AI Engineering Suite to Support Agile, Efficient Software Engineering

Deliverable D2.1

Systematic Protocol For AI Based Software Development

Project Number	**101189908**
Project Title	AI Engineering Suite to Support Agile Efficient Software Engineering
Project Acronym	AI4SWEng
Title of Deliverable	Systematic Protocol For AI Based Software Development
Version	V1.0
Due Date of Delivery	M12
Distribution level	PU

Work Package	**2**
Responsible Partner	IDE
Editor(s)	Pablo Gallegos(IDE)
Contributor(s)	Candela Jordano Bustos(IDE), Angel Fuentes(IDE), Victor Rosario(IDE), Jose Luis Barrera(IDE)
Reviewer(s)	Egemen Ener(BIT), Andrei Aksjonov(AI4), Marteyn Van Gasteren (ITCL)
Keywords	Privacy-by-Design, AI Engineering, Data Provenance, System Security, Traceability

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101189908.

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

Abstract: This document (D2.1) establishes the official methodology and foundational engineering reference for the AI4SWEng project, moving beyond standard compliance. It specifies the technical and functional requirements to embed robust Data Provenance, Privacy-by-Design, and comprehensive System Security into the platform's architecture throughout the data lifecycle. D2.1 proposes a holistic framework that unifies provenance, privacy, and security as symbiotic elements, serving as a critical technical bridge between regulatory governance and engineering implementation. This deliverable provides a definitive reference for developers and architects, laying the groundwork for an AI-engineering platform that is resilient, auditable, and inherently trustworthy.

Document Revision History

Version	Date	Author	Revision
v0.1	13/10/2025	IDE	Initial content structuring
v0.2	04/11/2025	IDE	Prefinal version
v0.3	11/11/2025	AI4	Reviewed
v0.4	15/11/2025	IDE	Document updated
v0.5	27/11/2025	IDE	Review completed
V1.0	28/11/2025	IDE	Final version completed

Disclaimer and Copyright

The information appearing in this document has been prepared in good faith and represents the views of the authors. Every effort has been made to ensure that all statements and information contained herein are accurate; however, the authors accept no statutory, contractual or other legal liability for any error or omission to the fullest extent that liability can be limited in law.

This document reflects only the view of its authors. Neither the authors nor the Research Executive Agency nor the European Commission are responsible for any use that may be made of the information it contains. The use of the content provided is at the sole risk of the user. The reader is encouraged to investigate whether professional advice is necessary in all situations.

No part of this document may be copied, reproduced, disclosed, or distributed by any means whatsoever, including electronic without the express permission of the AI4SWEng project Partners. The same applies for translation, adaptation or transformation, arrangement or reproduction by any method or procedure whatsoever. © Copyright

AI4SWEng Innovation Mission

AI4SWEng : AI Engineering Suite to Support Agile, Efficient Software Engineering	*AI4SWEng comprises an excellent chemistry of 15 Partners including experts and practitioners in Model-Driven Software Engineering and AI Engineering (particularly LLMs) from across Europe with five Use-Cases from Software Development, Healthcare, Cyber-Physical Systems involving multiarchitectural and resource-constrained systems and Next Generation Battery Management in EVs. The AI4SWEng mission is to significantly contribute to the transformation of agile software development in a sector turbocharged for design-by-contact, high efficiency, reliability, and securityprivacy compliant delivery. This is to exploit advanced AI-powered software development support to deliver high-quality reliable fast time-to-market solutions that are socially responsible, scalable, sustainable and lend themselves to audible ethical and regulatory compliance. AI4SWEng aims to deliver AI-powered support to software developers to reduce worker stress through the pain-points in agile SW development pipelines and higher productivity thus enhanced job satisfaction and creativity. Accordingly, using five Use-Cases the project will develop, deliver and validate a platform comprising a set of tools including the AISysDev and Test Automation Tool to support the software lifecycle end-to-end for both functional and nonfunctional deliverables. A host of tasks will be supported such as code generation, cross-compilation, advanced debugging and dynamic slicing, security-privacy-by-design, model optimisation for reliability, energy efficiency as well as model adaptation and maintenance. This also includes software quality support through data-driven insights for informed design decision-making and strategic software development planning. Additionally, with a commitment to user-centred design, AI4SWEng will provide structured design in advanced prompt engineering to support software developer teams to best exploit the capabilities of the platform.*

Table of Contents

AI4SWEng Innovation Mission 3
Abbreviations 6
Executive Summary 6
1. Introduction 7
1.1. Purpose of the deliverable 7

1.2. Objectives 8
1.3. Methodology 8
1.4. Structure of the document 8
1.5. Regulatory and standards context 9
2. Data Provenance Requirements 9
2.1. Definition and relevance 9
2.2. Requirements 10
2.3. Alignment with standards and regulations 11
3. Privacy Requirements 12
3.1. Principles and context 12
3.2. Requirements 13
3.3. Support for data subject rights 14
3.4. Alignment with standards and governance 15
4. Security Requirements 16
4.1. Threat modelling and risk assessment 16
4.2. Requirements 17
4.3. Alignment with standards and regulations 18
5. Cross-Cutting Concerns 18
5.1. Integration of privacy and security by design 18
5.1.1. Design-time integration 19
5.1.2. Tool-level enforcement 19
5.2. Interplay between provenance, privacy, and security 19
5.3. Ethical and legal implications 20
5.4. Lifecycle considerations 20
5.5. Risk mitigation strategies 21
6. Requirements Traceability Framework 22
6.1. Traceability matrix 22
6.2. Justification and auditability 24
7. Validation and Maintenance of Requirements 25
7.1. Requirements elicitation and review process 25
7.2. Stakeholder involvement and validation steps 26
7.3. Change management and version control 26
7.4. Alignment with WP3 and WP4 technical work 27
8. Conclusion 29

Abbreviations

AI	Artificial Intelligence
API	Application Programming Interface
CIA	Confidentiality, Integrity, Availability
CI/CD	Continuous Integration / Continuous Deployment
D	Deliverable
DMP	Data Management Plan
EU	European Union
FAIR	Findable, Accessible, Interoperable, Reusable
FL	Federated Learning
IPR	Intellectual Property Rights
GDPR	General Data Protection Regulation
GUI	Graphical User Interface
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
KIO	Key Innovative Outputs
LLM	Large Language Model
MFA	Multi-Factor Authentication
NLP	Natural Language Processing
PII	Personally Identifiable Information
RAG	Retrieval-Augmented Generation
RBAC	Role-Based Access Control
SBOM	Software Bill of Materials
SDLC	Software Development Lifecycle
STRIDE	Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, Elevation of Privilege
TAT	Test Automation Tool
UC	Use Case
WP	Work Package

Executive Summary

Functioning as the foundational engineering reference for the AI4SWEng project, D2.1 moves beyond standard compliance guidelines. It provides the official methodology for AI-based software development by specifying the technical and functional requirements to embed robust data security, traceability, and privacy-by-design principles into the platform's architecture and toolset throughout the entire data lifecycle. It operationalises the governance principles established in preceding deliverables, such as D1.4 (Data Management Plan & Report) and D1.5 (Innovation and IPR Management Plan), by translating them into concrete, implementable system requirements. Consequently, the document presents itself not merely as a list of rules but more significantly, as a technical bridge between regulatory compliance and engineering implementation in subsequent development and integration work packages (WPs).

This document actively builds a foundation of trust, privacy, and security through three core components. First, it establishes a rigorous Data Provenance protocol to maintain verifiable lineages for all assets, including datasets, models, and artefacts, thereby promoting reproducibility. Second, it embeds Privacy at a fundamental level by operationalising data protection by design and by default, achieved through data minimisation, anonymisation, and the safeguarding of GDPR data subject rights. Building on this foundation, Security requirements are detailed to ensure the Confidentiality, Integrity, and Availability (CIA) of data artefacts through structured threat modelling.

