



Privacy compliant health data as a service for AI development

Grant Agreement Number: 101095384

D2.2: Security and Privacy Measures

Deliverable Identifier:	D2.2
Deliverable Version:	v.1.0
Status	Final (F)
Work Package:	WP2 Requirements & Specifications
Task:	T2.3 Security and Privacy Requirements
Author(s) and Organisation:	Sofiane Lagraa (FJLU), Geoffroy Robin (FJLU), Moussa Ouedraogo (FJLU)
Peer Reviewer(s):	Andrei Kazlouski (UTU), Ileana Montoya Perez (UTU), Nick Shopland (NTU), Andy Burton (NTU)
Deliverable Due Date:	2024/10/31
Deliverable Submission Date:	2024/10/28
Dissemination Level:	PU: Public
Funding Authority:	European Commission
Funding Program:	Horizon Europe Health Work Programme 2021 – 2022
Topic:	HORIZON-HLTH-2022-IND-13-02
Rights:	PHASE-IV-AI Consortium

Document Control History

Version	Date	Edited by	Modification reason
v.0.1	2024/06/01	Sofiane Lagraa, Geoffroy Robin, FJLU + Security and Privacy Workshop	ToC
v.0.2	2024/06/15	Sofiane Lagraa, FJLU	1 st draft
v.0.3	2024/07/01	Sofiane Lagraa, FJLU	2 nd draft
v0.4	2024/07/16	Sofiane Lagraa, FJLU	3 rd draft + Requirements and Measures
v0.5	2024/07/31	Geoffroy Robin, FJLU	Comments + Requirements and Measures
v0.6	2024/08/12	Geoffroy Robin, FJLU	Comments
v0.7	2024/08/20	Sofiane Lagraa, FJLU	Corrections
v0.8	2024/09/10	Moussa Ouedraogo, FJLU	Comments + corrections
v0.9	2024/09/26	Sofiane Lagraa, FJLU	Double checks
v0.10	2024/09/27	Sofiane Lagraa, FJLU	Send to the reviewers
v0.11	2024/10/21	Geoffroy Robin, FJLU	Improvements and corrections from the review
v1.0	2024/10/23	Sofiane Lagraa, FJLU Geoffroy Robin, FJLU Moussa Ouedraogo, FJLU	Final version

Executive Summary

This deliverable outlines the security and privacy requirements and measures for the PHASE-IV-AI project, an effort focused on developing security and privacy-compliant health data as a service for AI development. The deliverable meticulously analyses relevant regulations, including GDPR, the Data Governance Act, and the AI Act, and incorporates best practices to ensure the protection of sensitive health information. It details a comprehensive set of requirements, categorised by system and theme, and proposes concrete measures and tools to address them. The deliverable emphasises a user-centric approach, integrating user stories and use cases to ensure that the project's security and privacy considerations are aligned with the needs and concerns of all stakeholders. By implementing these requirements and measures, the PHASE-IV-AI project aims to build trust, ensure compliance, and foster a secure and ethical environment for data handling and AI development in healthcare.

Disclaimer

The content of the publication herein is the sole responsibility of the publishers and it does not necessarily represent the views expressed by the European Commission or its services.

While the information contained in the documents is believed to be accurate, the authors(s) or any other participant in the PHASE-IV-AI consortium make no warranty of any kind with regard to this material including, but not limited to the implied warranties of merchantability and fitness for a particular purpose.

Neither the PHASE-IV-AI Consortium nor any of its members, their officers, employees or agents shall be responsible or liable in negligence or otherwise howsoever in respect of any inaccuracy or omission herein.

Without derogating from the generality of the foregoing neither the PHASE-IV-AI Consortium nor any of its members, their officers, employees or agents shall be liable for any direct or indirect or consequential loss or damage caused by or arising from any information advice or inaccuracy or omission herein.

Copyright message

©PHASE-IV-AI Consortium, 2023-2026. This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both. Reproduction is authorised provided the source is acknowledged.

Table of Contents

1. Introduction	8
1.1 Purpose of the Document	8
1.2 Reference Documents.....	8
1.3 Definitions	8
1.4 Structure of the Document.....	10
1.5 List of Acronyms.....	10
2. Analysis & Approach for the Systematic Elaboration of Security and Privacy Requirements	12
2.1 Methodological Approach for Specification of Security and Privacy Requirements and Controls	12
2.1.1 Leveraging PHASE IV AI Deliverables for Security and Privacy Requirements Discovery .	12
2.1.1.1 PHASE IV AI - User Stories, Use Cases and Related Processes	13
2.1.1.2 PHASE IV AI - Initial Technology Assessment and List	13
2.1.1.3 PHASE IV AI - Data Protection Impact Assessment.....	13
2.1.1.4 PHASE IV AI - Legal and Ethical Framework and Requirements	13
2.1.2 PHASE IV AI – AI Models and Services.....	13
2.1.3 Security and Privacy Requirements Discovery and Extraction	15
2.1.4 PHASE IV Purpose Analysis	15
2.1.5 Security and Privacy Measures.....	15
2.1.6 Data Protection, Digital Governance Regulations, and Key Standard	15
2.1.6.1 General Data Protection Regulation (GDPR).....	16
2.1.6.2 Data Governance Act.....	16
2.1.6.3 NIS2 Directive.....	16
2.1.6.4 E-Privacy Directive	16
2.1.6.5 Artificial Intelligence Act (AI Act)	17
2.1.6.6 EU Cybersecurity Act.....	18
2.1.6.7 ISO 27001.....	18
2.1.6.8 National Institute of Standards and Technology (NIST).....	18
2.1.7 A Research-Driven Approach for PHASE IV AI.....	19
2.2 Security and Privacy Requirements and Measures Template.....	19
3. Security and Privacy Requirements & Measures for AI models	21
3.1 Data and Model as a Service	21
3.1.1 Data as a Service.....	21
3.1.2 Model as a Service.....	21
3.2 AI Models-Based Data and Model as a Service	21

3.2.1	Federated Learning (FL).....	21
3.2.2	Variational AutoEncoder (VAE).....	22
3.2.3	Generative Adversarial Network (GAN).....	23
3.2.4	Diffusion Model (DM)	23
3.3	Security and Privacy Requirements and Measures for AI models	24
3.3.1	General Security and Privacy Requirements and Measures for AI models.....	24
3.3.1.1	REQUIREMENT #1: Human Agency and Oversight.....	24
3.3.1.2	REQUIREMENT #2: Technical Robustness and Safety.....	24
3.3.1.3	REQUIREMENT #3: Privacy and Data Governance.....	25
3.3.1.4	REQUIREMENT #4: Transparency.....	25
3.3.1.5	REQUIREMENT #5: Diversity, Non-discrimination and Fairness	27
3.3.1.6	REQUIREMENT #6: Social and Environmental Well-Being	28
3.3.2	Security and Privacy Requirements and Measures for Federated Learning.....	28
3.3.3	Common Security and Privacy Requirements and Measures for: DM, GAN, VAE.....	33
4.	Security and Privacy Requirements and Measures for Privacy-Enhancing Technologies.....	35
4.1	Privacy-Enhancing Technologies	35
4.1.1	Differential Privacy	35
4.1.2	Secure Multiparty Computation	35
4.1.3	Homomorphic Encryption	35
4.1.4	Characteristics and Similarities	36
4.2	Security and Privacy Requirements and Measures.....	36
4.2.1	Differential Privacy	36
4.2.2	Secure Multiparty Computation	38
4.2.3	Homomorphic Encryption	40
5.	Security and Privacy Requirements & Measures for User stories and Related Processes.....	42
5.1	Concept Definitions.....	42
5.2	Security and Privacy Requirements and Measures.....	44
6.	Security and Privacy Requirements & Measures for PHASE IV AI Platform	49
6.1	Data Handling in PHASE IV AI.....	49
6.2	European Health Data Space (EHDS)	49
6.3	Health Data Hub	49
6.4	Phase IV AI Platform	50
7.	Conclusion.....	59

Table of Figures

Figure 1. Security and Privacy Requirements and Measures Construction Methodology 12

Table of Tables

Table 1. List of Definitions	8
Table 2. List of Acronyms	10
Table 3. Template for Security and Privacy Requirements and Measures	19

1. Introduction

Security and privacy requirements and measures are the safeguards and protocols implemented to protect sensitive data and systems from unauthorised access, use, disclosure, disruption, modification, or destruction. These requirements define the specific security and privacy goals, while measures are the concrete actions taken to achieve those goals. This includes implementing access controls, encryption, data masking, security awareness training, incident response plans, and regular security audits. The goal is to ensure confidentiality, integrity, and availability of data and systems, while also respecting individual privacy rights and complying with relevant regulations.

1.1 Purpose of the Document

The purpose of this document is to define the security and privacy requirements and measures for the PHASE-IV-AI project. This document aims to:

- **Identify and document** the specific security and privacy needs of the project.
- **Prioritise** these requirements based on their criticality.
- **Propose concrete measures** and tools to address each requirement.
- **Ensure compliance** with relevant legal and ethical frameworks, including GDPR, Data Governance Act, NIS2, E-Privacy, AI Act, ISO 27001, and NIST (NIST Cybersecurity Framework (CSF)).
- **Foster trust** and transparency with stakeholders regarding data handling practices.

The proposed security and privacy requirements and measures will serve as a baseline for partners to develop or enhance their components by integrating these measures throughout their component life cycle.

1.2 Reference Documents

This deliverable is based on a variety of reference documents to provide the security and privacy requirements and measures, including PHASE IV AI deliverables like the Data Protection Impact Assessment (D1.5), Initial Technology Assessment and List (D2.5), User Stories, Usage Scenarios and Use Case Validation (D6.1), and Legal and Ethical Framework and Requirements (D2.3). It also references key regulations like GDPR, Data Governance Act, NIS2, E-Privacy, AI Act, EU Cybersecurity Act, ISO 27001, and NIST Cybersecurity Framework. These documents provide a comprehensive foundation for the security and privacy requirements and measures outlined in this deliverable, ensuring alignment with best practices, compliance with relevant regulations, and addressing the concerns of all stakeholders and used technologies.

1.3 Definitions

Table 1. List of Definitions

List of Definitions	
Accountability	Being answerable for one's actions and their outcomes. It can be legally enforced, like the GDPR's data protection rules, or be an ethical choice, such as tech firms avoiding facial recognition due to ethical concerns.
Accuracy	In AI, it refers to how well a model predicts unseen data. It is checked by comparing the model's predictions on a test dataset to the actual answers.
AI bias	Refers to systematic errors in AI systems that lead to unfair outcomes, favouring certain groups. It can arise from various factors including algorithm design, data

List of Definitions	
	handling, and societal biases. It's prevalent across platforms by perpetuating and amplifying existing inequalities and reinforcement of social biases.
AI developers	They are involved in all aspects of AI development, from conception to final use. This includes designing, programming, testing, and maintaining AI applications and components.
AI Explainability	AI explainability refers to methods for explaining the AI decision to both experts and non-experts. It means that the AI system's functions and operations can be explained in a non-technical manner to someone without specialised knowledge.
AI reliability	Refers to the consistent performance of an AI system, even when encountering new inputs that it hasn't been trained or tested on before.
AI Robustness	Refers to an AI system's technical stability in various contexts and its social adaptability, ensuring it considers its operating environment. This is vital to prevent unintended harm, even with good intentions. Robustness is one of the three key components for achieving Trustworthy AI.
AI systems	AI systems are human-designed software that perceive, interpret, reason, and act to achieve complex goals. AI encompasses all techniques such as machine learning, deep learning, and reinforcement learning.
Cybersecurity Education and Awareness	This involves training and educating individuals about potential security threats and how to protect their own and others' privacy.
Data governance	Data governance is to ensure high-quality data throughout the system lifecycle and implementing controls that align with business goals. It focuses on data availability, usability, consistency, integrity, and sharing. It involves setting up processes for effective data management across the organisation, including accountability for poor data quality.
Data poisoning	Data poisoning in AI systems refers to the injection of harmful data into the model's training set by an adversary, causing the system to learn incorrectly. This can significantly impact accuracy or even introduce a hidden backdoor, allowing the attacker to manipulate the system's behaviour.
Data Privacy	Data privacy, focuses on the rights of individuals to control how their personal information is collected, used, and shared. It involves ensuring that data is handled in compliance with legal and regulatory requirements, and that individuals' preferences and consent are respected in the management of their personal information.
Data Security	Refers to the measures and practices implemented to protect digital information from unauthorised access, corruption, or theft throughout its lifecycle. This includes the use of encryption, firewalls, access controls, and other technologies to safeguard data from cyber threats and breaches.
Data space Security and Privacy	This involves protecting data space-based data, applications, and infrastructures from threats, while also ensuring the privacy of user data.
End user	End user is the final user of the system, that ultimately uses or is intended to ultimately use the AI system. They are distinct from those who maintain or support the system, like IT experts, software professionals, or AI developers/designers.
Fairness	It encompasses concepts like equity, impartiality, egalitarianism, non-discrimination, and justice. It represents the ideal of treating individuals or groups equally, known as 'substantive' fairness. It also includes a procedural aspect, which is the capacity to seek and secure redress when individual rights are infringed.

