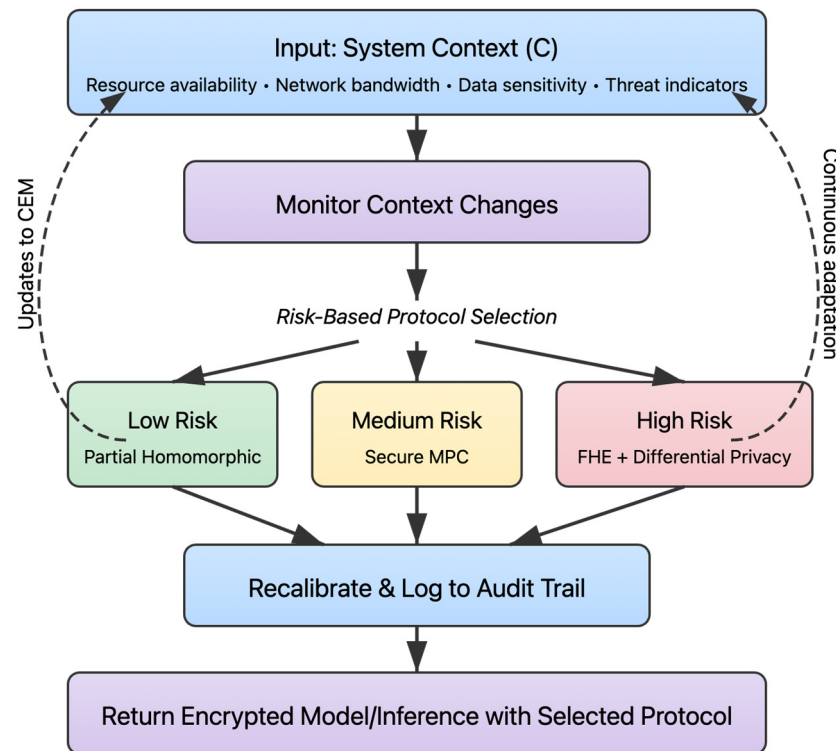


Figure 4. ACO-based adaptive cryptographic workflow.

6.3. Context-Aware Adaptations

One critical innovation in the ACO is the ability to incorporate contextual metadata—such as patient acuity levels, model drift signals, or emerging cyber threats—into its decision-making process [54]. This adaptability is especially vital in high-stakes clinical scenarios. For example, suppose an AI algorithm detects an unusual spike in membership inference attempts (logged by the CEM). In that case, the ACO can escalate from partial homomorphic encryption to fully homomorphic encryption in the following training round. Conversely, if an ICU system is overwhelmed, the ACO can relax specific encryption parameters temporarily to ensure near-instantaneous inference.

This orchestration is not simply protocol switching; the ACO can also alter parameters within a single cryptographic method. An example of a partially homomorphic scheme might adjust polynomial moduli or ciphertext expansion factors adaptively to sacrifice security for efficiency. Recent work has shown how subtle parameter tuning of cryptographic parameters can provide stunning computational improvements in federated learning, especially in patchy connectivity settings [55].

6.4. Operational Benefits and Challenges

By shifting cryptographic intensity only when necessary, computational resources are conserved during routine operations. At the same time, the system can rapidly intensify security measures if risk indicators surge. This is in accordance with the “privacy-on-

demand” principle and provides for continuous risk management as provided in the EU AI Act.

Nevertheless, there are a few problems. First, switching thresholds for switching encryption modes should be properly tuned such that they do not switch protocols too aggressively and intrusively or leak precious information. Second, having tight coordination among the ACO and local IT policies among different hospitals is difficult, mainly if hospitals are different in their use of certain encryption habits or have diverse compliance needs. Third, the interplay between encryption and explainability is an open research question. Advanced cryptographic methods can constrain the computation of fine-grained saliency maps or local feature attributions, which can hinder deep real-time interpretability.

6.5. Future Directions

In the future, the ACO can be augmented with its own machine learning algorithms, projecting future levels of risk or computational need from past trends. This predictive approach would allow the system to adjust encryption parameters in advance of a bottleneck or security problem, streamlining healthcare workflows and further securing patient data. Moreover, multi-modal integration—where the ACO orchestrates not just textual or tabular data but also genomic sequences, medical imaging, and streaming wearable data—could enable an even broader adoption of adaptive cryptographic techniques in modern healthcare ecosystems.