The disjointed application of provenance, privacy, and security is a fundamental weakness that leaves systems vulnerable. D2.1 directly confronts this issue by proposing a holistic framework that treats them as a single, unified system. The deliverable demonstrates that their effectiveness is symbiotic: accountability is made possible through the verifiable trail of provenance; privacy governance becomes meaningful only with that traceable data flow; and a robust security infrastructure guarantees the integrity of both. This integrated approach is what ultimately builds a resilient foundation for responsible data management.

With this foundational view established, the focus now shifts to implementation. The deliverable provides a comprehensive traceability matrix and validation framework, ensuring each requirement is verifiable and linked to its regulatory or technical origin. Consequently, D2.1 delivers a single, definitive reference framework for developers and architects in subsequent work packages, laying the groundwork for an AI-engineering platform that is technically advanced, resilient, auditable, and inherently trustworthy by design.

The authors of this report have used generative tools for the preparation of this document, aligned with the "Living guidelines on the RESPONSIBLE USE OF GENERATIVE AI IN RESEARCH" of the EC. Specifically, the authors have used generative AI tools for formatting purposes, grammar and correctness verification over the original content created by the authors. In the same line, the authors have checked that the content of the document does not include any plagiarism and that all the cited sources are correct.

1. Introduction

1.1. Purpose of the deliverable

The purpose of Deliverable D2.1 is to define the technical and functional requirements ensuring that the AI4SWEng suite handles data securely, traceably, and in compliance with privacy-by-design principles throughout its lifecycle.

This document serves as the primary engineering reference for embedding provenance, privacy, and security mechanisms into the architecture and tools of AI4SWEng. It complements the organisational and legal frameworks established in D1.4 (Data Management Plan and Report) and D1.5 (Innovation and IPR Management plan) by translating governance principles into concrete, implementable system requirements.

Whereas D1.4 and D1.5 define what the consortium must comply with, D2.1 defines how this compliance is achieved technically, specifying the mechanisms, interfaces, and verification steps to be implemented across the platform and its data pipelines. The document thus acts as a technical bridge between

regulatory compliance and engineering implementation in WP3 (Development of AI-Powered Software Development and Testing Tools) and WP4 (Implementation of AI-Powered Software Development and Testing Automation Tools).

1.2. Objectives

The main objective of D2.1 is to provide a comprehensive set of verifiable requirements that translate regulatory and ethical principles into a traceable framework for embedding provenance, privacy, and security in the design and implementation of AI4SWEng tools and data pipelines. Specifically, the deliverable:

- Defines data provenance requirements, ensuring that all datasets, models, and artefacts can be traced to their source, transformations, and responsible parties through verifiable metadata.
- Specifies privacy requirements that operationalise privacy-by-design and privacy-by-default principles, focusing on minimisation, anonymisation, consent management, and controlled data usage.
- Details security requirements guaranteeing confidentiality, integrity, and availability of data and software artefacts within distributed development and execution environments.
- Ensures traceability of all requirements across work packages, linking each to its origin (e.g., regulation, use case, or stakeholder input) and affected system components.
- Translate regulatory and ethical principles (e.g., GDPR [1], AI Act [2], FAIR principles [3]) into measurable technical controls and software-level constraints.

By doing so, the document provides a single point of reference for developers, architects, and compliance officers to ensure that all components of the AI4SWEng platform adhere to and enforce the same standards for data reliability, privacy protection, and cybersecurity.

1.3. Methodology

Requirements were elicited using a multi-source, iterative approach combining:

- Analysis of relevant EU regulations and international standards, including the GDPR, the EU AI Act, ISO standards ([4], [5], [6], [7], [8], [9]), and the W3C PROV [10] model.
- Use case (UC) analysis, extracting requirements from operational contexts such as software development optimisation, cybersecurity monitoring, and healthcare reporting.
- Consolidation and prioritisation through feedback loops, ensuring that each requirement is testable and traceable to its source.
- Cross-WP coordination, ensuring coherence between requirement definitions in WP2 and implementation planning in WP3 and WP4.

1.4. Structure of the document

This document is structured as follows:

- Section 1 introduces the purpose, objectives, methodology, and regulatory context of the document.
- Section 2 defines data provenance requirements, focusing on mechanisms for traceability, logging, and lineage tracking.

- Section 3 presents privacy requirements, covering data minimisation, pseudonymisation, consent, and rights management.
- Section 4 outlines security requirements, describing threat models, attack surfaces, and defensive controls for confidentiality, integrity, and availability.
- Section 5 integrates cross-cutting concerns, explaining how privacy and security are embedded throughout the design phase.
- Section 6 provides a traceability matrix linking each requirement to its origin, function, and relevant work package.
- Section 7 outlines the validation and maintenance process for requirements across the project lifecycle.
- Section 8 concludes with a synthesis of key findings and outlines the next steps for WP2 and subsequent implementation work packages.

1.5. Regulatory and standards context

This deliverable interprets and operationalises applicable regulatory and technical standards for the project:

- GDPR and related EU data-protection obligations for lawful processing, consent, data-subject rights and retention.
- EU AI Act risk-based requirements for high-risk AI systems, where applicable; ethics and explainability expectations.
- The FAIR Guiding Principles (ensuring data is Findable, Accessible, Interoperable, and Reusable) and associated metadata standards referenced in the DMP (DataCite [11], Dublin Core [12]), ensuring data findability, accessibility, interoperability, and reusability (FAIR).
- Provenance standards such as W3C PROV for metadata schemas and lineage representation.
- Cryptographic and transport standards referenced in the DMP and requirements (AES-256 [13] for data-at-rest; TLS 1.2+ [14] for data-in-transit).
- Security best practices and benchmarks (OWASP Top 10 [15], secure CI/CD scanning, integrity checks, Role-Based Access Control (RBAC) and Multi-Factor Authentication (MFA).
- Key ISO/IEC standards for information security management (e.g., ISO/IEC 27001), cloud security controls (e.g., ISO/IEC 27018), and a privacy framework for processing personal data (e.g., ISO/IEC 27701).

The specific application of these frameworks and their translation into technical controls are detailed in the alignment sections (Sections 2.3, 3.4, and 4.3) of D2.1.

2. Data Provenance Requirements

2.1. Definition and relevance

Data provenance is defined as the complete, verifiable record pertaining to the origin, lineage, and transformations of data throughout its lifecycle. Within the AI4SWEng framework, the function of provenance is to ensure that all information, whether used, generated, or transformed by the software

development tools, can be traced from input to output. This traceability is facilitated in a transparent, tamper-evident, and auditable manner.

In the context of AI-based software engineering, provenance supports four main goals:

- Trust and accountability, which is achieved by enabling the verification of authenticity for inputs, processes, and outputs, thereby supporting confidence in AI-generated code and analytical results.
- Reproducibility and validation, which allow for the reconstruction of workflows, datasets, and decisions, ensuring that outcomes can be replicated and verified under equivalent conditions.
- Compliance and auditability, which is demonstrated by establishing conformity with GDPR, EU AI Act, and cybersecurity requirements through the recording of processing activities and responsible entities.
- Integrity and security, which involves the detection of unauthorised alterations or corruption affecting datasets, model versions, and logs, facilitated by cryptographic assurance and continuous monitoring.

Within AI4SWEng, provenance serves as the foundation for explainability, traceability, and responsible automation throughout the SDLC. The resulting records generated from user interactions are formally linked to system-level logs and project governance mechanisms. This linkage is defined to enable continuous validation, secure collaboration, and technical accountability across all relevant WPs.

