Ref. Ares(2024)3146579 - 29/04/2024

Artificial Intelligence in Manufacturing leading to Sustainability and Industry5.0

Project Acronym:

AIMS5.0

Grant agreement no: 101112089

Deliverable no. and title	**D7.1 – Compilation of methods and sources evaluated for the guidelines for all industrial AI use-cases**	
Work package	WP7	Human Acceptance, Trust, and Ethics for Digital Workspaces
Task	T7.1 T7.2	Personas, Roles and Acceptance Dialogue with the society about ethics and legal norms
Subtasks involved	Not applicable	
Lead contractor	Infineon Technologies AG Thomas Gutt, mailto:thomas.gutt@infineon.com	
Deliverable responsible	PCL, Olga Kattan, olga.kattan@philips.com RSA, Markus Tauber, Yasin Ghafourian {markus.tauber,yasin.ghafourian}@researchstudio.at	
Version number	v1.0	
Date	29/04/2024	
Status	Final	
Dissemination level	Public	

Copyright: AIMS5.0 Project Consortium, 2024

Authors

Participant no.	Part. short name	Author name	Chapter(s)
	PCL	Olga Kattan	1-6
	RSA	Yasin Ghafourian	1-6
	RSA	Markus Tauber	1-6
	RSA	Belal Abu Naim	1-6
	HSZG	Fabian Lindner	1-6
	AIT	Erwin Schoitsch	2,3,5
	IFD	Germar Schneider	
	WUW	Gerald Reiner	3.1.1
	TIAG & TCAG	Anna Ryabokon	2,3
	TIAG & TCAG	Francesca Flamigni	2,3,4
	HUA	Konstantina Karathanasopoulou	2, 6
	HUA	George Dimitrakopoulos	2, 6
	SMRT	Erhan Tüfenkçi	
	TUBITAK	Gamze Büşra KAYA	
	BME	Pal Varga	
	BME	Mihály Héder	
	BME	Anikó Grad-Gyenge	

Document History

Version	Date	Author name	Reason
v0.1	06.02.2024	Olga Kattan	Initial Template
V0.1.1	07.03.2024	Olga Kattan Yasin Ghafourian Markus Tauber Belal Abu Naim Fabian Lindner Erwin Schoitsch Gerald Reiner Germar Schneider Anna Ryabokan Francesca Flamigni Konstantina Karathanasopoulou George Dimitrakopoulos	Incorporating content from paper to deliverable
v0.2	07.03.2024	Yasin Ghafourian, Olga Kattan	Input from the paper
v0.2	25.03.2024	Fabian Lindner	Update to Sections 3.1.1 and 6
V0.2.1	25.03.2024	Markus Tauber	Review and structure
v0.3	25.03.2024	Yasin Ghafourian	Latest inputs
v.04	29.03.2024		Changes accepted
v0.5	04.04.2024	Yasin Ghafourian	The initial Table from Survey in T7.2 was added
v0.6	10.04.2024	Olga Kattan	Review v0.5: Related tasks, minor formatting changes, addition of structure of the report.
V0.7	10.04.2024	Olga Kattan	Final Version before review
V0.8	11.04.2024	Yasin Ghafourian	Integrating the contributions from BME
V0.9	25.04.2024	Yasin Ghafourian	Feedback from reviewers applied
V1.0	26.04.2024	Yasin Ghafourian, Olga Kattan, Markus Tauber	Final Internal Revision

Executive Summary

To encourage the uptake of AI in industrial use cases, tools are required to support the engineering process throughout the life cycle of the AI application that is central to the use cases. Providing guidance on using AI in industrial setups is vital for creating trustworthy, reliable, and ethically compliant AI-based solutions. The related standardization landscape and available guideline repositories are large, scattered over the web, change rapidly, and are hard to keep up with. This limits the access and ease of use of these standards and guidelines. To address these limitations, we propose developing a self-assessment tool, empowered through AI algorithms and models such as Large Language Models, to improve the process of accessing and benefiting from those standards and guidelines.

This self-assessment tool will support various user groups while engineering their applications by identifying the most applicable guidelines according to the individual attributes of the specific user group. We argue that modeling specific attributes and mapping appropriate controls for self-assessments could be achieved by applying AI-based technologies. In this deliverable our ongoing efforts concerning the suggested supporting tools are outlined which offer a human-centric methodology. Additionally, we present initial results demonstrating how the needs of a particular user group can be accurately modeled. The results of this study are used for applications deploying AI in an industrial setting, enabling the two most important goals of Industry 5.0.

This deliverable is grounded in a peer-reviewed accepted paper [1].

Index Terms: Self-assessment Tool, AI Uptake, Industry 5.0, Personas, Artificial, human-centered workplaces

[1] Belal Abu-Naim et.al, A Self-assessment Tool to Encourage the Uptake of Artificial Intelligence in Digital Workspaces. Workshop on Management for Industry 5.0 (MFI5.0) during the IFIP/IEEE NOMS 2024 - Network Operations and Management 2024 Symposium (https://noms2024.ieee-noms.org/) [Accepted]

Table of Contents

1. Introduction 7
1.1 Objective and scope of the document 7
1.2 Structure of the deliverable report 8
2. Evaluation of Existing Guidelines and Relevant State of The Art 9
2.1 The Landscape of existing Guidelines 9
2.2 The Landscape of Standards and Regulations 10
2.3 Awareness and status of implementation of existing guidelines, regulations and standards for AIMS5.0 industrial use-cases 13
2.4 State of the art of relevant technologies and methods 16
2.5 Discussion 17
3. Self-assessment tool 19
3.1 Motivation 19
3.2 Steps towards the development of the self-assessment tool 19
3.2.1 Personas 20
3.2.2 Application of Standards, regulatory and guideline landscape 23
3.2.3 Large language models (LLM) 24
3.2.4 The Self-Assessment tool 25
3.3 Backcasting exercise (and strategic planning) in the Self-Assessment Tool 25
3.3.1 Overview 25
3.3.2 Implementation and Benefits 26
3.3.3 Integration and application 26
3.3.4 Gap Analysis and strategic planning 26
3.3.5 Implementation and continuous assessment 27
4. Outreach to the Community 28
4.1 Position Paper 28
4.2 Workshops 28
4.3 Cooperation with other projects and initiatives 28
5. Conclusions and Future Works 30
6. Appendices 36
6.1 Appendix A - Abbreviations 36
6.2 Appendix B - Codebook for interviews & surveys 36

List of figures

Figure 1: Recommendations for Ethical, Societal and Cybersecurity Concerns, overview 10
Figure 2: AI Standards with Ethical/Societal impact (subgroup of Trustworthiness), IoT Trustworthiness 11
Figure 3: UN SDGs in European Standardization Policy 12
Figure 4: „ Collaborative Safety“ - a holistic approach in context of humans, machines and operational environment 13
Figure 5: Large Language Models (LLM) Linking Standards and Guidelines 19

List of tables

Table 1: Status of Compliance with existing guidelines and regulations across AIMS50 industrial use cases based on an online survey 14
Table 2: Descriptive Sample Statistics 21
Table 3: Preliminary Personas of AI Usage in Industry 22
Table 4: Abbreviations 36

