Artificial Intelligence in Manufacturing leading to Sustainability and Industry5.0

Project Acronym:

AIMS5.0

Grant agreement no: 101112089

Deliverable no. and title	**D7.3 – Recommendations based on the validation of the generated consolidated guidelines regarding change management to Digital Workspaces**	
Work package	WP7	Human acceptance, trust and ethics for Digital Workspaces
Task	T7.4	Development of guidelines and input for regulations for Digital Workspaces
Subtasks involved	Not applicable	
Lead contractor	Infineon Technologies AG Thomas Gutt: thomas.gutt@infineon.com	
Deliverable responsible	AIT – Austrian Institute of Technology GmbH Christoph Schmittner, mailto: christoph.schmittner@ait.ac.at Erwin Schoitsch, mailto: erwin.schoitsch@ait.ac.at Abdelkader Shaaban, mailto: abdelkader.shaaban@ait.ac.at	
Version number	V2.0	
Date	14/05/2026	
Status	Final Version	
Dissemination level	Public	

Copyright: AIMS5.0 Project Consortium, 2026

Authors

Participant no.	Part. short name	Author name	Chapter(s)
14	AIT	Christoph Schmittner	1 & 2
14	AIT	Erwin Schoitsch	1 & 2
14	AIT	Abdelkader Shaaban	1, 2, & 3
40	PCL	Olga Kattan	3 & 6
19	RSA	Markus Tauber	3 & 5
19	RSA	Yasin Ghafourian	3 & 5
9	HSZG	Pascal Herfort	4
9	HSZG	Fabian Lindner	4
10	IFD	Germar Schneider	3 & 6
22	WUW	Gerald Reiner	4
22	WUW	Venugopal Ramadasu	4
32	HUA	Konstantina Karathanasopoulou	4
34	BME	Pal Varga	3 & 5
34	BME	Anikó Grad-gyenge	5
34	BME	Edit Tomasovszky	5
43	SIGNIFY	Marc de Samber	4
21	ViF	Martin Wifling	4

Document History

Version	Date	Author name	Reason
v0.1	08.10.2025	Abdelkader Shaaban	Initial Template
V1.1	10.04.2026	Abdelkader Shaaban	Consolidated and completed version
V1.2	06.04.2026	Olga Kattan	Additional points integrated
V1.3	07.04.2026	Abdelkader Shaaban	Version submitted for review
V1.4	20.04.2026	Abdelkader Shaaban	Updated version
V1.5	27.04.2026	Yasin Ghafourian	Executive summary, introduction, and Section 5
V1.6	4.05.2026	Abdelkader Shaaban	Consolidated and updated Version
V1.7	07.05.2026	Erwin Schoitsch	Standard-related information updates
V1.8	08.05.2026	Martin Wiling	Update section 4.6
V1.9	07.05.2026	Abdelkader Shaaban	Semi-final Version
V2.0	14.05.2026	Abdelkader Shaaban	Final Version

Publishable Executive Summary

This deliverable provides recommendations based on the validation of consolidated guidelines regarding change management for digital workspaces in the transition to Industry 5.0. Building on prior work motivating the need for secure, trustworthy, and ethically sound AI-based systems in safety-critical industrial settings (Yasin Ghafourian, 2024), and on persona-based insights into the backgrounds, expertise, and work environments of industrial AI stakeholders (Ghafourian Y, 2025).

It integrates an AI Assurance Framework combining regulatory landscape analysis, risk management using the Technology-Organisation-Environment (TOE) framework, trustworthiness assessment via the Assessment List for Trustworthy AI (ALTAI), and digital workspace evaluation employing the Unified Theory of Acceptance and Use of Technology (UTAUT) and the DeLone & McLean model for measuring system success.

The framework is validated through comparative assessment of an AI-driven chatbot against expert legal review across representative use cases (UC2 - AI Enhanced Value Chain, UC11 - AI-supported Industrial IoT for Indoor Food Production, UC12 - Human-Centred AI for the Optimization of Robust and Competitive Semiconductor Manufacturing Networks and Increased Sustainability, UC13 - Intelligent Sensors improving Robustness of Automated Wafer Transportation and Storage Systems, and UC14 - Speeding-up the Benefit of Automation by Virtual Commissioning through Digital Twins), demonstrating its effectiveness in identifying applicable governance instruments.

Findings indicate that the chatbot serves as a reliable support tool for early-stage compliance orientation and governance mapping (validated for UC2 & UC11 - 14 and demonstrated for all AIMS5.0 Use Cases), while the integrated framework provides a structured mechanism for moving from risk identification to trustworthiness controls, user acceptance, and operational success in AI-enabled digital workspaces.

Table of contents

1. Introduction 7
1.1 Objective and Scope of the Document 7
1.2 Structure of the Deliverable Report 7
2. Regulatory, Standardisation, and Compliance Landscape Relevant to the Change Management ... 9
2.1 GDPR, AI Act and AI-Standards, CRA 9
2.1.1 General Data Protection Regulation (GDPR) 9
2.1.1.1 GDPR Principles 9
2.1.1.2 GDPR and Relevant Supporting Standards 11
2.1.2 AI Act and AI Standards (CEN/CENELEC and ISO/IEC) 11
2.2 Other regulations: Machinery Directive, UNECE (Automotive), Medical Device Directive 14
2.3 CRA: Cyber Resilience Act 15
3. Integrated AI Assurance and Risk Management Framework for AIMS5.0 -Tools and Digital Workspaces 16
3.1 TOE Framework: Context Definition and Risk Identification 17
3.2 Dual-Branch Assessment Logic in the Proposed Model 18
3.2.1 ALTAI Framework for AI Tools in AIMS5.0 18
3.2.2 Digital Workspaces Evaluation Based on UTAUT and the DeLone and McLean Model . 20
3.2.2.1 UTAUT for User Acceptance 20
3.2.2.2 DeLone and McLean for System Success 20
3.3 Integrated Risk-to-Control and Evaluation Matrix 22
3.4 Compliance challenges for digital workspaces 24
3.5 Ethics Standards, Recommendations, and Legal Implications 25
4. Human-Centric Change Management 26
4.1 Worker Involvement, Acceptance, and Human-Centred Design and Tools 26
4.2 Socio-Technical Readiness and Workspace Assessment 26
4.3 Human-Centred Design (HCD) and Adaptive Tools 26
4.4 Impact of digital technologies on business task execution and user agency 27
4.5 Worker Involvement, Acceptance, and Human-Centred Design and Tools 27
4.5.1 Rating Questions on Social Influence, Trust, Usefulness, Ease of Use, and Facilitating Conditions 30
4.5.2 Demographic Profile of Respondents 32
4.5.3 Exploratory Evaluation and Results Analysis 32
4.5.4 Results of Open-Ended Questions 34
4.6 Gaps and Challenges in Worker-Centric Digital Transformation 35
4.6.1 Reconceptualising Acceptance and Trust in the Age of AI 35
4.6.1.1 The Black-Box Problem and the Verification Gap 35
4.6.1.2 Trust as a Prerequisite, Not a Consequence 35
4.6.1.3 Redefining IS Success 35
4.6.2 The Psychology of Trust and Bias: Implementation Hurdles 36
4.6.2.1 The Interpretability Gap 36
4.6.2.2 The Bias Mitigation Paradox 36
4.6.2.3 Anthropomorphism and Accountability 36
4.6.3 The Impact on Work Environment: Empowerment or Control? 37
4.6.3.1 The Paternalism-Assistance Continuum 37
4.6.3.2 Preserving Job Dimensions 37

4.6.4 Technical and Ethical Prerequisites: Moving Beyond the Checklist....37
4.6.4.1 The Trustworthiness-Trust Gap....37
4.6.4.2 Best Practices and Implementation Barriers....38
5. Validation of Consolidated Guidelines and Recommendations....39
5.1 Comparative Validation: AI Chatbot Outcomes vs. Legal Expert Assessment....41
5.1.1 Use Case 2: Servitization and service delivery enabled by Zero administration....41
5.1.2 Use Case 11: AI-supported Industrial IoT for Indoor Food Production....42
5.1.3 Use Case 12: Human-Centred AI for the Optimization of Robust and Competitive Semiconductor Manufacturing Networks....44
5.1.4 Use Case 13: Intelligent Sensors improving Robustness of Automated Wafer Transportation and Storage Systems....45
5.1.5 Use Case 14: Speeding up the Benefit of Automation by Virtual Commissioning through Digital Twins....46
5.2 Chatbot Output for all the Use Cases....47
6. Industrial Insights and Use Case Feedback....51
7. Conclusions....52
7.1 Integration of Framework, Standards, and Change Management....52
7.2 Chatbot Validation: Practical Insights and Limitations....52
7.3 Recommendations for Change Management in Digital Workspaces....53
8. References....54
9. Appendices....56
9.1 Appendix A – Abbreviations....56
9.2 Appendix B: List of ISO/IEC SC42 AI-Standards....57

List of figures

Figure 1 A high-level model of the AI Assurance Framework 17
Figure 2 Overall BIU Acceptance Meter 32
Figure 3 Average Score Distribution by Factor 33
Figure 4 Factor Correlation Matrix 33
Figure 5 An example of the Chatbot Outcomes for UC13 (Part 1) 41
Figure 6 An example of the Chatbot Outcomes for UC13 (Part 2) 41

List of tables

Table 1 Evaluation Matrix 23
Table 2 Gaps in Classical Trust and Acceptance Models 36
Table 3 Potential negative impacts of AI paternalism 37
Table 4 AIMS5.0 Use Case Clusters 39
Table 5 UC02 Comparing Chatbot Results and Expert Inputs 42
Table 6 UC11 Comparing Chatbot Results and Expert Inputs 43
Table 7 UC12 Comparing Chatbot Results and Expert Inputs 44
Table 8 UC13 Comparing Chatbot Results and Expert Inputs 45
Table 9 UC14 Comparing Chatbot Results and Expert Inputs 46
Table 10 Abbreviations 56

1. Introduction

1.1 Objective and Scope of the Document

The transition toward Industry 5.0 represents a fundamental paradigm shift, moving beyond the pure automation of Industry 4.0 to prioritize the synergy between human creativity and machine efficiency. The AIMS5.0 project leverages Artificial Intelligence, IoT, Semantic Web ontologies, and Machine Learning to support this transition by enabling human-centric workplace conditions and climate-friendly production (Maija BREQUE, 2021). Within this context, digital workspaces, which refer to all included and developed use cases throughout the project, are technology-enabled environments in which AI-driven tools, platforms, and decision-support systems integrate with human workers. These digital workspaces serve as the practical instantiation points where trustworthiness principles, governance frameworks, and user-centred design converge.

The present deliverable (D7.3) builds directly on the preceding deliverable D7.2 — Guidelines incl. checklist and self-assessment of how to apply AI in the European industry, and on the peer-reviewed publications produced by the AIMS5.0 consortium in the course of Tasks T7.1–T7.3. (Yasin Ghafourian, 2024) established the motivating rationale, articulating the need for secure, trustworthy, and ethically sound AI-based systems in safety-critical industrial settings. (Belal Abu Naim, 2024)then proposed the self-assessment tool concept to help industrial users and developers navigate the fragmented and rapidly evolving landscape of AI standards and guidelines. (Ádám Kővári, 2025) introduced the Retrieval-Augmented Generation (RAG) architecture of the chatbot that operationalises this concept for the EU AI Act, and (Ghafourian Y, 2025) complemented the tool-side work from the stakeholder side through a persona-based questionnaire of industrial AI practitioners, documenting the backgrounds, readiness levels, and compliance-awareness gaps (notably, ~49% of respondents being uncertain whether their organisations follow formal AI standards) that any such tool must address. Together, these four contributions define the conceptual, architectural, and stakeholder foundations of the present deliverable, which carries that line of work forward by evaluating the resulting framework against independent legal-expert review across representative AIMS5.0 use cases.

The deliverable has four main aims: (1) to map the regulatory, standardization, and compliance landscape relevant to AI and change management; (2) to present an integrated AI Assurance Framework that connects risk identification, trustworthiness assessment, and user acceptance evaluation; (3) to address human-centric change management, including socio-technical readiness and the investigation of worker involvement, acceptance, and human-centered design, while also examining the gaps and challenges in worker-centric digital transformation; and (4) to provide a comprehensive study through comparative chatbot analysis against expert legal review to derive actionable recommendations for organizations implementing AI in digital workspaces.

1.2 Structure of the Deliverable Report

This document is structured as follows. Section 2 provides the regulatory, standardisation, and compliance landscape relevant to change management in digital workspaces, covering the EU AI Act, CEN/CENELEC standards, ISO/IEC standards, and sector-specific regulations (Machinery Directive, UNECE vehicle regulations, Medical Device Directive). Section 3 presents the integrated AI Assurance Framework, detailing the Technology-Organization-Environment (TOE) risk identification layer, the Assessment List for Trustworthy Artificial Intelligence (ALTAI)

trustworthiness branch, and the digital workspace evaluation branch (Unified Theory of Acceptance and Use of Technology - UTAUT and DeLone & McLean model for Information System Success). Section 4 addresses human-centric change management, examining worker involvement, socio-technical readiness, human-centred design, and the psychological factors influencing AI adoption in digital workspaces. Section 5 presents the validation results, comparing chatbot outputs to expert assessments across UC2 & UC11–UC14 and documenting chatbot outputs across all AIMS5.0 use cases. Section 6 synthesises industrial insights and use case feedback. Section 7 concludes with recommendations based on the validation findings and consolidated guidelines for implementing change management in AI-enabled digital workspaces.

2. Regulatory, Standardisation, and Compliance Landscape Relevant to the Change Management

2.1 GDPR, AI Act and AI-Standards, CRA

AIMS5.0 addresses smart manufacturing and digitalisation of industry with the support of Artificial Intelligence. The change to high-level digitalisation of the manufacturing industry in several domains has an impact not only on the production process and supply chains, as indicated in the following chapters and the Use Cases, but also, as an important factor, on humans involved in the change to highly digitalised working environments. The EC has set up several Regulations to protect the interests of people involved and the public in general from negative impact, which have legally binding consequences and can be enforced by considerable financial and other legal penalties. They are valid not only for European companies but also for international companies of non-European origin which are bringing products and services to European markets, including online services. In the context of WP7. The most important ones are the GDPR (General Data Protection Regulations, protecting your personal data and privacy), the AI-Act (protecting people from critical AI-applications and their outcomes), and the Cybersecurity Resilience Act, protecting the public infrastructure, services and business.

Standards play an important role: EC Directives and regulations are referencing standards in accompanying documents published in the EC Journal, so called „Harmonized Standards“. They are either standards taken over from the ISO, IEC and ISO/IEC standardization landscape by the European SDOs (Standardization organizations) CEN, CENELEC or ETSI under a European Mandate and harmonized by referencing in an Annex ZA the articles, for which this standards provides a „presumption of conformance“, i.e. a supplier or applier of products and services can prove compliance to certain articles of the regulation. In some cases, new standards are developed by, e.g., CEN/CENELEC, to fulfil this functionality (e.g. CEN/CENELEC JTC21 for the AI-Act). Details are discussed in the following subchapters.

2.1.1 General Data Protection Regulation (GDPR)

2.1.1.1 GDPR Principles

The EU General Data Protection Regulation (GDPR, Regulation (EU) 2016/679) provides the legal framework for EU data protection, which is a fundamental right in the EU (article 8 of the EU Charter of Fundamental Rights). It is also part of the EEA agreement, i.e. it is also valid in Norway, Iceland and Liechtenstein (European Commission, 2016) (GDPR EU, 2026).

GDPR applies to anyone who processes the personal data of EU citizens or residents, or offers goods or services to such people, **even if the organization/person is not in the EU**. The **fines for violating the GDPR are very high**. There are two tiers of penalties, which max out at €20 million or 4% of global revenue (whichever is higher), plus data subjects have the right to seek compensation for damages.

There are seven protection and accountability principles outlined in Article 5.1-2:

1. **Lawfulness, fairness and transparency** — Processing must be lawful, fair, and transparent to the data subject.
2. **Purpose limitation** — You must process data for the legitimate purposes specified explicitly to the data subject when you collected it.
3. **Data minimisation** — You should collect and process only as much data as absolutely necessary for the purposes specified.
4. **Accuracy** — You must keep personal data accurate and up to date.

5. **Storage limitation** — You may only store personally identifying data for as long as necessary for the specified purpose.
6. **Integrity and confidentiality** — Processing must be done in such a way as to ensure appropriate security, integrity, and confidentiality (e.g. by using encryption).
7. **Accountability** — The data controller is responsible for being able to demonstrate GDPR compliance with all of these principles.

Article 6 lists the instances in which it's legal to process personal data. Avoid completely handling personal data, unless you can justify it with one of the following:

1. The data subject gave you specific, **unambiguous consent** to process the data. (e.g. They've opted in to your marketing email list).
2. Processing is necessary to execute or to prepare **to enter into a contract** to which the data subject is a party. (e.g. You need to do a background check before leasing property to a prospective tenant).
3. You need to process it **to comply with a legal obligation** of yours. (e.g. You receive an order from the court in your jurisdiction).
4. You need to process the data **to save somebody's life**. (e.g. Well, you'll probably know when this one applies).
5. Processing is necessary **to perform a task in the public interest** or to carry out some official function. (e.g. You're a private garbage collection company).
6. You have a **legitimate interest** to process someone's personal data. This is the most flexible lawful basis, though the "fundamental rights and freedoms of the data subject" always override your interests, especially if it's a child's data. (It's difficult to give an example here because there are a variety of factors you'll need to consider for your case. The UK Information Commissioner's Office provides helpful guidance here).