2.2. Requirements

Table 1: Data provenance requirements

ID	Requirement	Description	Validation criterion
RQ-PROV-01	Ingest and persist provenance metadata	The system must automatically ingest and persist provenance metadata for every Large Language Model (LLM) interaction, including prompt, model version, timestamp, and data source references	Verify metadata persistence in the database after a test interaction
RQ-PROV-02	Expose provenance API/GUI	A provenance API and/or GUI must enable users to query and visualise full data and code lineage for any issue or generated artefact	Query API for a known issue ID; metadata must match expected lineage
RQ-PROV-03	Store logs in tamper-proof format	Provenance logs must be append-only and cryptographically protected (e.g., checksums or hashes). The system must validate log integrity on a scheduled basis and raise alerts on anomalies	Simulate log modification; the system must detect and alert
RQ-PROV-04	Interoperable provenance representation	Provenance data must follow the W3C PROV-DM schema to ensure interoperability with external systems and compliance with open data standards	Validate metadata format against W3C PROV schema

RQ-PROV-05	Versioning and lineage traceability	The system must maintain a full version history of datasets, models, and outputs, allowing rollback or reconstruction of prior states	Inspect the version chain and verify the ability to reproduce past states
RQ-PROV-06	Provenance validation on data exchange	Any external data ingestion must include validation of provenance information (origin, license, consent status). Non-compliant datasets must be flagged	Test ingestion of a dataset lacking provenance; must trigger a warning
RQ-PROV-07	Machineactionable export	The system must support export of provenance data in open, machinereadable formats (JSON-LD, RDF) for external auditing.	Export a sample record and validate format compatibility.
RQ-PROV-08	Automated data lineage for AI models	Training and inference pipelines must automatically generate lineage entries describing datasets, hyperparameters, and execution environments.	Inspect generated lineage metadata after model training.

These requirements establish systematic traceability across the AI4SWEng lifecycle, linking technical, organisational, and legal dimensions of accountability.

2.3. Alignment with standards and regulations

The provenance framework for AI4SWEng is designed to satisfy both European regulatory requirements and international technical standards governing transparency, accountability, and data integrity in AI systems. Provenance metadata will provide verifiable evidence of processing activities and a chain of custody for data and artefacts. This evidence supports GDPR accountability obligations by documenting what was processed, when, by whom, and for what purpose. It also provides the audit data needed for breach investigations and to demonstrate that appropriate technical and organisational measures were implemented.

For AI-specific governance, provenance records are intended to fulfil the traceability and documentation requirements of the EU AI Act by recording datasets, model versions, training and inference events, and decision paths for high-risk AI components. The provenance chain will also align with the EU Data Act's [16] emphasis on secure, controlled data sharing by preserving an immutable history of accesses and transfers, thereby protecting data sovereignty when datasets are exchanged between parties or platforms.

Technically, provenance metadata will follow established schemas to ensure semantic interoperability and machine-readability. W3C PROV will be used as the canonical model for representing entities, activities and agents involved in data creation and transformation. Provenance storage and handling will adopt information security controls consistent with ISO/IEC 27001 [4] to protect the integrity, availability, and confidentiality of logs. Privacy-related provenance use cases will be implemented in line with ISO/IEC 27701 [5] so that provenance contributes to documented processing records without exposing unnecessary personal data.

Operationally, compliance with regulatory EU requirements will be ensured through schema validation, integrity verification, and automated auditing. Every provenance entry will be checked against the selected schema during ingestion. Integrity controls, such as cryptographic checksums and appendonly storage, will be executed on a scheduled basis, with anomaly detection triggering immediate alerts. Additionally, cross-referencing provenance entries with access-control logs will allow verification of authorised actions, and retention or deletion of provenance data will follow the lifecycle rules defined in the project's DMP.

Together, these regulatory alignments, standards-based representations, and operational safeguards establish a provenance infrastructure that is interoperable, auditable, and legally defensible, providing a technical backbone for trustworthy and compliant AI-driven software engineering within AI4SWEng.

The provenance system aligns with the project's DMP by supporting data lifecycle phases with the corresponding trace records:

Table 2: Provenance link to data lifecycle

Lifecycle stage	Provenance support
Creation	Capture origin metadata (source, actor, timestamp)
Processing and analysis	Record transformations and the responsible component
Storage	Maintain append-only logs and checksums
Sharing	Register access and transfer events
Archiving/deletion	Retain lineage summary and validate integrity before erasure

3. Privacy Requirements

3.1. Principles and context

Within the AI4SWEng framework, the defined privacy requirements serve to guarantee that all processing of personal and sensitive data is compliant with the principles of data protection by design and by default, in accordance with the mandates of GDPR Articles 5, 25, and 32. A core objective is the integration of privacy safeguards directly into the system architecture, APIs, and workflows, in lieu of reliance on post-hoc compliance procedures.

Diverse categories of information are processed by the platform, encompassing user inputs, software artefacts, logs, and datasets generated during the model training, testing, and deployment phases. It is recognised that certain datasets may contain personal or sensitive elements (e.g., user identifiers within commit logs or project metadata). The mitigation of these risks necessitates that the AI4SWEng framework ensures lawful processing, applies data minimisation and purpose limitation principles, and implements the technical enforcement of policies pertaining to consent, access control, and retention.

The integration of privacy into the technical core of the AI4SWEng framework is defined to serve three primary objectives:

- Legal compliance, ensuring that all operations involving personal data adhere to EU dataprotection frameworks and national transpositions.
- Trust and transparency, reinforcing user and stakeholder confidence that AI-assisted development operates under clear and auditable safeguards.
- Ethical alignment, supporting human-centric AI principles under the EU AI Act and the Ethics Guidelines for Trustworthy AI, where privacy and data governance are essential preconditions for trustworthy system behaviour.

These principles ensure that privacy protection is not an external constraint but an **intrinsic property of the system design**.

3.2. Requirements

Table 3: Privacy requirements

ID	Requirement	Description	Validation criterion
RQ-PRIV-01	GDPR compliance by default	The system should comply by default with applicable EU data protection frameworks, including GDPR, the EU AI Act, and the EU Data Act, enforcing explicit consent, data minimisation, and purpose limitation	Trigger first-use consent dialogue; verify system halts without consent, and processes stop upon withdrawal
RQ-PRIV-02	Redact and anonymise sensitive data	The system should automatically detect, classify, and redact sensitive information (e.g., Personally Identifiable Information (PII), credentials) using predefined privacy policies	Submit inputs containing PII; verify detection, redaction, and anonymisation before processing
RQ-PRIV-03	Limit personal data collection	The system should collect, and store only personal data categories explicitly required for operation	Inspect API endpoints and database structures; confirm absence of unnecessary fields
RQ-PRIV-04	Purpose tagging and enforcement	The system should ensure that personal data is used solely for the purpose for which it was collected, enforcing purpose tagging and alerts for violations	Verify purpose metadata tagging and confirm mismatched uses trigger alerts
RQ-PRIV-05	Delete intermediate artefacts	The system should delete intermediate or temporary data automatically once processing is complete or after a defined retention period	Verify deletion of temporary artefacts after task completion or defined expiry
RQ-PRIV-06	Controlled data access and rolebased segregation	Personal or sensitive data shall be accessible only to authorised roles with explicit permissions. Access control lists and authentication tokens shall enforce the principle of least privilege	Attempt unauthorised access and verify denial, alert generation, and logging of the event

RQ-PRIV-07	Audit logging for privacy operations	All privacy-relevant actions, such as consent updates, data deletions, anonymisations, and access requests, shall be logged with timestamps, user IDs, and operation types. Logs shall be immutable and auditable.	Inspect privacy logs and validate integrity and completeness
RQ-PRIV-08	Explicit consent management	Users must grant, review, and withdraw consent through a clear interface. Consent status must be logged and enforce data-processing halts on withdrawal	Simulate consent withdrawal; confirm all related processing stops and audit log records event

These requirements embed privacy-by-design into each development and operational layer, ensuring that personal data protection is verifiable, measurable, and consistently applied throughout AI4SWEng.