List of Definitions	
Identity and Access Management (IAM)	This involves ensuring that the right individuals have access to the right resources at the right times for the right reasons, while protecting user identities and maintaining their privacy.
Incident Response	This involves preparing for and responding to security incidents, including data breaches that could compromise privacy.
Network Security	This involves protecting a network infrastructure and the data it transports from unauthorised access, misuse, malfunction, modification, destruction, or improper disclosure.
Privacy by Design	This involves integrating privacy considerations into the design and operation of IT systems, networked infrastructure, and business practices.
Privacy Policies and Regulations	This involves understanding and complying with laws, regulations, and policies that govern data privacy and security, such as GDPR, HIPAA, etc.
Security by Design	This involves integrating security considerations into the design and operation of IT systems, networked infrastructure, and business practices.
Subject	A "Subject" refers to an individual or a group impacted by the AI system. This could be someone receiving benefits determined by an AI system, or the general public affected by technologies like facial recognition.
Trustworthy AI	It is defined by three key elements: (1) Lawfulness - it complies with all relevant laws and regulations, (2) Ethics - it respects and adheres to ethical principles and values, and (3) Robustness - it is reliable from both a technical and social standpoint, preventing unintentional harm even with good intentions. Trustworthiness applies not only to the AI system itself but also to all processes and actors involved in the system's life cycle.

1.4 Structure of the Document

This deliverable is structured to provide a comprehensive and logical overview of the security and privacy requirements and measures for the PHASE-IV-AI project. It begins with a detailed explanation of the methodology used for identifying and constructing the requirements and measures. The deliverable is divided into the specific security and privacy requirements and measures for various aspects of the project, including Data as a Service (DaaS) and Model as a Service (MaaS) (Section 3), privacy-enhancing technologies (Section 4), user stories (Section 5), and the PHASE-IV-AI platform (Section 6). Each section provides a clear description of the requirements, their rationale, and proposed measures or tools to address them. The document concludes with a summary of the key points and by emphasising the importance of stakeholder engagement, and research in the field of security and privacy. This structure ensures a clear and organised presentation of the information, making it easy for readers to understand the project's security and privacy considerations.

1.5 List of Acronyms

Table 2. List of Acronyms

List of Acronyms	
AI	Artificial Intelligence
DaaS	Data as a Service
EU	European Union
FL	Federated Learning
GDPR	General Data Protection Regulation
IT	Information Technology

List of Acronyms	
MaaS	Model as a Service
ML	Machine Learning
NIS	Network and Information Systems
NIST	National Institute of Standards and Technology
UC	Use Case
WP	Work Package

2. Analysis & Approach for the Systematic Elaboration of Security and Privacy Requirements

The Security and Privacy requirements and measures of PHASE IV AI encapsulate the collective expertise of project partners, addressing the critical needs, identified gaps, and ambitious goals for designing, implementing, and operating robust cybersecurity systems within healthcare IT environments. These guidelines meticulously outline specific security and privacy requirements, encompassing data protection, access control, encryption, and compliance with relevant regulations. This comprehensive approach ensures that PHASE IV AI prioritises the protection of sensitive patient data, fosters a secure and reliable healthcare IT ecosystem, and promotes responsible data handling practices.

2.1 Methodological Approach for Specification of Security and Privacy Requirements and Controls

Figure 1 shows an overview of the proposed methodology for discovering and constructing the security and privacy requirements and measures. The following sections depict the methodological approach by expanding on the activities that have helped towards the definition of Phase IV AI security and privacy requirements.

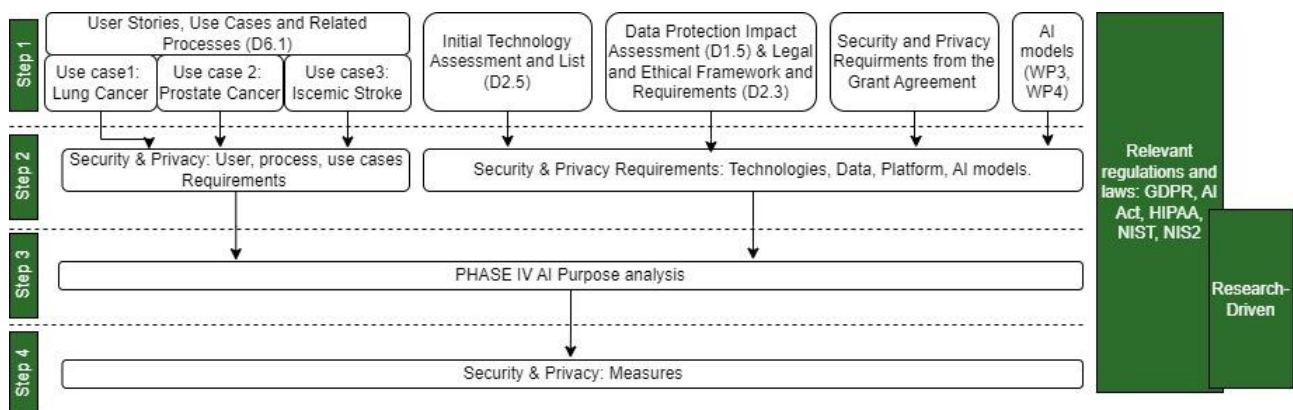


Figure 1. Security and Privacy Requirements and Measures Construction Methodology

2.1.1 Leveraging PHASE IV AI Deliverables for Security and Privacy Requirements Discovery

Leveraging PHASE IV AI Deliverables for Security and Privacy Requirements Discovery, as outlined in Step 1 of Figure 1, represents a collection of key deliverables, including initial technology assessments, data protection impact assessments, user stories and use cases, legal and ethical frameworks, and AI models and services as well as the grant agreement. These deliverables serve as inputs for the security and privacy requirements discovery and extraction. The process involves a careful analysis of these inputs, considering their potential impact on the project's purpose, security, and privacy.

This section provides a brief overview of the Phase IV AI deliverables, upon which the current deliverable is founded for extracting the security and privacy requirements.

2.1.1.1 PHASE IV AI - User Stories, Use Cases and Related Processes

The deliverable D6.1 focuses on demonstrating the use of synthetic datasets for three medical use cases: Lung Cancer, Prostate Cancer, and Ischemic Stroke. The document details how models and generated data will be validated, outlines plans for data sharing to enable secure collaboration, identifies legal and ethical risks with mitigation measures, and expresses user requirements as "user stories" to guide system development. This information is valuable for extracting security and privacy requirements, as it provides insights into data sharing protocols, legal and ethical considerations, and user needs.

2.1.1.2 PHASE IV AI - Initial Technology Assessment and List

The deliverable D2.5 outlines the initial technology list for the project, including contributions from partners across various work packages. The structured table details each technology's purpose, description, inputs/outputs, and potential applicability within the project. The initial assessment technology list as an outcome of the deliverable serves as a roadmap for defining project requirements and specifications. It is important to note that this is an initial overview, as various aspects are still under development. The list provides a foundation for collaboration and technological development to create solutions for the project's three use cases. The technology list can be a valuable resource for identifying and extracting security and privacy requirements and measures for the project.

2.1.1.3 PHASE IV AI - Data Protection Impact Assessment

The deliverable D1.5 provides an overview of data usage, processing, and subject rights within the PHASE IV AI project, focusing on the 6th month of the project. It emphasises a "privacy by design" approach, minimising data usage and relying on informed consent and secondary processing for scientific research purposes. The federated structure minimises direct contact with personal data. The assessment identifies potential risks, ranging from unauthorised data modification to natural disasters, and outlines mitigation strategies. These include encryption, access restrictions, regular audits, backup protocols, and disaster recovery plans. Despite challenges, PHASE IV AI remains committed to GDPR compliance and best practices, ensuring a resilient framework for data protection. Continuous evaluation and refinement of risk management strategies safeguard data subject interests and uphold the integrity of scientific research. This information is valuable for extracting security and privacy requirements and translating them into concrete technical measures. This ensures the project adheres to its ethical and legal obligations while safeguarding data subject rights and maintaining the integrity of scientific research.

2.1.1.4 PHASE IV AI - Legal and Ethical Framework and Requirements

The deliverable D2.3 outlines the legal and ethical guidelines for developing a privacy-compliant Health Data Hub. It analyses relevant EU regulations like GDPR, the Data Governance Act, and the proposed AI Act. The document also emphasises ethical principles for trustworthy AI development, referencing the European Commission's guidelines. It aims to ensure the project's compliance with data protection, security, and individual rights while promoting responsible AI development. This initial version provides a preliminary framework, that can be translated into technical security and privacy requirements and measures in this deliverable.

2.1.2 PHASE IV AI – AI Models and Services

Different technologies and AI models will be used in PHASE IV AI in both WP3 and WP4 as data-as-a-service and model-as-a-service, respectively. The list of technologies and AI models to be used, according to the Grant

Agreement and information provided by the partners responsible for the development of each service, are as follows:

- **Federated Learning:** A machine learning technique that trains models on decentralised data without sharing raw data, enhancing privacy. It addresses security concerns by limiting data exposure and enabling collaborative learning without compromising individual data¹.
- **Differential Privacy:** A technique that adds noise to data, protecting individual privacy while allowing for accurate statistical analysis. It ensures that the presence or absence of a single individual's data does not significantly affect the results, mitigating privacy risks².
- **Secure Multi-Party Computation (SMPC):** A cryptographic method enabling multiple parties to jointly compute a function on their private data without revealing the data itself. It ensures data confidentiality and integrity, preventing unauthorised access and manipulation³.
- **Generative Adversarial Network (GAN):** A type of deep learning model that consists of two competing neural networks: a generator and a discriminator. The generator learns to create realistic data samples, while the discriminator tries to distinguish between real and generated data. This adversarial training process leads to the generation of high-quality synthetic data. While GANs can generate realistic data, they pose privacy risks if used to create synthetic data that resembles real individuals. This can lead to identity theft and reputational damage⁴.
- **Diffusion Model (DM):** A generative model that learns to gradually add noise to data until it becomes pure noise, and then learns to reverse this process to generate new data. This approach enables the generation of high-quality samples with diverse features. Diffusion models can generate high-quality data but may also be used to create realistic fake content, raising concerns about misinformation and manipulation⁵.
- **Variational Autoencoder (VAE):** A generative model that learns a compressed representation of the data by encoding it into a lower-dimensional latent space. This latent space can then be used to generate new data samples by decoding them back into the original data space. VAEs can learn compressed representations of data, potentially exposing sensitive information if not properly secure. This can compromise privacy and lead to unauthorised data access⁶.
- **Bayesian Network:** A probabilistic graphical model that represents the relationships between variables using a directed acyclic graph. Each node in the graph represents a variable, and the edges represent dependencies between them. Bayesian networks are used for reasoning under uncertainty and making predictions based on observed data. Bayesian networks are not inherently tied to privacy or security concerns. However, their use in sensitive data analysis requires careful consideration of data anonymisation and access control to prevent privacy breaches⁷.

¹ McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. Communication-efficient learning of deep networks from decentralized data. 2017

² Dwork, C., McSherry, F., Nissim, K., & Smith, A. Calibrating noise to sensitivity in private data analysis. Theory of Cryptography Conference, 2006. 265-284

³ Goldreich, O., Micali, S., & Wigderson, A. How to play any mental game. Proceedings of the 19th Annual ACM Symposium on Theory of Computing. 1987. 218-229

⁴ Goodfellow, et al. Generative adversarial nets. Advances in Neural Information Processing Systems. 2014. 2672-2680

⁵ Ho, J., et al. Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems. 2020. 12840-12851

⁶ Diederik P. Kingma, Max Welling: Auto-Encoding Variational Bayes. ICLR 2014

⁷ Judea Pearl: Probabilistic reasoning in intelligent systems - networks of plausible inference. Morgan Kaufmann series in representation and reasoning, Morgan Kaufmann 1989, pp. I-XIX, 1-552

2.1.3 Security and Privacy Requirements Discovery and Extraction

Security and Privacy requirements discovery and extraction in PHASE IV AI project is the process of meticulously identifying, documenting, and analysing the specific security and privacy needs of the project. This involves understanding the project's context, identifying stakeholders, analysing existing policies, regulations, and PHASE IV AI deliverables and technologies. The process then focuses on extracting concrete security and privacy requirements, prioritising them based on their criticality, and documenting them clearly in a standardised format. This approach ensures the protection of sensitive health information, maintains patient trust, and complies with relevant legal and ethical frameworks, ultimately contributing to the success of PHASE IV AI and the advancement of healthcare innovation.

2.1.4 PHASE IV Purpose Analysis

The PHASE IV AI project purpose analysis, when considered alongside security and privacy requirements discovery and extraction, becomes a holistic process that ensures the project's ethical and responsible execution. It involves not only defining the project's "why" in terms of health outcomes, but also understanding the potential impact on individuals' privacy and security. This analysis involves identifying the types of technology being used and data being collected, processed, and stored, as well as the potential risks and vulnerabilities associated with these activities. By integrating security and privacy considerations into the purpose analysis, the PHASE IV AI project can ensure that their goals are achieved while maintaining ethical data handling practices and protecting patient trust while proposing adequate security and privacy measures.

2.1.5 Security and Privacy Measures

Security and privacy measures, when viewed through the lens of the PHASE IV AI project purpose analysis and security and privacy requirements discovery and extraction, become integral components of PHASE IV AI's responsible execution. They are not only technical safeguards, but rather proactive strategies designed to protect sensitive health information and ensure the project's alignment with its stated purpose. This involves implementing specific controls, such as data encryption, access control mechanisms, and robust authentication processes, all informed by the project's unique context, and prioritised requirements. By integrating security and privacy measures into the very fabric of the project, stakeholders can create a robust framework that safeguards patient data, fosters trust, and ultimately contributes to achieving the project's intended health outcomes.

2.1.6 Data Protection, Digital Governance Regulations, and Key Standard

A set of European regulations are used to create security and privacy requirements and measures such as GDPR, Data Governance Act, NIS2 Directive, E-Privacy Directive, AI Act.