As federated learning continues to evolve and healthcare systems become more digitized, the ACO approach underscores the importance of situational awareness and adaptability in privacy-preserving AI. By building cryptographic resilience to the evolving clinical and cyber threat environment, healthcare professionals can innovate and ensure patient safety in harmony with new regulatory requirements and ethical demands.

7. Theoretical Insights and Challenges

Although the common PPFL framework and Adaptive Crypto Orchestrator (ACO) offer effective answers to both compliance and privacy, theoretical foundations and practical implications are yet to be addressed. On these basic questions, which we have been addressing ourselves, we always seem to reach a point necessary for ongoing development and extensive inquiry to unfold. Such an initiative is crucial to ensure prolonged stability, fair fairness, and greater health system effectiveness powered by artificial intelligence technologies.

7.1. Balancing Privacy with Model Utility

A core theoretical concern is finding the optimal equilibrium between maintaining stringent privacy guarantees and preserving high model performance. Differential privacy and homomorphic encryption techniques can shield sensitive data, yet the additive noise or additional computation might degrade model accuracy or latency. Recent theoretical advances have examined ways to formalize this trade-off by modeling privacy–utility curves and proposing adaptive approaches that recalibrate encryption levels based on real-time metrics [56]. From a health perspective, if guarantees of privacy do begin to limit utility excessively, the system might miss occasional but vital clinical signs—to instead have to have a softer, domain-sensitive tradeoff.

7.2. Modular Cryptographic Architecture

Offering a suite of encryption and secure computation protocols within the PPFL framework maximizes flexibility and introduces architectural complexity. Each node in the federated network may independently select its preferred cryptographic module—whether fully homomorphic encryption or SMPC with partial encryption—potentially complicating cross-institutional interoperability. To make them function properly in concert, strict testing

and periodic standardization is required. These ensure against cryptographic “version mismatches”, which might compromise the system or escalate expenditure on communicating.

7.3. Continuous Monitoring and Risk Management

The EU AI Act stresses a lifecycle approach, requiring continuous monitoring and immediate risk mitigation. In practice, achieving this demands more than simply appending an audit log or generating routine compliance reports. Instead, systems must automate threat detection, referencing real-time model behavior (e.g., anomaly spikes in gradient distributions) and external threat intelligence feeds [57]. When combined with ACO, continuous monitoring can trigger protocol escalations on the fly, yet this level of automation also raises questions about potential false positives, human oversight, and fallback mechanisms to ensure uninterrupted clinical operations.

7.4. Enhancing Explainability

While explainable AI (XAI) tools—such as saliency maps, local explanations, or counterfactual analyses—are instrumental for clinician acceptance, encrypting model parameters complicates interpretability. In many privacy-preserving methods, the “black-box” nature of the encrypted computations obscures intermediate feature interactions. One potential solution is to employ safe enclaves or partially encrypted overlays to convey insights in understandable format while keeping data in their encrypted, raw format. Still, securely building such overlays is a current challenge in studies, particularly for deep learning models where interpretability is considerable in the first place.

7.5. Fairness and Bias Management

Ensuring equitable outcomes is not merely a design choice but a regulatory imperative in high-risk AI systems. Although differential privacy can inadvertently equalize noise distribution across populations, it does not inherently remove bias. Institutional data disparities—where specific demographics are over- or under-represented—may still skew model outputs despite encryption. Addressing fairness entails integrating bias detection tools, rebalancing strategies (e.g., data augmentation for minority cohorts), and post hoc error analysis. Critically, these efforts must be reconciled with cryptographic layers that might limit direct inspections of raw data or intermediate gradients.

7.6. Real-Time and Edge Scenarios

The advent of wearable devices and point-of-care diagnostic tools in clinics pushes the frontier of AI to resource-constrained edge environments. While federated learning can distribute training across such endpoints, the overhead of advanced cryptography may be infeasible for devices with limited computing power or intermittent connectivity. The ACO’s dynamic approach can mitigate some of these constraints by scaling encryption in line with available resources, but robust fallback protocols are essential if specific nodes cannot meet baseline security standards. Furthermore, the real-time inference is to exhibit latencies under a sub-second in such applications as emergencies, such as triage; such responsiveness in under-high-encrypt modes is something under development.