3.3. Support for data subject rights

The AI4SWEng suite includes **system-level functionalities** that enable, track, and audit the exercise of data-subject rights defined by the GDPR. These features are implemented through explicit APIs, administrative tools, and auditable logs, ensuring transparency and technical enforceability of user control.

Table 4: Data subject rights support

Right	Technical implementation	Verification method
Access	Provide user-facing interfaces and APIs to retrieve all personal data and associated metadata stored within the system	Query API using a user ID; confirm full and accurate export of associated data
Rectification	Allow authorised modification or correction of inaccurate or outdated records through secure update endpoints	Attempt update operation; verify that only authorised users can modify data
Erasure	Implement secure deletion workflows that use cryptographic erasure to delete all data linked to the user. Maintain logs of deletion events	Request deletion; verify all linked records are removed and log entry recorded
Restrict processing	Introduce flag-based suspension mechanisms to prevent further data use without deletion	Apply restriction flag; confirm processing operations are blocked
Data portability	Allow data export in interoperable, machinereadable formats (e.g., JSON, CSV) with full metadata	Execute export; verify structure complies with metadata schema
Object	Implement preference management to disable automated processing of user data or require explicit confirmation	Simulate objection event; confirm system stops automated data use

These mechanisms ensure compliance with regulatory obligations while reinforcing user trust and control. Integrating these rights into system APIs and user interfaces enables AI4SWEng to demonstrate accountability and transparency across all privacy-related operations.

3.4. Alignment with standards and governance

The privacy framework in AI4SWEng is defined to ensure full alignment with European data-protection legislation and international standards pertaining to information security and privacy management. This framework operationalises the legal obligations of the GDPR and associated EU acts by embedding the principles of lawfulness, fairness, transparency, data minimisation, integrity, and confidentiality directly into system design and operational processes.

In accordance with the GDPR, data protection by design and by default (Article 25 [17]) is mandated for AI4SWENG, ensuring that all processing of personal or sensitive data is limited to what is necessary and is adequately protected. Provenance and audit logs are utilised to provide documented evidence of compliance with the accountability principle; they also support the lawful exercise of data-subject rights. Furthermore, these privacy controls contribute to compliance with the EU AI Act, which mandates transparency, human oversight, and traceability for AI systems processing personal data, and with the EU Data Act, which governs secure and controlled data sharing across organisations.

From a technical standpoint, the implementation of privacy safeguards adheres to recognised international standards. The ISO/IEC 27701 privacy information management framework is used to guide the definition of roles, responsibilities, and documentation for controllers and processors. Concurrently, the ISO/IEC 29100 [6] privacy-framework principles inform the design of lawful processing mechanisms and purpose limitation. Security and access-control mechanisms are based upon ISO/IEC 27001 and 27018 [7], which ensures the implementation of encryption, identity management, and controlled sharing of cloud-stored data. Additionally, ENISA recommendations on data protection [18], in conjunction with ISO guidance such as ISO/IEC 27002 [8] for security controls and ISO/IEC 27018 for cloud PII protection, serve as key references identifying and protecting personally identifiable information. These standards complement the cybersecurity measures already outlined in the DMP.

Operational enforcement is realised through a combination of automated and procedural safeguards. All data schemas and APIs are subject to periodic validation to ensure that personal-data fields remain limited to their intended purpose. Encryption mechanisms (AES-256 for data at rest and TLS 1.2+ for data in transit) are utilised to secure storage and communication channels. Privacy-impact assessments (PIAs) are conducted as part of the internal audit process to identify and mitigate emerging risks. Finally, retention, anonymisation, and consent workflows are synchronised with the lifecycle management rules established in D1.4, guaranteeing that expired data is securely deleted or irreversibly anonymised.

Through this integrated framework, privacy protection within AI4SWEng is ensured not as a passive compliance activity but as an active, auditable, and technically verifiable property of the system. This approach provides legal defensibility under EU dataprotection law, technical interoperability with global standards, and operational assurance that personal data within the AI4SWEng framework is processed lawfully, minimally, and transparently.

4. Security Requirements

4.1. Threat modelling and risk assessment

Within the AI4SWEng framework, security is established as a fundamental component, with its objective being the safeguarding of the confidentiality, integrity, and availability (CIA) for all project data and system components. The system is designed to manage critical assets, comprising AI models, training data, code artefacts, and logs. Consequently, a compromise pertaining to these assets could result in data breaches, model tampering, or service disruption, which would undermine the trust and reproducibility of the project outcomes.

A structured threat modelling approach, based on STRIDE (Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, Elevation of Privilege), is applied to identify vulnerabilities and define preventive measures. The primary assets requiring protection include:

- Source code repositories and generated artefacts.
- Model parameters, training datasets, and associated metadata.
- Provenance and audit logs for traceability and compliance.
- User credentials, access tokens, and configuration secrets.
- APIs and communication channels between system components.
- Deployment environments and pipelines.

Table 5: Potential threats and mitigations

Threat	Description	Impact	Mitigation
Unauthorised access	Exploitation of weak credentials or inadequate access control	Disclosure or modification of confidential data and artefacts	Strong authentication, RBAC, encryption, and secret rotation
Data leakage	Accidental or malicious disclosure of sensitive information via logs, APIs, or generated content	Privacy breaches, GDPR violations, reputational damage	Data minimisation, output redaction, encryption, and access-limiting middleware
Tampering/data manipulation	Unauthorised modification of provenance records, datasets, or model weights	Loss of integrity and reproducibility	Integrity checks, tamper-proof storage, signed logs
Malicious code injection	Vulnerabilities introduced in automated code generation or dependencies	System compromise or propagation of vulnerabilities	Secure coding, code analysis, and dependency audits
Denial of Service (DoS)	Overloading system resources to cause unavailability	Unavailability of services or degraded performance	Rate limiting, load balancing, anomaly detection
Data poisoning/model corruption	Insertion of malicious data to influence model behaviour	Model bias, performance degradation, and	Dataset validation, source authentication,

		untrustworthy AI results	and continuous model monitoring
Supply chain compromise	Infiltration through third-party libraries, containers, or CI/CD dependencies	Compromised builds or hidden backdoors	Signed dependencies, software bill of materials (SBOM), and integrity validation

This model provides the basis for risk-based security engineering across the AI4SWEng lifecycle, ensuring consistent mitigation strategies from design to deployment.

4.2. Requirements

Table 6: Security requirements

ID	Requirement	Description	Validation criterion
RQ-SEC-01	Strong authentication and RBAC	The system should implement strong authentication (e.g., OAuth2 [19], SAML [20]) and fine-grained, rolebased access controls to protect sensitive data and artefacts	Attempt unauthorised access and verify denial, logging, and alerting
RQ-SEC-02	Encryption of data at rest	All data must be encrypted at rest (AES-256)	Inspect stored data; confirm encryption standards compliance
RQ-SEC-03	Encryption of data in transit	All data must be encrypted in transit (TLS 1.2+)	Inspect network traffic; confirm encryption standards compliance
RQ-SEC-04	Vulnerability scanning in CI/CD	The system should automatically scan generated code and artefacts for known vulnerabilities (e.g., OWASP Top 10) before deployment	Run pipeline with intentionally vulnerable code; verify deployment fails and alerts trigger
RQ-SEC-05	Integrity checks on artefacts	The system should apply digital signatures or checksums to ensure that all artefacts and code are authentic and unaltered	Validate the checksum or signature of deployed artefacts against source records
RQ-SEC-06	Anomaly detection and alerts	The system should detect unusual access patterns or behaviours and automatically alert administrators	Simulate anomalous login attempts; verify that alerts are generated
RQ-SEC-07	Logging and monitoring	All user activity and system events must be logged in protected, immutable records	Simulate log access and verify immutability and completeness of entries
RQ-SEC-08	Secure software supply chain (SBOM)	Every build and deployment must generate a Software Bill of Materials listing all dependencies with versions and licenses. Only signed and verified packages are allowed in production	Verify CI/CD pipeline blocks unsigned artefacts and that SBOMs are automatically published with each release

RQ-SEC-09	LLM safety controls	The system shall implement technical and procedural safeguards to mitigate risks associated with LLMs, including prompt injection, data leakage, harmful content generation, and model misuse.	Inject crafted prompts containing malicious payloads and verify they are blocked; test for data exfiltration via indirect prompts and ensure
			failure and alert generation

These controls collectively ensure that confidentiality, integrity, and availability are protected across the AI4SWEng architecture and that all system components can be audited and restored in the event of an incident.