This is very important for research projects with demonstrators/use cases where persons are involved and personal information (including behaviour, health data, surveillance of persons, even if anonymously), in particular if persons are involved not being partners in the project. Their consent has to follow strict rules:

- Consent must be "freely given, specific, informed and unambiguous."
- Requests for consent must be "clearly distinguishable from the other matters" and presented in "clear and plain language."
- Data subjects can withdraw previously given consent whenever they want, and you have to honour their decision. You can't simply change the legal basis of the processing to one of the other justifications.
- Children under 13 can only give consent with permission from their parent.
- You need to keep documentary evidence of consent.

The "privacy rights" of so-called "data subjects" are listed in short:

1. The right to be informed
2. The right of access
3. The right to rectification
4. The right to erasure
5. The right to restrict processing
6. The right to data portability
7. The right to object
8. Rights in relation to automated decision making and profiling.

Compliance with the GDPR rules has to be documented in case personal data is handled in a project.

2.1.1.2 GDPR and Relevant Supporting Standards

GDPR is not a standard and does not reference particular standards, but there are a number of **international standards, frameworks, and best practices** that help organisations demonstrate compliance.

A short overview without claiming completeness is provided in this subsection:

ISO/IEC 27000 family and related privacy standards:

- **ISO/IEC 27001**: The *most important* standard for GDPR alignment. Defines requirements for an **Information Security Management System (ISMS)**.
- **ISO/IEC 27002**: Provides detailed **security controls** (access control, encryption, logging, etc.).
- **ISO/IEC 27005**: Supports GDPR's **risk-based approach**.
- **ISO/IEC 27701**:2025 - Extends ISO 27001 specifically for **privacy management**, one of the **closest standards to GDPR compliance**.
- **ISO/IEC 29100**:2024 – Security techniques – Privacy framework.
- **ISO/IEC 29134**:2023 – Security techniques - Guidelines for privacy impact assessment.

AI-Standards in connection with GDPR issues:

In the following sections on AI-standards the standards to ensure fairness, transparency, explainability, accountability, governance, Ethical and lawful AI use (connection to the AI-Act) and to assess and mitigate risks and bias, have impact on personal data handled by AI-systems and add additional requirements, because AI introduces additional risk dimensions (profiling, automation, bias, explainability), so beyond classic ISO 27000 standards.

One of the ISO/IEC Standards on Big Data (an ISO/IEC SC42 topic) is developed by ISO/IEC SC27 (IT security, cybersecurity and privacy protection):

- **ISO/IEC 20547-4**:2020 Information technology — Big data reference architecture — Part 4: security and privacy.

CEN/CENELEC JTC21 standardization work is performed on **AI Privacy Protection, based on ISO/IEC 27701** (Privacy Information Management System standard providing the framework for GDPR compliance in AI contexts) and **ISO/IEC 29100** (Privacy framework standard establishing core principles for protecting personal data in AI systems). Future work is planned on **Data Extraction Protection** against unauthorised extraction of data from AI systems and on Coordinated, **integrated Governance**.

2.1.2 AI Act and AI Standards (CEN/CENELEC and ISO/IEC)

The European AI-Act (Regulation (EU) 2024/1689) and standardization:

The EC AI-Act regulates the human-centred, societal, and public interest issues of AI to protect citizens against the violation of their rights and misuse of powerful AI applications. Harmonized standards are the basis for suppliers and users to prove "presumption of conformity". The EC has sent 10 requests for standardisation to CEN/CENELEC for that purpose, which are now handled in CEN/CLC JTC 21, "Artificial Intelligence". Standardization members of AIMS5.0 are active on international and national levels in the mirror Committees, e.g., AIT in ASI K001 AG42 and ISO/IEC JTC1 SC42.

The **European Commission's AI Act** served as a foundational regulation for the request to CEN/CENELEC for harmonised standards, outlining ten critical requirements for **high-risk AI systems**. These include:

1. Risk management system for AI systems [Art. 9 AI-Act]

2. Governance and quality of datasets used to build AI systems [Art. 10 Data and data governance]
3. Record keeping - built-in logging capabilities in AI systems [Art. 11 Techn. doc., Art. 12 record-keeping]
4. Transparency and information to the users of AI systems [Art. 13 Transparency and user info]
5. Human oversight of AI systems [Art. 14 Human oversight]
6. Accuracy specifications for AI systems [Art. 15 Accuracy, robustness and cybersecurity]
7. Robustness specifications for AI systems [Art. 15 Accuracy, robustness and cybersecurity]
8. Cybersecurity specifications for AI systems [Art. 15 Accuracy, robustness and cybersecurity]
9. Quality management system for providers of AI systems [Art. 17 QM for providers]
10. Conformity assessment for AI systems [Art. 19 + Art. 43 Conformity Assessment]

These requirements highlight the need for harmonised standards to support compliance and ensure the safe and responsible deployment of AI technologies across Europe.

CEN/CENELEC Standards under development in response to the EC Standardisation Request:

These procedures are explicitly aligned with the forthcoming European Norms (ENs) of CEN/CENELEC JTC 21, expected to provide a presumption of conformity with the EU AI Act. These include:

- prEN 18229-1: AI trustworthiness framework – Part 1: Logging, transparency and human oversight
- prEN 18229-2: AI trustworthiness framework – Part 2: Accuracy and robustness
- prEN 18228: AI Risk Management
- prEN 18281: Artificial Intelligence – Evaluation methods for accurate computer vision systems
- prEN 18282: Artificial Intelligence – Cybersecurity specifications for AI systems
- prEN 18283: Artificial Intelligence – Concepts, measures and requirements for managing bias in AI systems
- prEN 18284: Artificial Intelligence – Quality and governance of datasets in AI
- prEN 18285: AI Conformity assessment framework
- prEN 18286: Artificial Intelligence – Quality management system for EU AI Act regulatory purposes
- prEN 18281: Artificial Intelligence – Evaluation methods for accurate computer vision systems

The project closely monitors the development of such norms and integrates their approach throughout the relevant activities.

International Standardization (ISO/IEC JTC1 SC42, Artificial Intelligence, CEN/CENELEC JTC21, ETSI SAI)

One of the "youngest" standardisation groups of ISO/IEC JTC1 (Information Technology) is SC 42, Artificial Intelligence. The work covers a broad range of Standards, Technical Reports and Technical Specifications (i.e., *ISO/IEC DIS 42105, Huuman oversight*). There are enormous efforts put into standardisation, often beyond existing (industrial) practice and state-of-the-art. By the end of April 2026, there are 41 standards published and 50 under development (including also a few amendments and maintenance of already published standards).

A basic standard of ISO/IEC JTC1 SC42 is ISO/IEC TR 24368:2022, "Overview of ethical and societal concerns". A few interesting standards addressing human aspects (impact), ethical and societal issues are listed here (some AIMS5.0 partners are involved in the standardization activities on national and/or international level), particularly in WG 3 (Trustworthiness) and WG 4 (AI Use cases, especially human-machine teaming, an upcoming topic since 2025):

A full list of ISO/IEC JTC1 SC42 AI standards as valid at submission of this document is provided in **Appendix B.**

Note 1: A TR has no requirements; it is rather an informal recommendation and study of an issue; a TS includes already normative requirements like a full standard.

Note 2: Standards under development ITALIC.

Ethical concerns, Governance, Human-machine teaming (use cases):

- **ISO/IEC TR 24368:2022** Information technology — Artificial Intelligence (AI) — Overview of ethical and societal concerns
- **ISO/IEC CD TS 22443** Information technology — Artificial intelligence— Guidance on addressing societal concerns and ethical considerations (under development)
- **ISO/IEC TR 24030:2024** Information technology — Artificial Intelligence — Use cases (Update of version 2021)
- **ISO/IEC 38507:2022** Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations
- **ISO/IEC TS 6254**:2025 Information technology — Artificial intelligence — Objectives and approaches for explainability of ML models and AI systems (to be published soon)
- **ISO/IEC DIS 42105** - Information technology — AI— Guidance for human oversight of AI system (under development)
- **Human-machine teaming, Use cases (under development)**
    - **ISO/IEC AWI TR 24030** Information technology — Artificial intelligence (AI) - Use cases
    - **ISO/IEC CD TR 42109** Information technology — Artificial intelligence - Use cases of human-machine teaming
    - **ISO/IEC PWI 25880** "Information technology — Artificial intelligence — Guidance for organizational implementation of human-machine teaming concept"

Controllability and bias:

- ISO/IEC TS 8200:2024 Information technology — Artificial intelligence — Controllability of automated artificial intelligence systems
- ISO/IEC TS 12791:2024 Information technology — Artificial intelligence — Treatment of unwanted bias in classification and regression machine learning tasks
- ISO/IEC TR 24027:2021 Information technology — Artificial Intelligence (AI) — Bias in AI systems and AI aided decision making

Trustworthiness and Robustness:

- **ISO/IEC TR 24028:2020** Information technology — Artificial Intelligence — Overview of trustworthiness in artificial intelligence
- **ISO/IEC TR 24029-1:2021** Artificial Intelligence — Assessment of the robustness of neural networks — Part 1: Overview
- **ISO/IEC 24029-2: 2023** Artificial Intelligence — Assessment of the robustness of neural networks — Part 2: Methodology for the use of formal methods
- **ISO/IEC DIS 24029-3** Artificial intelligence (AI) — Assessment of the robustness of neural networks — Part 3: Methodology for the use of statistical methods (under development)

Risk Management:

- **ISO/IEC 23894:2023** Information technology — Artificial Intelligence — Guidance on Risk Management
- **ISO/IEC 42005:2025** Information technology — Artificial intelligence — AI system impact assessment
- **ISO/IEC CD 42007** Information technology — Artificial intelligence — High-level framework and guidance for the development of conformity assessment schemes for AI systems (under development)
- **ISO/IEC AWI 25870** Artificial intelligence — Reporting framework for AI incidents (under development)

Functional Safety and AI:

- ISO/IEC TR 5469:2024 Artificial intelligence — Functional safety and AI systems
- As follow-up and replacement (under development):
    - **ISO/IEC CD TS 22440-1** – Artificial intelligence — Functional safety and AI systems — Part 1: Requirements
    - **ISO/IEC CD TS 22440-2** – Artificial intelligence — Functional safety and AI systems — Part 2: Guidance
    - **ISO/IEC CD TS 22440-3** - Artificial intelligence — Functional safety and AI systems — Part 3: Examples of application

Other Issues/Domains:

- **ISO/IEC TR 20226:2025** Information technology — Artificial intelligence — Environmental sustainability aspects of AI systems
- **ISO/IEC CD TR 18988**: Artificial intelligence — Application of AI technologies in health informatics (under development)
- **ISO/IEC AWI 22989-2** - Artificial intelligence —Concepts and terminology – Part 2: Healthcare (under development)

2.2 Other regulations: Machinery Directive, UNECE (Automotive), Medical Device Directive

High-level digitalisation of industry, as addressed in AIMS5.0, also covers general areas of interest that are not typically focused solely on AI and Cybersecurity, but also on machinery safety, partial use of devices from the medical device and rehab sector (such as exoskeletons), and mobility (automotive), as examples. If devices or attributes of parts in the processes address other areas, as in section 2.1, relevant directives, regulations and standards also apply. Some examples are provided here. If use cases lead to products, a thorough analysis of the potential regulations, directives and standards must be performed.

Scope and relevance of each directive/regulation:

- **Machinery**
    - Machinery Directive: REGULATION (EU) 2023/1230 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 June 2023 on machinery and repealing Directive 2006/42/EC of the European Parliament and of the Council and Council Directive 73/361/EEC (Commission, Regulation (EU) 2023/1230 of the European Parliament and of the Council of 14 June 2023 on machinery and repealing Directive 2006/42/EC of the European Parliament and of the Council and Council Directive 73/361/EEC, 2023).

- **Upcoming Machinery Regulation:**
  - EU Machinery Regulation (2023/1230/EU) (Services, 2025)

- **Automotive**
  - UNECE vehicle regulations (e.g., UN Regulations R155 & R156 for cyber security and software updates) and associated standards (e.g., ISO/SAE 21434).
    - UNECE Regulation No. 155 (R155) – Cybersecurity and Cybersecurity Management System (UNECE, 2021)
    - ISO/SAE 21434:2021 - Road vehicles — Cybersecurity engineering

- **Medical Devices Directive, as mentioned in:** eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31993L0042
  - This is related to the UC8 related to the exoskeleton using AI

2.3 CRA: Cyber Resilience Act

The European Cyber Resilience Act (CRA) introduces binding cybersecurity obligations for products with digital elements placed on the European Union market. It establishes a harmonised regulatory framework to ensure that software and hardware products, including their remote data processing functionalities, comply with essential cybersecurity requirements throughout their entire lifecycle. The Regulation applies to any product with digital elements, meaning software or hardware products whose remote data processing solutions are designed and developed by, or under the responsibility of, the manufacturer, and whose absence would prevent the product from performing one of its intended functions (European Commission (EC) - CRA Act, 2024).

The essential cybersecurity requirements set out in Annex I of the CRA can be directly linked to practical implementation experience generated within AIMS5.0. For example, in Use Case 11 (UC11 - AI-supported Industrial IoT for Indoor Food Production), a comprehensive cybersecurity framework was established using ThreatGet, a systematic threat modelling and risk assessment tool, explicitly aligned with the CRA's essential requirements.

The UC11 system deploys an organised architecture of interconnected components across field, edge, and cloud layers. The ThreatGet-based threat analysis covered the interconnected architecture of the UC, including field layer components, sensors, and actuators, as well as the remote monitoring and control environment, including network devices, servers, and data stores. The tool supports the investigation of potential threats that could be triggered by these interconnected components through the exploitation of security vulnerabilities, providing a clear vision for securing the UC architectural model by applying the appropriate set of security controls to mitigate such cyber risks.

Within AIMS5.0, the work conducted in UC11 represents an initial step towards automating compliance with the CRA's essential requirements through structured cybersecurity activities. The implementation insights derived from this work have further been used as input for the development of harmonised standards, as discussed in AIMS5.0 D3.3 - Final Architecture Concepts, Realisation, and Recommendations for Future Standardisation.

3. Integrated AI Assurance and Risk Management Framework for AIMS5.0 -Tools and Digital Workspaces

AIMS5.0 leverages new technologies from IoT, Semantic Web ontologies, Machine Learning and Artificial Intelligence to support the transition from Industry 4.0 to Industry 5.0 by enabling human-centric workplace conditions and climate-friendly production. In this context, the Assessment List for Trustworthy Artificial Intelligence (ALTAI) framework plays a key role by ensuring that these AI-driven advancements are developed and deployed in a trustworthy, transparent and resilient manner. As a baseline quality, safety and ethics framework for all AI tools within AIMS5.0, ALTAI embeds human oversight into AI control loops, enforces consistent safety and robustness principles, ensures transparency from the outset, and anticipates and mitigates risks to humans, society and the environment.

However, for AIMS5.0 it is not sufficient to assess only the trustworthiness requirements of the developed AI tools. Since the project also aims to demonstrate value in concrete digital workspaces, the assessment must additionally consider whether these solutions are accepted by users and whether they achieve successful and beneficial operation in practice. For this reason, the framework is extended with two complementary perspectives:

(1) The Unified Theory of Acceptance and Use of Technology (UTAUT) model (Wikipedia, 2026), which contributes the user-acceptance perspective by assessing factors such as performance expectancy, effort expectancy, social influence, and facilitating conditions;
(2) The DeLone & McLean Information Systems Success Model (DeLone WH, 2003), which evaluates system quality, information quality, service quality, user satisfaction, use, and net benefits.

While ALTAI provides the trustworthiness and risk-mitigation perspective for the AI tools, UTAUT and the DeLone & McLean model together address the practical effectiveness and user acceptance dimensions of the developed solutions in operational environments.

To connect these perspectives in a structured way, AIMS5.0 introduces an integrated mapping matrix that links the risks identified for each AI system and use case to the relevant ALTAI requirements and to the corresponding UTAUT and DeLone & McLean evaluation dimensions. This matrix provides a systematic mechanism for moving from risk identification to trustworthiness controls, use-case validation, and measurable outcomes, thereby supporting AIMS5.0 in addressing risks through structured, evidence-based solutions.

The parts of the proposed framework are described in detail in the following sections. Figure 1 illustrates this integrated assessment logic by showing how the Technology-Organization-Environment (TOE) based risk identification layer feeds into the ALTAI trustworthiness branch and the use-case evaluation branch based on UTAUT and the DeLone & McLean models, which are then combined through the mapping matrix into the AIMS5.0 AI Assurance Framework for structured risk mitigation, trustworthy AI, user acceptance, and operational success.