7.7. Cross-Border Implementation Complexities

The presence of diverse legal systems is a major challenge to health collaborations on an international scale. The AI Act in Europe has created general regulations; however, individual member states have their individual data protection regulations, leading to inconsistent obligations in regions. Such inconsistencies provide nimble systems to handle local cryptographic needs; such extra complexity makes design practices challenging and enhances resource pressures on federated systems. Future endeavors might explore

building standard, pan-border protocols to standardize disparate systems and improve collaborative clinical studies.

These theoretical insights and challenges underscore the fact that privacy-preserving healthcare AI is not a monolithic technical problem but an evolving domain shaped by legal stipulations, ethical mandates, and dynamic clinical practices. Ongoing research must converge on secure, adaptable, and context-aware solutions to ensure that AI's transformative potential in healthcare is realized responsibly and equitably.

8. Conclusions and Future Directions

This paper presents a comprehensive, attack-aware approach to designing high-risk AI systems for healthcare. We reviewed many data-centric and model-centric privacy attacks that threaten ML-based healthcare solutions. We highlighted data protection, integrity of models, and patient confidentiality by categorically grouping such attacks and presenting their application in clinical scenarios.

We then introduced our Privacy-Preserving Federated Learning (PPFL) framework, emphasizing critical aspects such as secure aggregation, compliance with EU AI Act stipulations, and an explainability module to maintain clinical trust. This framework employs several advanced cryptographic and privacy-enhancing technologies, including homomorphic encryption, secure multi-party computation, and differential privacy, to keep sensitive health data safe and sound for the life cycle of model training [58].

Additionally, we developed an Adaptive Cryptographic Algorithm (ACO) to optimize cryptography strategies in real time for resource economy and greater security. The ACO adaptively regulates cryptographic overhead in accordance with context such as computational constraints, clinical priority, and current threat. This “privacy on demand” paradigm provides a novel route to maintain high protection standards without imposing undue burdens on healthcare institutions or compromising patient care.

Although the integrated PPFL + ACO architecture offers a robust foundation, critical challenges remain. High-complexity encryption can produce latency that is problematic in urgent clinical scenarios, and cryptographic operations can further hinder the explainability of AI-based solutions. The governance of multi-institutional federated networks also calls for enhanced trust mechanisms—such as distributed ledgers or zero-knowledge proofs—to address liability, auditing, and regulatory variances across different regions. Furthermore, ongoing advancements in privacy-preserving techniques for generative AI and large language models underscore the breadth of emerging methods in this domain [59].

Future Research Directions

Lightweight Cryptographic Protocols: Future work could focus on designing or refining cryptographic protocols that strike a more efficient balance between security and computational overhead. For instance, partial homomorphic encryption schemes tailored to healthcare data distributions might achieve stronger privacy without markedly increasing inference latency. Additionally, hardware-based approaches like trusted execution environments (TEEs) warrant exploration to securely offload cryptographic computations.

Fairness-Driven Federated Learning: Addressing algorithmic fairness in scenarios securing privacy is challenging. Current efforts to ensure fairness often require direct inspection of raw data or model gradients, which may conflict with encryption strategies. New frameworks incorporating fairness metrics within the encryption or aggregation process could help mitigate disparities, especially when training data are geographically or demographically skewed.

Multi-Modal Healthcare Data Integration: The rise in wearable devices, genomic sequencing, and advanced imaging necessitates AI solutions to handle complex, multi-

modal inputs. Extending the PPFL + ACO approach to securely process these diverse data types could unlock new frontiers in personalized medicine. However, each modality introduces unique vulnerabilities; for example, property inference attacks on genomic data pose more significant ethical and privacy concerns than basic EHR data.

Transparent Model Updates and Versioning: Automated model updates in a federated environment risk inadvertently introducing regressions or biases if not adequately audited. The design of tamper-evident logs or integrated version control (potentially using blockchain-based registries) could offer traceability and rollback capabilities when anomalies are detected. Coupled with advanced governance models, these features further assure stakeholders that model evolution remains transparent and accountable.