4.3. Alignment with standards and regulations

The AI4SWEng security framework aligns with EU regulations and international standards to ensure the confidentiality, integrity, and availability of all project assets. It fulfils GDPR Article 25 obligations for data protection by design and supports EU AI Act requirements for system robustness and resilience.

Security governance and control implementation follow **ISO/IEC 27001** for **information security management systems, while ISO/IEC 27002** provides detailed **control objectives for access management, encryption, logging, and operational security**. Where cloud or hosted environments are involved, **ISO/IEC 27017** [9], for **security for cloud services**, and **ISO/IEC 27018**, for **protection of personal data in cloud computing**, can provide additional guidance.

Implementation relies on a combination of automated and procedural measures. Continuous vulnerability scanning and penetration testing will identify configuration or dependency weaknesses. Logs of security events will be maintained as immutable audit records, synchronised with provenance data to support forensic investigations. Encryption (AES-256 for data at rest, TLS 1.2 or higher for data in transit) and role-based access control form the baseline for protecting all communications and storage layers. Regular internal audits will verify conformance with ISO and ENISA guidance, ensuring that nonconformities are promptly corrected.

Through this integrated framework, AI4SWEng ensures its security posture is standards-compliant, riskbased, and verifiable, providing demonstrable assurance that all platform components maintain confidentiality, integrity, and availability in line with European cybersecurity expectations.

5. Cross-Cutting Concerns

5.1. Integration of privacy and security by design

The AI4SWEng suite is based on the key idea of privacy and security by design. This approach applies throughout the whole software development process. Privacy and security are not considered afterthoughts or simple compliance checks. Instead, they are integrated from the earliest stages of system design, requirements gathering, and tool development.

5.1.1. Design-time integration

The implementation of Privacy and Security by Design (GDPR Article 25) is handled systematically within the platform. This process begins by applying strict data minimisation and purpose limitation to all data handling. Non-personal and personal data are processed independently to maintain data integrity. Moreover, pseudonymisation methods (e.g., hashing, tokenisation, encryption) are used to protect user identities. An integrated consent management system that oversees the entire consent lifecycle is the foundation for the framework's compliance with GDPR Article 7 [21].

Security is addressed from the outset through threat modelling, which identifies attack vectors, exposed surfaces, and critical assets. This allows the implementation of a proactive security posture founded on multiple layers of defence, including the following key controls:

- Data and access security, such as robust encryption for both data at rest/in transit, strong authentication, and RBAC (e.g., OAuth2, SAML).
- Interface and API security: A multi-pronged approach is implemented to harden application interfaces and APIs, following key measures such as: ○ Enforce strong endpoint security policies. ○ Apply strict input validation on all data. ○ Conduct continuous vulnerability scanning.

5.1.2. Tool-level enforcement

The platform's operational tools exhibit the same dedication to privacy and security. The AI-SysDev Tool enhances security by enforcing Role-Based Access Control (RBAC) and data protection mechanisms, while the Test Automation Tool (TAT) includes capabilities for vulnerability scanning and fuzz testing in the development process. These tools, designed for environments with limited resources, can successfully employ privacy-protecting methods such as Federated Learning (FL) and local processing.

5.2. Interplay between provenance, privacy, and security

In complex AI-driven systems, trust and accountability are closely connected. They stem from the relationship among privacy, security, and provenance. These three elements work together to improve the system's ability to meet legal and ethical standards. They are independent: their linkage is what enables system-wide transparency and accountability.

Provenance is a key part of traceability. It allows the system to track where the data comes from, the changes it goes through, and who accesses it. This verifiable record is essential for confirming the proper use of personal data and ensuring it is used only for its intended purposes. The strength of privacy protections relies on this traceability. It helps enforce consent boundaries, safeguards the rights of data subjects, and lowers the chances of misuse or exposure of sensitive information.

In addition, security is significantly contingent upon provenance. It is much easier to spot unusual activities, verify the reliability of software components, and reduce potential threats when every action in the system is carefully recorded and linked to its source. Records of origin help piece together events and find weaknesses. This helps with investigations if a breach happens. At the same time, security measures such as encryption and access controls are necessary to ensure the reliability of provenance data. These measures ensure that traceability stays trustworthy and cannot be changed.

There is a strategic as well as technical connection among these three domains. The system protects against intrusions through security measures. Privacy ensures responsible data management, while

provenance offers visibility. Together, they form a strong basis for reliable AI-based software. In this system, each interaction is traceable, every risk is identifiable, and all decisions can be verified.

5.3. Ethical and legal implications

A primary goal of AI4SWEng is to ensure that creating AI-driven software is not only technically sound and quick, but also in line with the ethical and legal rules for responsible tech advances in Europe. The project turns general ideas such as fairness and honesty into real-tech needs. These needs can be put in place, checked, and reviewed throughout the software's life.

Legally, the platform must comply with rules such as the GDPR, the AI Act, and the FAIR principles. These set the rules for lawful data use, getting permission, and protecting fundamental rights, mainly in highrisk AI systems. AI4SWEng puts these duties into practice through several mechanisms (e.g., purpose tagging, consent management, and support for data subject rights). These are not just statements; they are part of the system's setup. They are shown in the platform's APIs, data structures, and access controls. The legal needs align with specific tech components, so the rules are not just stated; they are visible.

Ethical points are also key to how the platform is designed. Using LLMs, automatic code generation, and smart testing introduces risks, such as bias and unexpected outcomes. AI4SWEng addresses these risks by introducing mechanisms to track structured data and implement human controls. These let developers and users watch and step in at key points during AI-driven processes. These protections are critical to the project's UCs, so the platform is designed to anticipate and mitigate these risks.

The project acknowledges the evolving nature of ethical and legal frameworks. As regulations change and societal expectations shift, the platform is designed to be flexible. To help with this, the project has tracking and verification methods that keep it up to date with new standards. This forward-thinking plan ensures the system can evolve without losing its core promise of responsible design.

In conclusion, the ethical and legal effects of AI4SWEng are not just considered from the outside or checked at the end; they are part of the platform's core and manifest in every choice made. Ethical and legal ideas become system requirements, ensuring the platform supports not just quick, smooth creation but also trustworthy, socially beneficial, and legally compliant technology.

5.4. Lifecycle considerations

The AI4SWEng platform ensures that privacy, security, data origin, and ethical principles are incorporated into the entire software development process. Each phase includes mechanisms to uphold these principles, supported by verifiable controls.

- Requirements' definition: LLMs facilitate the transformation of natural language into machinereadable specifications. By employing systematic prompting techniques, pertinent information is discerned in accordance with ethical and legal standards, thereby allowing for traceability and verification from the initial stages.
- Architectural planning: Privacy and security are incorporated during the design phase through the utilisation of modular design frameworks. These frameworks mandate data minimisation, anonymisation, and purpose restriction. Each planning decision is documented with provenance metadata to facilitate subsequent audits and compliance verifications.
- Implementation and development: AI supports code generation, issue detection, and performance optimisation via various methods (e.g., reverse execution and predictive fault

assessment). Every created item is version-controlled and tied to its creation, ensuring everything is traceable and repeatable.