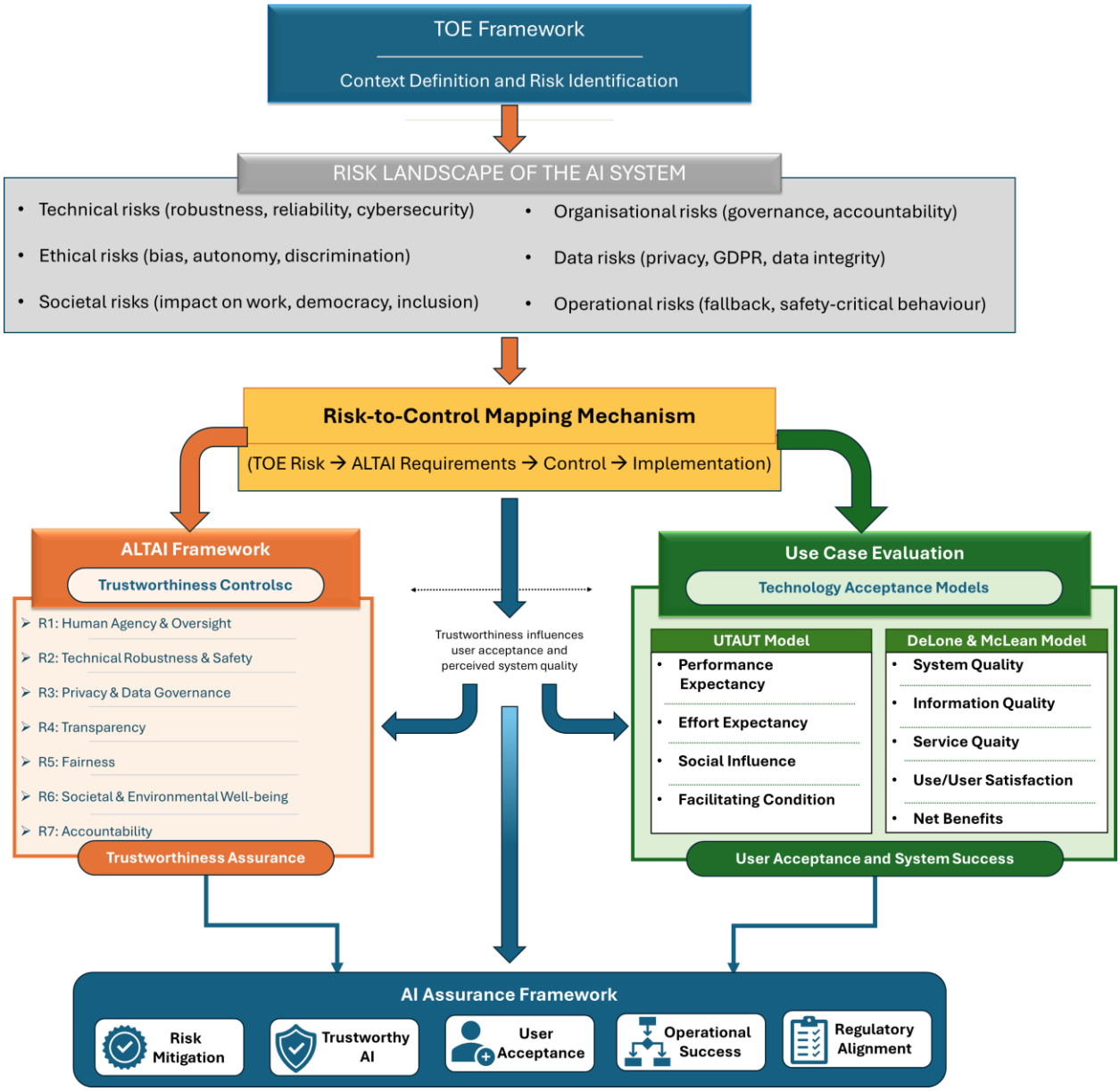

Figure 1 A high-level model of the AI Assurance Framework

3.1 TOE Framework: Context Definition and Risk Identification

The Technology–Organisation–Environment (TOE) framework is used in this model as the starting point for defining the evaluation context because, in the literature, it is described as a classical firm-level framework used to understand the determinants that influence the adoption of new technological innovation by a firm (Jere JN, 2020). It is built around three contexts: the technological, organisational, and environmental context of adoption. In AIMS5.0, this TOE-based contextual structure is applied to support the first phase of assessment and to identify the risk landscape of the AI system. In this sense, TOE is particularly suitable for structuring the initial assessment phase, because it provides a systematic way to examine the technological conditions, the organisational setting, and the external environment that may shape how risks emerge and how they need to be addressed. Accordingly, in the updated diagram the TOE block does not function as a trustworthiness or adoption model by itself; rather, it functions as the

project's risk-definition layer, from which the AI risk landscape is derived across technical, ethical, societal, organisational, data-related, and operational dimensions (Jere JN, 2020).

This role is fully consistent with the vision of AIMS5.0. The project leverages IoT, Semantic Web ontologies, Machine Learning and Artificial Intelligence to support the transition from Industry 4.0 to Industry 5.0 in a human-centric and climate-friendly direction. In this context, the TOE framework is proposed as a structured basis for determining which risks are relevant for each AI-enabled tool and digital workspaces. These include robustness and cybersecurity risks, risks related to bias and autonomy, governance and accountability risks, privacy and data integrity risks, and operational risks such as fallback and safety-critical behaviour. In the figure, this proposed integration is represented by the "Risk Landscape of the AI System "block, which captures the intended output of the TOE-based assessment phase within AIMS5.0, including ethical, technical, societal, data-related, operational, and organisational risks. In this way, the TOE-informed phase would provide the project with a structured and defensible basis for identifying the trustworthiness controls required and the dimensions of acceptance and success to be evaluated subsequently.

3.2 Dual-Branch Assessment Logic in the Proposed Model

The proposed model contains two branches of analysis because AIMS5.0 must address two distinct but interrelated questions. The first concerns whether the AI tools developed in the project are trustworthy and subject to appropriate controls. The second concerns whether the project's digital workspaces demonstrate user acceptance, effective use, and system success in practice.

For this reason, the diagram is structured into two branches following the TOE-derived risk landscape: a trustworthiness branch based on the ALTAI framework (Commission, Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment, 2020; Commission, Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment, 2020), and a use-case evaluation branch based on the UTAUT model (Wikipedia, 2026) and the DeLone and McLean Information Systems Success Model (DeLone WH, 2003).

This dual-branch structure is methodologically appropriate because it separates the assessment of tool-level trustworthiness controls from the evaluation of use-case-level adoption and system success, while allowing both branches to converge within a single integrated AIMS5.0 AI assurance framework. Within this structure, the Risk-to-Control Mapping Mechanism should be understood as the structured logic that links the risks identified through the TOE-based assessment to the relevant ALTAI requirements and, where applicable, to the user-acceptance and system-success dimensions to be evaluated at the use-case level. The model further adopts the design assumption that trustworthiness influences user acceptance and perceived system quality. This assumption is not presented as a universal causal claim, but as a rational integration principle underpinning the relationship between the trustworthiness and use-case evaluation branches of the proposed framework.

3.2.1 ALTAI Framework for AI Tools in AIMS5.0

The ALTAI framework consists of seven requirements that define what a Trustworthy AI system must demonstrate across its lifecycle. Each requirement includes controls, safeguards, or governance practices that must be assured through an assessment process. Below is a concise summary of all seven requirements (Commission, Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment, 2020):

REQUIREMENT 1: Human Agency and Oversight
AI systems must support, not undermine, human autonomy, and must always remain subject to meaningful human oversight.
This includes:
- Avoiding user confusion
- Preventing over-reliance
- Ensuring human-in-the-loop/on-the-loop/in-command oversight
- Ensuring the ability to stop or override the system

REQUIREMENT 2 - Technical Robustness and Safety
The system must be technically secure, resilient, reliable, and safe. This includes:
- Cybersecurity hardening
- Robustness to attacks (poisoning, evasion, inversion)
- Reliability and reproducibility
- Fail-safe and fallback capabilities
- Managing accuracy and error propagation

REQUIREMENT 3 - Privacy and Data Governance
Ensures protection of personal and sensitive data, GDPR alignment, and strong lifecycle data governance. This includes:
- DPIA / GDPR compliance
- DPO involvement
- Privacy-by-design
- Data minimisation

REQUIREMENT 4 - Transparency
AI systems must be understandable, traceable, and explainable. Includes:
- Traceability
- Ability to explain decisions
- Clear communication of purpose, limitations, risks
- Informing users when interacting with an AI system

REQUIREMENT 5 - Diversity, Non-Discrimination and Fairness
Ensures equitable outcomes, avoiding reinforcement of bias. Includes:
- Bias testing and monitoring
- Accessibility for all users, including those with disabilities
- Fairness metrics
- Stakeholder participation

REQUIREMENT 6 - Societal and Environmental Well-Being
AI must benefit society and avoid causing harm. Includes:
- Environmental impact assessment
- Impact on work, skills and workforce development
- Effects on democratic processes and societal cohesion

REQUIREMENT 7 - Accountability
Ensures strong governance, auditability, and risk management. Includes:
- Ethics board involvement
- Processes for reporting vulnerabilities and risks
- Redress mechanisms for affected users

The main value of the integration of the ALTAI framework is that it turns the risks identified through TOE into governed, auditable, and explainable control requirements. In practical terms, it enables AIMS5.0 to move from identifying that a risk exists, to deciding which

trustworthiness requirement is relevant, which control should be selected, and how that control should be implemented and evidenced. This is why the ALTAI sits at the centre of the project's trustworthiness-by-design logic.

3.2.2 Digital Workspaces Evaluation Based on UTAUT and the DeLone and McLean Model

3.2.2.1 UTAUT for User Acceptance

The growth of the e-commerce sector and emerging digital technologies drives the implementation of new technologies in organisations. However, market research suggests that the success rate of new technology adoption in organisations, whereby technologies bring the expected return on investment, is below 30 percent. To offer a comprehensive view of technology acceptance, Venkatesh et al. (Venkatesh V, 2003) introduced the Unified Theory of Acceptance and Use of Technology (UTAUT) in 2003. This model integrates key constructs from earlier technology acceptance theories (Marikyan M, 2021).

3.2.2.1.1 The Core UTAUT Model

The UTAUT theoretical model suggests that actual technology use is determined by a user's "behavioural intention". The likelihood of adopting a technology depends on the direct effect of four key constructs (Marikyan M, 2021):

- **Performance Expectancy:** According to Venkatesh **et al** (Venkatesh V, 2003)**, performance expectancy is defined as: "**The degree to which an individual believes that using the system will help him or her to attain gains in job performance". It is the strongest predictor of intention to use.
- **Effort Expectancy:** "The degree of ease associated with the use of the system", as introduced by **Venkatesh et al.** (Venkatesh V, 2003)**.**
- **Social Influence:** According to **Venkatesh et al.** (Venkatesh V, 2003)**, it is defined as "**The degree to which an individual perceives that important others believe he or she should use the new system".
- **Facilitating Conditions:** "The degree to which an individual believes that an organisation's and technical infrastructure exists to support the use of the system," according to **Venkatesh et al** (Venkatesh V, 2003)

The strength of these predictors on intention is moderated by four variables: age, gender, experience, and voluntariness of use.

In AIMS5.0, this means that the UTAUT part of the use-case branch can be used to assess whether the project's users perceive the AI-enabled solution as useful, manageable, socially supported, and adequately backed by organisational and technical infrastructure. This is important because the right-hand branch is not meant to duplicate the ALTAI assessment; rather, it is meant to determine whether the implemented AI solution is likely to be accepted and used within the target use cases.

3.2.2.2 DeLone and McLean for System Success

Information system success (ISS) has been a central focus in information systems and information technology research for several decades. Historically, diverse perspectives and fragmented measures have complicated comparisons across research. Recognising ISS as a key dependent variable, DeLone and McLean consolidated previous measures to propose a unified model in 1992 (DM1992). DeLone and McLean subsequently updated this model in 2003 (DM2003) to reflect the evolving role of information technology (Jeyaraj, 2020). The original DM1992

framework conceptualised Information System Success as a multidimensional construct comprising six core dimensions:

- **Information Quality (IQ)**
- **System Quality (SQ)**
- **User Satisfaction (US)**
- **System Usage (SU)**
- **Individual Impact (II)**
- **Organisational Impact (OI)**

The DeLone and McLean (1992) Information System Success (ISS) model explicitly defines nine key interconnections among its core dimensions to explain how systems create value. Rather than functioning as isolated metrics, these dimensions interact through specific functional pathways (Jeyaraj, 2020):

- **The Effects of Quality (Four relationships):** These foundational relationships depict how a system's technical and informational characteristics influence user behaviour and attitudes. Specifically, both System Quality (SQ) and Information Quality (IQ) are posited to directly affect a user's System Usage (SU) as well as their overall User Satisfaction (US). The pathways are defined as: SQ → SU, SQ → US, IQ → SU, and IQ → US.
- **The Mutual Effects (Two relationships):** The model highlights a reciprocal dynamic between how a user interacts with the system and how they feel about it. It establishes mutual effects between behaviour and attitude, where System Usage and User Satisfaction continuously reinforce one another (SU → US and US → SU).
- **The Effects on Individual Impact (Two relationships):** These pathways focus on the individual-level consequences of the system. The framework outlines that both System Usage and User Satisfaction independently lead to a measurable Individual Impact (II) on the user's performance (SU → II and US → II).
- **The Effects Between Impacts (One relationship):** Finally, the model culminates in a single, overarching relationship that bridges the gap between the individual and the enterprise. It proposes that Individual Impact ultimately drives broader Organisational Impact (OI), completing the chain from technical quality to enterprise-wide benefits (II → OI).

The Updated Model (DM2003)

Over time, DeLone and McLean noted that prior research overlooked the multidimensional and interdependent nature of ISS. Consequently, they revised the framework into the DM2003 model, introducing three major structural changes (DeLone WH, 2003):

- **Intention to Use (IU):** This variable was added as an alternative measure for system usage, to be applied when appropriate.
- **Service Quality (RQ):** This dimension was introduced to directly affect both system usage and user satisfaction.
- **Net Benefits (NB):** This construct consolidated and replaced the previous dimensions of "individual impact" and "organisational impact" from the 1992 model.

Both models have been extensively utilised and empirically supported in subsequent research evaluating diverse technologies across various contexts.

These dimensions are represented in the right-hand branch of the updated diagram. Within AIMS5.0, the DeLone and McLean component is proposed to serve to evaluate whether an AI-enabled use case delivers outputs of sufficient quality, whether the system performs adequately, whether the support and service environment is acceptable, whether the system is utilised, whether users are satisfied, and whether the use case generates meaningful benefits.

Furthermore, the DeLone and McLean Information Systems Success Model complements the UTAUT by extending the assessment beyond user acceptance and intended use to encompass system success in operational contexts. While UTAUT predicts the likelihood of user adoption and usage within AIMS5.0, the DeLone and McLean model evaluates whether the system achieves the expected quality, user satisfaction, and net benefits during actual operation.

The bottom block of the updated diagram combines the outputs of both branches into the **AI Assurance Framework** for AIMS5.0. This integrated framework formalises the project's assessment logic.

First, TOE provides the **context definition and risk identification**. Second, ALTAI transforms those risks into **trustworthiness controls and mitigation requirements** for the AI tools. Third, the use-case branch evaluates whether the resulting solutions are **accepted, used, and successful** in real organisational settings. Finally, the integrated framework combines these outputs into five project-level outcomes: **risk mitigation, trustworthy AI, user acceptance, operational success, and regulatory alignment**.

This integrated structure offers AIMS5.0 several advantages:

- It gives the project a clear way to move from **risk identification to structured mitigation**,
- it ensures that AI tools are assessed against an official trustworthiness framework rather than ad hoc criteria, and
- it makes it possible to evaluate whether technically and ethically controlled AI systems are also accepted and beneficial in practice.

In other words, the framework does not only ask whether the AI is trustworthy; it also asks whether it works in the use case, whether users accept it, and whether it produces measurable value.

3.3 Integrated Risk-to-Control and Evaluation Matrix

The goal of the matrix is to provide AIMS5.0 with a **structured decision and evidence mechanism** that links:

- the **risk categories identified through TOE**,
- the **relevant ALTAI requirements and controls**, and
- the **use-case evaluation criteria drawn from UTAUT and DeLone & McLean**.

In practical terms, the matrix helps the project answer five questions for each AI tool and use case:

1. What is the relevant risk (TOE Framework)?
2. Which ALTAI requirement addresses it?
3. What control or safeguard should be implemented?
4. Which acceptance or success dimensions must be checked at the use-case level?
5. What is the expected structured solution or assurance outcome?

The proposed matrix serves as an illustrative example of the practical operationalisation of the AIMS5.0 AI Assurance Framework. It demonstrates how risks identified through the TOE-based assessment can be systematically linked to relevant ALTAI requirements and to the corresponding UTAUT and DeLone and McLean evaluation dimensions. Accordingly, Table 1 Evaluation Matrix offers a structured example of the progression from risk identification to trustworthiness controls, use-case evaluation, and evidence-based assurance outcomes within the AIMS5.0 framework.

