Research Article

Beyond Algorithms: A Framework for Ethical Assessment of AI Applications in Industry

Enas Aref[1]

1. Western Michigan University, United States

Artificial intelligence (AI) is rapidly transforming industries and societies worldwide. While AI offers remarkable benefits, its development and deployment raise profound ethical concerns. Issues such as algorithmic bias, lack of transparency, privacy violations, and potential job displacement necessitate a comprehensive and proactive approach to ethical AI development. This paper presents a novel framework for the ethical assessment of AI applications. This framework moves beyond technical compliance towards an approach that is ethically driven. The cornerstone of the development of this framework are the ethical pillars of trustworthy AI established by the European Union (EU). These encompass Human agency and oversight, Technical Robustness and Safety, Privacy and Data Governance, Transparency, Diversity, Non-discrimination, and Fairness, Societal and Environmental well-being, and Accountability. Since this framework is developed to support industries, the assessment is further classified by a thematic structure to follow a human-centric approach and to ensure meeting organizational needs. These themes include Ethical Governance and Accountability, Operational Procedures and Security, Human-AI interaction, Data Management and Governance, and finally Stakeholder Engagement and Environmental Impact. To ensure adaptability and relevance, it considers the dynamic nature of AI development and the complex interplay of stakeholders within industrial settings. The proposed framework provides a structure for evaluating the ethical implications of AI application. It aids organizations in aligning AI systems integration with the organization's core values, building trust, and ultimately enhancing the overall benefits of AI adoption in the industrial sphere.

 doi.org/10.32388/3RHK4F

Introduction

The integration of artificial intelligence (AI) in various industries presents both opportunities and challenges. With numerous advantages to the use of AI including increased efficiency, productivity, and innovation (Flatscher et al., 2023), it is valid to predict an exponential spread of AI adoption across industries (Kelly et al., 2022). In the Business Trends and Outlook Survey data (BTOS) published by the United States Census Bureau, only 2.8% of the surveyed organizations in the "Manufacturing" sector indicated that they used AI in producing goods or services (*BTOS - Data*, 2024). According to the same survey, the percentage of organizations in the "Management of Companies" sector that reported the use of AI was 7.8% (*BTOS - Data*, 2024). Other industries reported a more elaborate use of AI in their operations including the financial industry (Flatscher et al., 2023; Kelly et al., 2022; Svetlova, 2022) as well as the healthcare industry (Goirand et al., 2021; Secinaro et al., 2021).

While the development of AI systems is still in progress, a significant amount of effort is dedicated to overcoming the challenges and implications of AI adoption through research, policies, and guidelines (Flatscher et al., 2023). One of the key challenges that has drawn a lot of attention is the lack of reflection upon ethical implications, risks, and challenges of adopting AI (Munoko et al., 2020; Secinaro et al., 2021; Svetlova, 2022). Ethics has been a central focus of philosophical study since the dawn of human consciousness, with seminal works such as Aristotle's Nicomachean Ethics laying the foundational theories that continue to influence our understanding of moral principles today. Aristotle described ethical virtue in his work that dates to 350 BC as "Ethical virtue is fully developed only when it is combines with practical wisdom" (Kraut, 2022). It is essential to point out that even prior to the era of AI**, e**thics was recognized as a crucial component to business excellence in conjunction with leadership, teamwork, effective communication, performance management and high productivity (Mercader et al., 2021). Recent research emphasized the crucial need to establish methods to "translate principles into practice" when developing and sustaining AI ethics (Mittelstadt, 2019) while considering the systemic effects of AI use (Svetlova, 2022).

This paper shows the assimilation of work on developing a framework to assess the ethics of AI applications in industry, specifically manufacturing. This paper consists of the following sections: Background, Framework Development, Framework Application, Discussion, and Conclusion.

Background

Artificial Intelligence Systems

The term Artificial Intelligence (AI) has been introduced by Alan Turing in the 1950s. Since then, AI has evolved and crossed the boarders of scientific laboratories to exist in the daily life of many people. Since the term AI has been coined, the definition of AI has changed to be as comprehensive and descriptive as possible of the levels of innovation, integration and collaboration with humas. Another important aspect of the definition of AI is its appropriateness for regulation (Bezerra et al., 2024). AI can be simply defined as the skill of a non-human system to simulate, or in other words mimic, human intelligence (Munoko et al., 2020). While the capabilities of AI systems are expansive, the fact is that each AI program works only to solve one critical problem that it was designed to solve (Plathottam et al., 2023). With that view in mind, the term artificial narrow intelligence (ANI) is sometimes used to describe those programs or AI elements even though those programs are structured based on several techniques (Plathottam et al., 2023). A definition for the general purpose AI system (GPAIS) has been introduced by the European Union (EU) in 2021 as "AI system able to perform generally applicable functions such as image/speck recognition, audio/video generation, pattern detection, question answering, translation, etc." (Council of the European Union, 2021).

Several AI systems are used in various industries. Each AI system achieves a different goal and is used in a distinct application based on the industry and the functional area. Each AI system brings a set of challenges as well. While the focus of this paper is on industrial manufacturing, it is worthy to explore AI applications in other industries. Exhibit 1 provided a comprehensive list of AI tools, applications and challenges for different industries.

Industry / Function	Tool	Application	Challenges
Manufacturing	Machine learning Cyber-physical systems IoT AI	Defect detection Quality Inspection Workplace Safety Operation optimization Automation Integrated systems management Real-time monitoring and control Predictive maintenance	Increased interoperability Data acquisition Data management Human resources Infrastructure Security Trust Implementation
Human Resources Management	AI-based hiring algorithm	Decision Making Enhancing recruitment efficiency	Bias in decision-making, Data security
Administrative	Automation tools	Streamlining administrative tasks	Adaptation by staff, Cost of implementation
Accounting and Financial	AI and big data	Fraud detection	Data security, Compliance with financial regulations
Leadership and Management	Decision support systems	Strategic decision making	Over-reliance on automated systems, Loss of human intuition in decisions
Research & Development	Computational modeling, simulation	Speeding up innovation	High computational costs, Complexity of models
Product Design	CAD and AI integration	Automated design adjustments	Balancing automation with creative input, Technology integration challenges, copyright and proprietary concerns

Exhibit 1. Summary of AI tools and systems in Industry.

doi.org/10.32388/3RHK4F

Ethics of Using AI Systems

The European Union (EU) recognized the importance of developing AI systems that are human-centric (European Commission, 2019). A similar viewpoint was adopted by the UNESCO when they proposed a human-rights approach to AI ethics (*Ethics of Artificial Intelligence | UNESCO*, 2022). Trustworthy AI emerged as a term to describe the characteristics of an AI system throughout its life cycle considering both technical and social perspectives. A trustworthy AI system should be lawful, ethical and robust (Diaz-Rodriguez et al., 2023). Trustworthy AI frameworks aim to manage risks associated with AI technologies and build trust among users, stakeholders, and society at large, however, there is still a lack of trust in these systems (Vianello et al., 2023). Organizations aiming to achieve trustworthy AI should prioritize aspects such as reliability, self-assessment, consideration of user interests and values, and information disclosure regarding the system's inner workings and objectives (Freiman, 2023).