Edge-Optimized Security: Bringing AI to low-power medical devices and remote monitoring systems accentuates the tension between strong encryption and real-time processing. Investigating specialized hardware accelerators, quantized encryption parameters, or custom ML architectures that maintain robust security on edge devices is a promising direction. Achieving sub-second inference in extreme cryptography is crucial in life-and-death applications such as detecting traumas or strokes.

Ultimately, the health sector is poised for paradigm-shifting transitions in which AI-driven insights reengineer patient treatment, accelerate diagnosis, and maximize operational efficiencies. However, the data that fuel these advanced models are among the most sensitive of any industry handles. We can foster a trustworthy AI ecosystem by developing integrated frameworks that combine federated learning, adaptable cryptographic orchestration, and rigorous compliance auditing. This environment is capable of delivering innovative clinical benefits while respecting patient privacy and adhering to strict regulations. In closing, this report only includes suggestions and resolutions, only up to launching, never to an endpoint. Collaboration between regulators, industry, clinical professionals, and investigators is imperative. Only through coordinated efforts, open dialog, and continuous refinement will we achieve the dual objectives of preserving patient welfare and advancing state-of-the-art AI in healthcare.

Author Contributions: Conceptualization, K.K.; methodology, K.K.; validation, K.K., G.F., and A.A.; formal analysis, K.K., P.R., and Y.K.; investigation, K.K.; writing—original draft preparation, K.K.; writing—review and editing, K.K., G.F., A.A., P.R., Y.K., and D.A.; visualization, K.K., G.F., A.A., P.R., and Y.K.; supervision, P.R., and D.A.; project administration, K.K., and D.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ML	Machine Learning
GDPR	General Data Protection Regulation
EU AI Act	European Union Artificial Intelligence Act
EHRs	Electronic Health Records
mHealth	Mobile Health
PHE	Partial Homomorphic Encryption
FHE	Fully Homomorphic Encryption
SMPC	Secure Multi-Party Computation
ESMPC	Enhanced Secure Multi-Party Computation

PPFL	Privacy-Preserving Federated Learning
FLC	Federated Learning Coordinator
SCL	Secure Computation Layer
CEM	Compliance and Explainability Module
ACO	Adaptive Crypto Orchestrator
GPU	Graphics Processing Unit
TPU	Tensor Processing Unit
PET	Privacy Enhancing Technologies
XAI	Explainable Artificial Intelligence
HL7	Health Level Seven
FHIR	Fast Healthcare Interoperability Resources
DICOM	Digital Imaging and Communications in Medicine
JSON	JavaScript Object Notation
XML	Extensible Markup Language
YAML	YAML Ain't Markup Language
CSV	Comma-Separated Values
REST	Representational State Transfer
SOAP	Simple Object Access Protocol
JWT	JSON Web Tokens
OAuth	Open Authorization
OIDC	OpenID Connect
TLS	Transport Layer Security
TEE	Trusted Execution Environment
GANs	Generative Adversarial Networks
API	Application Programming Interface

References

1. European Union. REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 13 June 2024—laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). *Off. J. EU* **2024**, *1689*, 1–144.
2. Kalodanis, K.; Rizomiliotis, P.; Anagnostopoulos, D. European Artificial Intelligence Act: An AI security approach. *Inf. Comput. Secur.* **2023**, *32*, 265–281. [\[CrossRef\]](#)
3. Wachter, S. The GDPR and the AI Liability Gap: How the Proposed EU AI Act Complements the GDPR. *Comput. Law Secur. Rev.* **2021**, *43*, 105567. [\[CrossRef\]](#)
4. Xu, L.; Jiang, C.; Wang, J.; Liu, K. A novel privacy-preserving federated learning framework for healthcare data. *IEEE Trans. Netw. Sci. Eng.* **2021**, *8*, 2050–2064.
5. Kalodanis, K.; Rizomiliotis, P.; Feretzakis, G.; Papapavlou, C.; Anagnostopoulos, D. High-Risk AI Systems—Lie Detection Application. *Future Internet* **2025**, *17*, 26. [\[CrossRef\]](#)
6. Mishra, P.; Sinha, R.; Eletreby, M.; Naveed, M.; Hicks, B. Delphi: A Cryptographic Inference Service for Neural Networks. In *Proceedings of the 29th USENIX Security Symposium (USENIX Security 2020)*; USENIX Association: Berkeley, CA, USA, 2020; pp. 2505–2522.
7. Parisi, L.; Zanella, M.; Gennaro, R. Efficient Hybrid Homomorphic Encryption for Large- 834 Scale Transformer Architectures. In *Proceedings of the 30th ACM Conference on Computer and 835 Communications Security (CCS)*, Copenhagen, Denmark, 23–30 November 2023.
8. Groen, P.; Kapitan, D.; Molengraaf, C.; Travkina, Y. *Privacy Enhancing Technologies Whitepaper: Developed by Centre of Excellence—Data Sharing and Cloud*; Centre of Excellence: Eindhoven, The Netherlands, 2023.
9. Rosenblatt, F. The Perceptron: A Probabilistic Model for Information Storage and Organization in The Brain. *Psychol. Rev.* **1958**, *65*, 386–408. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Li, Q.; Torres, M.A.; Chen, Y. Risk Assessment of AI-Driven Healthcare Systems under the EU AI Act: A Critical Review of Classification and Compliance. *Int. J. Healthc. Manag.* **2023**, *16*, 200–214.
11. Marques, S.; Giancaspro, M.; Dekker, T. Implementing the EU AI Act in Healthcare: A Roadmap for Risk Management, Data Governance, and System Reliability. *Healthcare* **2023**, *11*, 1250.