Table 1 Evaluation Matrix

TOE Risk Category	Typical AIMS5.0 risk description	Primary ALTAI requirement(s)	Typical ALTAI control / structured solution	UTAUT Model for UCs	DeLone & McLean Models for UCs	Main goal in AIMS5.0
Technical Risks	AI model failure, instability, cyberattack, unreliable outputs	R2 Technical Robustness & Safety	Security controls, robustness testing, fallback procedures, accuracy monitoring, reproducibility checks	Performance expectancy, facilitating conditions	System quality, service quality	Reduce technical failure risk and ensure dependable AI operation
Ethical Risks	Bias, discriminatory outcomes, interference with autonomy, over-reliance	R1 Human Agency & Oversight; R5 Diversity, Non-Discrimination and Fairness	Human oversight, override mechanisms, bias testing, fairness monitoring, inclusive design	Social influence, performance expectancy	User satisfacti on, net benefits	Ensure human-centred and fair AI use
Societal Risks	Negative effects on work, inclusion, democratic processes, social cohesion	R6 Societal & Environmental Well-Being	Social impact review, stakeholder engagement, workforce considerations, environmental and societal safeguards	Social influence, facilitating conditions	Net benefits, user satisfacti on	Prevent broader societal harm and support Industry 5.0 goals
Organisati onal Risks	Weak governance, unclear accountability , poor escalation or reporting	R7 Accountability	Governance mechanisms, auditability, reporting channels, ethics oversight, risk management procedures	Facilitating conditions	Service quality, net benefits	Establish accountable deployment and evidence-based governance
Data Risks	GDPR breaches, weak data governance, privacy harm,	R3 Privacy & Data Governance	DPIA, privacy-by-design, data minimisation, secure access, logging, DPO involvement	Facilitating conditions, effort expectancy	Informati on quality, system quality	Protect data, ensure lawful processing, and strengthen trust

	and integrity issues					
Transparency Risks	Black-box decisions, poor explainability, unclear system limitations	R4 Transparency	Explainability, traceability, logging, communication of purpose and limits	Effort expectancy, performance expectancy	Information quality, user satisfaction	Make AI decisions understandable and support informed use
Operational Risks	Safety-critical behaviour, weak fallback, poor support in real use	R2 Technical Robustness & Safety; R7 Accountability	Fail-safe design, stop procedures, monitoring, response mechanisms, operational governance	Facilitating conditions, effort expectancy	Service quality, system quality, net benefits	Ensure safe and sustainable operation in real use cases

3.4 Compliance challenges for digital workspaces

Compliance in AI-enabled digital workspaces is challenging largely because "what counts as compliant" is spread across many different kinds of documents, ethical guidelines, technical standards, and binding laws, and these sources evolve faster than most organisations can continuously track. In industrial settings, this creates a gap between high-level principles (trustworthy, safe, human-centric AI) and day-to-day engineering decisions, which can undermine both trust and adoption if teams cannot translate abstract requirements into concrete controls and evidence.

Relevant standards and guideline repositories are large, scattered across the web, frequently updated, and difficult to keep current with, while many static guideline documents are not easily searchable or accessible in a centralised way. This makes it hard for teams to confidently identify the right requirements for a specific use case at the moment decisions are made (e.g., during design, procurement, validation, deployment), and it increases the likelihood of missed obligations or inconsistent interpretations across projects.

On the regulatory side, the compliance burden increases because the landscape is multi-layered: non-binding frameworks (e.g., high-level principles), binding legislation (e.g., GDPR and the EU AI Act), technical standards (e.g., ISO/IEC, IEEE), and sector-specific rules all interact, and the applicable subset depends heavily on context and jurisdiction. The EU AI Act further complicates implementation by introducing risk-based obligations (especially for high-risk systems), demanding stronger practices in governance, transparency, documentation, and human oversight, and requiring teams not only to "do the right things" but also to show they did so with traceable evidence.

Finally, even when teams try to lean on AI assistance to navigate compliance, generic LLM tools can be too broad, may not reflect the latest texts, and often cannot incorporate proprietary or copyrighted standards that organizations rely on internally. This creates a challenge: developers still need a reliable way to retrieve the correct, current sources and map them to their system and stakeholder needs.

3.5 Ethics Standards, Recommendations, and Legal Implications

Across industrial domains, ethics standards, policy recommendations, and legal frameworks increasingly converge into a layered governance model for AI and data-driven systems. The UNESCO Recommendation on the Ethics of Artificial Intelligence provides a normative foundation rooted in human rights, sustainability and societal well-being, encouraging organisations to embed principles such as transparency, accountability and human oversight throughout the lifecycle of AI systems. These high-level principles are operationalised and enforced through binding regulations such as the General Data Protection Regulation (GDPR) and the AI Act, as well as through technical standards developed by international bodies like ISO and IEC. Within the European context, CEN and CENELEC play a critical intermediary role, translating both regulatory requirements and global standards into harmonised European standards that are directly aligned with EU legislation and market needs. Together, these instruments shape how industrial actors design, deploy, and manage AI developments and implementations, moving from voluntary ethical commitments toward enforceable compliance and verifiable best practices. Such regulations and standards are thoroughly described and discussed in Section 2.1 of this document.

The legal and operational implications for industry are substantial and increasingly strategic. The GDPR establishes strict requirements for data governance, consent, and individual rights, directly affecting how industrial data, particularly personal and operator-related data, is collected, stored and processed. The AI Act introduces a risk-based classification of AI systems, imposing obligations such as conformity assessments, documentation, and post-market monitoring for high-risk applications commonly found in industrial contexts. In this setting, standards from ISO and IEC provide globally recognised technical frameworks, while CENELEC adapt and refines these into harmonised standards that can confer a presumption of conformity with EU law. This significantly reduces interpretative uncertainty for organisations, offering concrete guidance on implementing requirements such as risk management, transparency, human oversight and system robustness. At the same time, it creates a dynamic compliance environment: companies must continuously monitor updates across regulatory texts and evolving standards ecosystems. The combined effect is a shift toward compliance-by-design and standardisation-by-design, in which adherence to ethical principles, legal mandates, and harmonised technical standards becomes integral to industrial innovation, risk management, and competitive positioning.

In short, ethical principles from UNESCO, binding regulations such as the GDPR and the AI Act and technical standards from ISO, IEC, CEN and CENELEC together form an integrated governance ecosystem that shapes how AI systems are designed, deployed and monitored in industry. These frameworks translate high-level values into enforceable requirements and harmonised technical practices, reducing ambiguity while increasing accountability and lifecycle oversight.

For industry, this means that developing and implementing AI is no longer just a technical challenge but a continuous, compliance-driven practice, where ethical alignment, regulatory adherence and standards integration are essential pre-requisites for market access, trust and sustainable innovation.

4. Human-Centric Change Management

4.1 Worker Involvement, Acceptance, and Human-Centred Design and Tools

The transition toward Industry 5.0 represents a fundamental paradigm shift, moving beyond the pure automation of Industry 4.0 to prioritise the synergy between human creativity and machine efficiency. In this evolving landscape, the success of AI-enabled hardware and software across the industrial value chain is no longer determined solely by technical specifications, but by the degree of worker involvement and the intentionality of human-centred design. As the AIMS 5.0 project seeks to bolster European digital sovereignty, it recognises that sustainable production is impossible without the active participation and trust of the workforce. By integrating the organisational readiness frameworks proposed by Naim et al. (2024) and the stakeholder-specific insights of Ghafourian et al. (2025), this section examines the critical factors influencing technology adoption. Through the lens of the QETAM (Quality, Evaluation, and Technology Acceptance Model) and empirical data gathered from the AIMS 5.0 Questionnaire, we analyse how the alignment of workspace readiness, adaptive design tools, and worker perception creates a resilient ecosystem where AI serves as an augmentative partner rather than a disruptive force.

4.2 Socio-Technical Readiness and Workspace Assessment

The transition toward AI-integrated industrial environments necessitates a shift from viewing technology in isolation to a holistic socio-technical perspective. As established by Abu Naim et al. (2024), the current landscape of AI standards and guidelines is fragmented and rapidly evolving, creating barriers to entry for European industries. To address this, the AIMS 5.0 project utilises a self-assessment methodology that categorises organisations into three distinct personas: Early Adopters, Followers, and Independent Learners. Each persona exhibits unique readiness levels and implementation risks; while Early Adopters focus on high-level resource efficiency, Independent Learners, often SMEs, prioritise niche applications but struggle with a lack of ready-to-use AI frameworks. The necessity of such a tool is further validated by the findings of Ghafourian et al. (Ghafourian Y, 2025), where survey data revealed that approximately 49% of industrial stakeholders are uncertain whether their organisations follow formal AI standards. By providing a dynamic, LLM-powered self-assessment interface, AIMS 5.0 bridges this communication gap, ensuring that "Digital Sovereignty" is built upon a foundation of informed, compliant, and ready workforces.

4.3 Human-Centred Design (HCD) and Adaptive Tools

Human-Centred Design (HCD) within the AIMS5.0 framework is not merely an ergonomic consideration but a strategic requirement for achieving the Industry 5.0 goals of worker well-being and resilient production. Our analysis of the Questionnaire on Worker Involvement highlights that successful AI deployment is contingent on addressing both "people-related" challenges, such as managing unrealistic expectations and mitigating fears of loss of autonomy, and "data-related" risks. Empirical evidence from persona-based studies suggests that integrating end-users early in the development phase is a primary success factor, with 24% of project participants citing it as a necessary mitigation measure. Consequently, the AIMS5.0 use cases focus on developing adaptive tools that prioritise transparency and "Explainability" (XAI). These tools aim to reduce operators' cognitive load by providing intuitive interfaces and manual override capabilities, ensuring that AI-enabled hardware serves as a supportive "cognitive partner" rather than a disruptive force on the shop floor.

4.4 Impact of digital technologies on business task execution and user agency

An assessment of the impact of digital technologies on the execution of business tasks was conducted among the members of the AIMS5.0 project. Semi-structured interviews were conducted, and the choice of technology that the participants used in their daily work was not restricted to what was developed within the AIMS5.0 project. While the interviews are ongoing, findings from the interviews conducted so far (n=9) nevertheless provide interesting insights.

Through the interviews, we gathered information from the participants about their career, education, their current role and associated responsibilities, digital tools used by them to accomplish their work-related tasks, and their experience working with AI and associated impacts. The interviewees cover both applicants and developers of AI in manufacturing and their daily tasks related to project administration and management, research and development, coding, teaching, and stakeholder communication and dissemination. To gain an understanding of the participants understanding on AI, a classification of AI was presented initially. We classified AI into rule-based and learning-based AI. While among practitioners there was not much interest in how AI gets classified, among researchers there were some differences in terms of AI classification. Furthermore, except for some researchers and developers, almost everyone referred to the use of LLMs most of the time.

Most of the participants viewed the agentic systems (like LLMs) as helpful assistants. While the general opinion was that the use of AI makes task execution efficient, in terms of effectiveness, there were differing opinions. None of the participants felt a loss in autonomy while using the AI tools and they did not feel losing any relevant competences to the AI except regarding some tedious tasks. In terms of social embeddedness, the participants did not report any loss of it.

In human-agentic AI systems, some viewed the human and AI capabilities as complementary and stated that each could benefit from the other, reflecting the framework of Baird and Maruping (Baird & Maruping, 2021) where the authors say that delegation happens in human-agentic AI systems due to their differing and complementary capabilities. Furthermore, keeping humans in the loop was seen as important, especially for critical tasks with significant consequences. This was especially emphasized while discussing the use of LLMs as the participants reported that the LLMs could hallucinate and so the user must account for it and must have control over the final decisions. These inputs suggest that while participants acknowledge the usefulness of agentic AI systems, and the changing dynamics in interaction between humans and these systems, they see no loss in agency when it comes to decision making. Rather the agency is shared between humans and the agentic systems depending on the task needs and capability requirements.

4.5 Worker Involvement, Acceptance, and Human-Centred Design and Tools

User acceptance of the tools developed throughout AIMS 5.0 represents a significant challenge to the overall impact of the project and the European Sovereignty we aim to foster. Throughout the project's duration, we have focused extensively on exploring User Acceptance, as these insights directly influence future development cycles and the scholarly publication of our results.

Within the framework of this deliverable, a final questionnaire was developed to gather comprehensive feedback. This instrument is designed to collect insights into workers' involvement during the development and deployment phases, their acceptance and perceived usability of AI-enabled digital systems, and the degree to which the implemented solutions adhere to human-centred design principles. The data collected will facilitate an evaluation of

how these technologies affect daily workflows, safety, and the general user experience, ultimately driving the refinement of solutions developed within the AIMS 5.0 ecosystem.

Below is the questionnaire distributed to the AIMS 5.0 Use Cases:

1. What is your age group?

18-24

25-34

35-44

45-54

55-64

65 or older

2 .Which gender do you identify most with?

Male

Female

Non Binary

Prefer not to say

3. Where is your workplace located? *

Austria

Germany

France

Greece

Netherlands

Latvia

Poland

Sweden

Spain

Hungary

Italy

Turkey

4. What is your highest educational qualification?

Secondary

Education

College

Bachelor's degree

Master's degree

Ph.D. or higher

Prefer not to say

5. What is your professional or academic field?

ICT / Computer Science / AI

Manufacturing / Industrial Operations

Research / Academia / Innovation

Management / Business / Administration

Other (please specify)

6. In which AIMS 5.0 Use Case are you primarily involved?

Use Case 1: Development of Green AI-Lifecycle Management Approaches

Use Case 2: AI Enhanced Value Chain

Use Case 3: Servitization and service delivery enabled by Zero administration

Use Case 4: AI based improved connection between production and logistics

Use Case 5: AI techniques at different layers in Machine Tool Domain

Use Case 6: Reconfigurable AI-based automated conveyer feeding by a robotic system

Use Case 7: Digital Automated Luminaire Manufacturing platform

Use Case 8: AI supporting and strengthening human and manufacturing cycles

Use Case 9: Energy-aware Scheduling of Batch Processing Machines in Wafer Fabs Use

Use Case 10: MES concept based on AI Improving Production Cycle: AI-Prod_MES Use

Use Case 11: AI-supported Industrial IoT for Indoor Food Production

Use Case 12: Human-Centred AI for the Optimization of Robust and Competitive Semiconductor Manufacturing Networks

Use Case 13: Intelligent Sensors improving Robustness of Automated Wafer Transportation and Storage Systems

Use Case 14: Speeding-up the Benefit of Automation by Virtual Commissioning through Digital Twins

Use Case 15: Productionized Machine Learning models at scale

Use Case 16: AI enhanced production and performance of RF transceivers for civilian SATCOM and radar applications

Use Case 17: Intelligent Contamination Management & Prescriptive Control Plans and Process Flows

Use Case 18: Using semantic Data Model as base for AI optimized economic & ecologic/societal supply chain improvements

Use Case 19: Extension of Digital Reference Ontology for Smart Sustainable Business Planning and incident handling

Use Case 20: Harmonization of Automated Digitized Equipment Control (AuDi- EC)

4.5.1 Rating Questions on Social Influence, Trust, Usefulness, Ease of Use, and Facilitating Conditions

Rating Questions on Social Influence, Trust, Usefulness, Ease of Use, and Facilitating Conditions

Social Influence (SI)

The extent to which organizational culture, colleagues, and supervisors shape employees' attitudes and intentions regarding the adoption and use of AIMS 5.0 tools and solutions

7. To what extent do your colleagues or supervisors positively influence your decision to use the new AIMS 5.0 tools and solutions?



Perceived Trust (PT)

The degree to which employees perceive the AIMS 5.0 tools as reliable, consistent, and trustworthy in their operation

8. To what extent do you believe the AIMS 5.0 tools behave consistently and predictably?



Perceived Usefulness (PU)

The degree to which employees believe that using the AIMS 5.0 tools enhances their job performance and overall work efficiency

9. To what extent do the AIMS 5.0 tools help you work more efficiently?



Perceived Ease to Use (PEU)

The degree to which employees perceive the AIMS 5.0 tools as easy to learn, understand, and operate

10. How easy is it for you to understand and operate the AIMS 5.0 tools?



Facilitating Conditions (FC)

The extent to which the organization provides the necessary resources, support, and infrastructure that enable employees to effectively use the AIMS 5.0 tools and solutions.

11. To what extent do you feel you have the necessary support and resources to use the AIMS 5.0 tools effectively?



General Open-Ended Questions

12. What benefits do you expect from using the AIMS 5.0 tools in your daily work, and what challenges do you anticipate or currently face when using them?

13. How do you think employee involvement could be strengthened in the design or implementation of the AIMS 5.0 Use Cases?

14. What concerns, if any, do you have about how the introduction of AIMS 5.0 technologies may affect your job role or job security?

15. To what extent do you feel that the implemented AIMS 5.0 solutions follow human-centred design principles and tools?

In this study, we employ the Quality, Evaluation, and Technology Acceptance Model (QETAM), an established methodology for understanding the behavioural intention to use (BIU) and adopt new technologies. QETAM builds on the fundamental principles of the Technology Acceptance Model (TAM) and integrates the systematic evaluation approach of Failure Mode and Effects Analysis (FMEA). It relies on measures such as Perceived Trust (PT), Perceived Usefulness (PU), Perceived Ease of Use (PEU), Social Influence (SI), and Facilitating Conditions (FC). This comprehensive model enables a deeper analysis of the multiple factors influencing BIU toward technology adoption in digital transformation contexts.

This study uses QETAM to understand how potential future users accept and utilise AI-enabled technologies designed in AIMS 5.0. The findings aim to inform user-centred design solutions and contribute to the successful integration of such innovations into everyday life.

The results derived from the QETAM-based User Acceptance Platform provide a comprehensive overview of end-users' Behavioural Intention to Use (BIU) AI-enabled tools designed throughout AIMS 5.0 use cases.

To maintain a cohesive flow for the deliverable, here is the demographic analysis presented in the following paragraphs in Section 4.5.2.

4.5.2 Demographic Profile of Respondents

The demographic data collected from the 20 respondents provides a clear profile of the user group involved in the AIMS 5.0 exploratory evaluation. The survey reached a mature and experienced demographic, with the majority of participants falling within the **25-34 (38.9%)** and **45-54 (33.3%)** age brackets. When combined with the **35-44** age group, nearly **90%** of the participants are in the prime stages of their professional careers. This suggests that the feedback gathered is grounded in substantial workplace experience and a practical understanding of industrial requirements.