The (EU) developed a framework to classify the risks associated with using AI systems. The framework consists of four different categories: low risk, limited risk, high risk, and unacceptable risk (*AI Act | Shaping Europe's Digital Future*, 2024). To evaluate the ethics of using AI systems in industry or manufacturing, several standards and tools have been developed. One significant standard is ISO 42001, which offers guidance for establishing, implementing, maintaining, and continually improving an AI management system within an organization(Golpayegani et al., 2023). Other frameworks, like the TAI-PRM, followed a human-centric approach to assess the ethical implication of AI systems (Vyhmeister & Castane, 2024). This framework blended ethical considerations and already established standards to "identify failure modes associated with AI artefacts" (Vyhmeister & Castane, 2024). Prior to those efforts, the European Commission's High-Level Expert Group of Artificial Intelligence developed the Assessment List For Trustworthy Artificial Intelligence (ALTAI) in 2020 (Commission et al., 2020). The ALTAI was developed based on the ethical guidelines established by the EU. The analysis conducted by Golpayegani et al. (2023) explained two characteristics of ALTAI. The first is that it does not cover several processes at the upper organizational governance level, the second is that ALTAI looks at the results of executed activities (Golpayegani et al., 2023).

Need for Ethics Tool

While new technological advancements are always encouraged by industry and developers for all the advantages it brings, such advancements are also meticulously examined on the ethical implication of

their adoption (D'Cruz et al., 2022). In the past 200 years, technological advancements brought along ethical implications (Nguyen et al., 2020), from the invention of the printing press (D'Cruz et al., 2022), to the most recent integration of AI technology in industry (Artkin, 2022). The Blueprint for an AI bill of Rights which was released by the White House in 2022 clarified the need for monitoring, auditing and conducting risk assessment before and after deployment of AI systems (OSTP, 2022). While several initiatives emphasized on the importance of the ethical development of AI systems (Mittelstadt, 2019), efforts are still in progress to properly define general purpose AI systems (GPAIS) in the first place (Gutierrez et al., 2023). Governance of AI systems in industry is a critical aspect that goes beyond mere legal compliance. Organizations are encouraged to proactively assess the risks associated with their AI systems, however assessment mechanisms are still lacking (Minkkinen et al., 2024).

A stakeholder-centered approach is essential for ensuring that AI is developed and utilized responsibly, considering the perspectives and interests of various stakeholders(Ademola, 2024). The need for ethical guidelines in AI is emphasized by the identification of important ethical considerations such as bias, data privacy, and labor market impact, highlighting the importance of a human-centered approach and engagement with local stakeholders (Huriye, 2023). Collaboration among diverse stakeholders, including governments, businesses, academia, and society, is vital to promote responsible and equitable AI practices, emphasizing the urgency of balancing technological advancement with ethical considerations (N. Li, 2023). Responsible AI practices require a multi-stakeholder ethical framework that involves identifying ethical issues and prescribing ethical principles for different functions of AI systems(Prikshat et al., 2022). Furthermore, the implementation of ethics in AI systems necessitates inclusive conversations among stakeholders to ensure transparency and ethical decision-making processes (Goirand, 2023).

The goal of this paper is to contribute to those initiatives of pioneering governance of AI deployment and usage in industry. This paper presents a robust framework to assess the ethical compliance of AI systems. This framework serves as an objective decision-making tool when considering procuring, deploying and /or using AI systems in industry specially manufacturing. As AI systems continue to develop, we are expecting an evolution in both their capabilities and applications. Also, the clarification of the characteristics of trustworthy AI systems will continue to develop (Golpayegani et al., 2023). This fact has been considered when developing this framework by designing it to be agile enough to adapt to those anticipated advancements. This work is established based on the

foundational knowledge on the ethics of AI, which makes the content of this framework an effective educational reference that can be used in training employees on the ethics of using AI systems.

Framework Development

The ethical development of AI systems in manufacturing requires a systemic, proactive perspective that considers the ethical implications and involves a diverse set of stakeholders to shape the ethics of AI in ways that align with their priorities. By engaging stakeholders, promoting transparency, and addressing ethical considerations, the manufacturing sector can harness the benefits of AI technology while upholding ethical standards.

As discussed earlier, the criteria for a trustworthy AI system must be considered in the full life cycle of the system(Diaz-Rodriguez et al., 2023). This includes the phases designing, development, deployment, and operation (B. Li et al., 2023). The framework presented in this paper focuses only on assessing the trustworthiness of AI systems in the phases of deployment and operation.

The foundation of the ethical guidelines considered in developing this framework were established based on the recommendations of the European Commission for "Trustworthy AI" (European Commission, 2019) and the UNESCO recommendations on the human-rights centered approach to the Ethics of AI (Diaz-Rodriguez et al., 2023; *Ethics of Artificial Intelligence | UNESCO*, 2022). The pillars considered in developing this framework are listed in Exhibit 2. While ALTAI was developed around the same pillars of trustworthy AI that are considered in this framework, there are subtle differences between the two frameworks. Specifically, this proposed framework is organized by themes, is agile to change and considers future advancements of AI systems and ensures stakeholder engagement.