1. Introduction

1.1 Objective and scope of the document

Digitalization and emerging technologies like the Internet of Things (IoT) and Artificial Intelligence (AI) significantly boost the modern industry's competitiveness, success, and sustain-ability by enhancing supply chain agility and resilience and promoting environmental sustainability and consciousness. Although digitalization presents numerous benefits and opportunities, the incorporation of AI in industrial digital workspaces and in production, particularly in European industries, raises issues related to acceptance, ethical and legal compliance, and trustworthiness, posing vast ethical risks for individuals and societies [1]. To address this, it is crucial to establish a comprehensive self-assessment framework that caters to various users in the industry. This framework serves as a foundational tool for the successful implementation of digitalization and AI in the process of digital transformation, aiming to create human-centered workplaces [2, 3] . Existing guidelines and documentation often fail to meet the specific requirements of certain users, hindering their practical application and eroding trust in AI in digital workspaces. Consequently, this adversely affects the uptake of AI in digital workspaces. Addressing these issues calls for developing various tools, including checklists, self-assessment tools, guidelines, and recommendations on how to apply AI in digital workspaces, which should also provide means for a holistic evaluation in various use cases. These developed self-assessment tools and guidelines must consider aspects such as change management and skills development, the role of current and in-development standards, and regulations in sustainability, ethical and legal boundaries, and the gender dimension. We will follow the hereinafter steps to develop the mentioned self-assessment tools and guidelines: (1) Develop personas and roles to derive the user groups and model their attributes. (2) Sketch the landscape of the available standards and guidelines. (3) Create a fine-tuned Large Language Model (LLM) targeting the requirements of the self-assessment tool. And (4) Use the created LLM as an underlying domain-specific model for the self-assessment tool. All these steps are not only crucial prerequisites to building the tool that supports engineering with a variety of relevant standards and regulations but also to improve trust and user acceptance, using AI-based systems and their full potential for boosting innovation, economic growth, and global competitiveness. Regarding (1), we present the first results on how to model requirements of typical roles in the AI engineering domain, deriving personas from structured interviews. We present some representative standards in (2) to map individual persona's attributes to

standards. Regarding (3) and (4), we propose an approach to fine-tune domain-specific use case knowledge using LLM and to structure the self-assessment tool to generate a recommendation w.r.t. standards and guidelines which is not a part of the current deliverable at hand but subject of forthcoming deliverables.

1.2 Structure of the deliverable report

This deliverable is structured into five main sections. Section 1, Introduction, introduces the project's objectives and scope. Section 2, Evaluation of Existing Guidelines, reviews the current landscape of AI guidelines, standards, and regulations, underscoring the need for ethical and legal alignment in AI applications. Section 3, Development of the Self-assessment Tool, elaborates on the methodology for developing a self-assessment tool tailored to diverse user needs in the AI industrial sector. Section 4, Outreach to the Community, describes initiatives aimed at discussing findings and engaging with the community through workshops and collaborations. Section 5, Conclusions and Future Works, summarizes the deliverable’s findings and outlines future research directions.

2. Evaluation of Existing Guidelines and Relevant State of The Art

2.1 The Landscape of existing Guidelines

Numerous "static" guidelines and documentation exist, none of which is accessible or searchable via central repositories. Acceptance of AI in an industry, besides technology uptake, hinges on ethics, legal norms, and trust. Huang et al. [1] provide a thorough outline of the existing work on the AI ethics field, including a summary and analysis of AI ethical issues, guidelines and principles, approaches to address these issues, and methods to evaluate the ethics of AI technologies, such as the UNs' recommendation on the AI ethics (adopted on 23rd of November, 2021) [4]. In other fields such as Cybersecurity, there exist already assessment methods. For example, Leszczyna [5], comprehensively identifies and analyzes various cybersecurity assessment methods described in scientific literature.

Existing work on ethical AI, which we should consider and build upon in the context of this work, includes but is not limited to the EC HLEG (High-Level Expert Group) publication titled 'Ethics Guidelines for Trustworthy AI' (Final report, April 2019, HLEG AI) [6] and the guideline for assessment, the HLEG "Trustworthy AI assessment list" [36]. Additionally, the IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems (AI/AS - April 2016) [7] with IEEE Ethically Aligned Design" [32] and the standard IEEE 7000-2021 "IEEE Standard Model Process for Addressing Ethical Concerns during System Design" , and Australia's AI Ethics Framework, developed by a steering committee of experts from CSIRO (The Commonwealth Scientific and Industrial Research Organization) in 2019 [8], provide valuable insights (see Figure 1) Notably, UNESCO's first-ever global standard on AI ethics – the 'Recommendation on the Ethics of Artificial Intelligence' – was adopted by all 193 Member States in November 2021. This historic agreement emphasizes the protection of human rights, transparency, fairness, and the importance of human oversight in AI systems [39]. Furthermore, the OECD AI Principles, adopted in May 2019, promote the use of innovative and trustworthy AI that respects human rights and democratic values. These principles set practical and flexible

standards for AI, aiming to stand the test of time [40].



Figure 1: Recommendations for Ethical, Societal and Cybersecurity Concerns, overview

Another important national document is from the German Ethics Commission (Deutscher Ethikrat, in German language) with recommendations and statements: "Humans and Machines – Challenges of Artificial Intelligence" (2023-03-20) [37]. Guidelines and ethical considerations are also provided from the "Institute for Ethics and AI" of TU Munich, e.g., Alexander Kriebitz, Research Brief "Protecting and Realizing Human Rights in the Context of Artificial Intelligence: A Problem Statement" [38].

2.2 The Landscape of Standards and Regulations

Ongoing standardization efforts by groups ISO/IEC JTC1 SC41 [9] (IoT) and SC42 [10] (AI) under the "trustworthiness" topic, such as ISO/IEC TR 24368 "AI - Overview of ethical and societal concerns" [11], and ISO/IEC 42001:2023 "international standard specifying requirements for establishing, implementing, maintaining, and continually improving an Artificial Intelligence Management System (AIMS) within organizations" [12] are also noteworthy. The AI Standards Hub has catalogued eleven standards that pertain to ethics[2]."Governance" (for organizations9 and topics like "Human oversight" are further topics of AI Standardization in ISO/IEC JTC1 SC42. The following figure (Figure2) provides some overview on IoT and AI Standards with respect to trustworthiness and ethics:

[2] https://aistandardshub.org/ai-standards-search/?_sf_s=ethic

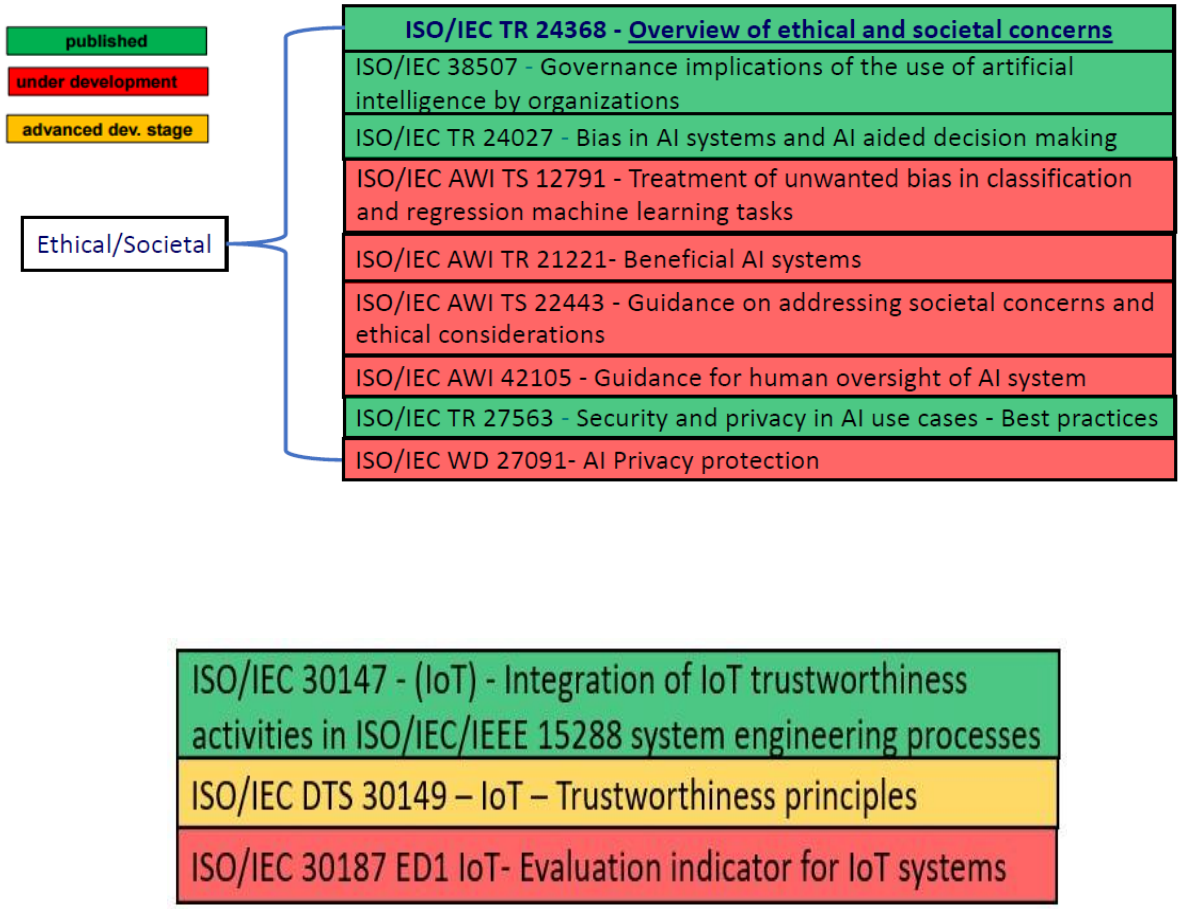

Figure 2: AI Standards with Ethical/Societal impact (subgroup of Trustworthiness), IoT Trustworthiness

Figure 2 shows AI standards with Ethical/Societal impact, specifically focusing on a subgroup of Trustworthiness in IoT. This figure illustrates the various standards that address ethical and societal considerations within the realm of IoT, contributing to the development of trustworthy IoT systems.

Ethical considerations may extend to issues like the sustainable development goals (SDG) approach as well. Ethical considerations within the framework of the Sustainable Development Goals (SDGs) are fundamental in ensuring that the goals contribute positively to societal and environmental outcomes while respecting human rights and promoting social justice. Ethical principles guide the prioritization and implementation of these goals to address global challenges such as poverty, inequality, climate change, and peace and justice comprehensively and fairly. For instance, considerations around equity can influence policies to ensure that advancements in sustainable practices do not disproportionately benefit the wealthy or further marginalize the disadvantaged. Moreover, ethics in transparency and accountability are vital in

maintaining public trust and participation in sustainable initiatives. Ethically driven approaches to the SDGs also emphasize the importance of sustainable economic practices that do not exploit labor or natural resources, thereby ensuring that development is both socially inclusive and environmentally sustainable. Thus, integrating ethical considerations into the SDGs leads to more holistic and effective solutions to global challenges.

"Green deal" aspects are covered by ISO/IEC AWI TR 20226 [3]- Environmental sustainability aspects of AI systems, a topic, that has also been taken up by CEN. Related to humans in the loop is also ISO/IEC AWI TR 21221[4] AI - Beneficial AI systems, and ISO/IEC TR 18988 - Application of AI technologies in health informatics.

The following figure (Figure 3) shows (source: CEN/CENELEC/UNE Webinar[5]) how European standards are following UN-SDG objectives, particularly issues of societal and ethical interests:



Figure 3: UN SDGs in European Standardization Policy

In the context of safety standardization within the IEC, which is also pertinent to the field of robotics, novel safety approaches are emerging, notably under the banner of "Collaborative Safety".

[3] ISO/IEC AWI TR 20226
[4] ISO/IEC AWI TR 21221
[5] https://www.cencenelec.eu/european-standardization/sustainable-development-goals-sdgs/

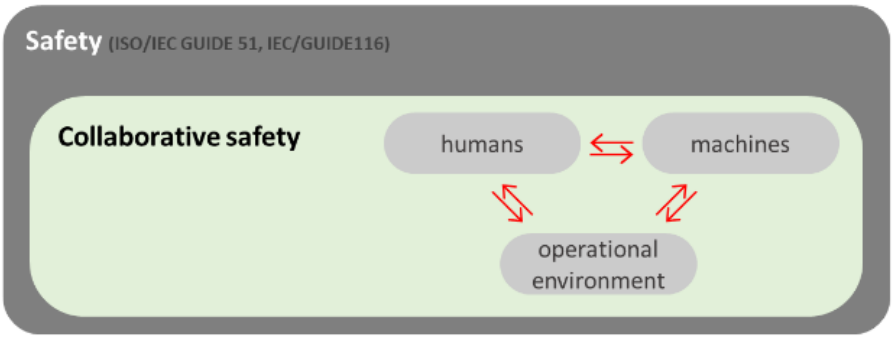

Figure 4: „ Collaborative Safety“ - a holistic approach in context of humans, machines and operational environment

The relevant EU legal framework includes but is not limited to the General Data Protection Regulation (GDPR) 2016/679/EU [13], the ePrivacy Directive 2002/58/EC [14], the relevant EU laws relating to Geo-Spatial Data (ACCESS Directive 2003/4/EC [15] and INSPIRE Directive 2007/2/EC [16]). The European AI framework (ECS SRIA Final 2021 [17]) is the foundation for understanding the impact of AI on stakeholders and users of business products and services. The Council of Europe's Committee on Artificial Intelligence (CAI)[6] is also working on a treaty concerning AI. The Committee of Ministers has tasked the CAI with establishing an international negotiation process and conducting work to elaborate a legal framework on the development, design, and application of AI systems[7].

2.3 Awareness and status of implementation of existing guidelines, regulations and standards for AIMS5.0 industrial use-cases

The European Commission issued a standardization request to the CEN/CENCELEC organization for 10 areas of artificial intelligence – these standards will become the backbone of the AI Act implementation. The CEN/CENELEC standards are often an adaptation of ISO/IEEE standards. In the meantime, at the European Court of Justice a decision was rendered, according to which all CEN harmonized standards should be made publicly available. This creates the conditions in which AIMS will be able to create a mapping of standards to AIMS use cases and find synergies. Table 1 demonstrated the current state of adherence to the existing AI guidelines and regulations. We have conducted an initial investigation via an online survey how AIMS5.0 use-cases already make use of new guidelines and are aware of recent developments in this field.

[6] https://www.coe.int/en/web/artificial-intelligence/cai
[7] https://www.gp-digital.org/the-council-of-europes-ai-treaty-perspectives-from-the-global-majority/

Table 1: Status of Compliance with existing guidelines and regulations across AIMS50 industrial use cases based on an online survey

Use Case #	Use Case Name	The binding regulations	Non-binding regulations	Standards	Guidelines
2	AI Enhanced Value Chain; UC7 - Digital Automated Luminaire Manufacturing platform	EU AI Act	-	-	▪ The Netherlands Code of Conduct for Research Integrity; ▪ The Guidelines on Ethical Data Handling in Research set by the Bernoulli Institute for Mathematics ▪ Computer Science and Artificial intelligence of the University of Groningen.
7	Digital Automated Luminaire Manufacturing platform	-	-	-	Assessment List for Trustworthy Artificial Intelligence (ALTAI)
8	AI supporting and strengthening human and manufacturing cycles	-	-	-	-
9	Energy-aware Scheduling of Batch Processing Machines in Wafer Fabs;	▪ European Data Act ▪ European Data Governance Act ▪ EU AI Act		ISO/IEC 22989:2022 - Artificial Intelligence — Artificial Intelligence Concepts and Terminology ISO/IEC 23053:2022 - Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML) ISO/IEC 27001:2013 - Information technology — Security techniques — Information security management systems — Requirements	Internal Guidelines