All these regulations pertain to the governance, protection, and management of data, particularly in digital formats and environments. They cover a range of topics including personal data protection, network and information systems security, e-privacy, artificial intelligence (AI), and health information privacy. Each regulation has its own specific focus and applies in different contexts.

2.1.6.1 General Data Protection Regulation (GDPR)

The General Data Protection Regulation (GDPR) is a comprehensive data protection law that applies to all organisations processing personal data of individuals within the European Union (EU)⁸. In PHASE IV AI, which involves data sharing and AI models, GDPR presents unique challenges due to the sensitive nature of healthcare data. GDPR mandates that handling the data in healthcare must adhere to principles of lawful, fair, and transparent processing, including data minimisation and robust security measures to protect personal data^{9,10,11}. Organisations must have a clear legal basis for collecting, storing, and using healthcare data, and they must be transparent about their data practices with individuals¹⁰. Furthermore, individuals have specific rights under GDPR, including the right to access, rectify, erase, and restrict the processing of their personal data, which must be respected when AI models are used in healthcare¹².

2.1.6.2 Data Governance Act

The Data Governance Act (DGA) is an EU regulation aimed at fostering the responsible use and sharing of data. It establishes a framework for data governance by promoting the creation of data spaces, which are collaborative ecosystems for data sharing and use. The DGA defines key concepts like data intermediaries, which act as trusted entities facilitating data sharing, and data altruism, encouraging individuals to contribute their data for the common good. It also introduces rules for data reuse, ensuring data is accessible and reusable for research and innovation. The DGA aims to empower individuals and organisations to control their data, fostering a more transparent and ethical data economy.

2.1.6.3 NIS2 Directive

The Network and Information Systems Security Directive (NIS2) is an EU regulation aimed at bolstering cybersecurity across critical infrastructure and essential services, including healthcare^{13,14}. This directive mandates a risk-based approach, requiring healthcare organisations to conduct thorough risk assessments and implement proportionate security measures to protect sensitive patient data and AI models used in diagnostics or treatment. NIS2 also strengthens reporting obligations, requiring healthcare providers to notify national authorities of cyber incidents impacting patient data or AI systems¹⁵. The directive emphasises supply chain security, ensuring the integrity of medical devices and AI software used in healthcare, and promotes a strong cybersecurity culture within healthcare organisations. Compliance with NIS2 is crucial for PHASE IV AI involving data and AI models, ensuring the safety and privacy of patient information and the reliable operation of critical medical systems¹⁶.

2.1.6.4 E-Privacy Directive

The ePrivacy Directive (Directive 2002/58/EC), also known as the Privacy and Electronic Communications Directive (PEC Directive), is a piece of EU legislation designed to protect the privacy of individuals in the

⁸ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

⁹ Article 4 of GDPR

¹⁰ Article 5 of GDPR: "Lawfulness, fairness and transparency", "Purpose limitation", "Data minimization", "Accuracy", "Storage limitation", "Integrity and confidentiality".

¹¹ Article 32 of GDPR: "Security of processing".

¹² Articles 15-22 of GDPR: "Rights of the data subject".

¹³ <https://www.enisa.europa.eu/topics/cybersecurity-policy/nis-directive-new>

¹⁴ https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI%282021%29689333

¹⁵ <https://nis2directive.eu/health/>

¹⁶ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2020%3A767%3AFIN>

electronic communications sector¹⁷. This directive is particularly relevant to the PHASE IV AI project involving AI models and health data due to its focus on data protection, consent, confidentiality, and security related to electronic communications. It covers the processing of personal data associated with electronic communications, including metadata and content. The directive mandates explicit consent for processing data for marketing purposes and requires providers to maintain confidentiality and implement appropriate security measures to protect personal data¹⁷.

2.1.6.5 Artificial Intelligence Act (AI Act)

The Artificial Intelligence Act (AI Act) is a regulation established by the EU that provides a common regulatory and legal framework for AI within the EU¹⁸. The Act classifies AI applications into four risk categories: unacceptable, high, limited, and minimal, plus an additional category for general-purpose AI¹⁸.

AI applications with unacceptable risks are prohibited, such as those deploying manipulative techniques to distort behaviour and cause significant harm. High-risk applications, which are expected to pose significant threats to health, safety, or fundamental rights of persons, **must comply with security, transparency, and quality obligations**. Limited-risk applications only have transparency obligations, while minimal-risk applications are not regulated. For general-purpose AI, transparency requirements are imposed, with additional evaluations for high-capability models. The majority of obligations fall on AI developers (providers) of high-risk systems. Users (deployers) of high-risk AI systems also have some obligations, though less than providers. This Act can apply extraterritorially to providers from outside the EU if they have users within the EU¹⁹.

In the context of the PHASE IV AI project, the used AI models and frameworks are classified as high-risk, so it needs to comply with the AI Act's security, transparency, and quality obligations.

There are 6 AI principles in AI act:

- **Human oversight:** AI models should be designed and operated in a way that allows for human oversight and intervention.
- **Technical robustness and safety:** AI models should be robust, reliable, and safe, minimising risks of harm to individuals and society.
- **Privacy and data protection:** AI models should respect privacy and data protection principles, ensuring the responsible collection, use, and storage of personal data.
- **Transparency:** The functioning and decision-making processes of AI models should be transparent, interpretable, and explainable (eXplainable AI), allowing users to understand how decisions are made.
- **Non-discrimination and fairness:** AI models should be designed and deployed in a way that avoids discrimination and promotes fairness, ensuring equal treatment for all individuals.
- **Societal and environmental well-being:** AI models should be developed and used in a way that benefits society and the environment, considering their potential impact on social structures, employment, and the natural world.

¹⁷ https://www.edps.europa.eu/sites/default/files/publication/dir_2002_58_en.pdf

¹⁸ High-level summary of the AI Act | EU Artificial Intelligence Act. <https://artificialintelligenceact.eu/high-level-summary/>.

¹⁹ Artificial intelligence act | Think Tank | European Parliament. https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI%282021%29698792

2.1.6.6 EU Cybersecurity Act

The EU Cybersecurity Act is a European regulation that introduces a harmonised system for the cybersecurity certification of ICT products, services, and processes²⁰. It aims to improve protection against threats to cybersecurity within the EU. The Act strengthens the EU Agency for cybersecurity (ENISA), giving it more resources and new tasks²¹. ENISA plays a key role in setting up and maintaining the European cybersecurity certification framework by preparing the technical groundwork for specific certification schemes. It is also in charge of informing the public about the certification schemes and the issued certificates. In the context of healthcare, the Act is particularly relevant as it helps ensure the cybersecurity resilience needed in the health sector^{22,23}. Compliance with the Act is crucial for ensuring the safety and privacy of patient information and the reliable operation of critical medical systems²⁴.

2.1.6.7 ISO 27001

ISO 27001 is an internationally recognised information security management system (ISMS) standard that provides a framework for organisations to establish, implement, maintain, and continually improve their information security practices^{25,26}. In the PHASE IV AI project, ISO 27001 plays a critical role in protecting sensitive patient data, particularly when dealing with electronic protected health information (ePHI). The standard mandates a risk-based approach to information security, requiring organisations to identify, assess, and mitigate potential threats to their data. This involves implementing a wide range of cybersecurity techniques, including access controls, encryption, data loss prevention, vulnerability management, and incident response plans. By adhering to ISO 27001, the PHASE IV AI project can demonstrate their commitment to data security, build trust with stakeholders, and comply with regulatory requirements, ultimately fostering a secure environment for conducting research and developing new innovations.

2.1.6.8 National Institute of Standards and Technology (NIST)

The National Institute of Standards and Technology plays a significant role in different sectors, especially in the healthcare sector. NIST's Health Information Technology (IT) program aims to improve the quality and availability of healthcare and reduce healthcare costs by enabling the establishment of an emerging health IT network that is correct, complete, secure, usable, and testable^{27,28}. This includes activities such as supporting the nation's health IT effort, pursuing the standards and measurement research necessary to achieve the goal of improving healthcare delivery through information technology²⁷. NIST also provides updated cybersecurity guidance for the healthcare industry to help protect patients' personal health information. For instance, the

²⁰ The EU Cybersecurity Act | EUR-Lex. <https://eur-lex.europa.eu/EN/legal-content/summary/the-eu-cybersecurity-act.html>

²¹ The EU Cybersecurity Act | Shaping Europe's digital future. <https://digital-strategy.ec.europa.eu/en/policies/cybersecurity-act>

²² Is the EU Healthcare Sector Cyber Healthy? The Conclusions of Cyber Europe 2022- ENISA. <https://www.enisa.europa.eu/news/is-the-eu-healthcare-sector-cyber-healthy-the-conclusions-of-cyber-europe-2022>.

²³ Good practices for the security of healthcare services — ENISA. <https://www.enisa.europa.eu/topics/critical-information-infrastructures-and-services/health/good-practices-for-the-security-of-healthcare-services>

²⁴ How European funded research can boost cyber resilience of hospitals. <https://digital-strategy.ec.europa.eu/en/events/how-european-funded-research-can-boost-cyber-resilience-hospitals>.

²⁵ <https://www.iso.org/standard/27001>

²⁶ <https://www.isms.online/iso-27001/>

²⁷ Health Information Technology (IT) | NIST. <https://www.nist.gov/healthcare>.

²⁸ Health IT at NIST - Program Overview | NIST. <https://www.nist.gov/programs-projects/health-it-nist-program-overview>

NIST Cybersecurity Framework for Improving Critical Infrastructure (NIST CSF) has been adopted by many healthcare organisations to uphold high security standards in an industry facing extreme cyber risk²⁹.

2.1.7 A Research-Driven Approach for PHASE IV AI

As PHASE IV AI is a research and innovation project, it is crucial to incorporate the most advanced and recent research papers on security and privacy measures for data protection and AI. This includes exploring cutting-edge techniques for data anonymisation, differential privacy, and secure multi-party computation, as well as examining the ethical implications of AI in healthcare and the development of robust governance frameworks to ensure responsible data usage. By leveraging the latest scholarship, we can ensure that our PHASE IV AI project adheres to the highest standards of data security and privacy, fostering trust and ethical practices in healthcare innovation.

2.2 Security and Privacy Requirements and Measures Template

The definition of requirements and measures for the PHASE IV AI project drew upon a diverse range of data sources and AI technologies. Discussions with partners, leveraging their experience and expertise, shaped the understanding of the cybersecurity landscape. Reference scenarios, user stories, pilot operations specifications, and key performance indicators (KPIs) were crucial in defining security and privacy requirements and measures from the user perspective. This approach was further enriched by incorporating the outcomes of the June 2024 PHASE IV AI security and privacy workshop, ensuring a comprehensive and user-centric approach to safeguarding data and privacy within the project.

To ensure a structured and comprehensive approach to defining requirements and measures for PHASE IV AI, we have adopted the VOLERE methodology³⁰. This methodology provides a systematic framework for discovering, communicating, and managing requirements throughout the project, ensuring traceability and clarity for all stakeholders. We have customised and adapted the VOLERE template to align with the specific activities and goals of PHASE IV AI, enabling agile refinement as the project progresses. This adapted methodology serves as a foundation for identifying and documenting each requirement and measure, as outlined in Table 3. By leveraging this structured approach, we aim to achieve a clear understanding of the security and privacy needs, and ensure their successful implementation.

Naming convention for Security and Privacy Requirements is SP-REQ-<Requirement No.> (e.g. SP-REQ-01). 'SP' stands for Security and Privacy, 'REQ' stands for Requirement, and <Requirement No.> is a unique number.

Table 3. Template for Security and Privacy Requirements and Measures

Field name	Details
ID	Give a unique ID for the security and privacy requirement
Title	Give a title/short name for the security and privacy requirement
Priority	One of the following MoSCoW method ³¹ : <ul style="list-style-type: none"> Must have (Essential): The system must implement the security and privacy requirement to be accepted.

²⁹ Compliance Guide: NIST CSF and the Healthcare Industry. <https://www.upguard.com/blog/compliance-guide-nist-csf-healthcare-industry>

³⁰ Robertson, J., & Robertson, S. (2017). *Volere Requirements Specification Template. Edition 10.1.*

³¹ Clegg, D., Baker, R.: CASE Method Fast-track: A RAD Approach. Addison-Wesley, Reading (1994)

Field name	Details
	<ul style="list-style-type: none"> • Should have (Desirable): The system should implement the security and privacy requirement: some deviation from the requirement as stated may be acceptable. • Nice to have (Could have): The system should implement the security and privacy requirement but may be accepted without it.
Category	Data, AI Model, Network, Application, Infrastructure and Management, Privacy Policies and Regulations, Security by Design, Privacy by Design
Description and Rationale	Specify the intention of the security and privacy requirement while giving a justification of this requirement.
Measures or Tools	Proposed measures to implement or tools to use for addressing the security and privacy requirement. The tools could be known (D2.5) or unknown and integrated in the next version of D2.5 and taken into consideration in D1.5.
Supporting materials	Provide a reference to documents and deliverables that demonstrate and clarify this requirement or assumption, specifically those related to AI.
Comments	

The proposed security and privacy requirements and measures will serve as a baseline for partners to develop or enhance their components by integrating these measures throughout their component life-cycle. Additionally, implementing these measures and mapping them to the PHASE IV AI components and architecture improvements in the upcoming deliverables is highly beneficial.

3. Security and Privacy Requirements & Measures for AI models

3.1 Data and Model as a Service

3.1.1 Data as a Service

Data as a Service (DaaS) refers to the provision of anonymous and synthetic health data through the Health Data Hub. This service ensures privacy by de-identifying real data or generating privacy-preserving synthetic data from real data, making it accessible for research and innovation while protecting patient confidentiality.