Ethical Pillar	AI System Requirement	Main Evaluation Criteria
Human agency and oversight	To empower human beings and allow them to make informed decisions	Fundamental rights- Human agency- Human oversight
Technical robustness and safety	To be resilient, secure, to ensure accuracy, and prevent unintentional harm	Resilience to attach and security- Fallback plan and general safety – Accuracy- Reliability and reproducibility
Privacy and data governance	To ensure full respect for privacy, quality and integrity of data, and data protection and adequate data governance mechanisms	Respect for privacy and data protection- Quality of integrity of data- Access to data
Transparency	The system, the model and the output decision must be transparent, traceable, and explainable	Traceability- Explainability- Communication
Diversity, non-discrimination, and fairness	To avoid unfair bias, foster diversity, and accessibility to all	Unfair bias avoidance- Accessibility and universal design- Stakeholder participation
Societal and environmental well-being	To monitor the effects of AI systems on human, social, and environmental interactions	Sustainable and environmentally friendly- Social impact
Accountability	To ensure responsibility for development, deployment and use of AI systems	Auditability- Trade-offs – Redress

Exhibit 2. Pillars of the Ethical Framework.

The evaluation criteria developed for the proposed framework are further classified in five themes beyond the seven ethical pillars. This facilitates the evaluation process, consolidates the conclusion, and identify areas where prompt actions are required. The five themes are: Ethical Governance and Accountability, Operational Procedures and Security, Human-AI interaction, Data Management and Governance, and finally Stakeholder Engagement and Environmental Impact. See Exhibit 3. for more details. Classifying the evaluation criteria into those proposed themes facilitates the evaluation

process and enhances the organization's potential to improve the areas that are nonconforming with the ethical requirements.

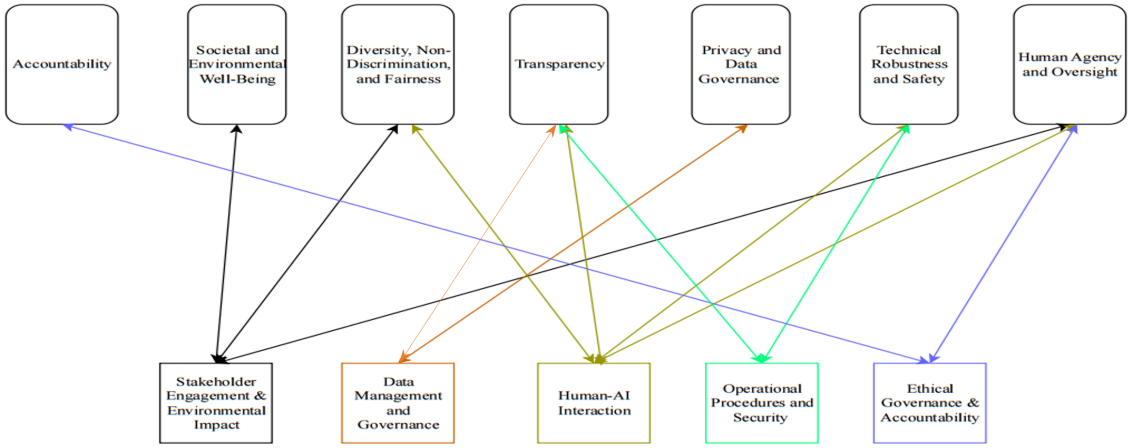


Exhibit 3. Cascading AI Pillars into Themes

The assessment is conducted by responding to a series of questions covering the main evaluation criteria listed in Exhibit 2. The detailed evaluation criteria along with all the questions are included in the Appendix. All questions are equally weighted, and each question can be answered by choosing one of two scores as shown in Exhibit 4. A score of one means that the requirement is met, a score of zero means that the requirement is not met. An option to answer as N/A is included for all questions, this option is selected when the criterion being investigated is not applicable A total score is populated automatically along with a percentage of the total possible score. The total possible score is calculated after excluding all the questions to which the assessor answered N/A.

One of the key strengths of this framework is that it is agile enough to adapt to future changes introduced by technology advancement in the development of AI systems as well as the changes in the functionality of applications of those systems. This framework is structured with future expansion in mind to accommodate for anticipated ethical requirements. This can be achieved by introducing additional ethical pillars, rearranging the evaluation into additional or different themes, introducing additional evaluation criteria or rendering certain questions obsolete and eliminating them from the assessment.

INDIVIDUAL QUESTION CRITERIA	SCORE
Observations indicate high risk - minimal standards, systems in place	0
Observations indicate low/minimal risk - standards/systems appear robust and effective	1

Exhibit 4. Individual Question Criteria

Framework Application

The execution of the framework is conducted through a four-stage process, commencing with the scoping and planning of the assessment and culminating in a definitive determination regarding the assessment's outcomes.

Stage 1: Team Formation and Training

A cross functional team is always encouraged to foster creativity, innovation and promote diversity (Aref, 2024). The team collectively should possess several skills including critical thinking (Aref & Fredericks, 2021) in addition to experience in their functional area within the organization. The team should ideally consist of members with diverse expertise to ensure a comprehensive assessment process. We suggest that the team members should collectively possess knowledge in the following areas: quality management (Aguiar et al., 2015), continuous improvement (Ahmed et al., 2021, p. 14001), human resources (Zografaki et al., 2017), risk management (Barafort et al., 2016), research and development (Chen et al., 2022), information technology (Wibowo & Ramli, 2022) and cyber security (Ben-Asher & González, 2015; Yusuff Marican et al., 2023), and technical skills relevant to the organization’s scope of work.

At the leadership level, digital leadership is highly encouraged as a construct to identify the role of the team leader(Singh & Pandey, 2024). While the concept of digital leadership has been introduced during the era of digital transformation, the definition introduced by Eberl and Drews (2021) provides a comprehensive view of the role and skills of the digital leader with emphasis on their role in leading innovation (Eberl & Drews, 2021).

To effectively run the assessment, the team must be trained in the technical and procedural aspects of the assessment. Two levels of training are encouraged, basic and specialized. The basic training

addresses the common concepts of AI, definitions, ethical considerations of using AI systems, the challenges and implications of using AI systems, and expectations and requirements of individuals when using AI systems. The basic training should be offered to every employee in the organization and is recommended to be made mandatory. The specialized training will be tailored for the assessment team that will run the ethical assessment of the AI systems. The specialized training must include information on the different purposes of the assessment and the scope of application of AI system. The trainees must learn as well how to collect data, the requirements to have the necessary clearance to collect and access the needed data, how the framework is structured and the area of focus of each question and assessment criterion. They also must be able to report the findings and make an informed decision.

Stage 2: Scoping and Planning

In this stage, the team explores and determines the scope of the assessment and plans their work. The purpose could be to conduct an on-going assessment of an AI system that is currently in use, to compare different systems in order to make informed decisions on strategic matters including procurement or adoption. The team will have access to the data needed for the assessment, completing the paperwork to document the process.