The participant base is highly specialized, ensuring that the evaluation of AI-enabled tools is technically informed. Nearly **45%** of respondents are active in **Research, Academia, and Innovation**, while **27.8%** originate from the **ICT, Computer Science, and AI** sectors. Crucially, **16.7%** of the pool represents **Manufacturing and Industrial Operations**, providing a vital "on-the-ground" perspective for the real-world deployment of the AIMS 5.0 solutions. This high level of expertise is further supported by the educational data, which shows that a combined **83.3%** of respondents hold either a **Ph.D. or higher (44.4%)** or a **Master's degree (38.9%)**.

In terms of geographic and gender distribution, the study reflects a strong Central European perspective. **Germany** accounts for the largest share of participants at **44.4%**, followed by a balanced distribution across **Austria, the Netherlands, and Latvia (each at 11.1%)**, with additional representation from Greece and France. Regarding gender, the respondent pool is predominantly **Male (72.2%)**, with **Females** representing **27.8%** of the sample. While this distribution is common within the ICT and manufacturing sectors, it highlights the specific professional lens through which the technology is being viewed.

4.5.3 Exploratory Evaluation and Results Analysis

This study employs an exploratory evaluation approach to understand how potential future users accept and utilize the AI-enabled technologies designed within AIMS 5.0. The findings are intended to inform user-centered design solutions and contribute to the successful integration of such innovations into daily operations. The results, derived from the QETAM-based User Acceptance Platform, provide a comprehensive overview of end-users' Behavioural Intention to Use (BIU) across the various AIMS 5.0 use cases.

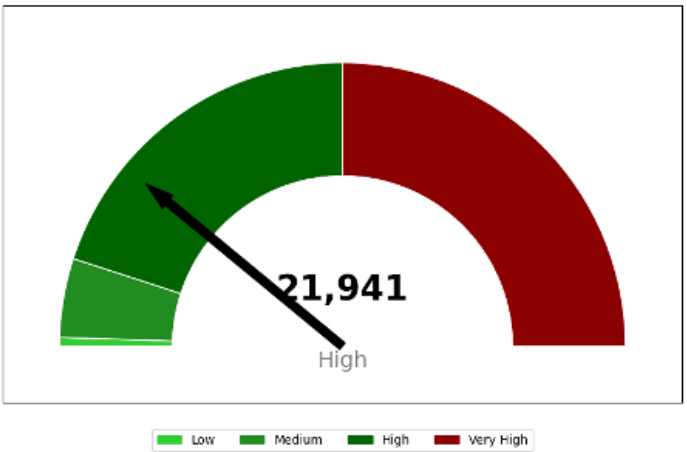

Figure 2 Overall BIU Acceptance Meter

The Overall BIU Acceptance Meter reveals a highly positive reception, with the acceptance level categorized as "High" based on a total score of 21,941. This suggests that the baseline for technology adoption is strong and that the AI-enabled tools are well-positioned for integration into the target environments. This high score is a critical indicator of the project's success in aligning its technical developments with the practical expectations and needs of the end-users.

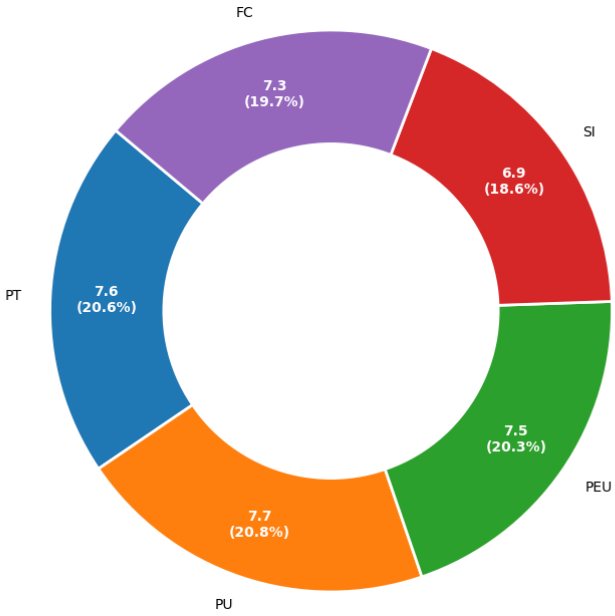

Figure 3 Average Score Distribution by Factor

An analysis of the Average Score Distribution by Factor shows a balanced performance across all QETAM dimensions, with scores ranging between 6.9 and 7.7. Perceived Usefulness (PU) emerged as the leading driver with a score of 7.7 (20.8%), closely followed by Perceived Trust (PT) at 7.6 (20.6%). These results indicate that users primarily value the functional benefits and reliability of the tools. While Social Influence (SI) recorded the lowest relative score at 6.9 (18.6%), it remains within a healthy range, suggesting that while the tools are personally valued, there is still room to enhance peer-to-peer advocacy and community-driven adoption.

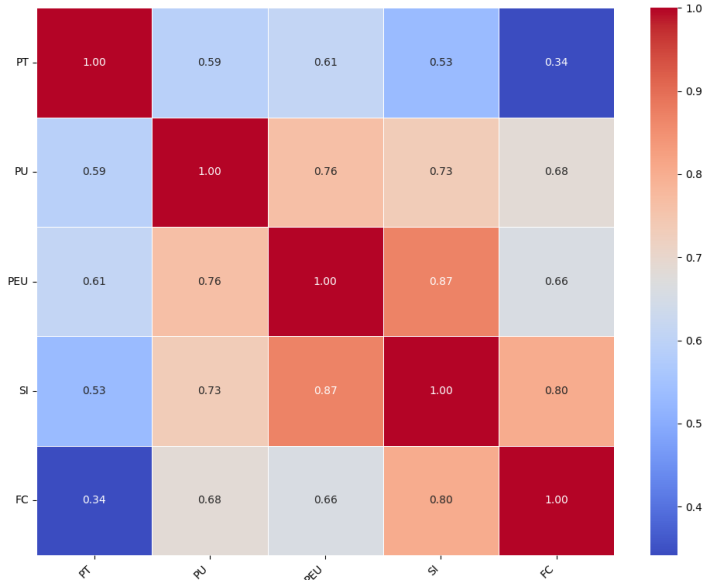

Figure 4 Factor Correlation Matrix

Finally, the Factor Correlation Matrix provides deeper insight into the dynamics of user acceptance. The most significant relationship was observed between Perceived Ease of Use (PEU) and Social Influence (SI), which share a very strong correlation of 0.87. This implies that as the technology becomes more intuitive and user-friendly, the positive influence of social and peer groups on adoption increases exponentially. Conversely, the weakest correlation was found between Perceived Trust (PT) and Facilitating Conditions (FC) at 0.34, suggesting that user trust is rooted in the inherent performance of the AI solutions rather than the external support structures provided. Together, these findings validate the human-centered design approach of AIMS 5.0 and provide a clear roadmap for future refinements

4.5.4 Results of Open-Ended Questions

The qualitative analysis of open-ended responses provides essential depth to the quantitative data, offering a nuanced view of participants' expectations, barriers, and design philosophy. When asked about anticipated benefits and challenges, respondents expressed a clear vision of how AIMS 5.0 tools could transform daily operations. The primary expectations center on significant improvements in operational efficiency, enhanced safety, and improved ergonomics. For example, in material handling contexts, users highlighted that AI-driven traffic management could optimise workflows and minimise downtime. The integration of advanced perception systems, such as radar and stereo cameras, was specifically noted for its potential to prevent collisions and reduce both cognitive and physical workload for workers. However, these benefits are tempered by concerns regarding initial technical complexity and the potential friction involved in human-machine coordination, particularly in high-noise factory environments.

Regarding strengthening employee involvement, a recurring theme in the feedback is the desire for a more participatory, continuous design process. Participants suggested that involvement should move beyond one-time surveys toward a continuous loop of engagement, such as "feedback sprints" in which workers can test prototypes in real time and see their suggestions implemented in subsequent iterations. There is a clear call for transparency concerning the logic behind specific AI behaviours. Respondents emphasised that for involvement to be meaningful, a bridge must be built between technical developers and shop-floor operators to ensure that the tools are perceived as helpful assistants rather than complex replacements.

In terms of job security and the evolution of professional roles, the feedback was notably optimistic. Rather than fearing total displacement, participants viewed the introduction of AI and autonomous systems through the lens of role evolution. Most respondents anticipate a shift toward higher-value tasks, such as system monitoring and the management of human-robot interactions. The primary concern raised was not the existence of the job itself, but the transition phase. This highlights a critical need for robust upskilling programs and clear communication strategies to help workers navigate the skills gap as repetitive manual tasks are replaced by autonomous mobile robots (AMRs) and AI-lifecycle management tools.

Finally, the assessment of Human-Centered Design (HCD) principles shows that the current solutions are highly valued for their focus on physical safety and reliable human detection. This adherence to HCD is likely the primary driver behind the high "Perceived Trust" scores noted in the quantitative analysis. However, the feedback also indicates room for improvement in consistent user experience (UX) and interaction clarity. Respondents suggested that the AI’s "intent" must be made more transparent to human workers and that systems should be refined to handle "edge cases" or unexpected environmental changes with less manual intervention. Collectively, these insights confirm that AIMS 5.0 is successfully aligning with the Industry 5.0 vision by treating users as expert collaborators in the pursuit of European Sovereignty and industrial efficiency.

4.6 Gaps and Challenges in Worker-Centric Digital Transformation

The widespread integration of Artificial Intelligence (AI) into information systems (IS) represents a transformative shift, but it also reveals significant gaps in our traditional understanding of technology adoption. While classical models such as TAM and UTAUT have served us well in the era of rule-based logic, they are often insufficient to capture the unique characteristics of AI, namely its opacity, unpredictability, and dynamic nature.

The following analysis is derived from a synthesis of current state-of-the-art literature and empirical insights gained from previous research initiatives (most notably the European project FACTS4WORKERS). While the current project's use cases adhered to a targeted risk-based approach focusing on technical robustness and immediate safety, this chapter addresses broader socio-technical gaps that often emerge during long-term AI integration. These insights are intended as a high-level recommendation for identifying challenges in future use-cases, particularly those that may fall outside the scope of structured assessment tools like the TOE framework or the ALTAI checklist.

4.6.1 Reconceptualising Acceptance and Trust in the Age of AI

The fundamental challenge in modernising technology acceptance is that we are moving from a world of "traceable tools" to a world of "black boxes".

4.6.1.1 The Black-Box Problem and the Verification Gap

For decades, information systems were based on predictable logic; internal workings were clear, and output was a direct result of programming. AI, particularly deep learning models, operates in ways that are often too complex for human comprehension.

- **The Transparency Gap**: Unlike traditional systems with linear logic, AI can make decisions without clear reasoning.
- **The Verifiability Challenge**: This opacity directly challenges classical notions of "System Quality" and "Information Quality" where reliability is typically assumed to be verifiable.
- **The Creators' Dilemma**: Even the creators of these algorithms may not fully understand the complex internal structures formed during the learning process.

4.6.1.2 Trust as a Prerequisite, Not a Consequence

In traditional frameworks, trust was often an implicit byproduct of a system's perceived ease of use and usefulness. If a system worked well, the user eventually trusted it.

- **The Structural Gap**: With AI, trust must be a core construct and a critical prerequisite for adoption.
- **The Initial Acceptance Barrier**: Because of AI's inherent opacity, a user must have a pre-existing level of trust in the algorithm's logic—even if unseen—before they can even begin to evaluate its usefulness.
- **The Concept of Growth**: Perceived Usefulness must now be expanded to include the belief that AI expands individual capabilities (growth) rather than just performing a fixed function.

4.6.1.3 Redefining IS Success

The core principles of IS success remain relevant, but their measurement must be expanded to address AI-specific challenges, as presented in Table 2.

Table 2 Gaps in Classical Trust and Acceptance Models

Classical IS Success Construct	The AI Challenge / Gap
Information Quality	Shifts from data accuracy to the trustworthiness of a dynamic, learning model's output.
System Quality	Moves beyond functionality to encompass technical robustness, security, and resilience against unforeseen inputs.
User Satisfaction	Now tied to "non-deskilling" and the feeling that the AI augments rather than replaces one's capabilities.

4.6.2 The Psychology of Trust and Bias: Implementation Hurdles

Psychology plays a far more central role in AI adoption than in previous technological shifts, yet several gaps exist in how we manage these psychological levers.

4.6.2.1 The Interpretability Gap

There is a common misconception that technical transparency (Explainable AI or XAI) is the primary driver of trust.

- **Outcome vs. Process**: Research shows that user trust is more driven by **outcome feedback** (the practical result) than by technical interpretability.
- **Practicality over Logic**: Users generally prioritise the question "Is this useful?" over "Can I explain this?".
- **The Expertise Divide**: A significant challenge exists in how information is delivered: experts require deep logic to satisfy their scepticism, while non-experts often find complex explanations confusing and prefer simple feedback.

4.6.2.2 The Bias Mitigation Paradox

Designers face a "Goldilocks" problem where the remedy for one human bias can exacerbate another.

- **Automation Bias**: This refers to an over-reliance on AI, or a "blind preference" for automated responses.
- **Algorithm Aversion**: This is the distrust of automated systems even when they are more accurate than human judgement.
- **The Paradox**: To reduce automation bias, developers might highlight errors. However, this often increases algorithm aversion, as users quickly lose confidence in a machine when it visibly errors.

4.6.2.3 Anthropomorphism and Accountability

- **The Fragility of Human-like AI**: Systems perceived as human-like (anthropomorphism) are more likely to be trusted initially, but this trust is fragile.
- **The Accountability Deficit**: A key challenge is ensuring clear responsibility for AI actions.
- **Moral Outrage Gap**: People tend to show less moral outrage toward algorithmic discrimination compared to human discrimination, which can lead to poor long-term evaluation of the organization.

4.6.3 The Impact on Work Environment: Empowerment or Control?

The way AI is implemented determines whether it fosters a state of "superagency" or leads to user rejection.

4.6.3.1 The Paternalism-Assistance Continuum

A major gap in worker-centric transformation is the risk of **AI Paternalism**.

- **The Paternalism Threat**: This occurs when an AI system makes decisions or takes actions without user input or consent.
- **Loss of Agency**: Paternalistic designs prioritize values that may not align with the user, leading to a feeling of lost control.
- **The Failure Path**: Paternalism is not just a flaw; it is a reliable path to project failure (**Lessons from Healthcare Failures**)
- **The Value of Autonomy**: User autonomy has a real, measurable impact on the success of an information system.

4.6.3.2 Preserving Job Dimensions

To avoid the pitfalls of paternalism, AI must preserve or enhance key job dimensions, as discussed in Table 3.

Table 3 Potential negative impacts of AI paternalism

Dimension	Potential Negative Impact (The Challenge)
User Autonomy	Algorithmic management reducing control over tasks; feeling of reduced agency.
Task Variety	Feeling of reduced meaning if the "engaging" or creative heart of the job is automated.
Perceived Competence	Over-reliance on AI potentially eroding critical thinking skills and domain expertise.

4.6.4 Technical and Ethical Prerequisites: Moving Beyond the Checklist

The final gap is the divide between a system being **trustworthy** (technical/ethical) and being **trusted** (psychological/social).

4.6.4.1 The Trustworthiness-Trust Gap

- **The Developer's Responsibility**: Trustworthiness is a technical concept—the system must be lawful, ethical, and robust.
- **The User's Belief**: Trust is a broader concept influenced by social dynamics and work processes beyond the developer's direct control.
- **Contextual Clashes**: A system can be technically perfect but fail if it erodes user autonomy or clashes with established social norms.

4.6.4.2 Best Practices and Implementation Barriers

Translating ethics into action involves several critical hurdles:

- **Interpretability by Design**: A major challenge is moving away from "tacked on" post-hoc explanations toward building interpretability into the model's architecture from the start.
- **The Fairness Burden**: Ensuring diverse data sets to reduce bias at the source is technically and logistically intensive.
- **The Accountability Gap**: Establishing traceability, ensuring decisions can be traced back to inputs, is increasingly difficult as models scale.
- **Human Oversight Maintenance**: Building continuous human review loops is essential to ensure AI remains an assistant, but this is often sidelined for the sake of efficiency.

5. Validation of Consolidated Guidelines and Recommendations

In this section, we present the results of our empirical study evaluating the self-assessment chatbot introduced and described in D7.2 - Guidelines incl. checklist and self-assessment of how to apply AI in the European industry. The chatbot is an AI-driven tool that helps AI developers and users assess whether their applications align with relevant ethical guidelines, legal regulations, and industry standards. Built on a Retrieval-Augmented Generation (RAG) approach and deployed on cloud infrastructure, it delivers real-time, context-aware support, helping users check their AI systems against the regulatory, standardisation, and ethics landscape in an efficient manner.

The design of this validation study deliberately builds on the preceding chapters of this deliverable. The curated knowledge base underpinning the chatbot mirrors the regulatory, standardisation, and compliance landscape mapped in Section 2 - the EU AI Act, CRA, GDPR, the CEN/CENELEC and ISO/IEC AI standards, and the sector-specific regulations (Machinery Directive, UNECE, MDR). The evaluation logic operationalises the Integrated AI Assurance Framework introduced in Section 3, in particular its TOE-based risk-identification layer and its dual-branch assessment combining ALTAI-based trustworthiness with UTAUT and DeLone & McLean for digital-workspace success. Finally, the choice of a legal-expert-in-the-loop comparison protocol reflects the human-centric change-management principles developed in Section 4 - worker involvement, human-centred design, and the trust and interpretability requirements that any AI compliance-support tool must meet to be accepted in real digital workspaces.