- Testing and validation: Automated testing uses methods such as fuzz testing and LLM-generated test cases to verify system functionality, ensuring reliability and dependability.
- Deployment and operation: To protect privacy, TinyML structures and FL approaches are used on low-resource devices. The system adjusts to its surroundings while still meeting performance, security, and legal needs.
- Maintenance and evolution: Ongoing surveillance, model deviation identification, and retraining protocols ensure the system remains precise, secure, and up to date. Change documentation and version control logs are preserved to facilitate transparency and enhance maintainability.
- Decommissioning and data disposal: All temporary and intermediate data are securely deleted, and retention policies are strictly enforced. Moreover, every deletion is logged to ensure full compliance with regulations such as GDPR.
- Governance and oversight: Technical safeguards, structured metadata, and human oversight are implemented across all phases to put ethics and legal principles into practice. The platform is created to support compliance from the start, helping with both internal controls and external examinations.

Instead of treating software development stages as separate, AI4SWEng combines them into a single framework that considers compliance and verifies that the software is technically sound, wellfunctioning, socially responsible, adaptable, and ready for what lies ahead.

5.5. Risk mitigation strategies

AI4SWEng takes a layered, proactive approach to risk mitigation, ensuring that privacy, security, and compliance risks are addressed throughout the software development lifecycle. These strategies, built into the architecture, tools, and workflows, can keep the system resilient, trustworthy, and compliant with regulations.

The methods used to deal with risks across the platform are explained below:

- Threat modelling and early risk identification: Identification of potential vulnerabilities through structured threat modelling during the code design phase.
- Privacy and security by design: As explained in Section 5.1, the platform includes both privacy- and security-by-design principles applied at both system and tool levels to ensure lawful data processing.
- Access control and authentication: Use of different authentication techniques (e.g., MFA), and secure identity protocols such as OAuth2 and SAML to prevent unauthorised access to the system and data.
- Encryption and data protection: To provide a solid foundation for security, all data must be encrypted, whether at rest (using AES-256) or in transit (utilising TLS 1.2+). This encryption should be rigorously managed through consistent practices, including the secure storage and handling of encryption keys, and regular verification of data integrity to prevent unauthorised modification.
- Secure development and CI/CD pipelines: Prevention of insecure deployments by integrating continuous vulnerability scanning and secure configuration management.
- Monitoring and anomaly detection: Real-time monitoring tools are deployed to detect unusual behaviour, and automated responses are configured to support rapid incident handling.

- Provenance and auditability: To support post-incident investigations, system interactions (e.g., data transformations and model outputs) are saved as metadata.
- Regulatory compliance controls: Mechanisms such as consent management are in place to enforce GDPR and the European AI Act obligations, and they are continuously evolving.
- Human oversight and ethical review: Human-in-the-loop mechanisms are integrated, particularly in high-risk AI contexts, to support oversight of critical decisions. Moreover, ethical review checkpoints are embedded into development workflows to ensure responsible innovation.

These strategies are designed not only to prevent and respond to risk, but also to create a growth environment where compliance, transparency, and accountability are integral to every technology solution. By embedding these security measures at the core of the platform, AI4SWEng fosters a culture of responsible design, ready to address both current and emerging issues.

6. Requirements Traceability Framework

6.1. Traceability matrix

Table 7 shows the Traceability Matrix, mapping each system requirement to its corresponding source, objective, and WP. The following attributes define the table's composition:

- *Source*: Specifies the origin of the requirement, such as regulatory frameworks (for instance, GDPR), technical standards (e.g., W3C PROV and OWASP), or stakeholder-driven needs.
- *Objective (Risk/Control)*: Outlines the requirement's primary function, indicating if it mitigates a risk or enforces a control.
- *Affected WP*: Designates the WP responsible for implementation, ensuring requirement traceability from design through to composition.

Additionally, Table 7 is utilised as a foundational tool, providing a mechanism for the alignment of system design with compliance obligations, stakeholder expectations, and project accountability. This tool enables structured verification and ensures that privacy, security, and provenance requirements are embedded by design within the AI4SWENG suite.

Table 7: Traceability matrix

Requirement	Source	Objective (Risk/Control)	Affected WP
Ingest and persist provenance metadata	W3C PROV, GDPR	Ensure data integrity and traceability	WP2
Expose provenance API/GUI	Stakeholders need	Enable transparency and auditability	WP3
Store logs in tamper-proof format	GDPR, Security standards	Preventing unauthorised modifications	WP4
Interoperable provenance representation	W3C PROV	Ensure interoperability	WP2
Versioning and lineage traceability	FAIR principles	Guarantee reproducibility and rollback of artefacts	WP3

Provenance validation on data exchange	GDPR, W3C PROV	Control the integrity and legitimacy of imported datasets	WP4
Machineactionable export	W3C PROV, EU Open Data Directive	Facilitate external auditing and reuse	WP4
Automated data lineage for AI models	AI Act	Ensure model accountability and explainability	WP3
Explicit Consent Management	GDPR	Respect for user rights and data control	WP3
GDPR compliance by default	GDPR	Legal compliance and privacy protection	WP3
Redact and anonymise sensitive data	GDPR	Prevent exposure of personal data	WP3
Limit personal data collection	GDPR	Data minimisation	WP2
Purpose tagging and enforcement	GDPR	Preventing misuse of personal data	WP4
Delete intermediate artefacts	GDPR	Limit data retention	WP4
Controlled data access and rolebased segregation	GDPR, Security standards	Enforce least-privilege access and separation of duties	WP3
Audit logging for privacy operations	GDPR, Security standards	Maintain accountability and auditability of privacy events	WP4
Strong authentication and RBAC	OAuth2, SAML	Prevent unauthorised access	WP3
Encryption of data at rest	AES-256	Protect data confidentiality	WP4
Encryption of data in transit	TLS 1.2+	Protect data confidentiality	WP4
Vulnerability scanning in CI/CD	OWASP Top 10	Preventing insecure deployments	WP3
Integrity checks on artefacts	GDPR, Security standards	Detect anomalies and ensure trust and authenticity	WP4
Anomaly detection and alerts	Stakeholders need	Detect and respond to threats	WP4
Logging and monitoring	GDPR, Security standards	Enable audit trails	WP4
Secure software supply chain (SBOM)	OWASP, Security standards	Prevent supply-chain compromise and dependency tampering	WP3

LLM safety controls	AI Act, GDPR	Ensure safe, privacypreserving, and robust operation of LLM components	WP3

6.2. Justification and auditability

The integrity, transparency, and regulatory alignment of the AI4SWEng suite are supported by the Traceability matrix presented in Table 7. By explicitly tracing each requirement to its source – whether technical standards, legal statutes, or operational needs articulated by stakeholders – the matrix validates that the system's development adheres to its foundational mandates and accountabilities.

The structured mapping provides a framework wherein meticulously defined objectives govern the development process, each corresponding to a distinct risk or compliance mandate. The Traceability matrix serves the dual function of elucidating strategic goals and operationalising them. The matrix enhances accountability by assigning each requirement to a specific WP. This systematic allocation is the mechanism that permits a more focused and rigorous approach across the implementation, testing, and verification phases, ensuring that objectives are effectively addressed.

In terms of auditability, the table provides a clear link between system functionality and the requirements it is intended to satisfy. This traceability is mandatory for both internal and external audits, regulatory reviews, and certification processes, thereby allowing auditors to systematically track decisions, verify the efficacy of protective measures, and ascertain the system's adherence to statutory and ethical obligations.