Use Case #	Use Case Name	The binding regulations	Non-binding regulations	Standards	Guidelines
				ISO/IEC 27701:2019 - Privacy information management system (PIMS) for processing personal data ISO/IEC 38500:2015 - Corporate governance of information technology ISO/IEC TR 24028:2020 Information technology — Artificial Intelligence — Overview of trustworthiness in AI	
10	MES concept based on AI Improving Production Cycle: AI-Prod_MES	▪ All Eu Mandatory regulations			Internal Guidelines
11	-	-	-	-	▪ Ethics guidelines for trustworthy AI from European Commission's HLEG on AI ▪ Internal Guidelines developed based on EU AI Act
12	-	-	-	-	-
13	-	-	-	-	▪ Guidelines issued by the European commission
14	-	-	-	-	-
15	-	▪ EU AI Act	-	-	-

2.4 State of the art of relevant technologies and methods

The above-mentioned documentation is provided in static documents, which are frequently updated. Hence AI-based technologies such as NLP or LLM can help to dynamically analyze the existing documents. They will receive the AI developers' personas in reference to their individual organizational and implementation features as input and will subsequently extract the applicable rules and regulations. In an agile AI driven system, this can be supported by continuous evaluation of the compatibility with such security standards i.e. IEC62443 as documented in [18]. There have been many studies exploring the power of LLMs in different domains. An example can be the research by Jeong [19] that investigates fine-tuning and leveraging domain-specific LLMs. Another useful study for this work is the survey by Xu et al. [20] that summarizes the recent works on LLMs for information extraction (IE). While several white papers of consultancies exist concerning implementations, risks, and benefits concerning AI in industry, some research literature using qualitative research methods [21, 22] is emerging on benefits and challenges of AI adoption [23, 24]. Nevertheless, applied works providing interventions to achieve successful AI implementation in practice are still scarce in this respect. Another important aspect of any technology development lies in its acceptance by the involved stakeholders, which defines its success. The same applies in the field of Industry5.0, where there is a need for guidelines to monitor, enhance and ensure AI acceptance. These guidelines should also address ethical considerations, legal limitations, as well as trustworthiness, forming a comprehensive framework for validation, human-centered use cases. To avoid duplicating efforts and take into account existing documentation beyond industrial systems, innovative AI-based approaches are necessary. These approaches can include the utilization of LLM or NLP, enabling the provision of tools and automated advice and guidance. By developing such automated guidance and support systems, gaps can be identified, which in turn allows for the creation of recommendations concerning regulation. These recommendations should consider change management and skills development as crucial factors.

The rate at which individuals are adopting new technologies is affected by various factors, such as the availability of the technology, convenience, consumer needs, trust, and other related aspects. As such, a lot of research has been carried out in the areas of technology acceptance modeling, resulting in the development of the Technology Acceptance Model (TAM) across generations of digital technology. Studies have explored TAM's application to AI-based products, and consulting frameworks like TOE (Technological, Organizational, and Environmental) to consolidate guidelines for decision-makers. In the context of contemporary society, successful integration and adoption of new technologies hinge on understanding and predicting user acceptance. TAM, a widely recognized model highlighted by Adams et al. [25],

focuses on perceived ease of use and usefulness. However, for cutting-edge concepts like AI, with inherent complexity and ethical considerations, TAM has been extended and adapted to address these complexities. The Unified Theory of Acceptance and Use of Technology (UTAUT) is a more recent and comprehensive model that integrates various factors influencing technology acceptance. UTAUT incorporates performance expectancy, effort expectancy, social influence, and facilitating conditions, providing a more nuanced understanding of user behavior [26].

As technology evolves, newer models like the Extended Technology Acceptance Model (eTAM) and the Innovation Diffusion Theory (IDT) are gaining prominence [27]. In conclusion, the importance of technology acceptance cannot be overstated, especially for innovative concepts like AI technologies. Ultimately, the EU AI Act is worth mentioning, which is a part of the Union's digital strategy and is the comprehensive law to promote and regulate AI utilization, addressing some risks related to human health, safety, and fundamental rights. This framework will support other legislation of AI usage, such as e.g. large online platforms such as the Digital Services Act (DSA) and Digital Markets Act (DMA) which are two regulatory frameworks established by the European Union. The above recognizes the need for comprehensive frameworks that consider not only perceived ease of use and usefulness but also social influence, ethical considerations, and contextual factors.

2.5 Discussion

As the AIMS5.0 project runs, the regulatory landscape is in profound movement supported by significant developments at both the European Union (EU) and global levels. The EU has been active in shaping the regulatory framework for Artificial Intelligence (AI) through its Artificial Intelligence Act. This legislation, considered the first binding worldwide horizontal regulation on AI, sets a common framework for the use and supply of AI systems within the EU. Globally, UNESCO has been collaborating with the EU to implement ethical guidelines on AI. In 2021, UNESCO's Member States adopted a Recommendation on the Ethics of Artificial Intelligence [39], which has been used by nearly 30 countries to establish national legislation. This recommendation provides a normative framework ensuring that AI respects fundamental freedoms and human rights.

The various standards, guidelines, and legal frameworks that have been discussed earlier in the section or document and mentioned by the industrial use cases texpress the will of society through its representatives. However, this will might not be expressed in a way that is directly implementable, hence creating and implementation gap. Understanding what actionable norms can be created is a major task.

The results presented in this section indicate that there is still a long way to go to ensure acceptance of innovative technologies in the context of the 5th industrial revolution. It is important that a deeper investigation of technology acceptance takes place, for ensuring a successful implementation and adoption of Industry5.0. Understanding and addressing the factors that drive technology acceptance within the workforce, including their perceived usefulness and ease of use, can lead to higher employee engagement, productivity, and overall success in the Industry5.0 paradigm.

3. Self-assessment tool

3.1 Motivation

As outlined in section 2, numerous static documents and guidelines exist, but there is also technology that makes the use of these standards and regulations easier when evaluating one's conformity with them. Therefore, we will create an AI-based self-assessment tool, which helps AI application developers dynamically link their specific user attributes to relevant standards and guidelines using AI technologies such as large language models. This approach ensures the development of trustworthy and ethically compliant AI solutions. The envisioned tool will have an scheme as depicted in Figure 5.

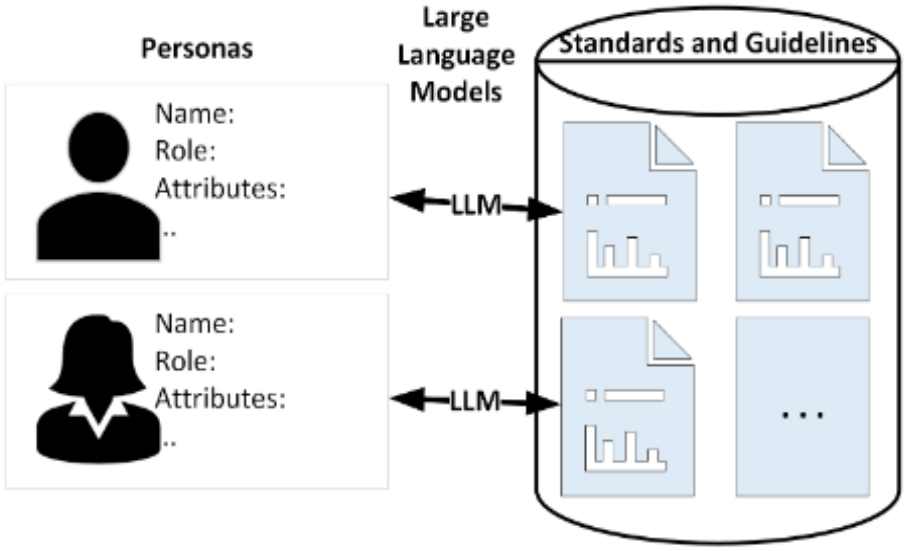

Figure 5: Large Language Models (LLM) Linking Standards and Guidelines of AI Application and Usage with Personas

3.2 Steps towards the development of the self-assessment tool

In this section, we provide insight into the four steps towards creating the self-assessment tool and guidelines. The self-assessment tool helps AI application developers by dynamically linking their specific user attributes to relevant standards and guidelines using AI technologies, such as large language models. This approach ensures the development of trustworthy and ethically compliant AI solutions.