We evaluate our LLM-based chatbot, which is grounded in a curated set of uploaded regulations, standards, and ethical guidelines rather than unrestricted internet access. The system is expected to read a natural-language AI use case description and identify which governance instruments (e.g., regulations, standards, and guidelines) are applicable. To assess and validate this capability, we selected four representative use cases, prepared their corresponding descriptions and asked legal experts within the AIMS5.0 project to independently determine which instruments apply to each use case. We then compare the chatbot's outputs to the experts' assessments.

We first list our representative use cases from AIMS5.0. These use cases 2, 11, 12, 13, and 14. We selected these use cases so that they can represent most of the AIMS5.0 use cases. We performed a qualitative clustering of all AIMS5.0 use cases based on (i) the primary role of AI in the socio-technical system (enterprise/platform governance, planning and scheduling decision support, shopfloor automation, equipment/process health and resilience, and digital infrastructure such as MES/MLOps/digital twins), and (ii) the dominant stakeholder group (e.g., planners/decision-makers, shopfloor operators, engineers) and governance surface (e.g., safety, cybersecurity, sector-specific regulation). The 20 AIMS5.0 use cases (UC1–UC20) are grouped into five clusters based on (i) the primary role of AI in the socio-technical system and (ii) the dominant stakeholder group and governance surface. The four use cases selected for chatbot validation in Section 5.1 (UC11–UC14), together with UC2, span all five clusters, making the validation sample representative of the full AIMS5.0 portfolio. The clusters are as follows in Table 4.

Table 4 AIMS5.0 Use Case Clusters

Cluster	What it captures (high level)	Use cases (representatives)
A. **Value chain, sustainability, enterprise semantics**	Lifecycle/value-chain perspectives, servitization, semantic models/ontologies	**UC1** Green AI-Lifecycle Management; **UC2** AI Enhanced Value Chain; **UC3** Servitization / Zero administration; **UC18** Semantic data model for supply chain; **UC19** Digital Reference Ontology extension

B. Planning, scheduling, and human-in-the-loop decision support	Decision support, planning/scheduling optimization, human-centered workflows	**UC4** Production-logistics connection; **UC8** Strengthening human & manufacturing cycles; **UC9** Energy-aware scheduling (wafer fabs); **UC12** Human-centred AI for semiconductor manufacturing networks
C. Shopfloor automation and robotics execution	Physical-world automation: robotics, feeding/handling, industrial IoT production setups	**UC6** Robotic conveyer feeding; **UC7** Automated luminaire manufacturing; **UC11** Industrial IoT for indoor food production
D. Equipment/process robustness, quality, and control	Robustness, monitoring, contamination/quality, equipment control harmonization	**UC5** AI in machine tool domain; **UC13** Robust wafer transport/storage sensors; **UC16** RF transceivers production/performance; **UC17** Contamination management & prescriptive control; **UC20** Digitized equipment control harmonization
E. Digital infrastructure, digital twins, and ML-at-scale	Backbone stuff: MES, digital twins/virtual commissioning, productionized ML	**UC10** AI-based MES concept; **UC14** Virtual commissioning via digital twins; **UC15** Productionized ML models at scale

We then selected UC2 & UC11–UC14 as a stratified set of representatives, each mapping to a distinct cluster: UC2 (AI enhanced value chain, Cluster A), UC11 (shopfloor IIoT in food production with explicit safety/compliance considerations- Cluster C), UC12 (human-centred decision support for semiconductor manufacturing networks, Cluster B), UC13 (robustness and monitoring of automated wafer transport/storage systems, Cluster D), and UC14 (virtual commissioning via digital twins, Cluster E). This design ensures diversity across domain, system-integration level, and compliance-relevant characteristics while keeping the evaluation scope tractable.

We have asked each use-case team to create a use case description that serves as evaluation stimuli. Each description is written in natural language and includes the core elements needed to assess applicability (e.g., application purpose, stakeholders, data types, deployment context, level of autonomy, and potential impact).

Here is the main question in the checklist in this regard that we have distributed among the use case teams: *"Please reflect on the following question and write a paragraph describing your current or future AI use-case or application. Include a high-level technical explanation of how the AI will function in your use-case. Additionally, address the ethical considerations outlined in the checklist below as part of your paragraph. These use-case stories will be used to help fine-tune and later test our self-assessment tool, which is being developed to improve ethical and legal compliance in AI projects."*

In the following subsections, we present the use case descriptions and the evaluation results per use case. For each case, we provide (i) the chatbot's output in the form of the most relevant documents retrieved from its knowledge base in response to the use case description, and (ii) the reference assessment produced by our legal expert team, which lists the standards and regulations that should be considered for that use case. The table is structured to enable a direct comparison between these two perspectives: it shows which expert-identified instruments are covered by the chatbot's retrieved documents, where the chatbot suggests additional instruments not identified by experts, and where expert-relevant instruments are missing from the chatbot output. An output of the self-assessment Chatbot, as an example for UC 13, is shown in Figure 5 and Figure 6 and (arbitrarily selected).

Based on the use case description provided and the knowledge base, the use case directly affects wafer fab operators, maintenance teams, AI and ML developers, and sensor manufacturers. Indirectly affected parties include production managers, end customers, and regulatory bodies ⓘ .
The regulations that this use case should consider are as follows:
Data Privacy and Security Regulations: Compliance with GDPR (General Data Protection Regulation) is essential if personal or sensitive data is involved in the anomaly detection process. The EU AI Act also emphasizes data protection and ethical AI usage ⓘ .
Wireless Communication Standards: Since data is transmitted via WiFi, adherence to wireless communication regulations is necessary to ensure secure and efficient data transmission ⓘ .
AI and Machine Learning Regulations: The EU AI Act outlines requirements for AI systems, especially in high-risk scenarios, ensuring transparency, ethical use, autonomy, adaptiveness, and inference capability ⓘ .
Health and Safety Standards: Ensuring that preventive maintenance solutions do not compromise worker safety within wafer fabs is critical. Adherence to health and safety regulations is essential ⓘ .
Environmental Regulations: Consideration of environmental impacts due to the deployment of new technologies like sensors should comply with relevant environmental regulations ⓘ .

Figure 5 An example of the Chatbot Outcomes for UC13 (Part 1)

Additional Regulations to Consider:
EU Machinery Directive: This directive applies to machinery used in industrial settings, including transportation systems in wafer fabs. It focuses on ensuring the safety of machinery, which would be relevant for the shuttles used in wafer fabs.
ISO Standards:
ISO 9001: Quality management systems standard that ensures consistent product quality and operational efficiency.
ISO 27001: Information security management standard that could apply due to data handling and storage involved in AI-driven anomaly detection.
IEC Standards:
IEC 61508: Functional safety of electrical/electronic/programmable electronic safety-related systems, which may be relevant for systems using sensors and AI for anomaly detection.
Cybersecurity Regulations: Given the use of WiFi for data transmission, compliance with cybersecurity standards such as NIST Cybersecurity Framework or ISO/IEC 27032 might be necessary to protect against data breaches and ensure secure communication.
Environmental Health and Safety (EHS) Regulations: These regulations ensure that industrial operations do not adversely affect employee health or the environment, which is pertinent when deploying new technologies like sensors.
Semiconductor Equipment and Materials International (SEMI) Standards: Specific standards set by SEMI for semiconductor manufacturing equipment could be applicable to ensure compliance with industry-specific requirements.
These additional regulations complement the previously identified ones, ensuring comprehensive compliance across various aspects of operations involving AI and machine learning in wafer fab transportation systems.

Figure 6 An example of the Chatbot Outcomes for UC13 (Part 2)

5.1 Comparative Validation: AI Chatbot Outcomes vs. Legal Expert Assessment

5.1.1 Use Case 2: Servitization and service delivery enabled by Zero administration

Use Case description: "*The aim of Use Case 2 (a) is to develop a method to detect defect-prone paths in manufacturing processes. These paths are identified by comparing an event log containing the cases related to defective products with an event log containing the cases related to non-defective products using the comparative process mining algorithm. This algorithm finds the paths which are composed of event attributes included in the*

event log, that are significantly more frequent for defective products than for non-defective products. By analysing these paths, potential causes of the defects can be identified, such as faulty equipment or misaligned machines, which provide process engineers with a more focused scope for their investigation. The output of the analysis is translated into actionable insights to instruct the operators to implement the corrective measures, increasing productivity (faster reaction time to implement corrective measure, less down time, less non-conformities) and reducing waste (from non-conformities).

The aim of Use Case 2 (b) is to improve the movements of robots used for product testing to make these movements more human like that they can be representative for the personal care routine of our customers (e.g. hair removal). Robot testing is currently performed for IPL (Intense Pulsed Light) devices. However, this robot testing is not human like enough, making it necessary to validate the results by (expensive) test panels. By improving robot movements, this validation step can be skipped, speeding both development time and market release of IPL devices. Additionally, a larger impact can be achieved when extending this approach to other products in the portfolio, optimizing testing protocols and data collection."

Evaluation Results (Table 5):

Table 5 UC02 Comparing Chatbot Results and Expert Inputs

#	Instrument	Legal experts	Chatbot	Alignment
Regulations				
1	EU AI Act	✓	✓	Match
2	GDPR (General Data Protection Regulation)	✓	–	Expert only
3	EU Directive 89/391/EEC (workplace health and safety)	✓	–	Expert only
4	UNESCO Recommendation on the Ethics of AI	–	✓	Chatbot only
Standards				
5	ISO 9001	✓	–	Expert only
6	ISO 22400	✓	–	Expert only
7	IEC 62264	✓	–	Expert only
8	ISO/IEC 27001	✓	–	Expert only
9	ISO/IEC 23894	✓	–	Expert only
10	IEEE 1849	✓	–	Expert only
11	ISO/IEC 62443	✓	–	Expert only
12	ISO/IEC TR 24028	–	✓	Chatbot only
Total congruences: 1 (regulations: 1; standards: 0)				

5.1.2 Use Case 11: AI-supported Industrial IoT for Indoor Food Production

Use Case description: "*This use case focuses on applying AI and machine learning to cultivate edible plants (e.g., chilies, lettuce) in a controlled growing environment. The target setting is an aquaponic vertical farm, where environmental conditions are continuously maintained within plant-specific ranges. Key parameters, such as water pH, CO2 concentration, humidity,*

and temperature, are measured via a network of sensors. Actuators (e.g., pumps, fans, and lighting controllers) then adjust the environment to keep these variables at their desired setpoints. AI supports autonomous operation through a closed-loop control architecture based on the MAPE-K model (Monitor–Analyze–Plan–Execute over a shared Knowledge base). The system monitors real-time sensor streams and analyses deviations from target ranges. The acceptable ranges for each plant type are derived from agronomic guidelines that are extracted and structured using Natural Language Processing (NLP). Based on the current state and these plant-specific requirements, the system plans corrective actions (e.g., increasing aeration, adjusting water circulation, modifying ventilation) and executes them via the available actuators, continuously repeating the cycle to stabilize and optimize growing conditions. To ensure resilient and trustworthy operation, integrated cybersecurity capabilities systematically analyse system assets, data flows, and control interactions to identify, assess, and investigate potential cyber-attacks and vulnerabilities, thereby establishing a structured and robust foundation for mitigating cyber risks“.

Evaluation Results (Table 6):

Table 6 UC11 Comparing Chatbot Results and Expert Inputs

#	Instrument	Legal experts	Chatbot	Alignment
Regulations				
1	EU AI Act	✓	✓	Match
2	EU Machinery Regulation (2023/1230)	✓	–	Expert only
3	EU NIS2 Directive (if critical infrastructure / large-scale plant)	✓	–	Expert only
4	EU Food Safety Law Regulation (EC) 178/2002	✓	–	Expert only
5	EU Water Framework Directive (2000/60/EC) (if large-scale plant)	✓	–	Expert only
6	HACCP Principles	✓	–	Expert only
7	EFSA Guidelines	✓	–	Expert only
8	GDPR (General Data Protection Regulation)	–	✓	Chatbot only
9	Digital Services Act	–	✓	Chatbot only
10	Data Act	–	✓	Chatbot only
11	AI Risk Management Framework	–	✓	Chatbot only
12	UNESCO Recommendation on the Ethics of AI	–	✓	Chatbot only
13	Regulation on the Free Flow of Non-Personal Data	–	✓	Chatbot only
14	Cyber Resilience Act (cybersecurity for products with digital elements)	–	✓	Chatbot only
Standards				
15	IEC 61508	✓	✓	Match
16	IEC 61511	✓	–	Expert only
17	ISO 13849	✓	–	Expert only
18	ISO/IEC 27001	✓	–	Expert only
19	IEC 62443	✓	–	Expert only
20	ISO/IEC 22989	✓	–	Expert only
21	ISO/IEC 23053	✓	–	Expert only
22	ISO/IEC 23894	✓	–	Expert only

#	Instrument	Legal experts	Chatbot	Alignment
23	ISO/IEC 42001	✓	–	Expert only
24	ISO 9001	✓	–	Expert only
25	ISO 30401	✓	–	Expert only
26	ISO 22000 (food safety management)	–	✓	Chatbot only
27	ISO 26262	–	✓	Chatbot only
Total congruences: 2 (regulations: 1; standards: 1)				

5.1.3 Use Case 12: Human-Centred AI for the Optimization of Robust and Competitive Semiconductor Manufacturing Networks

Use Case Description: "*Train regression-based ML models on historic data to be able to predict work-in-progress in front of work centers in advance and take respective countermeasures. Validation and development of an integrated transportation and production fab model. Implementation of digital processes to production lines to increase automation level. Synchronization and harmonization of wafer tests across different locations*"

Evaluation Results (Table 7):

Table 7 UC12 Comparing Chatbot Results and Expert Inputs

#	Instrument	Legal experts	Chatbot	Alignment
Regulations				
1	EU AI Act	✓	✓	Match
2	EU NIS2 Directive (if fab is essential/important entity)	✓	–	Expert only
3	EU Machinery Regulation (2023/1230)	✓	–	Expert only
4	EU Electromagnetic Compatibility Directive 2014/30/EU	✓	–	Expert only
5	ALCOA+ Principles	✓	–	Expert only
6	EU Ethics Guidelines for Trustworthy AI	–	✓	Chatbot only
7	UNESCO Recommendation on AI Ethics	–	✓	Chatbot only
8	GDPR (General Data Protection Regulation)	–	✓	Chatbot only
Standards				
9	ISO 9001	✓	✓	Partial match *
10	SEMI E10 / E79 / E90 / E142	✓	✓	Partial match *
11	IEC 62443	✓	✓	Partial match *
12	IEC 62061	✓	✓	Partial match *
13	IEC 62264 / ISA-95	✓	✓	Partial match *
14	IEC 61508	✓	✓	Partial match *
15	ISO/IEC 23894	✓	–	Expert only
16	ISO/IEC 23053	✓	–	Expert only
17	ISO/IEC 25012	✓	–	Expert only
18	ISO/IEC 5259 (series)	✓	–	Expert only
19	ISO 22400	✓	–	Expert only
20	ISO 19439 / 19440	✓	–	Expert only
21	ISO 15745	✓	–	Expert only

#	Instrument	Legal experts	Chatbot	Alignment
22	ISO/IEC 27001	✓	–	Expert only
23	ISO 10218	✓	–	Expert only
24	ISO/IEC 17025	✓	–	Expert only
25	ISO 8000	✓	–	Expert only
26	OSHA standards (workplace safety)	–	✓	Chatbot only
27	ISO/IEC TR 24028:2020	–	✓	Chatbot only
Total congruences: 7 (regulations: 1; standards: 6)				

5.1.4 Use Case 13: Intelligent Sensors improving Robustness of Automated Wafer Transportation and Storage Systems

Use Case Description: "*The main focus of this use case is the use of AI and machine learning methods to detect anomalies in a highly complex overhead transportation system in 300 mm wafer fabs. The SoA is based on prevented maintenance solutions for hundreds of shuttles which transports the production lots from tool to tool. If the any components in the shuttles e.g. belts or gear wheels are broken or damaged the production losses are dramatically. Therefore, the idea is to save time and money for preventive maintenance and use sensors, in this use case silicon microphones and ultrasonic sensors to detect the defects in the shuttles by anomaly detection e.g. via voice or visual control.*

The data are sent via WIFI from the edge device to the host and all data are collected and stored. Feature extraction of the audio signals is performed and the anomaly detection algorithms are already implemented and tested using unsupervised learning. The first predictive maintenance model (wear and tear model) was established. For UC13 only sound files or optical files for OHT shuttles are recorded. No personal or private data are created or stored. The use case will help to improve the work of the maintenance people and save a lot of time for much more value adding jobs. This use case offers a high scalability for many other tasks in the field of maintenance where predictive maintenance application could help to reduce the amount of motion task"

Evaluation Results (Table 8):

Table 8 UC13 Comparing Chatbot Results and Expert Inputs

#	Instrument	Legal experts	Chatbot	Alignment
Regulations				
1	EU AI Act	✓	✓	Match
2	EU Machinery Directive 2006/42/EC / EU Machinery Regulation (2023/1230)	✓	✓	Partial match †
3	GDPR (General Data Protection Regulation)	–	✓	Chatbot only
4	NIST Cybersecurity Framework	–	✓	Chatbot only
Standards				
5	IEC 61508	✓	✓	Match
6	ISO 10218-1 / ISO 10218-2	✓	–	Expert only
7	ISO/TS 15066	✓	–	Expert only
8	IEC 62061	✓	–	Expert only
9	ISO 13849-1 / -2	✓	–	Expert only