3.1.2 Model as a Service

Model as a Service (MaaS) enables the secure execution of machine learning workflows using secure multi-party computation (SMPC) techniques. This service allows researchers and developers to train and test models on data from multiple sources without compromising data privacy.

3.2 AI Models-Based Data and Model as a Service

3.2.1 Federated Learning (FL)

Federated learning is a collaborative machine learning technique that allows multiple participants to train a shared model without sharing their raw data. This is particularly useful in situations where data privacy is a concern, such as in healthcare. Federated learning addresses the challenge of training machine learning models on sensitive data that cannot be shared due to privacy regulations or competitive concerns. By keeping the data on the devices, federated learning enables collaborative model training while preserving data privacy.

Federated Learning offers a multitude of benefits that revolutionise machine learning methods. One of the key advantages is *Data Privacy*³². It allows protection of sensitive data by ensuring it remains on the user's device, thereby enhancing security. Another significant benefit is *Collaboration*³³. Federated Learning enables multiple parties to collectively train a model on their combined data without the need to share the data itself, fostering a collaborative environment while maintaining data privacy. It also offers *Scalability*^{34,35}, as it can efficiently train models on large datasets that are distributed across numerous devices, overcoming the limitations of traditional centralised methods. Lastly, Federated Learning allows for *Personalisation*^{36,37}. It enables the development of personalised models that are trained on individual user data, resulting in more accurate and tailored predictions. Thus, Federated Learning presents a promising avenue for machine learning, balancing data privacy with collaborative learning, scalability, and personalisation.

Federated learning offers a promising approach to training machine learning models on decentralised data without compromising user privacy. However, it is not without its potential risks. While the raw data itself

³² Harmandeep Kaur, Veenu Rani, Munish Kumar, Monika Sachdeva, Ajay Mittal, Krishan Kumar: Federated learning: a comprehensive review of recent advances and applications. *Multim. Tools Appl.* 83(18): 54165-54188 (2024)

³³ Yashothara Shanmugarasa, Hye-young Paik, Salil S. Kanhere, Liming Zhu: A systematic review of federated learning from clients' perspective: challenges and solutions. *Artif. Intell. Rev.* 56(S2): 1773-1827 (2023)

³⁴ Syed Zawad, Feng Yan, Ali Anwar: Local Training and Scalability of Federated Learning Systems. *Federated Learning* 2022: 213-233

³⁵ Yong Zhou, Yuanming Shi, Haibo Zhou, Jingjing Wang, Liqun Fu, Yang Yang: Towards Scalable Wireless Federated Learning: Challenges and Solutions. *CoRR abs/2310.05076* (2023)

³⁶ Royson Lee, Minyoung Kim, Da Li, Xinchu Qiu, Timothy M. Hospedales, Ferenc Huszar, Nicholas D. Lane: FedL2P: Federated Learning to Personalize. *NeurIPS* 2023

³⁷ Koji Matsuda, Yuya Sasaki, Chuan Xiao, Makoto Onizuka: An Empirical Study of Personalized Federated Learning. *CoRR abs/2206.13190* (2022)

remains on individual devices, the model updates shared during training can inadvertently leak sensitive information. In fact, the key privacy risks associated with federated learning are described as follows:

- **Membership Inference Attacks:** An attacker can potentially infer which devices contributed data to the training process by analysing the model updates. This could reveal sensitive information about individuals, such as their participation in a medical study or their political affiliation^{38,39,40}.
- **Model Inversion Attacks:** By analysing the model's predictions, an attacker could potentially reconstruct the training data, even if it was never shared directly^{41,42}. This could lead to the exposure of sensitive personal information.
- **Data Poisoning Attacks:** A malicious actor could inject poisoned data into the training process, potentially manipulating the model's behaviour or causing it to learn biased or inaccurate information. This could have serious consequences for the model decisions, predictions, and applications^{43,44}.

3.2.2 Variational AutoEncoder (VAE)

A variational autoencoder (VAE) is a type of generative neural network that learns a compressed representation of input data, called a *latent space*, and can generate new data similar to the training data. VAEs achieve this by using a probabilistic approach, where the encoder maps input data to a probability distribution in the latent space, and the decoder reconstructs the input data from a sample drawn from this distribution. This probabilistic nature allows VAEs to generate diverse outputs and handle uncertainty in the data. However, VAEs can pose significant security and privacy risks^{45,46}. For example, the latent space representation can be vulnerable to adversarial attacks. Additionally, the training data used to train VAEs may contain sensitive information, which could be leaked through the latent space or the generated data⁴⁷. To mitigate these risks, researchers have proposed various security and privacy measures, including differential privacy, and data obfuscation⁴⁷. These techniques aim to enhance the robustness and privacy of VAEs while preserving their generative capabilities.

³⁸ Hao Sui, Xiaobing Sun, Jiale Zhang, Bing Chen, Wenjuan Li: Multi-level membership inference attacks in federated Learning based on active GAN. *Neural Comput. Appl.* 35(23): 17013-17027 (2023)

³⁹ Gergely Dániel Németh, Miguel Angel Lozano, Novi Quadrianto, Nuria Oliver: Addressing Membership Inference Attack in Federated Learning with Model Compression. *CoRR abs/2311.17750* (2023)

⁴⁰ Gongxi Zhu, Donghao Li, Hanlin Gu, Yuxing Han, Yuan Yao, Lixin Fan, Qiang Yang: Evaluating Membership Inference Attacks and Defenses in Federated Learning. *CoRR abs/2402.06289* (2024)

⁴¹ Seunghyeon Shin, Mallika Boyapati, Kun Suo, Kyungtae Kang, Junggab Son: An empirical analysis of image augmentation against model inversion attack in federated learning. *Clust. Comput.* 26(1): 349-366 (2023)

⁴² Ruihan Wu, Xiangyu Chen, Chuan Guo, Kilian Q. Weinberger: Learning To Invert: Simple Adaptive Attacks for Gradient Inversion in Federated Learning. *UAI 2023: 2293-2303*

⁴³ Subhash Sagar, Chang-Sun Li, Seng W. Loke, Jinho Choi: Poisoning Attacks and Defenses in Federated Learning: A Survey. *CoRR abs/2301.05795* (2023)

⁴⁴ Vale Tolpegin, Stacey Truex, Mehmet Emre Gursoy, Ling Liu: Data Poisoning Attacks Against Federated Learning Systems. *ESORICS* (1) 2020: 480-501

⁴⁵ Xiang Li, Shihao Ji: Defense-VAE: A Fast and Accurate Defense Against Adversarial Attacks. *PKDD/ECML Workshops* (2) 2019: 191-207

⁴⁶ Matthew Willetts, Alexander Camuto, Tom Rainforth, Stephen J. Roberts, Christopher C. Holmes: Improving VAEs' Robustness to Adversarial Attack. *ICLR 2021*

⁴⁷ Abhishek Singh, Ethan Garza, Ayush Chopra, Praneeth Vepakomma, Vivek Sharma, Ramesh Raskar: Decouple-and-Sample: Protecting Sensitive Information in Task Agnostic Data Release. *ECCV* (13) 2022: 499-517

3.2.3 Generative Adversarial Network (GAN)

A Generative Adversarial Network (GAN) is a type of deep learning model that consists of two competing neural networks: a generator and a discriminator. The generator learns to create new data that resembles the training data, while the discriminator learns to distinguish between real and generated data. However, GANs also present significant security and privacy concerns⁴⁸. For example, the training data used to train GANs may contain sensitive patient information, which could be leaked through the generated data or the model's internal representations. Additionally, GANs can be vulnerable to adversarial attacks, where malicious actors can manipulate the model's input or internal parameters to generate unintended or harmful outputs. To address these concerns, researchers have explored techniques like differential privacy, adversarial training, and data obfuscation. These strategies aim to enhance the robustness and privacy of GANs while preserving their generative capabilities⁴⁸.

3.2.4 Diffusion Model (DM)

Diffusion models are a class of generative AI models that learn to generate new data by gradually adding noise to existing data and then learning to reverse this process. They work by first corrupting the training data with increasing levels of noise until it becomes indistinguishable from random noise. Then, the model learns to reverse this process, starting from pure noise and gradually removing noise to generate new data that resembles the original training data. However, they also present security and privacy concerns. For instance, the training data used to train diffusion models may contain sensitive information, which could be leaked through the generated data or the model's internal representations^{49,50,51}. Additionally, diffusion models can be susceptible to adversarial attacks, where malicious actors can manipulate the model's input or internal parameters to generate unintended or harmful outputs^{52,53}. To address these concerns, researchers are exploring techniques like differential privacy⁵⁴, adversarial training⁵⁵, and data obfuscation⁵⁶ to enhance the robustness and privacy of diffusion models. These strategies aim to prevent the leakage of sensitive information and mitigate the risks of adversarial attacks while preserving the generative capabilities of diffusion models⁵⁷.

⁴⁸ Zhipeng Cai, Zuobin Xiong, Honghui Xu, Peng Wang, Wei Li, Yi Pan: Generative Adversarial Networks: A Survey Toward Private and Secure Applications. *ACM Comput. Surv.* 54(6): 132:1-132:38 (2022)

⁴⁹ Thomas Cilloni, Charles Fleming, Charles Walter: Privacy Threats in Stable Diffusion Models. *CoRR abs/2311.09355* (2023)

⁵⁰ Xinjian Luo, Yangfan Jiang, Fei Wei, Yuncheng Wu, Xiaokui Xiao, Beng Chin Ooi: Exploring Privacy and Fairness Risks in Sharing Diffusion Models: An Adversarial Perspective. *CoRR abs/2402.18607* (2024)

⁵¹ Derui Zhu, Dingfan Chen, Jens Grossklags, Mario Fritz: Data Forensics in Diffusion Models: A Systematic Analysis of Membership Privacy. *CoRR abs/2302.07801* (2023)

⁵² Boyang Zheng, Chumeng Liang, Xiaoyu Wu, Yan Liu: Understanding and Improving Adversarial Attacks on Latent Diffusion Model. *CoRR abs/2310.04687* (2023)

⁵³ Hondamunige Prasanna Silva, Lorenzo Seidenari, Alberto Del Bimbo: DiffDefense: Defending Against Adversarial Attacks via Diffusion Models. *ICIAP (2) 2023: 430-442*

⁵⁴ Tsai, Yu-Lin et al. "Differentially Private Fine-Tuning of Diffusion Models." (2024).

⁵⁵ Zekai Wang, Tianyu Pang, Chao Du, Min Lin, Weiwei Liu, Shuicheng Yan: Better Diffusion Models Further Improve Adversarial Training. *ICML 2023: 36246-36263*

⁵⁶ Luca Piano, Pietro Basci, Fabrizio Lamberti, Lia Morra: Latent Diffusion Models for Attribute-Preserving Image Anonymization. *CoRR abs/2403.14790* (2024)

⁵⁷ Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Wentao Zhang, Bin Cui, Ming-Hsuan Yang: Diffusion Models: A Comprehensive Survey of Methods and Applications. *ACM Comput. Surv.* 56(4): 105:1-105:39 (2024)

3.3 Security and Privacy Requirements and Measures for AI models

There are many overlapping security and privacy requirements and measures across the different neural network models that may be used in the PHASE IV AI project, such as Federated Learning, Diffusion Models, GANs, and VAEs. The common security and privacy requirements and measures are guided by the AI Act principals and are described as follows:

3.3.1 General Security and Privacy Requirements and Measures for AI models

3.3.1.1 REQUIREMENT #1: Human Agency and Oversight

Field name	Details
ID	SP-REQ-01
Title	Human-in-the-Loop Design
Priority	Should have
Category	AI model
Description and Rationale	The AI model SHOULD incorporate human oversight and intervention at key stages of its operation, ensuring that human judgment and decision-making remain central.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Human Review of Critical Decisions: Implement mechanisms for human review of critical decisions made by the AI model, especially those with significant consequences. • Human-in-the-Loop Feedback Mechanisms: Allow for human feedback and input to improve the AI model's performance and address potential biases. • Human Control over AI model: Ensure that humans retain ultimate control over the AI model, with the ability to override or modify its decisions when necessary.
Supporting materials	
Comments	

3.3.1.2 REQUIREMENT #2: Technical Robustness and Safety

Field name	Details
ID	SP-REQ-02
Title	Security Measures and Testing
Priority	Should have
Category	Data, AI model
Description and Rationale	Robust security measures SHOULD be implemented to protect the AI model from unauthorised access, data breaches, and other security threats.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Implement strong access control mechanisms to restrict access to sensitive data and model.
Supporting materials	
Comments	

3.3.1.3 REQUIREMENT #3: Privacy and Data Governance

Field name	Details
ID	SP-REQ-03
Title	Privacy-by-Design
Priority	Should have
Category	Data, AI model, Privacy-by-Design
Description and Rationale	Privacy considerations SHOULD be integrated into the design and development of the AI system from the outset, ensuring that privacy is protected by default.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Privacy-by-Design Principles: Implement privacy-enhancing technologies and practices throughout the development lifecycle. • Data Minimisation: Collect and process only the data strictly necessary for the AI model's functionality. • Anonymisation and Pseudonymisation: Employ techniques to anonymise or pseudonymise data where possible, minimising the risk of identifying individuals.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-04
Title	Data Protection Impact Assessment (DPIA)
Priority	Should have
Category	Data, AI model
Description and Rationale	A DPIA SHOULD be conducted to assess the risks to individuals' privacy and data protection associated with the AI model.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • DPIA Framework: Develop a comprehensive DPIA framework that covers all stages of the AI model's lifecycle. • Risk Identification and Mitigation: Identify potential risks to privacy and data protection and implement appropriate mitigation measures. • Documentation and Reporting: Document the DPIA process, findings, and mitigation measures taken.
Supporting materials	D1.5, D2.3
Comments	

3.3.1.4 REQUIREMENT #4: Transparency

Field name	Details
ID	SP-REQ-05
Title	Data Traceability and Auditability
Priority	Should have
Category	Data, AI model

Field name	Details
Description and Rationale	The AI model SHOULD be designed to allow for tracing the origin and usage of data throughout its lifecycle. This includes tracking data sources, transformations, and usage in model training and decision-making.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> • Data Lineage Tracking: Implement a system to track the lineage of data used for training and testing the AI system. • Data Provenance Management: Establish mechanisms to document the source, transformations, and usage of data throughout its lifecycle. • Auditable Logs: Maintain detailed logs of data access, modifications, and usage to facilitate audits and investigations.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-06
Title	Model Traceability and Explainability
Priority	Should have
Category	Data, AI model
Description and Rationale	The AI model SHOULD be designed to allow for tracing the model's development, training, and deployment, as well as understanding the rationale behind its decisions.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> • Model Versioning: Implement version control for AI models, including documentation of changes and updates. • Model Explainability Techniques: Utilise techniques like feature importance analysis, partial dependence plots, and SHAP values to explain the model's predictions. • Decision Rationale Documentation: Record the key factors and data points considered by the AI model when making decisions or recommendations.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-07
Title	Output Quality Assessment
Priority	Should have
Category	Data, AI model
Description and Rationale	The AI model SHOULD be continuously monitored and evaluated to assess the quality of its outputs, including accuracy, consistency, and reliability.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> • Performance Metrics: Implement metrics to evaluate the accuracy, precision, recall, and F1-score of the AI system's outputs. • Quality Control Procedures: Establish procedures for reviewing and validating the AI system's outputs, including manual checks and expert review.