Step-by-Step Description

1) Personas Development: Identifies target user groups through interviews and questionnaires, creating detailed personas that guide the tool's customization.
2) Survey of Standards and Guidelines: Maps the current landscape of AI standards and guidelines to keep the tool's recommendations up-to-date and relevant.
3) Large Language Model (LLM) Fine-Tuning: Adjusts a large language model to interpret user attributes effectively and provide accurate compliance recommendations.
4) Development of the Self-Assessment Tool: Integrates all components into a functional tool that offers personalized guidance for AI system implementation according to current standards.

As the first step and towards the development of the self-assessment tool, we need to understand user needs. AI development is a broad field, and developers come from diverse backgrounds with varying levels of experience. Personas help us identify these differences. By creating detailed personas, we can get to know the target users and understand their specific needs and challenges when searching for regulations and guidelines.

Furthermore we research the existing national and international standards bodies (IEEE, ISO, IEC) relevant to AI development, and Identify specific regulations and guidelines related to different AI use cases (data privacy, safety, bias).

The landscape of AI regulations and guidelines is constantly evolving (As outlined in section 2). Surveying this landscape ensures that the tool points its users to the most current and relevant resources. Additionally, different industries and AI use cases have specific regulatory requirements. Surveying the landscape allows the self-assessment tool to understand these nuances and tailor its recommendations to the user's specific context.

3.2.1 Personas

The main objective of this step is to develop personas and roles. In this reporting period we have defined two groups of participants in the project work: 1) facilitators, and 2) stakeholders and contributors. The facilitators are the ones who design, conduct, and evaluate questionnaires as well as desktop research. The stakeholders and contributors are experts who participate in the questionnaires and semi-structured interviews. To establish an initial set of personas in the landscape of industrial AI usage, we developed a semi-structured interview guide. This guide comprised general questions concerning interviewees' personal and

organizational backgrounds, their understanding of AI, its application, and the benefits of using AI in manufacturing. Additionally, we explored challenges, risks, solutions, and open topics related to AI usage. The questions, as well as codes and categories, were derived from the existing literature on AI adoption in the industry [23, 24] and were refined iteratively throughout the research process.

In September 2023, we interviewed 27 experts from two large European research and development (R&D) projects (AIMS5.0 and arrowhead fPVN) in the field of AI, representing various stakeholders (Table 2). The interview questions are provided under Appendix B. We present the findings from these interviews in this deliverable in Table 3 and elaborate on them more in detail thereafter.

Table 2: Descriptive Sample Statistics

Type of organization	
Enterprise	55.6 %
Research institution	37.0 %
Other (NGO)	3.7 %
N/A	3.7 %

Size of organization	
Large	70.4 %
Medium	14.8 %
Small	7.4 %
Micro	3.7 %
N/A	3.7 %

Country of organization	
Austria	33.3 %
Germany	25.9 %
Hungary	11.1 %
Spain	7.4 %
Sweden	7.4 %
Other (1 each)	11.1 %
N/A	3.7 %

Position of interviewee	
Manager (R&D)	37.0 %
Technical expert	25.9 %
Researcher	18.5 %
Other (freelance)	3.7 %
N/A	14.8 %

Table 3: Preliminary Personas of AI Usage in Industry

	Organization & Involvement	Application Needs	Benefits	Acceptance & Implementation	Risks & Mitigation	Interview IDs
Early Adopter / Avantgarde	Mainly large enterprise or research institute, already involved in (several) AI R&D projects and further project proposals	In quality management, healthcare and production as well as network systems, telecommunications, material handling and human resources	Human support, time, material, resource and cost efficiency, increased performance, throughput and quality	Differences in AI acceptance between knowledge, white- and blue-collar workers; top-down or both-ways implementation, while the decision power lies with different stakeholder groups	Data and computational power issues, cybersecurity threats, insufficient training, over-reliance on AI, the mismatch between efforts and benefits; mitigated by training, adherence to standards and regulations, risk assessments, non-AI fallback solutions	1-10, 13, 14, 18, 20-21, 24-26
Follower	Mainly small or medium-sized enterprises or research institutes, involved in some R&D projects	Improve decision-making and material handling in farming and mechatronics	Performance improvements as decision-making and planning support in production & logistics	General openness towards AI; both top-down and bottom-up implementation with top- and middle-level managers deciding	Data issues, false expectations and lack of competences among stakeholders are mitigated by training	12, 16, 19, 22-23
Independent Learner	Mainly small or medium-sized enterprise or research institute, involved in few R&D projects, but actively looking for new partners and opportunities to tackle own ideas	Very specific applications, where few to no ready-to-use AI approaches exist, e.g. in automotive or aerospace	Increased time, material, and resource efficiency, e.g. by using LLMs for ideation	Great openness towards AI, but not for its own sake; rather a bottom-up implementation, but the final decision is with the CEO (only few hierarchical levels)	Biased data, unrealistic expectations, and lack of supervision are known risks with planned interventions by increasing human oversight	11, 15, 17, 27

The interview transcripts have been coded by three independent coders, achieving a net inter-coder reliability between 72 and 90 percent for three randomly selected interviews. Based on the analyses from the coders, three personas were distinguished in the first run of investigation: early adopter/avantgarde, follower, and independent learner. The rationale that differentiates early adopters and followers is based on the Bass diffusion models to forecast the diffusion of innovative technologies [28, 29]. Table 3 describes and distinguishes personas based on organizational characteristics and implementations, their needs when using AI, achieved or desired benefits, employee acceptance of AI, and challenges, risks, and mitigation measures related to AI. While early adopters are mainly large companies, involved in several R&D projects concerning AI, followers and independent learners are mostly represented by small or medium-sized enterprises with less R&D experience in AI. However, all of them share similar expected benefits and risks of using it. Distinctively, early adopters have often more concerns about human factors, such as incompetencies, too high expectations, and thus low acceptance of using AI. Cybersecurity threats are also of high relevance for the large companies. Nevertheless, they also have already more measures in place to mitigate such risks, focusing on using on premise training of ML models. Notably, independent learners actively seek new partnerships to address their specific application needs due to the limited availability of ready-to-use AI approaches, while concerns regarding human acceptance play an inferior role at this stage. Although some of the interviewees stressed that AI should not be implemented gratuitously and/or wherever possible, and equally powerful non-AI solutions should be preferred to AI, there were no real skeptics or hesitance against AI solutions.

3.2.2 **Application of Standards, regulatory and guideline landscape**

This step is concerned with locating and integrating the existing ethical and legal norms into product development. We have already done and initial investigation (see section 2) in this reporting period. In the following reporting periods we will fine tune it in line with the development of the LLM. As technology evolves faster than legal constraints, it is essential to stay updated on existing frameworks, e.g., Austria's "Bundesministerium für Klimaschutz" entitled Artificial Intelligence Mission Austria 2030 [30]. End users can gain prompt access to evolving standards and guidelines once they appear by feeding them to the self-assessment tool.

Other organizations of interest are EurAI [8] (European Association for Artificial Intelligence), with national organizations as members, e.g. ASAI, the Austrian Society for Artificial Intelligence (see members of EurAI [9]). A comprehensive summary is provided in section 2.