12. ISO 13485; Medical Devices—Quality Management Systems—Requirements for Regulatory Purposes. International Organization for Standardization: Geneva, Switzerland, 2016.
13. Zhang, W.; Li, X.; Chen, Y.; Majeed, M.; Haq, I.U. Artificial Intelligence and Patient Confidentiality: Reviewing Privacy Challenges in Large Healthcare Databases. *Sensors* **2023**, *23*, 3456.
14. Zhu, L.; Liu, Z.; Han, S. Deep Leakage from Gradients: Understanding Inference Attacks in Collaborative Learning. *IEEE Trans. Inf. Forensics Secur.* **2021**, *16*, 408–421.
15. Mothukuri, V.; Maamar, Z.; Kolici, V.; Khan, S.; Srivastava, G. Federated Learning-based Clinical Data Processing for Privacy-Preserving Healthcare Systems. *IEEE Internet Things J.* **2023**, *10*, 2785–2797.
16. Kim, D.; Park, C.; Shin, S.; Hong, M. Performance Evaluation of Homomorphic Encryption for Deep Neural Networks in Cloud Environments. *IEEE Access* **2023**, *11*, 45012–45028.
17. Schneider, T.; Kales, D.; Angelova, Z.; Holzer, A.; Richter, B. Advances in Secure Multiparty Computation Protocols for Distributed Artificial Intelligence. *Inf. Sci.* **2023**, *646*, 321–345.
18. Feretzakis, G.; Verykios, V.S. Trustworthy AI: Securing Sensitive Data in Large Language Models. *AI* **2024**, *5*, 2773–2800. [[CrossRef](#)]
19. Sharma, S.; Li, T.; Zhao, W. Exploring re-identification attacks on anonymized healthcare data using cross-attribute correlation. *IEEE Access* **2022**, *10*, 100501–100512.
20. Nguyen, H.; Tsikerdekis, M.; Choi, J.; Tawalbeh, L.A. Evaluating de-identification strategies for high-dimensional datasets in healthcare. *Future Internet* **2022**, *14*, 321.
21. Zhu, P.; Lan, T.; Feng, Z. Automated text re-identification in anonymized clinical notes: A deep learning approach. *Artif. Intell. Med.* **2023**, *137*, 102460.
22. Zhang, X.; Chen, L.; Zhu, T. Reconstruction Attacks on Privacy-Preserving Machine Learning Models in Edge–Cloud Systems. *Sensors* **2023**, *23*, 4567.
23. Zhu, L.; Liu, Z.; Han, S. Deep Leakage from Gradients. *Adv. Neural Inf. Process. Syst.* **2019**, *33*, 14774–14784.
24. Yang, D.; Xu, W.; Chen, Y. Real-time dataset reconstruction using generative models: Implications for federated learning. *Inf. Sci.* **2022**, *610*, 428–444.
25. Park, S.; Qiu, X.; Kinyua, J.; Luo, J. Balancing privacy and accuracy in medical imaging: A study of encryption and noise-injection methods. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 3830–3842.