Stage 3: Assessment and Analysis

The team embarks on a detailed exploration of each AI application area, engaging deeply with the assessment questions. Each team member, selected based on their expertise aligned with the theme area, leads the evaluation in their respective domain, ensuring a thorough and expert-driven assessment. The evaluation process is collaborative, requiring consensus on answers to ensure a unified understanding and approach to the ethical implications of the AI systems assessed. This collaborative approach not only enhances the depth of the evaluation but also fosters a shared responsibility among team members to uphold ethical standards. Further assessment by an external assessor is encouraged. However, this is only applicable upon further standardization of the framework and when external assessors are trained in using this evaluation technique. Additional assessment by an external assessor may be required if the in-house team that is assigned to run the assessment lacks one or more of the key personnel with the required functional or technical skills.

Once the evaluation of all thematic areas is complete, the team conducts a comprehensive analysis of the results. This analysis involves identifying areas where the AI systems do not meet ethical requirements, as well as highlighting sections where the assessment was deemed not applicable (N/A). The team calculates the total score and critically examines compliance across the different functional areas, themes, and ethical pillars. This stage is crucial for transparently documenting the ethical posture of AI applications and preparing for any necessary interventions. See Exhibit 5. for an example of scoring summary by theme and the corresponding radar chart. An example of the scoring summary for the same assessment by ethical pillar is included in the Appendix.

Stage 4: Decision Making and Mitigation

The decision-making process is guided by the scoring system established in the framework. The overall compliance is determined based on a scoring scheme detailed in Exhibit 6. If the score per pillar and per theme is greater than 85%, the AI system is deemed ethically acceptable with no further action required. Scores between 65% and 85% indicate conditional acceptance. In these cases, the organization must develop and execute a detailed plan to address and resolve the identified ethical shortcomings to achieve full compliance. A score below 65% signals significant ethical concerns, necessitating an immediate halt in the use of the AI system. Urgent actions must be taken to rectify these violations.

Artificial Intelligence Ethics Assessment

Scoring Summary per Theme

	Criteria	Organization Self Assessment	External Assessment	Target %
		Overall	Overall	
1.0	Ethical Governance and Accountability	94.12%	0.00%	100.00%
2.0	Operational Procedures and Security	84.21%	0.00%	100.00%
3.0	Human-AI interaction	22.86%	5.71%	100.00%
4.0	Data Management and Governance	0.00%	0.00%	100.00%
5.0	Stakeholder Engagement and Environmental Impact	22.22%	0.00%	100.00%
	Total Score (%)	44.21%	2.08%	

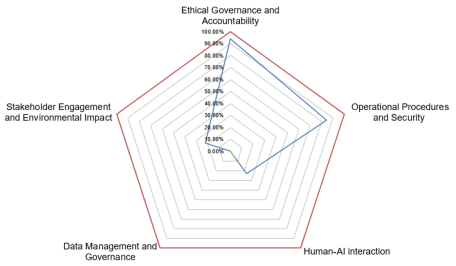


Exhibit 5. Scoring Summary by Theme

The mitigation plans are tailored to the severity of the ethical issues identified and are prioritized based on the strategic goals of the organization and the impact of the violations. The mitigation process involves assembling a dedicated team tasked with re-evaluating the AI systems and implementing solutions. This team is responsible for crafting remedial actions that not only address

immediate issues but also align with the long-term ethical deployment of AI technologies across industries.

By structuring the decision-making and mitigation processes around quantitative thresholds and expert consensus, the framework ensures that decisions are both data-driven and contextually informed. This approach not only helps in maintaining the integrity of AI applications but also supports organizations in navigating the complex ethical landscape associated with modern AI technologies.

OVERALL SCORING CRITERIA	
The overall score is calculated based on the numeric percentage of the sum of the scores for individual questions.	
Individual questions have the same weighting.	
The overall score will be downgraded based on the results of the scores for the starred Ethical Requirements questions.	
IMPORTANT: If any Core Ethical Requirements questions are rated 1 (red), the final audit score will be RED.	
Unacceptable: Overall Score < 65%: Unacceptable, **Major Ethical Noncompliance**, Immediate ethical development plan required	
Conditionally Acceptable: Overall Score >= 65% < 85%: **Conditionally acceptable**, ethical development plan required	
Acceptable: Overall Score >= 85% and no red scores: Acceptable, no ethical development plan required	

Exhibit 6. Overall Scoring Criteria

Discussion

This work presents a novel framework developed to assess the ethical considerations of AI systems used in industry. At its core, this framework adopts the ethical pillars of trustworthy AI which were developed by the EU (European Commission, 2019). These pillars along with other ethical guidelines established the basis of several other initiatives to either develop AI systems, assess their trustworthiness, or standardize the ethics of AI systems (Lewis et al., 2020). However, at an organization level, understanding the ethical requirements of AI systems was found to be fragmented (Agbese et al., 2023). The more the ethical pillars and requirements of AI systems are directly linked to the functional areas of an organization, the easier it is for the employees to comprehend the requirements and perform accordingly. Those functions may include financial viability (Agbese et al., 2023), stakeholder management (Lewis et al., 2020), and the technical capabilities of the system (Chen et al., 2022).

The significance of this framework lies in its ability to translate ethical principles into actionable criteria that are relevant to organizations. This is achieved by arranging the assessment by themes that are directly related to strategic management goals, stakeholder concerns, and the functional areas within the organization. The evaluation process is divided into four stages, starting with forming and training a team to run the assessment. The following stages include the series of actions needed to run and report the assessment and finally derive a decision that can be translated into an action plan. One of the critical challenges in developing this framework was ensuring its relevance across different types of AI applications and its adaptability to future technological advancements. To address this, the framework was designed to be dynamic, allowing for updates and modifications as new ethical standards emerge and as AI technologies evolve. This adaptability is crucial for maintaining the framework's effectiveness in promoting ethical AI practices as the landscape of AI continues to change rapidly.

The use of ethical evaluation frameworks for AI offers a multitude of benefits that are essential for the ethical advancement and implementation of artificial intelligence. We are proposing this framework to industry and practitioners as well to test and adopt. This framework helps in the early identification and mitigation of potential ethical risks associated with AI systems. Additionally, AI ethical evaluation frameworks generally, and this one specifically, significantly improve the transparency and credibility of AI systems. Moreover, evaluating the ethical implication of AI systems plays a crucial role in promoting the responsible implementation of AI across various sectors. By setting ethical standards and recommended practices, this framework assists organizations in creating and deploying AI systems that are aligned with societal values and ethical standards.