To deploy a product including AI technology, there are AI standards and the EC AI-Act as well as GDPR to consider. One requirement is to maintain an AI Management System within the organization using such tools as addressed here. ISO/IEC 42001, AI Management system, specifies requirements for establishing, implementing, maintaining, and continually improving an Artificial Intelligence Management System (AIMS) within organizations. It is designed for entities providing or utilizing AI-based products or services, ensuring responsible development and use of AI systems (from ISO/IEC 42001 [41]).
The European AI-Act was approved by the European parliament by March 13, 2024. The AI Act will enter into force 20 days after publication in the Official Journal (expected in May or June 2024). Most of its provisions will become applicable two years after the AI Act's entry into force. Harmonized standards will become available (derived after a specific process called HAS. The HAS process refers to a procedure used within the European Union to develop and recognize standards that support European legislation, ensuring that products, services, or processes comply with the EU's health, safety, and environmental protection legislation) and will be referenced in accompanying publications of the EC.

3.2.3 **Large language models (LLM)**

In this step, we will focus during the next reporting periods on preparing the LLMs to make them qualified to provide the required support for various personas based on their attributes. One way to achieve this preparation is by fine-tuning the LLMs using task-dependent datasets. There are several fine-tuning strategies using different types of datasets to enhance the capabilities of the LLMs, such as [31] fine-tuning with Human Annotated datasets, fine-tuning with LLM-Generated datasets, and fine-tuning with Real-World datasets. In this work, we will mainly focus on fine-tuning using LLM-Generated datasets by prompting ChatGPT to generate the relevant guidelines, standards, or actions tailored to the defined personas inputs, thereby enhancing the benefits and acceptance of AI in manufacturing (Fig. 5).

The use of LLMs in addressing personal data underlies the GDPR rules and the AI-Act (upcoming, not yet in force at the moment). There are guidelines around, e.g. from Research Institute AG

[8] https://eurai.org/
[9] https://eurai.org/members

& CoKG, Digital Human rights Center, Vienna: „Recommendations for use of "Large Language Models" (LLM) such as ChatGPT, Google Bard und BingAI" [10].

Standards are under development:

- IEEE SA P3419 "Standard for Large Language Model Evaluation" (criteria for the evaluation of Large Language Models (LLMs) and extending to multimodal models)
- The **IEEE P3394** standard, titled **"Standard for Large Language Model Agent Interface,"** which defines natural language interfaces that facilitate communication between **Large Language Model (LLM)** applications, agents, and human users.
- Related standardization in ISO/IEC SC42:
- ISO/IEC AWI TR 23281: Overview of AI tasks and functionalities related to natural language processing,
- ISO/IEC AWI 23282: Evaluation methods for accurate natural language processing systems.

Nevertheless, the already published standards addressing ethical aspects have to be considered, although they are not particularly addressing LLMs.

3.2.4 The Self-Assessment tool

In this step, we will focus in the next reporting periods on the architecture of our self-assessment tool. Developing an LLM-based tool requires integrating several modules, such as the profiling module, the memory module, the planning module, and the action module [20, 31]. Next is to integrate the outcome of the previous steps into the development of the self-assessment tool. The first step served as a preparation step to define target users and their requirements, while the second step provided the required materials for the LLMs fine-tuning intended for the third step. The outcome of the third step is the self-assessment tool underlying fine-tuned LLM capable of fulfilling the requirements of the various stakeholders.

3.3 Backcasting exercise (and strategic planning) in the Self-Assessment Tool

3.3.1 Overview

One of the methods that we are currently evaluating for applying in the development of our self-assessment tool is Backcasting. Backcasting is a strategic planning approach that involves

[10] https://researchinstitute.at/wp-content/uploads/2023/07/LLM-Guidelines-1.pdf

defining a desired future outcome and methodically tracing the steps required to realize it from the present standpoint. This technique allows for the detailed envisioning of a target future scenario, followed by a reverse analysis to identify the necessary actions and milestones to achieve that scenario.

3.3.2 Implementation and Benefits

Employing backcasting requires a forward-thinking mindset, ensuring that current strategies are aligned with long-term objectives thus offering a clear roadmap for progression towards the envisioned future as the successful achievement of goals set out in a project like ours. By adopting backcasting, we embrace a proactive mindset, prioritizing future planning and vision over mere reactions to present-day circumstances. This approach is crucial for integrating AI into manufacturing in a way that aligns with sustainability objectives and overall project ambitions.

3.3.3 Integration and application

The integration of backcasting into the development of a self-assessment tool can provide significant benefits for AI application developers. This approach allows developers to anticipate potential side effects and externalities of their innovations on society at large, extending beyond their primary domain. By incorporating backcasting, developers can work backwards from a desired future state of regulatory compliance and societal impact, identifying necessary steps and adjustments in the present. This proactive strategy ensures alignment with existing regulations and guidelines, fostering responsible and ethical AI development.

Applying the method as described in [35], the journey starts with a thorough assessment of the present scenario, where the current capabilities of AI in manufacturing are closely examined alongside the regulatory environment, existing challenges, and technological gaps. This critical evaluation serves as the foundation for the entire process, establishing a clear understanding of the starting point. It is planned to apply this approach in the development process carried in the next reporting periods.

3.3.4 Gap Analysis and strategic planning

Following this, a gap analysis is essential, pinpointing the discrepancies between today's reality and the future we aim to achieve. This involves identifying the barriers that obstruct the path to this future and the opportunities that could accelerate the journey. It's a stage that demands

a deep dive into what needs to change, covering technological needs, policy adjustments, and societal shifts.

With the gaps laid out, the process then naturally flows into the development of transition pathways. These pathways are not just steps but a series of strategic initiatives and actions, conceptualized by working backward from the envisioned future. They encompass the necessary technological innovations, policy reforms, and shifts in industry practices required to bridge the current state with the desired outcomes.

Strategic action planning follows, where the pathways translate into a concrete plan. This phase is about prioritizing actions based on their potential impact, feasibility, and the timeline for implementation. It's a dynamic blueprint, ready to evolve as the process moves forward, embodying the essence of backcasting as a flexible and forward-looking approach.

3.3.5 **Implementation and continuous assessment**

The implementation of this plan marks the beginning of a continuous cycle of action, assessment, and adaptation. Recognizing the iterative nature of backcasting, the strategy is revisited and revised in response to new developments, technological breakthroughs, and shifts in societal values or regulations.

Integral to this entire process is the engagement with stakeholders across the spectrum—from industry players and policymakers to academia and the public. Their involvement ensures that the plan is not only robust but also inclusive, reflecting a wide array of perspectives and securing broad-based support for the envisioned changes.

Monitoring progress towards the envisioned future is as critical as the initial planning. This ongoing evaluation ensures that the strategy remains aligned with the goal, making adjustments as necessary to navigate through the unpredictable landscape of technological innovation and societal change.

4. Outreach to the Community

In this section we detail various outreach activities to underscore the significance of these activities within the project. It is crucial to recognize that our engagement with the community are foundational to the project's success, reflecting our commitment to meaningful interaction and impact.

4.1 Position Paper

This deliverable is grounded in the peer-reviewed paper accepted at the 5th Workshop on Management for Industry 5.0 (MFI5.0) to take place during the IFIP/IEEE NOMS 2024 - Network Operations and Management 2024 Symposium. This external validation was crucial for obtaining a structured and authoritative assessment of our approach.

4.2 Workshops

The work included in this deliverable has also been discussed with experts in the field at several workshops and conferences, listed below:

- Deep Tech Workshop, Sep 27, Sep 28, Saint Pölten
- Arrowhead fPVN - AIMS 5.0 Wworkshops in LTU (Luleå University of Technology)- Jan 23- Jan 25, 2024
- European Conference on EDGE AI Technologies and Applications (EEAI), Oct 17 – Oct 19, Athens

4.3 Cooperation with other projects and initiatives

With respect to synergies in standardization between related projects, particularly AIMS5.0, AI4CSM, A-IQ Ready and PowerizeD on AI and Cybersecurity/safety challenges and standardization, presentations were given or Posters/Roll-up and discussions during poster sessions at the following occasions:

- DECSoS Workshop at Safecomp 2023 in Toulouse (2023-09-19): AIMS5.0 and related projects have been presented and have been part of the Plenary discussion on human-centred AI and societal impact.