The framework serves as an operational, dynamic reference integral to impact analysis, change management, and ongoing development. It not only furnishes stakeholders with the clarity needed for confident interaction with the platform's architecture but also allows the project to evolve without deviating from its core compliance commitments.

7. Validation and Maintenance of Requirements

7.1. Requirements elicitation and review process

The elicitation and review of requirements in AI4SWEng are conceived as a continuous, structured, and multi-source process that ensures alignment among the platform's technical development, regulatory obligations, stakeholder expectations, and the operational needs of the four use cases. This process is not limited to the project's initial phases. Still, it is designed to evolve iteratively throughout the platform's lifecycle, enabling the progressive refinement and validation of requirements in response to emerging insights, technological constraints, and contextual changes.

To establish a solid foundation for AI4SWEng, requirements are defined using a dual approach. On the one hand, an exhaustive review of key regulatory frameworks, such as the GDPR, the EU AI Act, and ISO standards, is conducted to identify legal and technical obligations. On the other hand, real use cases are analysed in detail across several sectors (software development, healthcare, etc.), allowing the extraction of functional requirements.

The integrity and relevance of the requirements were ensured through a rigorous process. First, iterative feedback cycles were implemented with key stakeholders (technical partners, use case experts, endusers, and compliance officers) to consolidate and prioritise. Subsequently, each requirement underwent an analysis of feasibility, testability and traceability to its origin. Final formalisation was carried out through structured documentation that specifies, for each requirement, a unique identifier, its functional description, and explicit validation criteria.

The requirements review process is fully integrated into the project governance structure and synchronised with its key milestones. Formal checkpoints have been scheduled in alignment with project milestones to reassess the validity and relevance of the requirements. This reassessment is based on an analysis of technical progress, stakeholder feedback, and developments in the applicable regulatory framework. To ensure consistency between the specification and implementation, reviews are carried out collaboratively by WP2, WP3 and WP4.

The requirements management methodology ensures their viability and testability through two complementary processes:

- Validation mechanisms: Each requirement is assigned to a specific set of validation mechanisms to ensure that it is tangible and verifiable. These mechanisms include, among others, simulation scenarios, prototype testing, and regulatory compliance audits.
- Prioritisation criteria: At the same time, requirements are prioritised following a multi-criteria assessment based on three principal axes:
    - Criticality: Its importance to the core functionality of the system. ○ Complexity: The estimated technical effort required for its implementation. ○ Strategic impact: Its contribution to cross-cutting objectives such as security, privacy, and explainability.

The methodology employed establishes a rigorous foundation for the design, implementation, and continuous improvement of an AI-based software engineering platform that is reliable, auditable, and compliant with current regulations. This robustness stems from a requirements framework designed to be inherently dynamic, capable of continuously integrating new knowledge, stakeholder feedback, and updates to the regulatory environment.

7.2. Stakeholder involvement and validation steps

To ensure that requirements remain consistently aligned with project objectives, validation in AI4SWEng takes a co-creation approach. This approach is not limited to an initial consultation; it actively involves stakeholders in the ongoing evaluation and refinement of requirements throughout their lifecycle, ensuring they align with the changing needs of use cases (UCs) and the technical roadmap.

The project's stakeholder ecosystem comprises technical partners (WP3 and WP4), UC experts, internal or external legal and compliance officers, and the end-users and their representatives. The participation of these actors is channelled through a structured framework combining bilateral consultations, thematic workshops, and inter-WP coordination meetings. This multi-stakeholder interaction approach is designed to ensure comprehensive information gathering, covering both high-level expectations and specific operational constraints, with a particular focus on privacy, security, and explainability.

To guarantee the effective integration of stakeholder input, AI4SWEng employs a traceable feedback mechanism. Each requirement is linked to its source – whether a regulatory obligation, a UC, or a stakeholder contribution – and systematically evaluated based on its technical feasibility, implementation cost, and strategic relevance. Such traceability is not only essential for auditability but also provides an objective framework for prioritisation and conflict resolution when the requirements of different actors or domains conflict.

The validation process phases are synchronised with the project milestones and integrated into the WP2 governance structure. At each scheduled checkpoint, those responsible for WP2, WP3 and WP4, with the participation of relevant stakeholders, jointly review the consolidated requirements. These reviews focus on verifying the consistency of the requirements with the system architecture, their testability, and their alignment with the project's ethical and legal commitments. If the requirements need to evolve due to technical constraints or regulatory changes, stakeholders are consulted again to confirm the validity of the proposed changes.

This structured and iterative validation process ensures that the requirements framework remains robust, context-appropriate and responsive to stakeholder expectations. It also reinforces the project's commitment to transparency, accountability, and co-creation, which are essential to the successful deployment of AI-based software engineering tools in sensitive and high-impact domains.

7.3. Change management and version control

A DevOps-oriented methodology is employed within the AI4SWEng framework for the management of changes and the control of versions. This methodology is defined to guarantee that all modifications pertaining to requirements and system components are executed in a traceable, verifiable, and technically sound fashion. In lieu of conventional document-centric change request processes, the initiative integrates change tracking seamlessly within the software development lifecycle. This integration is facilitated via the defined modular architecture, established continuous integration (CI) workflows, and infrastructure-as-code (IaC) principles.

Modifications to specifications, arising from stakeholder input, technical limitations, or compliance revisions, are managed through version-controlled repositories. These repositories are configured to facilitate collaborative editing, branching, and rollback functionalities. Furthermore, these repositories are interconnected with automated testing and deployment workflows, thereby enabling the prompt identification of discrepancies and regressions. Each specification maintains a formal association with

its relevant implementation artefacts (e.g., architectural designs, code components, testing scenarios), ensuring comprehensive traceability throughout the development lifecycle.

The modular framework of Key Innovative Output (KIO) 1, which is comprised of containerised components and microservices, supports independent updates and controlled dissemination or modification. This architecture is defined to enable the progressive enhancement of specific functionalities (e.g., code generation, automated testing, and reverse execution) without compromising overall system integrity. Moreover, the incorporation of Retrieval-Augmented Generation (RAG) and optimised LLMs into KIO7 facilitates the agile adjustment of requirements and code artefacts in response to changing data and contextual conditions, while concurrently preserving a transparent history of alterations.

Verification of modifications is integrated directly into the DevOps pipeline through automated unit, integration, and compliance assessments. These automated processes are enhanced by milestoneoriented evaluations involving WP2, WP3, and WP4. This dual approach is designed to ensure that enhancements are not only technically robust but also remain consistent with stakeholder requirements and regulatory guidelines. Surveillance and logging systems provide real-time insights into system performance, enabling proactive anomaly detection and remediation as required.

Change management within the AI4SWEng framework is not approached as an isolated administrative function. Rather, it constitutes a fundamental component of the agile, automated, and traceable development ecosystem. This defined approach ensures the platform can adapt to emerging insights and evolving requirements, while simultaneously maintaining the integrity, dependability, and auditability mandated for a high-assurance AI engineering framework.

7.4. Alignment with WP3 and WP4 technical work.

The requirements engineering and validation processes in AI4SWEng are closely interconnected with the technological advancements executed in WP3 (Development of AI-Powered Software Development and Testing Tools) and WP4 (Implementation of AI-powered Software Development and Testing Automation Tools). This synchronisation guarantees that the advancing requirements are not only technically viable but also shape and are shaped by the execution and assessment of the project's fundamental innovations.

WP3 focuses on advancing the AI-SysDev Tool and its associated components, including requirements elicitation, architectural design, code synthesis, and debugging. The requirements gathered and enhanced in WP2 serve as a foundation for the design and execution of these tools, primarily through the NLP-to-Formal Requirements (KIO3) and Formal Requirements-to-Code Architectural Planning (KIO4) modules. These tools undergo validation against the established requirements to confirm their ability to facilitate traceable, explainable, and auditable software development methodologies.