#	Instrument	Legal experts	Chatbot	Alignment
10	ISO 12100	✓	–	Expert only
11	ISO/IEC 22989	✓	–	Expert only
12	ISO/IEC TR 24028	✓	–	Expert only
13	ISO/IEC TR 24368	✓	–	Expert only
14	ISO/IEC 23053	✓	–	Expert only
15	ISO/IEC 42001	✓	–	Expert only
16	IEC 62443	✓	–	Expert only
17	NIST SP 800-series (e.g., SP 800-53, NIST AI risk guidance)	✓	–	Expert only
18	ISO 9001	–	✓	Chatbot only
19	ISO 27001	–	✓	Chatbot only
20	ISO/IEC 27032	–	✓	Chatbot only
Total congruences: 3 (regulations: 2; standards: 1)				

5.1.5 Use Case 14: Speeding up the Benefit of Automation by Virtual Commissioning through Digital Twins

Use Case Description: "*Our AI application focuses on predicting travel times for a fleet of mobile, free-ranging transport robots operating in an industrial environment. Accurate travel time predictions are essential for optimizing overall fleet performance, since key questions such as "When will a robot reach its destination?" and "Could there be potential conflicts between robots?" depend on this information. However, these predictions involve significant uncertainty due to factors such as interactions between robots and the influence of specific layout characteristics. Our solution estimates travel times by analysing the current system status, including the layout, locations, and planned destinations of all vehicles. To achieve this, we use historical robot log files (timestamps and locations) collected from various layouts as our data foundation. We would like to know if there are key safety standards for deploying AI in industrial automation and robotics*"

Evaluation Results (Table 9):

Table 9 UC14 Comparing Chatbot Results and Expert Inputs

#	Instrument	Legal experts	Chatbot	Alignment
Regulations				
1	EU AI Act	✓	✓	Match
2	EU Machinery Directive 2006/42/EC and CE marking	✓	–	Expert only
3	ETSI (relevant standards)	✓	–	Expert only
4	GDPR (General Data Protection Regulation)	✓	–	Expert only
5	UNESCO Recommendations on AI Ethics	–	✓	Chatbot only
Standards				
6	IEC 61508	✓	✓	Match ‡
7	ISO/IEC TR 24028:2020	✓	✓	Match
8	ISO 12100	✓	–	Expert only
9	IEC 62061	✓	–	Expert only

#	Instrument	Legal experts	Chatbot	Alignment
10	ISO 13849-1 / -2	✓	–	Expert only
11	ISO 10218 family / ISO/TS 15066	✓	–	Expert only
12	ISO/IEC 22989	✓	–	Expert only
13	ISO/IEC 42001	✓	–	Expert only
14	IEC 62443	✓	–	Expert only
15	ISO/IEC 27001	✓	–	Expert only
16	IEEE 802.11	✓	–	Expert only
17	OSHA Standards	–	✓	Chatbot only
Total congruences: 3 (regulations: 1; standards: 2)				

5.2 Chatbot Output for all the Use Cases

As it was observed before, the expert-based comparison provides an important reason to place confidence in the chatbot's results. Across the representative use cases, the chatbot retrieved documents that broadly aligned with the types of instruments identified by the legal expert team, especially for cross-cutting regulations, ethical guidance, and standards that were available in its knowledge base. This does not mean that the chatbot replaces legal or compliance review, but it does indicate that the system can serve as a reliable and practical support tool for identifying a substantial share of the governance instruments relevant to a given AI use case. As a result, we progressed with documenting the chatbot's output across all use cases.

The following table (Table 10) presents the output of the self-assessment chatbot across the AIMS5.0 use cases, providing a consolidated overview of the regulations, standards, and ethical guidance documents retrieved when each use case is described in natural language. The table should be understood as a structured first-pass compliance aid that helps map industrial AI scenarios to potentially applicable governance documents. The consolidated table therefore serves two purposes. First, it documents the chatbot's outputs across the wider set of AIMS5.0 use cases and highlights recurring governance patterns, such as the frequent appearance of the EU AI Act, cybersecurity-oriented frameworks, and trustworthiness-related standards. Second, when interpreted in light of the expert validation, it offers evidence that the chatbot's recommendations are sufficiently grounded to support early-stage compliance orientation.

Table 10 Chatbot Outcomes Across All AIMS5.0 UCs

Use Case 1	**Regulations**	• EU AI Act • AI Risk Management Framework (AI RMF 1.0) • Data Act • UNESCO Recommendation on the Ethics of Artificial Intelligence
	Standards	• ISO 50001:2018 - Energy Management Systems • ISO/IEC TR 24028:2020 • Overview of Trustworthiness in Artificial Intelligence
Use Case 2	**Regulations**	• UNESCO Recommendation on the Ethics of Artificial Intelligence • EU AI Act
	Standards	• ISO/IEC 62443 Standards • ISO/IEC TR 24028
Use Case 3	**Regulations**	• Data Act • EU AI Act • UNESCO Recommendation on the Ethics of Artificial Intelligence

		• Artificial Intelligence Risk Management Framework (AI RMF 1.0)
	Standards	• ISO/IEC TR 24028:2020
Use Case 4	**Regulations**	• EU AI Act • Cyber Resilience Act • Universal Guidelines for Artificial Intelligence • Artificial Intelligence Risk Management Framework (AI RMF 1.0)
	Standards	• ISO/IEC TR 24028 • IEC 62443
Use Case 5	**Regulations**	• Cyber Resilience Act • EU AI Act
	Standards	• IEC 62443 Industrial Security Standards • ISO/IEC TR 24028:2020 - Trustworthiness in AI Systems
Use Case 6	**Regulations**	• EU AI Act • Data Act • UNESCO Recommendations on Ethics of Artificial Intelligence
	Standards	• ISO 18646-2:2019 • ISO 26262 and IEC 61508 • TÜV SÜD IEC 62443 Industrial Security Standards
Use Case 7	**Regulations**	• Cyber Resilience Act • EU AI Act • Universal Guidelines for Artificial Intelligence
	Standards	• ISO/IEC TR 24028 • TÜV SÜD IEC 62443 Industrial Security Standards:
Use Case 8	**Regulations**	• EU AI Act • Universal Guidelines for Artificial Intelligence • Guidelines on AI System Definition
	Standards	• ISO/IEC TR 24028
Use Case 9	**Regulations**	• Artificial Intelligence Risk Management Framework (AI RMF 1.0)
	Standards	• ISO 50001:2018 - Energy Management Systems • TUV-SUD-IEC 62443 Industrial Security Standards • ISO/IEC TR 24028:2020 - Information Technology
Use Case 10	**Regulations**	• Artificial Intelligence Risk Management Framework (AI RMF 1.0) • Universal Guidelines for Artificial Intelligence • EU AI Act
	Standards	• IEC 62443 Industrial Security Standards • ISO 50001
Use Case 11	**Regulations**	• UNESCO Recommendation on the Ethics of Artificial Intelligence • EU AI Act; GDPR (General Data Protection Regulation); • Digital Services Act; • Data Act; • AI Risk Management Framework; • UNESCO Recommendation on the Ethics of Artificial Intelligence; • Regulation on the Free Flow of Non-Personal Data • Cybersecurity requirements for products with digital elements and amending Regulation
	Standards	• ISO 22000 for food safety management • ISO 26262 • IEC 61508
	Regulations	• EU AI Act EU • Ethics Guidelines for Trustworthy AI

Use Case 12		• UNESCO Recommendation on AI Ethics • GDPR (General Data Protection Regulation)
	Standards	• ISO standards related to manufacturing and quality management (e.g., ISO 9001) • SEMI standards for semiconductor manufacturing • OSHA (Occupational Safety and Health Administration) standards for workplace safety • IEC standards for electrical equipment and automation systems • ISO/IEC TR 24028:2020
Use Case 13	**Regulations**	• GDPR (General Data Protection Regulation) • EU AI Act • NIST Cybersecurity Framework
	Standards	• ISO 9001 • ISO 27001I • EC 61508 • ISO/IEC 27032
Use Case 14	**Regulations**	• General Data Protection Regulation (GDPR) • EU AI Act • UNESCO Recommendations on AI Ethics
	Standards	• Occupational Safety and Health Administration (OSHA) Standards • ISO/IEC TR 24028:2020 • ISO 61508
Use Case 15	**Regulations**	• EU AI Act • Artificial Intelligence Risk Management Framework (AI RMF 1.0) • Universal Guidelines for Artificial Intelligence • Cyber Resilience Act
	Standards	• ISO/IEC TR 24028:2020
Use Case 16	**Regulations**	• EU AI Act • Cyber Resilience Act • Guidelines on the Definition of an Artificial Intelligence System • Artificial Intelligence Risk Management Framework (AI RMF 1.0) • Universal Guidelines for Artificial Intelligence
	Standards	
Use Case 17	**Regulations**	• Cybersecurity Requirements for Products with Digital Elements • EU AI Act
	Standards	• IEC 62443 Industrial Security Standards • ISO/IEC TR 24028:2020 - Trustworthiness in Artificial Intelligence
Use Case 18	**Regulations**	• EU AI Act • Cyber Resilience Act • Universal Guidelines for Artificial Intelligence
	Standards	• IEC 62443 Industrial Security Standards
Use Case 19	**Regulations**	• AI Risk Management Framework (AI RMF 1.0) • EU AI Act • General Data Protection Regulation (GDPR) • Supply Chain Security Regulations
	Standards	• ISO 27001 • The ISO/IEC TR 24368:2022 • ISO 14001

Use Case 20	**Regulations**	• Cyber Resilience Act
	Standards	• IEC 62443-2-4 • IEC 62443-4-2 • IEC 62443-4-1 • IEC 62443-3-3

6. Industrial Insights and Use Case Feedback

In this section, we collected the feedback of all AIMS5.0 Use Cases through a dialogue with the corresponding representatives.

As stated earlier in this document, the industry faces the fact that developing and implementing AI is no longer just a technical challenge but a continuous, compliance-driven discipline, where ethical alignment, regulatory adherence, and standards integration are essential prerequisites for market access, trust and sustainable innovation. Through this study and the development of the chatbot, we intend to help bridge this challenge and enable smooth AI development and implementation thereof.

In general, and when looking into the chatbot output summarised in Table 10 for all AIMS5.0 Use Cases, the chatbot response sufficiently aligns with experts' feedback, taking into consideration the border limitations such as the availability of documents to be input to the chatbot. It provides a clear and structured overview of the relevant regulatory and ethical landscape, and it is helpful in identifying the applicability of the General Data Protection Regulation and the AI Act to the specific AIMS5.0 use case. When applicable, the classification of the system as potentially high-risk is particularly valuable, as it highlights early on the need for compliance measures such as documentation, risk assessment, and human oversight. The inclusion of ethical guidance from UNESCO and references to standards bodies like ISO, IEC, CEN and CENELEC also supports a comprehensive understanding of the broader governance framework.

However, from an implementation perspective, the guidance remains somewhat high-level and requires further operationalisation. For example, a more concrete direction is needed on how to integrate the applicable regulations, standards and ethical guidelines into the specific application. Identifying specific risks and providing actionable metrics and validation approaches would help to properly assess such risks and be able to mitigate them effectively.

Overall, the chatbot output is a strong starting point for understanding compliance obligations and ethical considerations, but should be complemented with domain-specific guidance, internal policies and possible expert review to ensure full and efficient implementation. The chatbot is equipped to further improve and expand in such directions.

7. Conclusions

7.1 Integration of Framework, Standards, and Change Management

This deliverable has presented a comprehensive examination of the regulatory, standardization, and ethical landscape relevant to change management in AI-enabled digital workspaces, grounded in an integrated AI Assurance Framework and supported by empirical validation.

The regulatory and standardisation landscape is mapped in Section 2 demonstrates the increasing convergence of EU legislation (AI Act, GDPR), CEN/CENELEC and ISO/IEC standards, and sector-specific requirements (Machinery Directive, UNECE, Medical Device Directive) into a layered governance model. This complexity underscores a critical finding: organisations cannot rely on a single compliance checklist; instead, they must develop adaptive processes capable of identifying applicable instruments within their specific technological, organisational, and environmental context.

In Section 3, the AI Assurance Framework is proposed as a solution to investigate risks and translate them into solutions through integrated models. By combining TOE-based risk identification, ALTAI trustworthiness requirements for AI-developed tools, and digital workspace evaluation through UTAUT and DeLone & McLean models, the framework creates a coherent pathway from risk identification to measurable outcomes across five dimensions: risk mitigation, trustworthy AI, user acceptance, operational success, and regulatory alignment.

Section 4's examination of human-centric change management, including socio-technical readiness, worker involvement, and the psychological factors underlying AI adoption, reveals that technical trustworthiness alone is insufficient. Organisations must simultaneously address the gaps and challenges in worker-centric digital transformation, ensuring that AI implementation fosters worker empowerment rather than perceived control, and that human-centered design is embedded at every stage of development and deployment.

7.2 Chatbot Validation: Practical Insights and Limitations

To assess the practical effectiveness of the proposed framework and support consolidated guidelines, Section 5 presented an empirical validation study of an AI-driven chatbot tool. This validation process suggests that the self-assessment chatbot can function as a practical first-pass compliance tool: given a natural-language use case description, it retrieves and surfaces governance-relevant resources from its curated knowledge base in a way that broadly aligns with the types of instruments identified by the legal expert team.

Across the selected representative use cases (UC2, UC11–UC14), the chatbot returned documents that map well to expert expectations, especially for cross-cutting requirements (e.g., widely applicable regulatory and ethics-oriented instruments) and for standards and guidelines that are actually present in its corpus.

Importantly, these results should be interpreted in light of the chatbot's deliberately constrained knowledge setting. The system does not have unrestricted access to external repositories or the full universe of standards, many of which are paywalled or otherwise not practically ingestible at scale. This simulates a real use case of the chatbot being deployed in a company with access to universal and public regulations, as well as confidential regulatory documents and internal standards. Despite these limitations, the chatbot was still able to retrieve a share of the relevant resources within the boundaries of what it has access to, and

its retrieved evidence frequently overlapped with the instruments identified by experts. This indicates that, when the knowledge base contains the necessary sources, the RAG-based approach can effectively locate and prioritise them based on a use case description.

At the same time, remaining mismatches between chatbot outputs and expert lists point to predictable boundaries of the approach. Differences can stem from (i) knowledge-base coverage gaps (instruments missing due to access constraints such as paywalls), (ii) granularity and naming mismatches (the same requirement referenced at different levels of detail), and (iii) underspecified use case descriptions, where applicability depends on contextual factors not stated explicitly. These findings reinforce that the tool is best positioned as decision support, not as a replacement for legal review.

7.3 Recommendations for Change Management in Digital Workspaces

Based on the integrated findings from the regulatory landscape analysis, the AI Assurance Framework, human-centric change management examination, and chatbot validation, the following recommendations are proposed for organisations implementing AI in digital workspaces:

1. **Adopt a Structured, Context-Aware Compliance Approach:** Organisations should use a TOE-informed methodology to identify applicable regulations, standards, and ethical guidelines within their specific technological, organisational, and environmental context, rather than attempting to comply with a universal checklist.
2. **Integrate ALTAI Trustworthiness Requirements into AI Development:** The seven ALTAI requirements (Human Agency and Oversight, Technical Robustness and Safety, Privacy and Data Governance, Transparency, Diversity and Non-Discrimination, Societal and Environmental Well-Being, and Accountability) should be embedded as mandatory design and governance criteria throughout the AI system lifecycle, not treated as post-hoc compliance checks.
3. **Prioritise Human-Centric Change Management:** Change management strategies must address both worker acceptance and perceived control. Organisations should invest in socio-technical readiness assessments, engage workers early in development phases, and design adaptive tools that prioritise transparency and explainability (XAI) to foster worker empowerment rather than resistance.
4. **Evaluate System Success Beyond Trustworthiness:** Use the DeLone and McLean framework to assess whether AI-enabled digital workspaces achieve expected quality, user satisfaction, and net benefits in operational contexts. This evaluation should be conducted in parallel with trustworthiness assessment to ensure that technical controls do not undermine practical success or user adoption.
5. **Leverage Governance Support Tools with Appropriate Guardrails:** Organisations can benefit from AI-driven tools (such as the validated chatbot) to support early-stage compliance orientation and governance mapping, provided that such tools are treated as decision support mechanisms, not replacements for expert legal review. Internal knowledge bases should be curated to reflect both public regulations and organisational-specific standards.
6. **Establish Continuous Monitoring and Adaptive Governance:** Given the rapid evolution of AI standards and regulations, organisations should implement processes for continuous monitoring of applicable governance instruments and periodic reassessment of digital workspace compliance, rather than treating compliance as a one-time activity.

8. References

Ádám Kővári, Y. G. (2025). Let's Have a Chat with the EU AI Act. *NOMS 2025-2025 IEEE Network Operations and Management Symposium*. Retrieved from https://ieeexplore.ieee.org/document/11073655/authors#authors.

Baird, A., & Maruping, L. M. (2021). The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts. *MIS Quarterly*. Retrieved from https://misq.umn.edu/misq/article-abstract/45/1/315/1855/The-Next-Generation-of-Research-on-IS-Use-A?redirectedFrom=fulltext.

Belal Abu Naim, Y. G. (2024). A Self-assessment Tool to Encourage the Uptake of Artificial Intelligence in Digital Workspaces. *NOMS 2024-2024 IEEE Network Operations and Management Symposium*. Retrieved from https://ieeexplore.ieee.org/document/10575125/authors#authors.