- "Edge AI Standardisation Activities"; Poster and a presentation in Session C02 of the European Conference on EDGE AI Technologies and Applications – EEAI, **Athens,** October 17-20th, 2023. (Acknowledgements: A-IQ Ready, AIMS5.0, AI4CSM).
- 2 Roll-Up Poster presentations on "Synergies in Standardization between related projects", (PowerizeD, AI4CSM, A-IQ Ready, AIMS5.0), at the PowerizeD Project Workshop and General Assembly in Vienna, AIT, November 7-8, 2023. (EU-Projekt Veranstaltung).
- Roll-Up Poster presentation on "Synergies in Standardization between related projects", (PowerizeD, AI4CSM, A-IQ Ready, AIMS5.0), at the ARROWHEAD fPVN – AIMS5.0 Workshop at the University of Technology, Lulea, Sweden, January 24-25, 2024.
- "Society 5.0 – Sustainability as a Key Challenge for Society and Technology – Technology driven, but human-centred", presentation at "Society 5.0 – Sustainable Technology Trends and Society", Vienna, CON.ECT Informunity /Future Network, Jan. 30, 2024.
- Additionally, at the review meetings of PowerizeD and A-IQ Ready in March 2024, the Synergies-Roll-up was presented and collaborative activity were high-lighted.

In addition, we established initial contacts with HEU project, AIRedgio 5.0 (AI REDGIO 5.0 | EU project (airedgio5-0.eu)). The project aims at exploiting the benefit of Artificial Intelligence in manufacturing domain with more than 20 experiments with manufacturing SMEs that will be conducted. We have discussed the possibility of expanding the audience of the surveys that will be produced in the different tasks of WP7 as AI REDGIO would provide a much broader community for the studies.

5. Conclusions and Future Works

This deliverable delineates our ongoing endeavors to formulate checklists and guidelines, manifesting as a self-assessment tool designed to promote the adoption and uptake, on behalf of the various stakeholders involved, of trustworthy, ethically sound, and legally compliant AI-based technologies within the industrial sector, contributing to human-centered workplaces. As the first step toward the development of this tool, we have interviewed experts from large to small manufacturing companies associated with industrial AI, with the objective of creating an initial collection of personas within the domain of industrial AI utilization. The preliminary personas allow first distinctions of goals, requirements, benefits, and concerns of different stakeholder groups applying AI in industry, which will be elaborated on in the ongoing work. While the personas at the current state represent mainly organizational levels, a further step will also be the assessment of attitudes of individual white- and blue-collar workers to achieve more specific insights and derive more targeted tool support for implementing and using AI in industry, building upon previous works of the partner [34]. To this aim, we plan to survey stakeholders involved in industrial AI use cases using questionnaires. This will help us gather a more extensive range of personas, including their attributes and observed legal/ethical guidelines. This acquired knowledge will be pivotal in the training of AI algorithms intended for the system design of the self-assessment tool. Subsequently, we aim to construct a regulation-aware AI model, leveraging Large Language Models (LLMs) and fine-tuning it based on the standards, regulations, and ethical guidelines obtained also through dialogues with society. This regulation-aware model will serve as the nucleus of our self-assessment tool, guiding industrial AI users to develop more normative AI applications by providing them with applicable guidelines and regulations.

Moreover, in the course of our future activities within the project we will also investigate how technology acceptance models can be appropriately adapted and become oriented towards AI tools and methods, so that they can efficiently foresee technology uptake, as well as in general support the change management towards digital workspaces.

- Structured questionnaire (T7.1)
- Standardization and Guideline landscape (T7.2)
- System design for tool support (T7.4, LLM)
- Align self-assessment with TAM

Through collaboration between WP7 and partners in the project, the following contributions to the legal landscape are also foreseen:

- Identify relevant EU legislation and assess the risks, particularly concerning the developments covered by the project. In the context of the study, we search the potential for the legal application of each solution in detail, with particular attention to the conditions for the lawful use of artificial intelligence, ensuring adequate data protection and the protection of human rights. The analysis aims to support the project, exploring how the technological objectives can be achieved in the broadest possible way while maintaining a legal framework.
- Analyze and evaluate intellectual property issues of the use cases involved in the project. Copyright law and patent law issues are of significant importance when applying AI. Use cases are to examine from the point of view of existing regulation (first of all relevant provisions of AI Act, Digital Services Act and Digital Markets Act), revealing the eventual regulatory gaps regarding this field. Intellectual property law has a human-based approach towards the outputs of AI, so this is the starting point when evaluating of the industrial application of AI. We highlight challenges regarding use cases' compliance with intellectual property law.
- Address data protection legal issues and consider risks in the light of GDPR during product development. We examine the civil law and data protection law risks of specific technical developments, in particular camera systems to be developed with artificial intelligence. In the case of developments affecting employees, we analyse employment law aspects and make proposals in accordance with the labour legislation in force.

References

[1] Huang C, Zhang Z, Mao B, Yao X. An Overview of Artificial Intelligence Ethics. IEEE Transactions on Artificial Intelligence 2023; 4(4): 799–819 [https://doi.org/10.1109/TAI.2022.3194503]

[2] Keil S, Lindner F, Moser J, Weth R von der, Schneider G. Competency Requirements at Digitalized Workplaces in the Semiconductor Industry. In: Keil S, Lasch R, Lindner F, Lohmer J, editors. Competency Requirements at Digitalized Workplaces in the Semiconductor Industry; 2020. Cham: Springer International Publishing; 88–106.

[3] Leyer M, Richter A, Steinhüser M. 'Power to the workers': Empowering shop floor workers with worker-centric digital designs. International Journal of Operations & Production Management 2019; 39(1): 24–42 [https://doi.org/10.1108/IJOPM-05-2017-0294]

[4] Recommendation on the Ethics of Artificial Intelligence; 2022. Available from: URL: https://goo.by/jgJSlx [cited 2024 March 10]

[5] Leszczyna R. Review of cybersecurity assessment methods: Applicability perspective. Computers & Security 2021; 108: 102376.

[6] Ethics guidelines for trustworthy AI; 2024. Available from: URL: https://t.ly/1qhNY [cited 2024 March 10]

[7] IEEE Standards Association. IEEE SA - The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems; 2023. Available from: URL: https://standards.ieee.org/industry-connections/ec/autonomous-systems/ [cited 2024 March 10]

[8] Dawson D, Schleiger E, Horton J, et al. Artificial Intelligence: Australia's Ethics Framework. Australia: Data61 CSIRO; 2019.

[9] ISO/IEC JTC 1/SC 41 - Internet of things and digital twin; 2024. Available from: URL: https://www.iso.org/committee/6483279.html [cited 2024 March 10]

[10] ISO/IEC TR 24028:2020; 2024. Available from: URL: https://www.iso.org/standard/77608.html [cited 2024 March 10]

[11]ISO/IEC TR 24368:2022; 2024. Available from: URL: https://www.iso.org/standard/78507.html [cited 2024 march 10]

[12]International Organization for Standardization. ISO/IEC 42001:2023; 2023.

[13]Regulation - 2016/679 - EN - gdpr - EUR-Lex; 2024. Available from: URL: https://goo.by/nnvBAG [cited 2024 March 10]

[14]Directive-2002/58-EN-eprivacy directive - EUR-Lex; 2024. Available from: URL: https://t.ly/-qvk5 [cited 2024 March 10]

[15]Directive - 2003/4 - EN - EUR-Lex; 2024. Available from: URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=cele:A32003L0004 [cited 2024 March 10]

[16]Directive - 2007/2 - EN - EUR-Lex; 2024. Available from: URL: https://eur-lex.europa.eu/eli/dir/2007/2/oj [cited 2024 March 10]

[17]ECS-SRIA 2021 | EPoSS; 2024. Available from: URL: https://www.smart-systems-integration.org/publication/ecs-sria-2021 [cited 2024 March 10]

[18]Bicaku A, Schmittner C, Rottmann P, Tauber M, Delsing J. Security Safety and Organizational Standard Compliance in Cyber Physical Systems. Infocommunications Journal 2019; XI(1): 2–9.