Future work will focus on empirically testing the framework in various industrial settings to refine its criteria and ensure its effectiveness in real-world applications. Additionally, expanding the framework's scope to include case studies and sector-specific adaptations will enhance its utility and provide deeper insights into the challenges and opportunities associated with implementing ethical AI systems.

Conclusion

This paper presented a robust framework for the ethical assessment of AI systems in industrial settings, emphasizing adaptability and stakeholder engagement. By anchoring on the ethical pillars of trustworthy AI, the framework offers a systematic approach to evaluate AI systems beyond mere

compliance, focusing on a human-centric and ethical integration. This framework is designed to be dynamic, accommodating rapid technological changes and varying industrial needs, ensuring its applicability across different sectors. It facilitates organizations in aligning AI deployment with ethical standards and societal values, fostering trust and ethical responsibility in AI applications. Future developments will aim to refine the framework through empirical testing and to expand its scope to address sector-specific challenges, thereby enhancing its effectiveness and relevance in promoting responsible AI practices globally.

Appendices

Artificial Intelligence Ethics Assessment
1.0 Human agency and oversight

	Criteria	Organization Self Assessment	External Assessment	Theme
1.0	**Fundamental rights**			
1.1	**Are the fundamental rights of**	ç		Ethical Governance and Accountability
1.2	Are there **action plans** to ensure risks can be reduced or justified as necessary in order to respect the rights and freedoms of others	ç		Ethical Governance and Accountability
1.3	Is there a clear identification where **AI system affect human autonomy** by interfering with the (end) user's decision-making process in an unintended way?	ç		Ethical Governance and Accountability
1.4	**Potential trade-offs made between the different principles and rights** are identified and documented?	ç		Ethical Governance and Accountability
1.5	Is there a mechanisms in place to **receive external feedback** regarding AI systems use and thier impact?	ç		Stakeholder Engagement and Environmental Impact
	Human agency			
1.6	Did you consider the task allocation between the AI system and humans for meaningful interactions and appropriate human oversight and control?	ç		Human-AI interaction
1.7	AI system enhance or augment human capabilities and support individuals in making better choices in accordance with their goals ?	ç		Human-AI interaction
1.8	Safeguards in place to prevent overconfidence in or over reliance on the AI system for work processes?	ç		Human-AI interaction
1.9	Are the users given the knowledge and tools to comprehend and interact with AI systems	ç		Human-AI interaction
	Human oversight			
1.10	Did you consider the appropriate level of human control for the particular AI system and use case?	ç		Human-AI interaction
1.11	Did you deploy mechanisms and measures to ensure human control or oversight are in place?	ç		Ethical Governance and Accountability
1.12	Did you put measures in place to enable audit and to remedy issues related to governing AI autonomy?	ç		Ethical Governance and Accountability
1.13	Specific mechanisms of control and oversigh in place for self-learning or autonomous AI systems?	ç		Ethical Governance and Accountability
1.14	Detection and response mechanisms established to assess whether something could go wrong with self-learning or autonomous AI system	ç		Ethical Governance and Accountability
1.15	Stop button or procedure is in place to safely abort an operation where needed to abort the process entirely, in part, or delegate control to a human?	ç		Ethical Governance and Accountability
	Actual Score	ç	é	
	Total Potential Score	ç	ç	
	1.0 - Management Percentage	**100.00%**	**0.00%**	

doi.org/10.32388/3RHK4F

Artificial Intelligence Ethics Assessment
2.0 - Technical robustness and safety

	Criteria	Organization Self Assessment	External Assessment	Comments
2.0	**Resilience to attack and security:**			
2.1	Is there a process for regularly updating and patching the AI system and its dependencies?	Ç		Operational Procedures and Security
2.2	Are security vulnerabilities assessed and prioritized for remediation?	Ç		Operational Procedures and Security
2.3	Is there a documented inventory of all system components and their configurations?	é		Operational Procedures and Security
2.4	Are access controls (authentication and authorization) implemented for users and system components?	>fi		Operational Procedures and Security
2.5	Are default credentials changed, and unnecessary services disabled?	Ç		Operational Procedures and Security
2.6	Are data retention policies in place to ensure unnecessary data is not stored?	Ç		Operational Procedures and Security
2.7	Is there an intrusion detection system (IDS) in place to monitor system activity?	é		Operational Procedures and Security
2.8	Is there a defined incident response plan in case of a security breach?	Ç		Operational Procedures and Security
2.9	Are users and administrators trained in security best practices and aware of potential threats?	Ç		Human-AI interaction
	Fallback plan and general safety			
2.10	Did you ensure that your system has a sufficient fallback plan in place if the AI system cannot answer your questions or provide assistance?	Ç		Operational Procedures and Security
2.11	Are you aware of your responsibilities as a user in ensuring safe and ethical interactions with the AI system?	é		Human-AI interaction
2.12	In case the AI system provides incorrect information or behaves unexpectedly, do you know how to report the issue to us?	é		Human-AI interaction
2.13	Do you know what to do if the AI system provides harmful or inappropriate content?	é		Human-AI interaction
2.14	Is there safety guidelines for using the AI system in place?	Ç		Operational Procedures and Security
	Accuracy			
2.15	Did you assess what level and definition of accuracy would be required in the context of the AI system and use case?	Ç		Operational Procedures and Security
2.16	Did you assess how accuracy is measured and assured?	Ç		Operational Procedures and Security
2.17	Did you verify what harm would be caused if the AI system makes inaccurate predictions?	Ç		Operational Procedures and Security
2.18	Did you put in place ways to measure whether your system is making an unacceptable amount of inaccurate predictions?	Ç		Operational Procedures and Security
2.19	Did you put in place a series of steps to increase the system's accuracy?	Ç		Operational Procedures and Security
	Reliability and reproducibility			
2.20	Does the AI system rely on high-quality and diverse data sources?	Ç		Operational Procedures and Security
2.21	Are there mechanisms in place to handle noisy or incomplete data?	Ç		Operational Procedures and Security
2.22	Did you put in place verification methods to measure and ensure different aspects of the system's reliability and reproducibility?	Ç		Operational Procedures and Security
2.23	Does the AI system provide clear error messages or explanations when it cannot provide a reliable output?	Ç		Operational Procedures and Security
2.24	Is there a continuous monitoring system in place to detect anomalies or deviations in the AI system's performance?	é		Operational Procedures and Security
2.25	Are there guidelines or instructions provided to users on how to replicate the AI system's results?	>fi		Operational Procedures and Security
	Actual Score	Çã	é	
	Total Potential Score	+ò	+›	
	2.0 - Waste Reduction Percentage	**73.91%**	**0.00%**	