Field name	Details
	<ul style="list-style-type: none"> Error Reporting and Analysis: Implement mechanisms to track and analyse errors or discrepancies in the AI model's outputs, identifying potential issues and areas for improvement.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-08
Title	Decision Logging and Transparency
Priority	Should have
Category	Data, AI model
Description and Rationale	The AI system SHOULD record and log its decisions or recommendations, providing a transparent record of its activities. This information should be accessible for auditing, analysis, and user understanding.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> Decision Logging: Implement a system to log all decisions made by the AI model, including the relevant data points, model parameters, and rationale behind the decision. Decision History: Maintain a comprehensive history of AI model decisions, allowing for analysis of patterns, trends, and potential biases. User Access to Logs: Provide users with appropriate access to decision logs, enabling them to understand the AI model's activities and rationale.
Supporting materials	
Comments	

3.3.1.5 REQUIREMENT #5: Diversity, Non-discrimination and Fairness

Field name	Details
ID	SP-REQ-09
Title	Data Diversity and Representativeness
Priority	Should have
Category	Data, AI model
Description and Rationale	Depending on the targeted problem such as disease diagnosis and detection, the data used to train and test the AI model SHOULD be diverse and representative of the population it will be used on. This includes considering factors like age, gender, ethnicity, socioeconomic status, and geographic location.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> Data Collection: Actively seek out data from diverse populations. Data Augmentation: Use techniques to artificially increase the diversity of the training data. Data Sampling: Ensure representative sampling of diverse populations in the data used for training and testing.
Supporting materials	

Field name	Details
Comments	

Field name	Details
ID	SP-REQ-10
Title	Bias Detection and Mitigation
Priority	Should have
Category	Data, AI model
Description and Rationale	The AI model SHOULD be regularly tested for bias, and mechanisms should be in place to mitigate any identified biases. This includes using publicly available technical tools and methods to analyse the data, model, and performance for potential biases.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Bias Detection Tools: Utilise bias detection tools. • Bias Mitigation Techniques: Implement techniques like re-weighting, adversarial debiasing, and counterfactual fairness to mitigate identified biases. • Regular Bias Audits: Conduct regular audits to assess the potential for bias in the AI model throughout its lifecycle.
Supporting materials	
Comments	

3.3.1.6 REQUIREMENT #6: Social and Environmental Well-Being

Field name	Details
ID	SP-REQ-11
Title	Environmental Impact Assessment
Priority	Should have
Category	AI model
Description and Rationale	The AI model's development and deployment SHOULD be assessed for their potential environmental impact, including energy consumption, and resource usage.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Life Cycle Assessment: Conduct a life cycle assessment to evaluate the environmental impact of the AI model throughout its lifecycle, from development to deployment. • Energy Efficiency Optimisation: Implement strategies to minimise the AI model's energy consumption, such as using efficient algorithms and hardware. • Sustainable Data Management: Adopt sustainable data management practices, including data storage optimisation, data deduplication, and data archiving.
Supporting materials	
Comments	

3.3.2 Security and Privacy Requirements and Measures for Federated Learning

Field name	Details
ID	SP-REQ-12
Title	Data Privacy
Priority	Must have
Category	Data, AI model
Description and Rationale	The data subject's data of PHASE IV AI MUST be protected from unauthorised access, disclosure, modification, or deletion. In fact, protecting user data is essential for maintaining trust and ensuring compliance with privacy regulations. Unauthorised access or disclosure could lead to identity theft or reputational damage. However, data modification or deletion could compromise the integrity of the federated learning process and lead to inaccurate or biased models.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> • Implement strong encryption for data at rest and in transit. • Use secure storage mechanisms, such as hardware security modules (HSMs). • Enforce strict access controls and role-based permissions. • Regularly audit and monitor data security practices.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-13
Title	Model Privacy
Priority	Must have
Category	Data, AI model
Description and Rationale	The PHASE IV AI model updates MUST not reveal sensitive information about the underlying data. In fact, FL model updates can inadvertently leak information about the training data, even if the raw data itself is never shared. This could expose sensitive personal information. Protecting model privacy is crucial for ensuring user privacy and preventing potential harm.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> • Employ differential privacy techniques to add noise to model updates. • Use secure aggregation protocols to prevent individual contributions from being identified. • Explore homomorphic encryption to enable computations on encrypted data. • Implement secure model sharing mechanisms to prevent unauthorised access.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-14
Title	Membership Privacy

Field name	Details
Priority	Must have
Category	Data, AI model
Description and Rationale	Data subjects SHOULD not be identifiable from their participation in the federated learning process. In fact, membership privacy ensures that data subjects can participate in federated learning without fear of being identified or targeted, where data subject anonymity is crucial for protecting privacy and preventing discrimination.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> • Use anonymisation techniques to remove personally identifiable information (PII) from data. • Implement secure device registration and authentication protocols. • Avoid collecting or storing any information that could be used to identify individual users.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-15
Title	Model Security
Priority	Must have
Category	Data, AI model
Description and Rationale	Models MUST be protected from malicious attacks that could compromise their accuracy or integrity. In fact, models trained using federated learning can be vulnerable to attacks that could manipulate their behaviour or cause them to learn biased or inaccurate information. Protecting the model from malicious attacks is crucial for ensuring the trustworthiness and reliability of the federated learning process.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> • Implement robust security measures to prevent unauthorised access to models. • Use secure model deployment and execution environments. • Regularly monitor models for anomalies and potential attacks.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-16
Title	Data Integrity
Priority	Must have
Category	Data, AI model
Description and Rationale	Data used for training MUST be accurate, complete, and consistent. In fact, data integrity is essential for ensuring the accuracy and reliability of the federated learning

Field name	Details
	models. Inaccurate or incomplete data can lead to biased or inaccurate models, which could have serious consequences in the developed applications.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Implement data validation and verification procedures. • Use tamper-proof logging mechanisms to track data provenance. • Regularly monitor data quality and integrity.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-17
Title	Transparency and Accountability
Priority	Should have
Category	Data, AI model
Description and Rationale	Data subjects SHOULD be informed about how their data is used in the federated learning process. In fact, transparency and accountability are essential for building trust with users and ensuring ethical data practices. Data subjects should have clear and concise information about how their data is collected, used, and shared. They should also have the right to control their data and opt-out of participation if they choose.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Provide clear and concise privacy policies that explain data collection, usage, and sharing practices. • Offer users control over their data, including the ability to opt-in or opt-out of participation.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-18
Title	Poisoning Attacks on federated learning models
Priority	Nice to have
Category	Data, AI model
Description and Rationale	Poisoning attacks in federated learning can significantly degrade the performance of the global model by introducing malicious updates. In PHASE IV AI it would be NICE to have an implementation of defence mechanisms against malicious actors targeting both data and the federated learning model. Indeed, malicious actors (clients) contributing to the federating learning process can inject poisoned data (data poisoning) into the training process, leading to compromised model accuracy and biased predictions. This can occur when malicious clients upload tampered data weights, potentially gaining control over multiple clients' local models and manipulating the global model with their crafted data. The malicious party also can modify the model updates (model poisoning) directly before sending them to the central server for

Field name	Details
	aggregation. This enables them to inject malicious parameters into the global model, poisoning its integrity and functionality.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Local Node Measures <ul style="list-style-type: none"> ○ Data Sanitisation: Ensure that the data used for training is clean and free from anomalies. This can involve outlier detection and removal⁵⁸. ○ Robust Training Algorithms: Use robust training algorithms that are less sensitive to malicious data points⁵⁹. ○ Behaviour Monitoring: Monitor the behaviour of the local model updates. If a node's updates deviate significantly from the norm, it could be an indication of poisoning⁶⁰. • Server Node Measures <ul style="list-style-type: none"> ○ Anomaly Detection: Implement anomaly detection mechanisms to identify and exclude suspicious updates from the aggregation process⁶⁰. ○ Reputation Systems: Maintain a reputation score for each node based on the quality of their updates. Nodes with consistently poor performance can be excluded from the training process⁶¹. ○ Behaviour Attestation: Use behaviour attestation techniques to verify that the updates from each node are consistent with expected behavior⁶⁰.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-19
Title	Membership Inference Attacks on federated learning models
Priority	Nice to have
Category	Data, AI model
Description and Rationale	In PHASE IV AI it would be NICE to have an implementation against membership inference attacks on federated learning models. Indeed, attackers utilise the global model to infer the presence of specific data points in the training set. This violates individual privacy by revealing information about the data contributors.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Adding noise (Differential Privacy Techniques) to the training data or model updates to protect individual privacy. This makes it difficult for attackers to infer the presence of specific data points in the training set. • Using algorithms that are designed to protect individual privacy, such as federated learning with differential privacy. These algorithms ensure that the model does not reveal sensitive information about individuals.
Supporting materials	

⁵⁸ <https://arxiv.org/abs/2308.11333>
⁵⁹ Amirhossein Reiszadeh, Farzan Farnia, Ramtin Pedarsani, Ali Jadbabaie: Robust Federated Learning: The case of Affine Distribution Shifts. NeurIPS 2020

⁶⁰ <https://ieeexplore.ieee.org/stampPDF/getPDF.jsp?tp=&arnumber=10309113&ref=>
⁶¹ <https://arxiv.org/abs/2011.10464>

Field name	Details
Comments	

By taking into consideration the requirements and implementing these defence mechanisms and solutions, PHASE IV AI can build secure and trustworthy Federated Learning systems that protect individual privacy whilst unlocking the potential of collaborative AI.

3.3.3 Common Security and Privacy Requirements and Measures for: DM, GAN, VAE.

There are many overlapping security and privacy requirements and measures across the different neural network models that will be used in the PHASE IV AI project, such as DMs, GANs, and VAEs. The common security and privacy requirements and measures are as follows:

Field name	Details
ID	SP-REQ-20
Title	Data Security by Design and Privacy by Design
Priority	Nice to have
Category	Data, AI model, Security by design, Privacy by Design
Description and Rationale	In PHASE IV AI it would be NICE to have an incorporation of security measures into the design and development of the model from the beginning. The model could be used for the synthetic health data generation or the prediction of disease/diagnosis.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> • Use secure development practices to minimise vulnerabilities. • Implement security testing throughout the development lifecycle. • Regularly update the model and its security measures.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-21
Title	Model trustworthiness
Priority	Should have
Category	AI model
Description and Rationale	PHASE IV AI SHOULD ensure the model is developed and used in a fair, transparent, and accountable manner, by implementing measures to mitigate bias and discrimination in the model's outputs. It should ensure the model's outputs are interpretable and understandable by humans.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> • Use diverse and representative training data to minimise bias. • Implement fairness auditing techniques to assess and mitigate bias. • Develop explainability methods to understand the model's reasoning. • Consider using differential privacy techniques to protect individual data in the training process.

Field name	Details
Supporting materials	D1.5, D2.3, D6.1
Comments	

4. Security and Privacy Requirements and Measures for Privacy-Enhancing Technologies

In this section, we describe the security and privacy requirements and measures for privacy-enhancing technologies that will be implemented in PHASE IV AI, particularly in WP4. Each of these solutions addresses security and privacy concerns, though they come with their own advantages and disadvantages. Additionally, each solution has specific requirements and measures that need to be met.

4.1 Privacy-Enhancing Technologies

This section proposes an overview of the proposed privacy-enhancing technologies in PHASE IV AI.

4.1.1 Differential Privacy

Differential privacy is a framework for ensuring privacy in data analysis by injecting controlled noise into the data, making it difficult to identify individual data points while preserving the overall statistical properties. It guarantees that the output of an algorithm remains essentially unchanged even if a single individual's data is added or removed from the dataset. This is achieved by adding carefully calibrated noise to the data or the results, making it impossible to identify or infer information about any specific individual. Differential privacy provides strong privacy guarantees while still allowing for the extraction of valuable insights from the data⁶².