WP4 is tasked with creating the AI-powered Test Automation Tool (KIO11) and incorporating validation protocols throughout the software development lifecycle. Criteria for testability, compliance, and performance define the framework's scope and functionality. Insights derived from WP4's validation processes are communicated back to WP2 to enhance the requirements and maintain ongoing coherence with the system's operational dynamics.

The modular framework of the AI Engineering Suite (KIO1) enhances this alignment by enabling iterative development and validation. Each KIO is tailored to meet specifications and is assessed through

application-specific scenarios in WP5 and WP6. This approach guarantees that the requirements are both theoretically robust and practically validated within real-world applications.

Regular coordination meetings between these three WPs, as well as joint milestones reviews, ensure that requirements are continuously updated to reflect technical progress, stakeholder feedback, and regulatory changes. This integrated approach guarantees that the requirements framework remains a living artefact, evolving in tandem with the project's technical trajectory.

8. Conclusion

D2.1 establishes the fundamental protocol for integrating compliance, trust, and resilience within the AI4SWEng suite. Utilising a systematic, multi-faceted approach, the document converts legal, ethical, and operational principles into a unified technical specification framework that guides the development and execution of AI-powered software engineering tools.

The requirements outlined in this deliverable encompass the entire range of system considerations. Provenance requirements shall be implemented to render all data, models, and artefacts traceable, auditable, and tamper-resistant, which supports the objectives of transparency and reproducibility. Concurrently, privacy requirements are defined to uphold GDPR principles, such as data minimisation, consent management, and the facilitation of data subject rights, and to integrate privacy-by-design into the platform architecture. Furthermore, security specifications shall be established to address threat modelling, access control, encryption, vulnerability assessment, and anomaly detection, thereby guaranteeing the confidentiality, integrity, and availability of system components.

The requirements for provenance, privacy, and security shall not be implemented as standalone protocols; rather, they constitute a set of interconnected systems designed to be mutually reinforcing. Provenance, within this framework, facilitates privacy enforcement by providing a structured methodology for tracking data utilisation and imposing defined consent limitations. Concurrently, security protocols are mandated to safeguard the authenticity of provenance documentation and to guarantee that all established privacy protections remain intact and uncompromised. Collectively, these components establish a "trust triad" that provides the necessary foundation for the platform's ethical and legal defensibility.

A fundamental contribution of D2.1 lies in incorporating these specifications across all stages of the software development lifecycle, encompassing requirements gathering and architectural design, as well as deployment, ongoing maintenance, and eventual decommissioning. The platform is architected to ensure adherence and accountability via organised metadata, automated verification, and human supervision. This lifecycle framework ensures that reliability is not merely a fixed characteristic, but an evolving quality sustained throughout the system's progression.

Additionally, the deliverable presents a Traceability Matrix. This matrix provides a formal mapping, associating each defined requirement with its specific origin (e.g., regulatory, technical, or stakeholderderived), its corresponding risk/control objective, and the accountable Work Package (WP). The function of this matrix is to facilitate systematic verification, support both internal and external auditing processes, and ensure all platform components maintain alignment with the primary objectives of dependability, interpretability, and compliance with established legal standards.

The requirements outlined in D2.1 constitute the technical foundation for the development and integration tasks subsequently planned in WP3 and WP4. These specifications are utilised within WP3 to guide the architectural design of the AI-SysDev Tool, which encompasses modules for NLP-to-formal requirement translation, system architecture planning, and code generation. Concurrently, WP4 employs these requirements to establish validation protocols, automated testing strategies, and performance metrics for the Test Automation Tool. The modular framework of the AI Engineering Suite (KIO1) ensures that each component is developed and validated in accordance with these defined requirements, thereby facilitating iterative enhancements and ensuring practical applicability

In summary, D2.1 delivers a robust, implementable, and sustainable methodology for AI-driven software engineering. This methodology ensures that the AI4SWEng suite is developed in adherence not only to advanced technological specifications but also to principles of ethical responsibility, legal compliance, and social accountability. The foundational integration of compliance and reliability within the system architecture establishes a reference standard for accountable AI engineering, defined by its provisions for transparency, auditability, and future preparedness.

References

[1] "Regulation (EU) 2016/679 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)," May 2016. [Online]. Available: https://eurlex.europa.eu/eli/reg/2016/679/oj/eng.

[2] "Regulation (EU) 2024/1689 on artificial intelligence (AI Act)," July 2024. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.

[3] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al., "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific Data,* vol. 3, no. 1, p. 160018, 2016.

[4] International Organization for Standardization, "Information security, cybersecurity and privacy protection — Information security management systems — Requirements (ISO/IEC 27001:2022)," 2022. [Online]. Available: https://www.iso.org/es/norma/27001.

[5] International Organization for Standardization, "Information security, cybersecurity and privacy protection — Privacy information management systems — Requirements and guidance (ISO/IEC 27701:2025)," 2025. [Online]. Available: https://www.iso.org/es/norma/27701.

[6] International Organization for Standardization, "Information technology — Security techniques — Privacy framework (ISO/IEC 29100:2024)," 2024. [Online]. Available: https://www.iso.org/es/contents/data/standard/08/59/85938.html.

[7] International Organization for Standardization, "Information security, cybersecurity and privacy protection — Guidelines for protection of personally identifiable information (PII) in public clouds acting as PII processors (ISO/IEC 27018:2025)," 2025. [Online]. Available: https://www.iso.org/es/norma/27018.

[8] International Organization for Standardization, "Information security, cybersecurity and privacy protection — Information security controls (ISO/IEC 27002:2022)," 2022. [Online]. Available: https://www.iso.org/es/contents/data/standard/07/56/75652.html.

[9] International Organization for Standardization, "Information technology — Security techniques — Code of practice for information security controls based on ISO/IEC 27002 for cloud services (ISO/IEC 27017:2015)," 2015. [Online]. Available: https://www.iso.org/standard/43757.html.

[10] P. Missier, K. Belhajjame and J. Cheney, "The W3C PROV Family of Specifications for Modelling Provenance Metadata," in *Proceedings of the 16th International Conference on Extending Database Technology*, ACM, 2013, pp. 773-776.

[11] DataCite, "DataCite Metadata Schema," [Online]. Available: https://schema.datacite.org/.

[12] International Organization for Standardization, "Information and documentation. The Dublin Core Metadata Element Set (ISO 15836-1:2017).," 2017. [Online]. Available: https://www.iso.org/standard/71339.html.

[13] National Institute of Standards and Technology, “Advanced Encryption Standard (AES),” 2001.

[14] T. Dierks and E. Rescorla, “The Transport Layer Security (TLS) Protocol Version 1.2 (RFC 5246),” 2008.

[15] Open Worldwide Application Security Project, “OWASP Top 10,” 2021. [Online]. Available: https://owasp.org/Top10/.

[16] “Regulation (EU) 2023/2854 on harmonised rules on fair access to and use of data (Data Act),” 2023. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2023/2854/oj/eng.

[17] L. A. Bygrave, “Article 25 data protection by design and by default,” in *The EU General Data Protection Regulation (GDPR)*, 2020.

[18] European Union Agency for Cybersecurity, “Data Protection Engineering,” 2022.

[19] D. Hardt, “The OAuth 2.0 Authorization Framework (RFC 6749),” 2012.

[20] B. Campbell, C. Mortimore and M. B. Jones, “Security Assertion Markup Language (SAML) 2.0 Profile for OAuth 2.0 Client Authentication and Authorization Grants (RFC 7522),” 2015.

[21] *Article 7 GDPR. Conditions for consent*.