[19]Jeong C. Fine-tuning and Utilization Methods of Domain-specific LLMs. arXiv preprint arXiv:2401.02981 2024.

[20]Xu D, Chen W, Peng W, et al. Large Language Models for Generative Information Extraction: A Survey. arXiv preprint arXiv:2312.17617 [cs.CL] 2023 [https://doi.org/10.48550/arXiv.2312.17617]

[21]Kuckartz U. Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. 4th ed. Beltz Verlagsgruppe 2018.

[22]Mayring P. Qualitative Content Analysis. Forum: Qualitative Social Research 2000; 1(2) [https://doi.org/10.17169/fqs-1.2.1089]

[23]Kutz J, Neuhüttler J, Spilski J, Lachmann T. Implementation of AI Technologies in manufacturing - success factors and challenges. The Human Side of Service Engineering 2022; 62(62) [https://doi.org/10.54941/ahfe1002565]

[24]Somer P, Thalmann S. Risk Management of AI in Industry: A Literature Review. In: Risk Management of AI in Industry: A Literature Review; 2023.

[25]Dennis A. Adams, R. Ryan Nelson, Peter A. Todd. Perceived Usefulness, Ease of Use, and Usage of Information Technology: A Replication. MIS Quarterly 1992; 16(2): 227–47.

[26]Ilias Panagiotopoulos, George Dimitrakopoulos. An empirical investigation on consumers' intentions towards autonomous driving. Transportation Research Part C: Emerging Technologies 2018; 95: 773–84 [https://doi.org/10.1016/j.trc.2018.08.013]

[27]Rogers EM, Singhal A, Quinlan MM. Diffusion of innovations. In: An integrated approach to communication theory and research. Routledge 2014; 432–48.

[28]Bass FM. A new product growth for model consumer durables. Management science 1969; 15(5): 215–27.

[29]Jérôme Massiani, Andreas Gohs. The choice of Bass model coefficients to forecast diffusion for innovative products: An empirical investigation for new automotive technologies. Research in Transportation Economics 2015; 50: 17–28 [https://doi.org/10.1016/j.retrec.2015.06.003]

[30]Artificial Intelligence Mission Austria 2030: Shaping the Future of Artificial Intelligence in Austria. Ministry of Transport, Innovation, and Technology; 2019. Available from: URL: http://tinyurl.com/5btkdub6.

[31]Wang L, Ma C, Feng X, et al. A Survey on Large Language Model based Autonomous Agents. Available from: URL: http://arxiv.org/pdf/2308.11432.pdf.

[32] IEEE Standards Association. „Ethically aligned Design", v2. https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf

[33] IEEE Standards Association. IEEE 7000-2021 - - IEEE Standard Model Process for Addressing Ethical Concerns during System Design

[34] Lindner, Fabian; Przybysz, Kazimierz A.; Schneider, Germar; Keil, Sophia: WORKERS' ATTITUDES TOWARDS DIGITAL TRANSFORMATION AND PERCEIVED WELL-BEING AT WORK. A CORRELATIONAL STUDY IN THE SEMICONDUCTOR INDUSTRY. Under review

[35]Dreborg KH. Essence of backcasting. Futures 1996; 28(9): 813–28.

[36] EC High-Level Expert Group on AI: "Trustworthy AI Assessment List", July 17th, 2020; (https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment)

[37] Grunwald A. Humans and Machines – Challenges of Artificial Intelligence. Opinion of German Ethics Council. Institut für Technikfolgenabschätzung und Systemanalyse (ITAS), Karlsruher Institut für Technologie (KIT)

[38] Kriebitz A. Protecting and Realizing Human Rights in the Context of Artificial Intelligence: A Problem Statement. [Online]. Available: PDF. Accessed: April 24, 2024.

[39] **UNESCO.** *Recommendation on the Ethics of Artificial Intelligence*. [Online].

[40] OECD. (2019). *OECD Principles on Artificial Intelligence*. Retrieved from OECD AI Principles

[41] ISO/IEC. (2023). *Artificial intelligence — Management system* (ISO/IEC 42001:2023)

6. Appendices

6.1 Appendix A - Abbreviations

Table 4: Abbreviations

Abbreviation	Meaning
AIMS5.0	Artificial Intelligence in Manufacturing leading to Sustainability and Industry5.0
Arrowhead fPVN	Arrowhead flexible Production Value Network
EC HLEG	European Commission's High-Level Expert Group
SDG	Sustainable development goal
(e)TAM	(Extended) Technology Acceptance Model
TOE	Technological, Organizational, and Environmental
GDPR	General Data Protection Regulation
DSA	Digital Services Act
DMA	Digital Markets Act
UTAUT	The Unified Theory of Acceptance and Use of Technology
EurAI	European Association for Artificial Intelligence
CSIRO	The Commonwealth Scientific and Industrial Research Organization

6.2 Appendix B – Interview questions to inquire "The Landscape of AI in production"

Setting the Scene - Organizational Background

The first set of questions helps us to gain insights about your organizational background and your current role.

1. **Could you please describe your organization?** [*Remark for interviewer: does the interviewee work in a company, research institution, etc.*]

[If not mentioned already, also ask:]

2. **Where is the organization / base of operation located?** (Country)
3. **How many employees does the organization have** (roughly)**?**
4. **What does the organization manufacture? (if not manufacturing yourself .. which topics do you support in use cases in collaboration with partners)**
5. **What is your current role in the organization?**

Setting the Scene - Artificial Intelligence

6. **What do you personally understand by artificial intelligence?**
7. **How do you distinguish between AI applications? Do you consider AI operated by hyperscalers (e.g., ChatGPT, MS365) or do you use AI on-premise?** (Note: e.g., differentiate between standardized tools and internally developed customized tools)

Utilization of AI and types of AI

The following questions concern the type and utilization of AI.

[If the interviewee works in industry]: Please answer the following questions with regards to the use of AI in production specifically.

[If the interviewee works in any other organization]: For the following questions, please refer to use cases, cooperation, and other applications in AI you are involved in.

8. **How and in which areas of production, or in areas related to production, does the organization utilize AI technologies? (if not manufacturing yourself .. which topics do you support in use cases in collaboration with partners)**
9. **Which types of AI are being used in the organization´s production processes?**
10. **Are there specific examples of projects or applications where AI has been particularly successful?** (Remark for Interviewer: Focus on AI in the production process of organizations)

Implementation of AI in the Production Process

11. **Are AI applications rather initiated top-down or bottom-up?**
12. **Who makes decisions about which types of AI are being used in the organization?**
13. **How have AI technologies (in production) been integrated into the business processes?**
14. **What challenges have you encountered in implementing AI in the organization?**
15. **How did you overcome these challenges or what solutions were found?**

AI in practice

16. **What are the benefits of AI applications in the production process in the organization?**
17. **How is AI contributing to the profitability of the organization?**
18. **How is AI contributing to the sustainability of the organization?**
19. **Which risks do you see when using AI in the production process of the organization?** (E.g. data protection, ethical decision making, technical risks**) if not manufacturing yourself .. which topics do you support in use cases in collaboration with partners**
20. **How are potential risks handled in the organization?**

Outlook

21. **Do you plan any enhancements to existing AI projects in the organization´s production process?**
22. **Do you plan new AI-related projects in the organization´s production process?**
23. **Is there anything else you would like to add or that is important to understanding the importance of AI in the production process?**
24. **Are there any other resources or contacts you can recommend for learning more about AI in the production process?**