doi.org/10.32388/3RHK4F

Artificial Intelligence Ethics Assessment

3.0 - Privacy and data governance

	Criteria	Organization Self Assessment	External Assessment	Theme
3.0	**Respect for privacy and data Protection:**			
3.1	The AI system has provided clear information about how the data will be used.			Data Management and Governance
3.2	Did you assess the type and scope of data in your data sets (for example whether they contain personal data)?			Data Management and Governance
3.3	The AI system asked for your consent before collecting your data?			Data Management and Governance
3.4	Are you aware of the types of data the AI system collects about you? (e.g., personal information, browsing history, location data, etc.)			Data Management and Governance
3.5	Do you have the option to control what data the AI system collects about you?			Data Management and Governance
3.6	Is there a clear process for you to request access to your data or request its deletion?			Data Management and Governance
	Quality and integrity of data			
3.7	Did you align your system with relevant standards (for example ISO, IEEE) for daily data management and governance?			Data Management and Governance
3.8	Did you establish oversight mechanisms for data collection, storage, processing and use?			Data Management and Governance
3.9	How are data labels and annotations created?			Data Management and Governance
3.10	Did you put in place processes to ensure the quality and integrity of your data?			Data Management and Governance
3.11	Do you track and measure data quality using relevant metrics (e.g., accuracy, completeness)?			Data Management and Governance
	Access to data			
3.12	Protocols, processes and procedures are in place and implemented to manage and ensure proper data governance ?			Data Management and Governance
3.13	Did you assess who can access users' data, and under what circumstances?			Data Management and Governance
3.14	Did you ensure that these persons are qualified and understand the details of data protection policy?			Data Management and Governance
3.15	Did you ensure an oversight mechanism to log when, where, how, by whom and for what purpose data was accessed?			Data Management and Governance
	Actual Score	ć	ć	
	Total Potential Score	Ç	Ç	
	3.0 - Internal Auditing Percentage	**0.00%**	**0.00%**	

doi.org/10.32388/3RHK4F

4.0 - Transparency

	Criteria	rganization Self Assessmen	External Assessment	Theme
	Traceability			
4.1	The AI system provides clear documentation regarding its purpose, capabilities, and limitations			Human-AI interaction
4.2	The system provides a comprehensive list of the data sources and datasets used for training			Human-AI interaction
4.3	The system conducts regular audits and assessments of its traceability and fairness.			Human-AI interaction
4.4	There are clear channel for users to escalate concerns or complaints			Human-AI interaction
	Explainability			
4.5	Do you have a clear understanding of the purpose and objectives of the AI system you are interacting with?			Human-AI interaction
4.6	Did you assess to what degree the system's decision influences the organisation's decision-making processes?			Human-AI interaction
4.7	Do you have a clear understanding of the data sources and types of data used by the AI system?	Ç	Ç	Human-AI interaction
4.8	Did you assess how the AI system makes decisions or predictions?	>fi	>fi	Human-AI interaction
4.9	Can you easily access detailed explanations when you need them while interacting with the AI system?			Human-AI interaction
	Communication			
4.10	Did you communicate this clearly and intelligibly to the intended audience?			Human-AI interaction
4.11	Did you establish processes that consider users' feedback and use this to adapt the system?			Human-AI interaction
4.12	Did you communicate around potential or perceived risks, such as bias?			Human-AI interaction
4.13	Did you consider communication and transparency towards other audiences, third parties or the general public?			Human-AI interaction
4.14	Did you clearly communicate characteristics, limitations and potential shortcomings of the AI system?			Human-AI interaction
	Actual Score	Ç	Ç	
	Total Potential Score	Çò	'	
	4.0 - Customer PPAP Percentage	**7.69%**	**12.50%**	

doi.org/10.32388/3RHK4F

Artificial Intelligence Ethics Assessment
5.0 - Diversity, non-discrimination and fairness

	Criteria	rganization Self Assessmen	External Assessment	Theme
	Unfair bias avoidance			
5.1	Have you implemented ongoing monitoring for potential bias in your AI system's predictions or decisions?			Human-AI interaction
5.2	Have you documented all steps taken to address bias in your AI system?			Human-AI interaction
5.3	Did you establish clear steps and ways of communicating on how and to whom such issues can be raised?			Human-AI interaction
5.4	Did you consider others, potentially indirectly affected by the AI system, in addition to the (end)-users?			Human-AI interaction
5.5	Did you ensure an adequate working definition of "fairness" that you apply using the AI systems?			Human-AI interaction
5.6	Did you establish mechanisms to ensure fairness in your AI systems?			Human-AI interaction
	Accessibility and universal design			
5.7	Did you take the impact of your AI system on the potential user audience into account	Ç	Ç	Human-AI interaction
5.8	Did you evaluate the accessibility features and usability of AI system.			Human-AI interaction
5.9	Did you assess whether there could be persons or groups who might be disproportionately affected by negative implications?			Human-AI interaction
5.10	Did you get feedback from other teams or groups that represent different backgrounds and experiences?			Stakeholder Engagement and Environmental Impact
5.11	The documentation for the AI system includes accessibility instructions.			Human-AI interaction
5.12	The AI system encourages users to provide feedback on accessibility issues.			Human-AI interaction
5.13	There is a visible commitment to improving accessibility based on user feedback			Human-AI interaction
5.14	Is it representative of the wider population, considering also of other groups who might tangentially be impacted?			Human-AI interaction
	Stakeholder participation			
5.15	Did you consider a mechanism to include the participation of different stakeholders in the AI system's use?	é	é	Stakeholder Engagement and Environmental Impact
5.16	Did you ensure informing and involving impacted workers and their representatives in advance?	>fi	>fi	Stakeholder Engagement and Environmental Impact
5.17	Stakeholders were informed about the data sources			Stakeholder Engagement and Environmental Impact
5.18	Stakeholders were consulted on ethical guidelines and principles for the AI system			Stakeholder Engagement and Environmental Impact
	Actual Score	Ç	Ç	
	Total Potential Score	Çë	Çë	
	5.0 - Corrective / Preventive Action Percentage	**5.88%**	**5.88%**	