4.1.2 Secure Multiparty Computation

Secure multiparty computation (SMPC) is a cryptographic technique that enables multiple parties to jointly compute a function on their private data without revealing their individual inputs to each other. This means that healthcare institutions, researchers, or even patients themselves can collaborate on analysing sensitive data, such as medical records, without compromising privacy. For example, hospitals could jointly identify disease trends or develop new treatments without sharing individual patient information. MPC ensures that the computation is performed securely, and the results are only revealed to authorised parties^{63,64}.

4.1.3 Homomorphic Encryption

Homomorphic encryption (HE) is a revolutionary cryptographic technique that allows computations on encrypted data without decrypting it. This means sensitive healthcare data, like patient records, can be analysed and processed while remaining secure and confidential. HE offers several benefits for PHASE IV AI, including enhanced privacy, improved security, facilitated data sharing and collaboration, and the ability to perform advanced analytics on encrypted data. However, challenges remain, such as computational overhead, limited functionality, and implementation complexity. Despite these challenges, HE holds immense potential to transform how healthcare data is managed and utilised, leading to improved patient care and research advancements^{65,66,67}.

⁶² Dwork, C. (2011). Differential Privacy. In: van Tilborg, H.C.A., Jajodia, S. (eds) Encyclopedia of Cryptography and Security. Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-5906-5_752

⁶³ Andrew Chi-Chih Yao: Protocols for Secure Computations (Extended Abstract). FOCS 1982: 160-164

⁶⁴ Lindell, Yehuda & Pinkas, Benny. (2008). Secure Multiparty Computation for Privacy-Preserving Data Mining. IACR Cryptology ePrint Archive. 2008. 197. 10.29012/jpc.v1i1.566.

⁶⁵ Craig Gentry: A fully homomorphic encryption scheme. Stanford University, USA, 2009

⁶⁶ Zvika Brakerski, Vinod Vaikuntanathan: Efficient Fully Homomorphic Encryption from (Standard) LWE. SIAM J. Comput. 43(2): 831-871 (2014)

⁶⁷ Shai Halevi, Victor Shoup: Algorithms in HElib. CRYPTO (1) 2014: 554-571

4.1.4 Characteristics and Similarities

Security and privacy similarities common to Secure Multiparty Computation (MPC), Homomorphic Encryption, and Differential Privacy include:

- **Privacy-Preserving Data Analysis:** All three techniques enable data analysis to be carried out on sensitive data while preserving privacy. They allow for computations and insights to be derived from data without compromising individual privacy.
- **Data Confidentiality:** They all prioritise data confidentiality by ensuring that sensitive information is not exposed during processing. This is achieved through various mechanisms like encryption, noise addition, and secure computation protocols.
- **Compliance with Data Protection Regulations:** They can be employed to comply with data protection regulations like GDPR, HIPAA, and NIS2, by ensuring that data is handled securely and ethically, respecting individual rights and minimising privacy risks.
- **Data Integrity and Availability:** They all aim to maintain the integrity and availability of data during processing. They often include mechanisms to detect and prevent data tampering, ensuring that the results of analysis are reliable and trustworthy.
- **Flexibility and Adaptability:** They can be adapted to different data analysis tasks and scenarios. They can be integrated into various data processing pipelines and workflows, enabling secure and privacy-preserving analysis across different applications.
- **Focus on Data Protection:** They are all designed with a strong focus on data protection and privacy. They prioritise safeguarding sensitive information while enabling valuable insights to be derived from data.

4.2 Security and Privacy Requirements and Measures

This section proposes the security and privacy requirements and measures of the differential privacy, secure multiparty computation, and homomorphic encryption.

4.2.1 Differential Privacy

For creating the security and privacy requirements and their measures for the differential privacy, we use the guidelines provided by *NIST SP 800-226 ipd*⁶⁸ for evaluating differential privacy guarantees.

Field name	Details
ID	SP-REQ-22
Title	Data Security During Computation
Priority	Nice to have
Category	Data

⁶⁸ <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-226.ipd.pdf>

Field name	Details
Description and Rationale	In PHASE IV AI, it would be NICE to protect data from exposure during processing and analysis.
Measures or Tools	<p>The following measures and techniques are suggested to be combined with Differential Privacy to enhance data security and privacy:</p> <ul style="list-style-type: none"> • Use Secure Multi-Party Computation (MPC) techniques to allow computation on encrypted data without decryption. • Utilise Fully Homomorphic Encryption FHE to perform computations directly on encrypted data. • Leverage secure hardware enclaves (e.g., Intel SGX, AMD SEV) to protect data during computation.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-23
Title	Privacy Budget Management
Priority	Must have
Category	Data
Description and Rationale	PHASE IV AI MUST carefully manage the total privacy loss (ϵ) across all analyses of a dataset in order to ensure sufficient privacy protection. Allocating a privacy budget in differential privacy is like dividing a limited resource among different analyses of the same dataset. Each analysis consumes a portion of the overall privacy budget, represented by the privacy parameter ϵ . The goal is to distribute this budget strategically, ensuring that no single analysis consumes too much privacy, while still allowing for meaningful insights from the data. This process involves considering the sensitivity of each query, the importance of the insights it provides, and the overall privacy goals for the dataset.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Implement mechanisms to track and allocate privacy budget for each query. • Use adaptive privacy budgets that adjust dynamically based on query complexity and data sensitivity. • Prioritise queries based on their importance and potential privacy risks.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-24
Title	Side Channel Protection
Priority	Should have
Category	Data
Description and Rationale	PHASE IV AI SHOULD prevent information leakage through unintended channels, such as timing, error messages, or query execution times.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Implement algorithms that execute in constant time, regardless of data values.

Field name	Details
	<ul style="list-style-type: none"> Handle errors gracefully and avoid revealing sensitive information through error messages. Obfuscate query execution times and other side channels. Regularly audit systems for potential side channel vulnerabilities.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-25
Title	Transparency and Accountability
Priority	Should have
Category	Data
Description and Rationale	PHASE IV AI SHOULD ensure transparency about the privacy guarantees provided by the system and maintain accountability for privacy practices
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> Provide clear and concise documentation of the differential privacy framework used, including privacy parameters (ϵ, δ), unit of privacy, and threat model. Implement mechanisms for auditing and monitoring privacy practices to ensure compliance with regulations and best practices. Disclose information about privacy guarantees and practices to users in a clear and understandable manner.
Supporting materials	
Comments	

4.2.2 Secure Multiparty Computation

Secure Multiparty Computation is a cryptographic technique that allows several parties to collaboratively compute a function on their private data without disclosing the data itself. Each party's input is divided into shares using secret sharing techniques, ensuring that no single party can reconstruct the original input without collaborating with others. It guarantees data confidentiality and integrity, safeguarding against unauthorized access and tampering. To define the security and privacy requirements and their measures for secure multiparty computation, we use the guidelines provided by NIST⁶⁹.

Field name	Details
ID	SP-REQ-26
Title	Confidentiality of Data and Computations
Priority	Should have
Category	Data

⁶⁹ <https://csrc.nist.gov/presentations/2023/mpts2023-day1-talk-mpc-apps>

Field name	Details
Description and Rationale	SMPC in PHASE IV AI SHOULD keep data inputs and intermediate computations confidential to each participant.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> Distributing data and computations across multiple parties, ensuring no single party has access to the full information. Using protocols designed to ensure confidentiality throughout the computation process.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-27
Title	Ensuring Accurate and Trustworthy Computations
Priority	Should have
Category	Data
Description and Rationale	SMPC in PHASE IV AI SHOULD correctly perform the computation and the output should be trustworthy.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> Ensuring that each party holds a valid share of the secret. Allowing parties to prove the correctness of their computations without revealing any private information. Byzantine Fault Tolerance: Handling malicious participants who may try to disrupt the computation.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-28
Title	Availability
Priority	Nice to have
Category	Data, Application, Infrastructure
Description and Rationale	In PHASE IV AI it would be NICE to provide resilient computation to failures and continue even if some parties are unavailable.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> Designing protocols that can tolerate a certain number of failures. Deploying multiple servers to ensure availability even if some servers fail. Allowing parties to communicate asynchronously, reducing the impact of network delays or outages.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-29
Title	Fairness
Priority	Should have
Category	Data, Application
Description and Rationale	SMPC in PHASE IV AI SHOULD receive the output of the computation simultaneously from all participants, preventing any party from gaining an unfair advantage.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> • Designing protocols that ensure fairness by preventing any party from prematurely learning the output. • Allowing parties to commit to their inputs before the computation begins, preventing them from changing their inputs after seeing the results of others.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-30
Title	Privacy-Preserving Computations
Priority	Should have
Category	Data, Application
Description and Rationale	SMPC in PHASE IV AI SHOULD not reveal any sensitive information from the computation about the participants' inputs beyond what is necessary to produce the output.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> • Adding noise to the data to protect individual privacy while still allowing for meaningful analysis using differential privacy. • Aggregating data from multiple parties in a way that preserves individual privacy.
Supporting materials	
Comments	

4.2.3 Homomorphic Encryption

While Homomorphic Encryption (HE) offers a promising approach to mitigate privacy risks, it is important to note that it introduces its own set of security and privacy considerations that need to be addressed. The security and privacy requirements and measures for homomorphic encryption are as follows.

Field name	Details
ID	SP-REQ-31
Title	Security of encryption keys
Priority	Should have
Category	Data, infrastructure

Field name	Details
Description and Rationale	PHASE IV AI SHOULD securely manage encryption keys to prevent unauthorised access and use.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Implement secure key management systems to generate, store, and manage encryption keys. • Store encryption keys in dedicated hardware security modules for enhanced protection. • Secure access to key management systems using multi-factor authentication.
Supporting materials	D2.5.
Comments	

5. Security and Privacy Requirements & Measures for User stories and Related Processes

The deliverable D6.1 User stories, usage scenarios and use case validation, serves as a comprehensive plan for the PHASE IV AI project, outlining the project's goals, objectives, and approach to validating its solutions in real-world settings. It details three specific use cases focused on lung cancer, prostate cancer, and ischemic stroke, including the datasets to be used, the potential AI models to be developed, and the strategies for data sharing and security. The deliverable also presents a collection of user stories gathered from stakeholder workshops, representing the needs and concerns of various individuals and organisations involved in the project. These user stories are categorised by theme, system, and priority, providing valuable insights for guiding the development of the PHASE IV AI health data hub and its services as well as the identification of the security and privacy requirements.

5.1 Concept Definitions

The deliverable D6.1 defined the user stories based on the following concepts:

- **Data Governance:** This describes the user stories for ensuring responsible and ethical use of health data within the project. The user stories highlight the importance of data governance in building trust, ensuring compliance, and protecting patient privacy while facilitating the use of health data for research and development.
- **Data Federation:** describes the importance of collaboratively analysing data from multiple sources without physically moving or sharing the raw data itself. The user stories emphasise the desire for secure, collaborative data access and analysis, highlighting the importance of data in healthcare.
- **Risk Prediction:** describes the process of using data and algorithms to estimate an individual's likelihood of developing a specific disease, for example lung cancer. The user stories illustrate how different stakeholders, including medical researchers, clinicians, and healthcare managers, are interested in utilising risk prediction tools for various purposes, such as patient stratification, and data-driven decision support.
- **Detection/Diagnosis:** refers to the process of identifying and confirming the presence of a disease (e.g., lung cancer) in individuals. This involves using various tools and techniques, such as chest X-rays, CT scans, and AI algorithms, to analyse patient data and identify potential anomalies or signs of disease. The goal is to achieve early detection, which allows for earlier diagnosis and potentially better treatment outcomes. The user stories highlight how different stakeholders, including clinicians, healthcare managers, and survivors, are interested in improving detection and diagnosis methods to achieve earlier intervention and reduce delays in treatment.
- **Data Gathering:** refers to the process of collecting and integrating relevant data from various sources to support disease detection, diagnosis, and risk prediction. This involves gathering data from different healthcare providers, systems, and databases, including primary care records, imaging scans, and patient demographics. The goal is to create a comprehensive and unified dataset that can be used to train and validate AI models, improve risk assessment, and enhance early detection efforts. The user stories highlight the need for data harmonisation, data sharing, and automated data extraction to ensure that data is collected efficiently and effectively from a broad base, ultimately leading to more robust and reliable results.