26. Taghavi, M.; Zhang, Y.; Wang, K. Latent feature leakage in deep neural networks: A review on property inference attacks. *Neural Comput. Appl.* **2022**, *34*, 18675–18686.
27. Ma, Q.; Li, L.; Kairouz, P.; Truex, S. Gradient-based property extraction in federated learning: A multi-institutional case study. In Proceedings of the 29th ACM Conference on Computer and Communications Security (CCS), Los Angeles, CA, USA, 7–11 November 2022; pp. 1421–1433.
28. Liu, R.; Fang, M.; Chen, G. Mitigating latent feature inference in healthcare AI: A fairness-aware framework. *Sci. Rep.* **2023**, *13*, 12789.
29. Wang, Z.; Mukherjee, A.; Pathak, S. Efficient decision boundary approximation for model extraction attacks in healthcare ML. *IEEE Trans. Inf. Forensics Secur.* **2022**, *17*, 1345–1358.
30. Jin, Y.; Jiang, Z.; Parra, J. Membership inference under realistic conditions: A federated learning perspective. *Pattern Recognit.* **2022**, *127*, 108593.
31. Yan, K.; Rowe, M.; Li, F.; Shah, P. GAN-assisted model inversion: Reconstructing imaging data from classification probabilities. *IEEE Trans. Med. Imaging* **2022**, *41*, 1475–1486.
32. Silva, A.; Wang, N.; De Cristofaro, E. Towards a universal shadow model approach: Transfer learning for membership and property inference in clinical AI. *ACM Symp. Appl. Comput. (SAC)* **2023**, *1*, 828–837.
33. Chang, R.; Xie, C.; Chen, M.; Wu, X. Comprehensive Survey on Adversarial Attacks and Defenses in Deep Learning: Theory, Tools, and Developments. *Sensors* **2023**, *23*, 3456.
34. Anjum, M.S.; Hassan, N.; Zhang, X. Gradient-based adversarial strategies for medical image classification: A systematic review. *Comput. Biol. Med.* **2022**, *148*, 105877.
35. Li, G.; Zhang, S.; Chen, T.; Wei, X. Multi-modal adversarial attacks in smart healthcare: Challenges and solutions. *IEEE Internet Things J.* **2023**, *10*, 12566–12579.
36. Song, C.; Shejwalkar, V.; Shokri, R.; Ristenpart, T. Analyzing Memorization in Learning from Bounded Data. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 26358–26369.
37. Zhao, L.; Feng, C.; Tang, Y. Large-scale memorization in clinical language models: Identifying and quantifying sensitive data leakage. *Trans. Assoc. Comput. Linguist.* **2022**, *10*, 2112–2125.
38. Choi, H.; Mirjalili, S.; Nguyen, P.; Turk, M. Latent Backdoor Threats in Transfer Learning-Based Medical AI Systems. In Proceedings of the Annual Symposium on Neural Information Processing Systems (NeurIPS), New Orleans, LA, USA, 28 November–9 December 2022; pp. 1–13.