Artificial Intelligence Ethics Assessment
6.0 - Societal and environmental well-being

	Criteria	rganization Self Assessmen	External Assessment	Theme
6.0	**Sustainable and environmentally friendly AI**			
6.1	Did you establish mechanisms to measure the environmental impact of the AI system's deployment and use (for example the type of energy used by the data centres)?	Ç		Stakeholder Engagement and Environmental Impact
6.2	Did you ensure measures to reduce the environmental impact of your AI system's life cycle?			Stakeholder Engagement and Environmental Impact
6.3	Did you evaluatem the impact of the AI system throughout thr life cycle on energy consumption and reduce its carbon footprint and resource consumption (e.g., server farms, rare materials, etc.)??			Stakeholder Engagement and Environmental Impact
	Social impact			
6.4	Did you evaluate the impact on the broader society, other sentient beings and the environment throughout the AI system's life cycle			Stakeholder Engagement and Environmental Impact
	Actual Score	Ç	é	
	Total Potential Score	Ÿ	Ÿ	
	6.0 - Engineering / Testing Percentage	**25.00%**	**0.00%**	

doi.org/10.32388/3RHK4F

Artificial Intelligence Ethics Assessment
7.0 - Accountability

	Criteria	Organization Self Assessment	External Assessment	Theme
	Auditability			
7.1	Are AI systems independently audited and evaluated?	ć		Ethical Governance and Accountability
7.2	Evaluation by internal and external auditors are planned and conducted reguraly	Ç		Ethical Governance and Accountability
7.3	Action plans are in place for the findings out of of the internal and external audits	Ç		Ethical Governance and Accountability
	Trade-offs			
7.4	In situations in which no ethically acceptable trade-offs can be identified, the deployment and use of the AI system is not proceeded in that form	Ç		Ethical Governance and Accountability
7.5	Any decision about which trade-off to make are reasoned and properly documented.	Ç		Ethical Governance and Accountability
7.6	The decision-maker must be accountable for the manner in which the appropriate trade-off is being made	Ç		Ethical Governance and Accountability
7.7	The decision-maker must continually review the appropriateness of the resulting decision to ensure that necessary changes can be made to the system	Ç		Ethical Governance and Accountability
	Redress			
7.8	Do you have an accessible mechanism to ensure adequate redress is deployed if unject adverse impact occurs?	Ç		Ethical Governance and Accountability
	Actual Score	ë	ć	
	Total Potential Score	'	'	
	7.0 - Program Management Percentage	**87.50%**	**0.00%**	

Assessment questions organized by Theme

		Ethical Governance and Accountability		
1	1.1	Are the fundamental rights of	Ç	ć
2	1.2	Are there action plans to ensure risks can be reduced or justified as necessary in order to respect the rights and freedoms of others	Ç	ć
3	1.3	Is there a clear identification where AI system affect human autonomy by interfering with the (end) user's decision-making process in an unintended way?	Ç	ć
4	1.4	Potential trade-offs made between the different principles and rights are identified and documented?	Ç	ć
5	1.11	Did you deploy mechanisms and measures to ensure human control or oversight are in place?	Ç	ć
6	1.12	Did you put measures in place to enable audit and to remedy issues related to governing AI autonomy?	Ç	ć
7	1.13	Specific mechanisms of control and oversigh in place for self-learning or autonomous AI systems?	Ç	ć
8	1.14	Detection and response mechanisms established to assess whether something could go wrong with self-learning or autonomous AI system	Ç	ć
9	1.15	Stop button or procedure is in place to safely abort an operation where needed to abort the process entirely, in part, or delegate control to a human?	Ç	ć
10	7.1	Are AI systems independently audited and evaluated?	ć	ć
11	7.2	Evaluation by internal and external auditors are planned and conducted reguraly	Ç	ć
12	7.3	Action plans are in place for the findings out of of the internal and external audits	Ç	ć
13	7.4	In situations in which no ethically acceptable trade-offs can be identified, the deployment and use of the AI system is not proceeded in that form	Ç	ć
14	7.5	Any decision about which trade-off to make are reasoned and properly documented.	Ç	ć
15	7.6	The decision-maker must be accountable for the manner in which the appropriate trade-off is being made	Ç	ć
16	7.7	The decision-maker must continually review the appropriateness of the resulting decision to ensure that necessary changes can be made to the system	Ç	ć
17	7.8	Do you have an accessible mechanism to ensure adequate redress is deployed if unject adverse impact occurs?	Ç	ć
		Actual Score	Ç	ć
		Total Potential Score	Çë	Çë
		7.0 - Program Management Percentage	**94.12%**	**0.00%**

		Operational Procedures and Security		
18	2.1	Is there a process for regularly updating and patching the AI system and its dependencies?	Ç	é
19	2.2	Are security vulnerabilities assessed and prioritized for remediation?	Ç	é
20	2.3	Is there a documented inventory of all system components and their configurations?	é	é
21	2.4	Are access controls (authentication and authorization) implemented for users and system components?	>fi	é
22	2.5	Are default credentials changed, and unnecessary services disabled?	Ç	é
23	2.6	Are data retention policies in place to ensure unnecessary data is not stored?	Ç	é
24	2.7	Is there an intrusion detection system (IDS) in place to monitor system activity?	é	é
25	2.8	Is there a defined incident response plan in case of a security breach?	Ç	é
26	2.10	Did you ensure that your system has a sufficient fallback plan in place if the AI system cannot answer your questions or provide assistance?	Ç	é
27	2.14	Is there safety guidelines for using the AI system in place?	Ç	é
28	2.15	Did you assess what level and definition of accuracy would be required in the context of the AI system and use case?	Ç	é
29	2.16	Did you assess how accuracy is measured and assured?	Ç	é
30	2.17	Did you verify what harm would be caused if the AI system makes inaccurate predictions?	Ç	é
31	2.18	Did you put in place ways to measure whether your system is making an unacceptable amount of inaccurate predictions?	Ç	é
32	2.19	Did you put in place a series of steps to increase the system's accuracy?	Ç	é
33	2.20	Does the AI system rely on high-quality and diverse data sources?	Ç	é
34	2.21	Are there mechanisms in place to handle noisy or incomplete data?	Ç	é
35	2.22	Did you put in place verification methods to measure and ensure different aspects of the system's reliability and reproducibility?	Ç	é
36	2.23	Does the AI system provide clear error messages or explanations when it cannot provide a reliable output?	Ç	é
37	2.24	Is there a continuous monitoring system in place to detect anomalies or deviations in the AI system's performance?	é	é
		Actual Score	Ç	é
		Total Potential Score	Ç½	÷é
		7.0 - Program Management Percentage	**84.21%**	**0.00%**