Commission, E. (2020). *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment*. Retrieved from Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-Assessment: https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment.

Commission, E. (2023). *Regulation (EU) 2023/1230 of the European Parliament and of the Council of 14 June 2023 on machinery and repealing Directive 2006/42/EC of the European Parliament and of the Council and Council Directive 73/361/EEC*. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02023R1230-20230629.

DeLone WH, M. E. (2003). The DeLone and McLean model of information systems success: a ten-year update. *Journal of management information systems*.

European Commision. (2016). *Legal framework of EU data protection*. Retrieved from European Commision: https://commission.europa.eu/law/law-topic/data-protection/legal-framework-eu-data-protection_en.

European Commission (EC) - CRA Act. (2024). *Regulations (EU) No 168/2013 and (EU) No 2019/1020 and Directive (EU) 2020/1828 (Cyber Resilience Act)*. Official Journal of the European Union.

GDPR EU. (2026). *What is GDPR, the EU's new data protection law?* Retrieved from GDPR EU: https://gdpr.eu/what-is-gdpr/.

Ghafourian Y, L. F. (2025). Understanding Stakeholders of Industrial AI: Insights from a Persona-Based Questionnaire. *InNOMS 2025-2025 IEEE Network Operations and Management Symposium*.

Jere JN, N. N. (2020). organisation and environment framework analysis of information and communication technology adoption by small and medium enterprises in Pietermaritzburg. *South African Journal of Information Management*.

Jeyaraj, A. (2020). DeLone & McLean models of information system success: Critical meta-review and research directions. *International Journal of Information Management*.

Maija BREQUE, L. D. (2021). *Industry 5.0 - Towards a sustainable, human-centric and resilient European industry*. Retrieved from https://op.europa.eu/en/publication-detail/-/publication/468a892a-5097-11eb-b59f-01aa75ed71a1/language-en.

Marikyan M, P. P. (2021). Unified theory of acceptance and use of technology.

Services, N. D. (2025). *NTT DATA – Gear up for change: The EU Machinery Regulation*. Retrieved from https://services.global.ntt/en-us/insights/blog/gear-up-for-change-the-eu-machinery-regulation-is-coming: https://services.global.ntt/en-us/insights/blog/gear-up-for-change-the-eu-machinery-regulation-is-coming?utm_source=chatgpt.com

UNECE. (2021). *UN Regulation No. 155 on Cybersecurity and Cybersecurity Management System for Vehicles*. https://unece.org/sites/default/files/2023-02/R155e%20%282%29.pdf.

Venkatesh V, M. M. (2003). acceptance of information technology: Toward a unified view1.

Wikipedia. (2026). *Wikipedia*. Retrieved from Unified theory of acceptance and use of technology: https://en.wikipedia.org/wiki/Unified_theory_of_acceptance_and_use_of_technology

Yasin Ghafourian, M. T. (2024, October). Engineering Secure, Trustworthy, and Ethically Sound AI-Based Computer Systems. *ERCIM News*. Retrieved from https://ercim-news.ercim.eu/en139/r-i/engineering-secure-trustworthy-and-ethically-sound-ai-based-computer-systems.

9. Appendices

9.1 Appendix A – Abbreviations

Table 10 Abbreviations

Abbreviation	Meaning
AI	Artificial Intelligence
AIMS5.0	Artificial Intelligence in Manufacturing leading to Sustainability and Industry5.0
ALTAI	Assessment List for Trustworthy Artificial Intelligence
ASI	Austrian Standards Institute
AWI	Approved Work Item
CD	Committee Draft
CEN/CENELEC	European Committee for Standardization / European Committee for Electrotechnical Standardization
DeLone & McLean	DeLone and McLean Information Systems Success Model
DPIA	Data Protection Impact Assessment
DPO	Data Protection Officer
EU	European Union
GDPR	General Data Protection Regulation
HCD	Human-Centred Design
II	Individual Impact
IIoT	Industrial Internet of Things
IoT	Internet of Things
IQ	Information Quality
IS	Information Systems
ISS	Information System Success
IU	Intention to Use
ISO/IEC	International Organization for Standardization / International Electrotechnical Commission
ISO/SAE	International Organization for Standardization / Society of Automotive Engineers
JTC	Joint Technical Committee
LLM	Large Language Model
MES	Manufacturing Execution System
MLOps	Machine Learning Operations
NB	Net Benefits
OI	Organisational Impact
PWI	Preliminary Work Item
QETAM	Quality, Evaluation, and Technology Acceptance Model

RAG	Retrieval-Augmented Generation
R155	UNECE Regulation No. 155 – Cybersecurity and Cybersecurity Management System for Vehicles
R156	UNECE Regulation No. 156 – Software updates for vehicles
RQ	Service Quality (Request Quality)
SME	Small and Medium-sized Enterprise
SQ	System Quality
SU	System Usage
TAM	Technology Acceptance Model
TOE	Technology-Organization-Environment
TR	Technical Report
TS	Technical Specification
UC	Use Case
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNECE	United Nations Economic Commission for Europe
US	User Satisfaction
UTAUT	Unified Theory of Acceptance and Use of Technology
XAI	Explainability (Explainable AI)

9.2 Appendix B: List of ISO/IEC SC42 AI-Standards

Standards and projects under SC 42 (2026-04-20): AI (36+50) and Big Data (5)

Artificial Intelligence (36) published (RED: published 2026, violet: published 2025)

- ISO/IEC 4213:2022 Information technology — Artificial Intelligence — Assessment of machine learning classification performance.
- ISO/IEC 5259-1:2024 Artificial Intelligence — Data quality for analytics and ML — Part 1: Overview, terminology, and examples.
- ISO/IEC 5259-2:2024 Artificial Intelligence — Data quality for analytics and ML — Part 2: Data quality measures.
- ISO/IEC 5259-3:2024 Artificial Intelligence — Data quality for analytics and ML — Part 3: Data quality management requirements and guidelines.
- ISO/IEC 5259-4:2024 Artificial Intelligence — Data quality for analytics and ML — Part 4: Data quality process framework.
- **ISO/IEC 5259-5:2025** Artificial intelligence — Data quality for analytics and machine learning (ML) — Part 5: Data quality governance framework.
- ISO/IEC 5338:2023 Information technology — Artificial intelligence — AI system life cycle processes.
- ISO/IEC 5339:2024 Information Technology — Artificial Intelligence — Guidelines for AI applications.
- ISO/IEC 5392:2024 Information technology — Artificial intelligence — Reference architecture of knowledge engineering.
- ISO/IEC TR 5469:2024 Artificial intelligence — Functional safety and AI systems

- **ISO/IEC TS 6254:2025** Information technology — Artificial intelligence — Objectives and approaches for explainability of ML models and AI systems.
- ISO/IEC 8183:2023 Information technology — Artificial intelligence — Data life cycle framework.
- ISO/IEC TS 8200:2024 Information technology — Artificial intelligence — Controllability of automated artificial intelligence systems.
- ISO/IEC TS 12791:2024 Information technology — Artificial intelligence — Treatment of unwanted bias in classification and regression machine learning tasks.
- **ISO/IEC 12792:2025** Information technology — Artificial intelligence — Transparency taxonomy of AI systems.
- ISO/IEC TR 17903:2024 Information technology — Artificial intelligence — Overview of machine learning computing devices.
- **ISO/IEC TR 20226:2025** Information technology — Artificial intelligence — Environmental sustainability aspects of AI systems.
- **ISO/IEC TR 21221:2025** Information technology - Artificial intelligence - Beneficial AI systems.
- ISO/IEC 22989:2022 Information technology - Artificial intelligence — Artificial Intelligence Concepts and Terminology.
- ISO/IEC 23053:2022 Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML).
- ISO/IEC 23894:2023 Information technology — Artificial Intelligence — Guidance on Risk Management.
- ISO/IEC TR 24027:2021 Information technology — Artificial Intelligence (AI) — Bias in AI systems and AI aided decision making.
- ISO/IEC TR 24028:2020 Information technology — Artificial Intelligence — Overview of trustworthiness in artificial intelligence.
- ISO/IEC TR 24029-1:2021 Artificial Intelligence — Assessment of the robustness of neural networks — Part 1: Overview.
- ISO/IEC 24029-2: 2023 Artificial Intelligence — Assessment of the robustness of neural networks — Part 2: Methodology for the use of formal methods.
- ISO/IEC TR 24030:2024 Information technology — Artificial Intelligence — Use cases (**Withdrawn version 2021).**
- ISO/IEC TR 24368:2022 Information technology — Artificial Intelligence (AI) — Overview of ethical and societal concerns.
- ISO/IEC TR 24372:2021 Information technology — Artificial Intelligence — Overview of computational approaches for AI systems.
- ISO/IEC 24668:2022 Information technology — Artificial intelligence — Process management framework for big data analytics.
- ISO/IEC TS 25058:2024 Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Guidance for quality evaluation of artificial intelligence (AI) systems.
- ISO/IEC 25059:2023 Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI-based systems.
- ISO/IEC 38507:2022 Information technology — Governance of IT — Governance implications of the use of artificial intelligence by organizations.
- ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system.
- **ISO/IEC 42005:2025** Information technology — Artificial intelligence — AI system impact assessment.
- **ISO/IEC 42006:2025** Information technology — Artificial intelligence — Requirements for bodies providing audit and certification of artificial intelligence management systems.

- **ISO/IEC TS 42119-2: 2025** Artificial intelligence —Testing of AI – Part 2: Overview of testing AI systems.

Other responsibility:

- ISO/IEC AWI TS 29119-11:2020- Software and systems engineering — Software testing — Part 11: Guidelines on the testing of AI-based systems (under responsibility of ISO/IEC JTC1 SC 7).

Big Data (5) published, plus one under ISO/IEC JTC1 SC27 (Cybersecurity)

- ISO/IEC 20546:2019 Information technology - Big data - Overview and vocabulary
- ISO/IEC 20547-1:2020 Information technology — Big data reference architecture — Part 1: Framework and application process.
- ISO/IEC TR 20547-2:2018 Information technology — Big data reference architecture — Part 2: Use cases and derived requirements
- ISO/IEC 20547-3:2020 Information technology — Big data reference architecture — Part 3: Reference architecture.
- ISO/IEC 20547-4:2020 Information technology — Big data reference architecture — Part 4: security and privacy (**developed** by **ISO/IEC JTC1 SC27,** IT security, cybersecurity and privacy protection).
- ISO/IEC TR 20547-5:2018 Information technology — Big data reference architecture — Part 5: Standards roadmap.

Under Development (47) (change of status included: DTS, DTR, DIS, FDIS)

Currently under publication: Change of status (to CD, DIS, etc. marked too in colour!)

Note: DTS, DTR, FDIS expected to be published soon! (RED: new 2026; violet new 2025)

- ISO/IEC CD 4213 Artificial intelligence – Performance measurement for AI classification, regression, clustering and recommendation tasks.
- ISO/IEC **TR** 5259-6 Data quality for analytics and machine learning (ML) — Part 6: Visualization framework for data quality.
- ~~ISO/IEC AWI TS 17847: Information technology — Artificial intelligence — Verification and validation analysis of AI systems~~ - replaced by 42119-3.
- ISO/IEC **CD** TR 18988: Artificial intelligence — Application of AI technologies in health informatics.
- ISO/IEC **CD** TS 22440-1 – Artificial intelligence — Functional safety and AI systems — Part 1: Requirements.
- ISO/IEC **CD** TS 22440-2 – Artificial intelligence — Functional safety and AI systems — Part 2: Guidance.
- **ISO/IEC CD TS 22440-3** - Artificial intelligence — Functional safety and AI systems — Part 3: Examples of application.
- ISO/IEC CD TS 22443 Information technology — Artificial intelligence— Guidance on addressing societal concerns and ethical considerations.
- ISO/IEC AWI 22989-2 - Artificial intelligence —Concepts and terminology – Part 2: Healthcare.
- ISO/IEC 22989:2022/DAmd1 - Information technology – Artificial intelligence — Artificial Intelligence Concepts and Terminology – Amendment 1.
- **ISO/IEC 22989:2022/AWI Amd2** - Information technology – Artificial intelligence — Artificial Intelligence Concepts and Terminology – Amendment 2.
- ISO/IEC 23053:2022/DAmd1 - Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML) – Amendment 1.

- **ISO/IEC 23053:2022/AWI Amd2** - Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML) – Amendment 2.
- ISO/IEC CD TR 23281 – Artificial intelligence — Overview on AI tasks and functionalities related to natural language processing.
- ISO/IEC CD 23282 - Artificial intelligence — Evaluation methods for accurate natural language processing systems.
- ISO/IEC **DIS** 24029-3 Artificial intelligence (AI) — Assessment of the robustness of neural networks — Part 3: Methodology for the use of statistical methods.
- **ISO/IEC AWI TR 24030** Information technology — Artificial intelligence (AI) — Use cases
- **ISO/AWI TR 24492** Standardization needs for the application of artificial intelligence in the financial services industry.
- ISO/IEC DIS 24970 Artificial intelligence – AI system logging.
- ISO/IEC **DIS** 25029 Artificial intelligence – AI-enhanced nudging.
- **ISO/IEC AWI 25058** Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Measurement and guidance for quality evaluation of AI systems.
- ISO/IEC DIS 25059 Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality models for AI-based systems.
- ISO/IEC AWI TS 25223 Information technology – Artificial intelligence – Guidance and requirements for uncertainty quantification in AI systems.
- ISO/IEC **CD** TS 25258 Information technology – Artificial intelligence – Hybrid AI inference framework for AI systems.
- ISO/IEC AWI TR 25523 Information technology – Artificial intelligence – Overview of data profiles for analytics and ML.
- **ISO/IEC AWI TS 25566** Terminology and concepts for domain engineering of AI systems
- **ISO/IEC WD TS25568 -** Information technology — Artificial Intelligence — Guidance on addressing risks in generative AI systems.
- **ISO/IEC AWI TS 25569** - Artificial Intelligence — Implementation guidance on de-identification of data used in Machine Learning (ML).
- **ISO/IEC AWI TS 25570** - Information Technology — Artificial Intelligence — Reliability assessment of AI systems.
- **ISO/IEC AWI TS 25571** - Artificial Intelligence — Example template for documenting ethical issues of an AI system.
- **ISO/IEC AWI 25589** Information technology — Artificial intelligence — Framework for human-machine teaming.
- **ISO/IEC AWI 25590** - Information technology — Artificial intelligence — Guidance for output data quality of generative AI applications.
- **ISO/IEC AWI 25623** Artificial intelligence — Machine learning (ML) model description framework.
- ~~**ISO/IEC AWI 25651** - Information technology — Artificial intelligence — Guidance on the implementation of ISO/IEC 42001 (AI management system)~~ - **replaced by ISO/IEC AWI 42003.**
- **ISO/IEC AWI 25704** Artificial Intelligence — Process assessment model for AI system life cycle processes.
- **ISO/IEC AWI TS 25864** Information technology – Artificial Intelligence (AI) – Resilience assessment of AI systems.
- **ISO/IEC AWI 25870** Artificial intelligence — Reporting framework for AI incidents.
- ISO/IEC AWI 25872-1 Artificial intelligence — Knowledge enhancement for pretrained machine learning models — Part 1: Framework.
- **ISO/IEC AWI TS 26312** Information technology — Artificial intelligence — Identification and treatment of bias in AI by healthcare organizations.

- **ISO/IEC AWI TS 26320** Information technology — Artificial intelligence — Corpus development and maintenance for natural language processing systems (NEW April 2026).
- **ISO/IEC AWI 42003** - Information technology — Artificial intelligence — Guidance on the implementation of ISO/IEC 42001 (AI management system).
- ISO/IEC **CD** 42007 Information technology — Artificial intelligence — High-level framework and guidance for the development of conformity assessment schemes for AI systems.
- ISO/IEC **DIS** 42102 - Information technology — Artificial intelligence — Taxonomy of AI system methods and capabilities.
- ISO/IEC **CD** TR 42103 Information technology — Artificial intelligence — Overview of synthetic data in the context of AI systems.
- ISO/IEC DIS 42105 - Information technology — AI— Guidance for human oversight of AI system.
- ISO/IEC **TR** 42106 Information technology — Artificial intelligence — Overview of differentiated benchmarking of AI system quality characteristic.
- ISO/IEC **CD** TR 42109 Information technology — Artificial intelligence —Use cases of human-machine teaming.
- ISO/IEC **CD** TS 42111 Information technology — Artificial intelligence —Guidance on light-weight AI systems.
- ISO/IEC DTS 42112 Information technology — Artificial intelligence —Guidance on machine-learning model training efficiency optimization.
- ISO/IEC DTS 42119-3.**2** Artificial intelligence —Testing of AI - Part 3: Verification and validation analysis of AI systems.
- **ISO/IEC AWI TS 42119-7** Artificial intelligence — Testing of AI — Part 7: Red teaming.
- **ISO/IEC AWI TS 42119-8** Artificial intelligence — Testing of AI — Part 8: Quality assessment of prompt-based text-to-text systems that utilize generative AI.

NEW PWI (Preliminary Work ITEM): "Governance and management of human oversight of AI systems**"**

NEW PWI: ISO/IEC PWI 25880 "Information technology — Artificial intelligence — Guidance for organizational implementation of human-machine teaming concept"

NEW NWIP: Artificial Intelligence - Framework for structured multi-agent interaction in AI systems (MAS) (April 20, 2026)