39. Jones, K.; Lyu, X.; Martin, A. Multi-institutional Approaches to AI-Driven Medical Imaging: A Comprehensive Systematic Review. *Radiol. Artif. Intell.* **2022**, *4*, e210123.
40. Chen, Y.; Barlow, R.; Kim, S.; Freedman, S. Examining Real-World Clinical Validations of AI for Cardiac Prognosis: Opportunities and Gaps. *J. Am. Coll. Cardiol.* **2023**, *81*, 1234–1245.
41. Kessler, C.; Pohl, S.; Gan, B. Ethical Implications of AI in Healthcare under the EU AI Act Framework. *Comput. Law Secur. Rev.* **2022**, *46*, 105721.
42. Cross, J.; Choma, M.; Onofrey, J. Bias in medical AI: Implications for clinical decision-making. *PLoS Digit. Health* **2024**, *3*, e0000651. [\[CrossRef\]](#)
43. MedicaI. HL7 vs. DICOM: Decoding the Backbone of Healthcare IT. Available online: <https://blog.medicaI.io/en/hl7-vs-dicom/> (accessed on 20 March 2024).
44. Delgado, H.; Kumar, R.; Yuan, T.; Rutherford, K. Radiomics meets multi-modal data: Applications of advanced AI to cross-domain healthcare analytics. *IEEE Trans. Biomed. Eng.* **2023**, *70*, 250–261.
45. Teo, Z.L.; Jin, L.; Li, S.; Miao, D.; Zhang, X.; Yan Ng, W.; Tan, T.F.; Lee, D.M.; Chua, K.J.; Heng, J.; et al. Federated machine learning in healthcare: A systematic review on clinical applications and technical architecture. *Cell Rep. Med.* **2024**, *5*, 101419. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Li, X.; Wang, H.; Qiu, Y.; Wu, F. Dynamic Scheduling for Federated Learning in Heterogeneous Healthcare Environments. *IEEE Trans. Netw. Sci. Eng.* **2022**, *9*, 4837–4849.
47. Chen, W.; Hong, L.; Shafiq, B.; Gera, M. Modular Encryption Approaches in Collaborative Healthcare Systems: A Resource-Aware Triage for Privacy Protection. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 3952–3962.
48. Zhou, M.; Xing, W.; Guo, F.; Jiang, Q. Adaptive Encryption in Healthcare Federated Learning: Balancing Latency and Security. *IEEE Trans. Emerg. Top. Comput.* **2022**, *10*, 1291–1303.
49. Narkedimilli, S.; Venkata Sriram, A.; Raghav, S. FL-DABE-BC: A Privacy-Enhanced, Decentralized Authentication, and Secure Communication for Federated Learning Framework with Decentralized Attribute-Based Encryption and Blockchain for IoT Scenarios. *arXiv* **2024**, arXiv:2410.20259.
50. Song, C.; Wang, Z.; Peng, W.; Yang, N. Secure and Efficient Federated Learning Schemes for Healthcare Systems. *Electronics* **2024**, *13*, 2620. [\[CrossRef\]](#)
51. Dong, X.; Randolph, A.D.; Weng, C.; Kho, N.A.; Rogers, M.J.; Wang, X. Developing High Performance Secure Multi-Party Computation Protocols in Healthcare: A Case Study of Patient Risk Stratification. *AMIA Jt. Summits Transl. Sci. Proc.* **2021**, 200.
52. Al Badawi, A.; Yusof, M.F. Private pathological assessment via machine learning and homomorphic encryption. *BioData Min.* **2024**, *17*, 33. [\[CrossRef\]](#)
53. Pentyala, S.; Railsback, D.; Maia, R.; Dowsley, R.; Melanson, D.; Nascimento, A.; De Cock, M. Training Differentially Private Models with Secure Multiparty Computation. *arXiv* **2022**, arXiv:2202.02625.
54. Yusuf, T.; Brannigan, J.; Toreini, E.; Ingram, J. Context-Aware Cryptographic Protocols for AI Systems in Hospitals: A Cybersecurity Perspective. *ACM Trans. Priv. Secur. (TOPS)* **2023**, *26*, 1–28.
55. Park, H.; Chen, J.; Chen, Q. Dynamic Parameter Tuning for Partial Homomorphic Encryption under Intermittent Connectivity. *Future Gener. Comput. Syst.* **2022**, *134*, 232–243.
56. Chen, L.; Roy, S.; Madurai Elavarasan, R.; Zhang, H. Adaptive Privacy-Utility Trade-off Modeling in Federated Healthcare AI: A Survey of Methodologies. *ACM Comput. Surv.* **2023**, *55*, 1–36.
57. Chen, G.; Pham, Q.; Zhang, Y. Real-Time Threat Intelligence in Healthcare AI: Anomaly Detection and Automated Response under the EU AI Act. *IEEE Trans. Inf. Forensics Secur.* **2022**, *17*, 4320–4335.
58. Froelicher, D.; Troncoso-Pastoriza, J.R.; Raisaro, J.L.; Cuendet, M.A.; Sousa, J.S.; Cho, H.; Hubaux, J.P. Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption. *Nat. Commun.* **2021**, *12*, 5910. [\[CrossRef\]](#) [\[PubMed\]](#)
59. Feretzakis, G.; Papaspyridis, K.; Gkoulalas-Divanis, A.; Verykios, V.S. Privacy-Preserving Techniques in Generative AI and Large Language Models: A Narrative Review. *Information* **2024**, *15*, 697. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.