- **Accuracy/Quality:** refers to the reliability and trustworthiness of the data used the use cases. The user stories highlight the importance of identifying and addressing inaccuracies in primary care data, as well as validating synthetic data against real-world data to ensure its reliability, and achieving high-quality data that can support accurate diagnoses, improve risk assessment, and ultimately lead to better patient outcomes.
- **Data Operations:** refers to the processes involved in managing, manipulating, and transforming data to support disease detection, and diagnosis. This encompasses tasks such as data cleaning, de-identification, data harmonisation, data augmentation, and data labelling. The goal is to ensure that the data is accurate, consistent, and usable for downstream analysis and model development. The user stories highlight the need for robust data operations to address challenges like data quality, and privacy, and the need for specific data formats for AI algorithms.
- **Missing data:** refers to the absence of information about individuals in different use cases. This can occur due to various reasons, such as individuals not participating in screening programs, incomplete medical records, or individuals being disengaged from healthcare systems. The user stories highlight the need to address missing data. The goal is to identify and understand the characteristics of the population with missing data to improve engagement, follow-up, and ultimately, the effectiveness of screening programs.
- **Healthcare efficiency:** refers to optimising the use of resources, time, and effort in disease detection and diagnosis. This involves streamlining workflows, reducing unnecessary interventions, and utilising technology to improve accuracy and speed. The user stories describe the desire for AI-powered tools to automate tasks, prioritise patients for screening, and provide rapid results, ultimately reducing the burden on healthcare providers and improving patient care.
- **End User Delivery:** pertains to the presentation and accessibility of risk prediction information to patients and healthcare providers. This involves making the information understandable, accessible, and actionable for both groups. The user stories describe the need for user-friendly interfaces, clear communication of risk scores, and the ability for patients to access their own results. The goal is to ensure that both patients and clinicians can use the information provided by the system to make informed decisions about screening, treatment, and overall health management.
- **Exploitation:** refers to the use of data and technology for practical applications and commercial purposes. This involves making data accessible to researchers, companies, and other stakeholders to develop new tools, products, and services that can improve detection, diagnosis, and treatment of cancers and diseases. The user stories highlight the need for clear guidelines and mechanisms for data access and use, ensuring that data is used responsibly and ethically.
- **Inclusion:** refers to ensuring that prediction models and interventions are fair and equitable for all individuals, regardless of their demographics or socioeconomic status. The user stories describe the need for data that reflects the real distribution of the population, as well as the importance of addressing potential biases in algorithms to prevent discrimination.
- **Academic:** refers to the use of data and technology for research, education, and knowledge advancement, for example, in the field of cancers. This involves using data to train AI models, develop new algorithms, and conduct research studies to improve understanding of cancer, its causes, and its treatment. The user stories highlight the use of synthetic data for training, educational purposes, and simulation studies, as well as the desire to publish research findings to advance the field.

5.2 Security and Privacy Requirements and Measures

The following security and privacy requirements and measures target the described categories in the last section for user stories, usage scenarios and use case validation.

Field name	Details
ID	SP-REQ-32
Title	GDPR Article 5: Lawfulness, Fairness, and Transparency
Priority	Must have
Category	Data, AI model, Infrastructure
Description and Rationale	The data of PHASE IV AI MUST be processed lawfully, fairly, and transparently.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Data Processing Agreements (DPAs): Where necessary, establish clear DPAs with all involved parties outlining data processing purposes, legal basis, and data subject rights. • Privacy Notices: Provide clear and concise privacy notices to data subjects detailing how their data is collected, used, and protected. • Transparency in Data Use: Document and make transparent the specific purposes for which data is used, including for AI model development and validation.
Supporting materials	D1.5, D2.3, D6.1.
Comments	It targets the following categories in the deliverable D6.1: Data Governance, Data Federation, Risk Prediction, Detection/Diagnosis, Data Gathering, Data Operations, Missing Data, Healthcare Efficiency, End User Delivery, Exploitation, Inclusion, Academic.

Field name	Details
ID	SP-REQ-33
Title	GDPR Article 6: Lawfulness of Processing
Priority	Must have
Category	Data, AI model, Infrastructure
Description and Rationale	The PHASE IV AI processing MUST be based on a lawful basis, such as consent, contract, or legitimate interest.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Conduct a thorough assessment of the legitimate interest for processing health data for AI development, balancing the interests of data subjects with the public benefit. • Obtain explicit and informed consent from data subjects for the processing of their data, where applicable.
Supporting materials	D1.5, D2.3, D6.1.
Comments	It targets the following categories in the deliverable D6.1: Data Governance, Data Federation, Risk Prediction, Detection/Diagnosis, Data Gathering, Data Operations,

Field name	Details
	Missing Data, Healthcare Efficiency, End User Delivery, Exploitation, Inclusion, Academic.

Field name	Details
ID	SP-REQ-34
Title	GDPR Article 32: Security of Processing
Priority	Must have
Category	Data, AI model, Infrastructure, Management, Privacy Policies and Regulations.
Description and Rationale	PHASE IV AI MUST implement appropriate technical and organisational measures to protect personal data against unauthorised processing and accidental loss, destruction, or damage.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Encrypt data at rest and in transit using robust encryption algorithms. • Implement strong access control mechanisms to limit access to data based on need-to-know principles. • Use secure data storage and processing environments with physical and logical security measures. • Conduct regular security assessments and penetration testing to identify and mitigate vulnerabilities.
Supporting materials	D1.5, D2.3, D6.1.
Comments	It targets the following categories in the deliverable D6.1: Data Governance, Data Federation, Risk Prediction, Detection/Diagnosis, Data Gathering, Data Operations, Missing Data, Healthcare Efficiency, End User Delivery, Exploitation, Inclusion, Academic.

Field name	Details
ID	SP-REQ-35
Title	GDPR Article 17: Right to Erasure ("Right to be Forgotten")
Priority	Must have
Category	Data, AI model, Infrastructure, Management, Privacy Policies.
Description and Rationale	PHASE IV AI MUST provide a mechanism so that the data subjects have the right to have their personal data erased under certain circumstances.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Data Deletion: Implementing processes to permanently delete personal data from all systems, including backups. • De-identification: Anonymising data so that it can no longer be linked to the individual. • Access Control: Restricting access to personal data to prevent unauthorised processing. • Data Mapping: Keeping an updated map of where personal data is stored to ensure all copies are erased. • Notification Systems: Informing other controllers and processors to delete any copies or links to the personal data.
Supporting materials	D1.5, D2.3, D6.1.

Field name	Details
Comments	It targets the following categories in the deliverable D6.1: Data Governance, Data Federation, Risk Prediction, Detection/Diagnosis, Data Gathering, Data Operations, Missing Data, Healthcare Efficiency, End User Delivery, Exploitation, Inclusion, Academic.

Field name	Details
ID	SP-REQ-36
Title	GDPR Article 20: Right to Data Portability
Priority	Must have
Category	Data, AI model, Infrastructure, Management, Privacy Policies.
Description and Rationale	PHASE IV AI MUST provide a mechanism so that the data subjects have the right to receive their personal data in a portable format.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> Develop mechanisms to allow data subjects to receive their data in a commonly used format (e.g., CSV, JSON) for transfer to other services.
Supporting materials	D1.5, D2.3, D6.1.
Comments	It targets the following categories in the deliverable D6.1: Data Governance, Data Federation, Risk Prediction, Detection/Diagnosis, Data Gathering, Data Operations, Missing Data, Healthcare Efficiency, End User Delivery, Exploitation, Inclusion, Academic.

Field name	Details
ID	SP-REQ-37
Title	Data Governance Act
Priority	Should have
Category	Data, AI model, Infrastructure, Management, Privacy Policies.
Description and Rationale	PHASE IV AI SHOULD ensure the responsible use and governance of data, including for AI development.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> Establish a comprehensive data governance framework that aligns with the Data Governance Act, including roles, responsibilities, and processes for data management. Implement processes for data quality management, ensuring data accuracy, completeness, and consistency. Develop standardised data sharing agreements that comply with the Data Governance Act, including provisions for data access, use, and security.
Supporting materials	D1.5, D2.3, D6.1.
Comments	It targets the following categories in the deliverable D6.1: Data Governance, Data Federation, Risk Prediction, Detection/Diagnosis, Data Gathering, Data Operations, Missing Data, Healthcare Efficiency, End User Delivery, Exploitation, Inclusion, Academic.

Field name	Details
ID	SP-REQ-38
Title	Cybersecurity enhancements
Priority	Should have
Category	Data, AI model, Infrastructure, Management, Privacy Policies.
Description and Rationale	PHASE IV AI SHOULD enhance cybersecurity requirements for essential services such as the detection/diagnosis, data federation and governance.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Conduct thorough risk assessments to identify and prioritise cybersecurity threats and vulnerabilities. • Develop and implement a comprehensive incident response plan to address cybersecurity incidents effectively. • Implement continuous security monitoring and logging to detect and respond to suspicious activity. • Provide regular security awareness training to all personnel involved in data handling.
Supporting materials	D1.5, D2.3, D6.1.
Comments	It targets the following categories in the D6.1: Data Governance, Data Federation, Risk Prediction, Detection/Diagnosis, Data Gathering, Data Operations, Missing Data, Healthcare Efficiency, End User Delivery, Exploitation, Inclusion, Academic.

Field name	Details
ID	SP-REQ-39
Title	AI act-based AI model development
Priority	Should have
Category	AI model
Description and Rationale	PHASE IV AI SHOULD establish rules for the development and deployment of AI systems, including requirements for transparency, accountability, and risk mitigation.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Conduct a comprehensive risk assessment of AI systems used for health data processing, identifying potential risks and implementing mitigation measures. • Ensure that AI systems are transparent and explainable, allowing users to understand how decisions are made. • Maintain human oversight of AI systems to ensure ethical and responsible use.
Supporting materials	D1.5, D2.3, D6.1.
Comments	It targets the following categories in the deliverable D6.1: Data Governance, Data Federation, Risk Prediction, Detection/Diagnosis, Data Gathering, Data Operations, Missing Data, Healthcare Efficiency, End User Delivery, Exploitation, Inclusion, Academic.

Field name	Details
ID	SP-REQ-40

Field name	Details
Title	Pseudonymisation/Anonymisation implementation
Priority	Should have
Category	Data
Description and Rationale	PHASE IV AI SHOULD implement appropriate pseudonymisation or anonymisation techniques to protect sensitive data. This is critical for all categories, especially Data Governance, Data Federation, and Data Gathering.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Employ robust pseudonymisation and anonymisation techniques. This involves replacing directly identifiable information with unique, non-sensitive identifiers, effectively masking personal details while preserving data integrity and usability. • Carefully select and implement appropriate techniques based on the specific data categories and their associated risks, ensuring a balance between data utility and privacy protection.
Supporting materials	D1.5, D2.3, D6.1.
Comments	It targets the following categories in the D6.1: deliverable Data Governance, Data Federation, Risk Prediction, Detection/Diagnosis, Data Gathering, Data Operations, Missing Data, Healthcare Efficiency, End User Delivery, Exploitation, Inclusion, Academic.

Field name	Details
ID	SP-REQ-41
Title	Data retention implementation
Priority	Must have
Category	Data
Description and Rationale	PHASE IV AI MUST implement data retention policies that comply with GDPR requirements, ensuring data is not kept longer than necessary.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Comply with GDPR regulations and ensure data privacy by implementing strict data retention policies across all categories. These policies will clearly define the duration for which specific data types will be retained, ensuring that data is only kept for as long as necessary to fulfil its intended purpose. • Regularly review and update these policies to reflect changes in legal requirements and organisational needs. This proactive approach will minimise the risk of data breaches and its consequence while ensuring that non-sensitive and sensitive information is not stored beyond its legitimate use, promoting data security and individual privacy.
Supporting materials	D1.5, D2.3, D6.1.
Comments	It targets the following categories in the deliverable D6.1: Data Governance, Data Federation, Risk Prediction, Detection/Diagnosis, Data Gathering, Data Operations, Missing Data, Healthcare Efficiency, End User Delivery, Exploitation, Inclusion, Academic.

6. Security and Privacy Requirements & Measures for PHASE IV AI Platform

6.1 Data Handling in PHASE IV AI

The security and privacy requirements and measures have been created from both DPIA (D1.5) and Initial Technology assessment and List (D2.5) following these steps:

1. **Identifying the relevant data protection laws and regulations:** We started by identifying the relevant data protection laws and regulations that apply to the project, such as the General Data Protection Regulation (GDPR).
2. **Analysing the PHASE IV AI project context:** We analysed the project context to identify the types of data being collected and processed, the purposes of processing, the data subjects involved, and the potential risks to data subjects.
3. **Drawing on best practices:** We looked at best practices in data protection to identify additional requirements and measures that could be implemented to further enhance data protection.
4. **Using security and privacy technical knowledge of data protection:** We studied both security and privacy technical knowledge of data protection to develop a comprehensive list of requirements and measures that address the specific needs of the project.
5. **Consulting with data protection experts involved in PHASE IV AI:** We consulted with the data protection experts in our team to ensure that the understanding of the requirements and measures was accurate and up-to-date.

6.2 European Health Data Space (EHDS)

The European Health Data Space (EHDS) is a comprehensive framework designed to facilitate the secure and responsible sharing of health data within the EU. It aims to empower citizens by granting them access to their personal health data, enhance healthcare through seamless cross-border access for healthcare professionals, and foster research and innovation by providing a secure platform for data reuse⁷⁰. The EHDS achieves this through a secure infrastructure, interoperability rules, clear access guidelines, robust data governance, and diverse participation. It is divided into three main sections: MyHealth@EU for primary use in patient care, HealthData@EU for secondary use in research and innovation, and a certification section for ensuring the security and interoperability of digital health services. By creating a truly integrated European health data ecosystem, the EHDS aims to improve patient care, drive innovation, and empower citizens to take control of their health information⁷⁰.

6.3 Health Data Hub

The Health Data Hub will be developed as a part of PHASE IV AI, and serves as a central platform for facilitating the secondary use of health data. It will act as a market for exchanging information about data requirements and availability, providing access to Data as a Service (DaaS) and Model as a Service (MaaS), and ensuring the quality and trustworthiness of that exchanged data. The Health Data Hub will promote GDPR compliance and foster collaboration among stakeholders in the healthcare ecosystem.

⁷⁰ https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en

6.4 Phase IV AI Platform

The Phase IV AI Platform, encompassing the Health Data Hub, Model as a Service (MaaS), and Data as a Service (DaaS), will form a secure and collaborative ecosystem for utilising health data for research and innovation. The Health Data Hub acts as a central marketplace for discovering and acquiring data, ensuring data quality and GDPR compliance. DaaS will provide access to anonymised and synthetic health data, safeguarding patient privacy. MaaS will empower researchers to train and test machine learning models on data from multiple sources without compromising privacy through secure multi-party computation. This integrated system will facilitate ethical data sharing, accelerate research, and drive the development of data-driven healthcare solutions.

The security and privacy requirements and measures are described as follows:

Field name	Details
ID	SP-REQ-42
Title	Data Minimisation
Priority	Should have
Category	Data, Network, Application, Infrastructure and Management.
Description and Rationale	The PHASE IV AI platform SHOULD process only the minimum amount of personal data necessary to achieve the project's objectives for both EHDS and the complete system.
Measures or Tools	<p>The following measures and tools are suggested:</p> <ul style="list-style-type: none"> • Implement federated learning to minimise data transfer and processing on central servers. • Develop specific data models that only include the essential information for each use case. • Purpose specification: Clearly define the purpose for data collection and ensure that only data relevant to that purpose is collected. • Data collection limitation: Use techniques like checkboxes or dropdown menus instead of freeform text to limit the amount of data collected. • Data anonymisation and pseudonymisation: Replace personal identifiers with pseudonyms or anonymise data to reduce the risk of identification. • Automated data deletion: Implement automated processes to delete data that is no longer needed after a certain period. • Regular audits: Conduct periodic reviews and audits to ensure that only necessary data is retained, and any excess data is deleted.
Supporting materials	D6.1
Comments	

Field name	Details
ID	SP-REQ-43
Title	Purpose Limitation
Priority	Should have
Category	Data, Network, Application, Infrastructure and Management.
Description and Rationale	The PHASE IV AI platform SHOULD process data only for the specific, explicit, and legitimate purposes defined in the project's scope.
Measures or Tools	The following measures and tools are suggested:

Field name	Details
	<ul style="list-style-type: none"> Clearly document the intended use of data in data sharing agreements, consent forms, and technical documentation. Regularly review data processing activities to ensure they remain aligned with the original purpose.
Supporting materials	D1.5, D2.3, D6.1.
Comments	

Field name	Details
ID	SP-REQ-44
Title	Data Security
Priority	Should have
Category	Data, Network, Application, Infrastructure and Management.
Description and Rationale	The PHASE IV AI platform SHOULD implement technical and organisational measures to protect data from unauthorised access, use, disclosure, alteration, or destruction.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> Encrypt data at rest and in transit using strong encryption algorithms. Implement robust access controls, including multi-factor authentication and role-based access. Use secure storage solutions, such as encrypted cloud storage or dedicated secure processing environments. Regularly conduct security audits and vulnerability assessments. Develop and implement incident response plans to address data breaches or security incidents.
Supporting materials	D2.5, D6.1.
Comments	

Field name	Details
ID	SP-REQ-45
Title	Data Integrity
Priority	Should have
Category	Data, Network, Application, Infrastructure and Management.
Description and Rationale	The PHASE IV AI platform SHOULD ensure the accuracy, completeness, and consistency of data throughout its lifecycle.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> Establish data quality checks and validation processes. Implement data governance mechanisms to monitor and ensure data integrity. Use data lineage tracking to document the origin and transformations of data. Implement data versioning and change management processes. Employ digital signatures and hashing algorithms to ensure data integrity.
Supporting materials	D2.5, D6.1.
Comments	

Field name	Details
ID	SP-REQ-46
Title	Data Subject Rights
Priority	Must have
Category	Data, Network, Application, Infrastructure and Management.
Description and Rationale	The PHASE IV AI platform MUST ensure data subjects can exercise their rights under the GDPR, including access, rectification, erasure, restriction, and data portability.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> • Provide clear and accessible information to data subjects about their rights and how to exercise them. • Develop and implement procedures for handling data subject requests efficiently and effectively. • Establish a designated contact point for data subjects to address their concerns.
Supporting materials	D1.5, D2.5, D6.1.
Comments	

Field name	Details
ID	SP-REQ-47
Title	Transparency
Priority	Should have
Category	Data, Network, Application, Infrastructure and Management.
Description and Rationale	The PHASE IV AI platform SHOULD provide clear and transparent information to data subjects, data holders, and other stakeholders about the project's data processing activities.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> • Take into consideration the PHASE IV AI data management plan (DMP) outlining data processing activities. • Provide clear and concise information about the PHASE IV AI's purpose, data sources, and intended use of data in consent forms and data sharing agreements. • Develop a user-friendly interface for the Health Data Hub that provides information about data access, usage, and security.
Supporting materials	D1.5, D2.3, D2.5, D6.1.
Comments	

Field name	Details
ID	SP-REQ-48
Title	Accountability
Priority	Must have
Category	Data, Network, Application, Infrastructure and Management.
Description and Rationale	The PHASE IV AI platform MUST be compliant with the applicable data protection regulations, including the GDPR, and adhere to established ethical principles.

Field name	Details
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Maintain detailed records of data processing activities, including data sources, purpose, recipients, and security measures. • Conduct regular internal audits to assess compliance with data protection regulations and ethical guidelines.
Supporting materials	D1.5, D2.3.
Comments	

Field name	Details
ID	SP-REQ-49
Title	Trustworthy AI
Priority	Must have
Category	AI model
Description and Rationale	The PHASE IV AI platform MUST be compliant with the EU AI Act requirements for high-risk AI systems, including risk management, data governance, transparency, and human oversight.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Conduct a thorough risk assessment for high-risk AI systems, including potential biases and unintended consequences. • Implement robust data governance practices to ensure the quality, integrity, and representativeness of training data. • Develop and deploy AI systems with appropriate human oversight mechanisms to mitigate risks and ensure ethical use. • Provide clear and transparent information to users about the AI system's capabilities, limitations, and potential impacts.
Supporting materials	D1.5, D2.3. D6.1
Comments	

Field name	Details
ID	SP-REQ-50
Title	Protecting the Health Data Hub: Measures and Best Practices
Priority	Nice to have
Category	Infrastructure
Description and Rationale	In the PHASE IV AI platform, it would be NICE TO HAVE measures to protect the Health Data Hub and its underlying infrastructure from cyberattacks and other security threats.
Measures or Tools	<p>By adhering to relevant cybersecurity standards, such as ISO 27001 and NIST Cybersecurity Framework, the following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Implement robust security controls, including firewalls, intrusion detection systems, and anti-malware software. • Conduct regular penetration testing and vulnerability assessments. • Develop and implement incident response plans to address cyberattacks or security breaches.

Field name	Details
Supporting materials	D6.1
Comments	

Field name	Details
ID	SP-REQ-51
Title	Data Protection for Electronic Communications: Adhering to the E-Privacy Directive
Priority	Should have
Category	Infrastructure
Description and Rationale	The PHASE IV AI platform SHOULD protect electronic communications data, including email, and messaging, in accordance with the E-Privacy Directive.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Use encryption and other security measures to protect electronic communications data. • Obtain explicit consent from data subjects for the processing of electronic communications data. • Implement measures to prevent the interception or monitoring of electronic communications without lawful authorisation.
Supporting materials	D1.5, D2.3.
Comments	

Field name	Details
ID	SP-REQ-52
Title	Data Encryption
Priority	Should have
Category	Infrastructure
Description and Rationale	All data in transit and at rest within the PHASE IV AI platform SHOULD be encrypted using standard algorithms.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Utilise SSL/TLS for secure communication channels. • Implement robust encryption algorithms for data storage (e.g., AES-256).
Supporting materials	D1.5, D2.3.
Comments	

Field name	Details
ID	SP-REQ-53
Title	Access Control and Authorisation
Priority	Should have
Category	Infrastructure

Field name	Details
Description and Rationale	Access to data and functionalities within the PHASE IV AI platform SHOULD be strictly controlled and authorised based on user roles and responsibilities.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> • Implement a robust Identity and Access Management (IAM) system. • Employ Role-Based Access Control (RBAC) to restrict access based on user roles. • Implement multi-factor authentication for sensitive operations.
Supporting materials	D1.5, D2.3.
Comments	

Field name	Details
ID	SP-REQ-54
Title	Data Security and Privacy by Design
Priority	Should have
Category	Infrastructure, Security by Design, Privacy by Design
Description and Rationale	Security and privacy considerations SHOULD be integrated into all stages of the PHASE IV AI development lifecycle, from design and development to deployment and maintenance.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> • Conduct thorough security and privacy risk assessments throughout the development process. • Implement security testing and penetration testing to identify vulnerabilities.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-55
Title	Data Governance and Compliance
Priority	Must have
Category	Infrastructure
Description and Rationale	The PHASE IV AI platform MUST adhere to all relevant data protection and security regulations, including GDPR, NIS2 Directive, Cyber Resilience Act, AI Act and ISO 27001.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> • Establish a comprehensive data governance framework. • Conduct regular audits and assessments to ensure compliance.
Supporting materials	D1.5, D2.3.
Comments	

Field name	Details
ID	SP-REQ-56

Field name	Details
Title	Auditing and Monitoring
Priority	Should have
Category	Infrastructure
Description and Rationale	The PHASE IV AI platform SHOULD have mechanisms for auditing user activities, data access, and security events. This information should be logged and monitored to detect potential security breaches and unauthorised access.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> • Implement comprehensive logging and monitoring systems. • Establish clear audit trails for data access and modification. • Develop incident response plans to handle security breaches effectively.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-57
Title	Data Security in the Health Data Space
Priority	Must have
Category	Infrastructure
Description and Rationale	Data exchange between PHASE IV AI platform and the external health data spaces MUST be compliant with relevant regulations.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> • Utilise secure communication protocols (e.g., FHIR) for data exchange. • Implement data access control mechanisms for external data spaces. • Ensure compliance with relevant data protection regulations for cross-border data transfers.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-58
Title	Transparency and Accountability
Priority	Must have
Category	Infrastructure
Description and Rationale	Users MUST be informed about how their data is collected, processed, and used within the PHASE IV AI platform. They MUST also have the right to access, rectify, and erase their data.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> • Develop clear and concise privacy policies and terms of use. • Provide users with transparent information about data processing activities. • Implement mechanisms for users to exercise their data rights (e.g., access, rectification, erasure).
Supporting materials	

Field name	Details
Comments	

Field name	Details
ID	SP-REQ-59
Title	Compliance with Regulations
Priority	Must have
Category	Data, AI model
Description and Rationale	The platform MUST comply with all applicable privacy and data protection regulations. In fact, compliance with regulations is essential for ensuring legal and ethical data handling. Failure to comply with regulations could result in fines, penalties, and reputational damage.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> • Conduct regular compliance assessments. • Engage with legal and privacy experts to ensure compliance with relevant regulations. • Maintain appropriate documentation to demonstrate compliance efforts.
Supporting materials	
Comments	

Field name	Details
ID	SP-REQ-60
Title	Right to Erasure
Priority	Must have
Category	Data, AI model, Security by design
Description and Rationale	In Phase IV AI, individuals MUST have the ability to have their personal data erased.
Measures or Tools	The following measures and techniques are suggested: <ul style="list-style-type: none"> • Implement procedures for individuals to request data erasure. • Erase data promptly and securely upon request.
Supporting materials	D1.5, D2.3.
Comments	

Field name	Details
ID	SP-REQ-61
Title	Zero Trust Security
Priority	Nice to have
Category	Infrastructure

Field name	Details
Description and Rationale	The PHASE IV AI framework would benefit from adopting a Zero Trust security model, assuming that no user or device can be trusted by default. This approach would be NICE TO HAVE for enhancing security.
Measures or Tools	<p>The following measures and techniques are suggested:</p> <ul style="list-style-type: none"> • Implement strong authentication and authorisation mechanisms. • Utilise micro-segmentation to isolate sensitive data and applications. • Continuously monitor and assess security posture.
Supporting materials	
Comments	

7. Conclusion

This deliverable outlines a comprehensive set of security and privacy requirements and measures for the PHASE-IV-AI project. It covers various aspects of data security, AI model security, privacy-preserving techniques, and compliance with regulations. The document includes:

- 61 security and privacy requirements categorised by system and theme, such as data security, AI model security, privacy-preserving techniques, compliance with regulations, and user stories.
- Several proposed measures and tools to address each requirement, including encryption, access controls, data governance frameworks, risk assessments, auditing, and monitoring mechanisms.
- Detailed analysis of relevant regulations and their implications for the project.
- Integration of user stories and use cases to ensure a user-centric approach to security and privacy.

The security and privacy requirements and their measures have been prioritised by the FJLU team according to the VOLERE methodology. We assigned a system category to each requirement such as: Data, AI Model, Infrastructure. We also grouped the requirements into themes such as: data governance, AI model development, PHASE IV AI platform security.

The PHASE-IV-AI project is committed to safeguarding data and protecting user privacy. This deliverable provides a robust framework for developing and implementing security and privacy measures that align with relevant regulations and best practices. By adhering to these requirements and measures, the project can build trust with stakeholders, ensure compliance, and foster a secure and ethical environment for data handling and AI models development. The deliverable emphasises the importance of:

- Continuous monitoring and evaluation of security and privacy measures to adapt to evolving threats and regulations for both AI, data, and the developed PHASE IV AI platform.
- Proactive engagement with stakeholders to address their concerns and ensure security, privacy, and transparency.
- Ongoing research and development of innovative security and privacy techniques to enhance data protection.

By implementing these principles, the PHASE-IV-AI project can contribute to the advancement of responsible and ethical AI in healthcare, ultimately benefiting patients, organisations, and society. The proposed security and privacy requirements and measures will help develop and enhance the partners' components by integrating these measures throughout the project life cycle. Additionally, implementing these measures, and mapping them to the PHASE IV AI components and architecture improvements in the upcoming deliverables is recommended.