doi.org/10.32388/3RHK4F

		Human-AI interaction		
38	1.6	Did you consider the task allocation between the AI system and humans for meaningful interactions and appropriate human oversight and control?	Ç	é
39	1.7	AI system enhance or augment human capabilities and support individuals in making better choices in accordance with their goals ?	Ç	é
40	1.8	Safeguards in place to prevent overconfidence in or over reliance on the AI system for work processes?	Ç	é
41	1.9	Are the users given the knowledge and tools to comprehend and interact with AI systems	Ç	é
42	1.10	Did you consider the appropriate level of human control for the particular AI system and use case?	Ç	é
43	2.9	Are users and administrators trained in security best practices and aware of potential threats?	Ç	é
44	2.11	Are you aware of your responsibilities as a user in ensuring safe and ethical interactions with the AI system?	é	é
45	2.12	In case the AI system provides incorrect information or behaves unexpectedly, do you know how to report the issue to us?	é	é
46	2.13	Do you know what to do if the AI system provides harmful or inappropriate content?	é	é
47	4.1	The AI system provides clear documentation regarding its purpose, capabilities, and limitations	é	é
48	4.2	The system provides a comprehensive list of the data sources and datasets used for training	é	é
49	4.3	The system conducts regular audits and assessments of its traceability and fairness.	é	é
50	4.4	There are clear channel for users to escalate concerns or complaints	é	é
51	4.5	Do you have a clear understanding of the purpose and objectives of the AI system you are interacting with?	é	é
52	4.6	Did you assess to what degree the system's decision influences the organisation's decision-making processes?	é	é
53	4.7	Do you have a clear understanding of the data sources and types of data used by the AI system?	Ç	Ç
54	4.8	Did you assess how the AI system makes decisions or predictions?	>fi	>fi
55	4.9	Can you easily access detailed explanations when you need them while interacting with the AI system?	é	é
56	4.10	Did you communicate this clearly and intelligibly to the intended audience?	é	é
57	4.11	Did you establish processes that consider users' feedback and use this to adapt the system?	é	é
58	4.12	Did you communicate around potential or perceived risks, such as bias?	é	é
59	4.13	Did you consider communication and transparency towards other audiences, third parties or the general public?	é	é
60	4.14	Did you clearly communicate characteristics, limitations and potential shortcomings of the AI system?	é	é
54	5.1	Have you implemented ongoing monitoring for potential bias in your AI system's predictions or decisions?	é	é
55	5.2	Have you documented all steps taken to address bias in your AI system?	é	é
56	5.3	Did you establish clear steps and ways of communicating on how and to whom such issues can be raised?	é	é
57	5.4	Did you consider others, potentially indirectly affected by the AI system, in addition to the (end)- users?	é	é
58	5.5	Did you ensure an adequate working definition of "fairness" that you apply using the AI systems?	é	é
59	5.6	Did you establish mechanisms to ensure fairness in your AI systems?	é	é
60	5.7	Did you take the impact of your AI system on the potential user audience into account	é	é
61	5.8	Did you evaluate the accessibility features and usability of AI system.	Ç	Ç
62	5.9	Did you assess whether there could be persons or groups who might be disproportionately affected by negative implications?	é	é
64	5.11	The documentation for the AI system includes accessibility instructions.	é	é
65	5.12	The AI system encourages users to provide feedback on accessibility issues.	é	é
67	5.13	There is a visible commitment to improving accessibility based on user feedback	é	é
67	5.14	Is it representative of the wider population, considering also of other groups who might tangentially be impacted?	é	é
		Actual Score	'	÷
		Total Potential Score	ò›	ò›
		7.0 - Program Management Percentage	**22.86%**	**5.71%**

		Stakeholder Engagement and Environmental Impact		
87	1.5	Is there a mechanisms in place to receive external feedback regarding AI systems use and thier impact?	Ç	é
88	5.10	Did you get feedback from other teams or groups that represent different backgrounds and experiences?	é	é
89	5.15	Did you consider a mechanism to include the participation of different stakeholders in the AI system's use?	é	é
90	5.16	Did you ensure informing and involving impacted workers and their representatives in advance?	>fi	>fi
91	5.17	Stakeholders were informed about the data sources	é	é
92	5.18	Stakeholders were consulted on ethical guidelines and principles for the AI system	é	é
93	6.1	Did you establish mechanisms to measure the environmental impact of the AI system's deployment and use (for example the type of energy used by the data centre	Ç	é
94	6.2	Did you ensure measures to reduce the environmental impact of your AI system's life cycle?	é	é
95	6.3	Did you evaluatem the impact of the AI system throughout thr life cycle on energy consumption and reduce its carbon footprint and resource consumption (e.g., serv	é	é
96	6.4	Did you evaluate the impact on the broader society, other sentient beings and theenvironment throughout the AI system's life cycle	é	é
		Actual Score	÷	é
		Total Potential Score	½	½
		7.0 - Program Management Percentage	**22.22%**	**0.00%**

doi.org/10.32388/3RHK4F

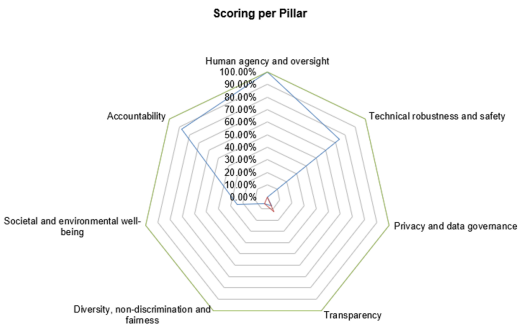
		Data Management and Governance		
72	3.1	The AI system has provided clear information about how the data will be used.	ć	ć
73	3.2	Did you assess the type and scope of data in your data sets (for example whether they contain personal data)?	ć	ć
74	3.3	The AI system asked for your consent before collecting your data?	ć	ć
75	3.4	Are you aware of the types of data the AI system collects about you? (e.g., personal information, browsing history, location data, etc.)	ć	ć
76	3.5	Do you have the option to control what data the AI system collects about you?	ć	ć
77	3.6	Is there a clear process for you to request access to your data or request its deletion?	ć	ć
78	3.7	Did you align your system with relevant standards (for example ISO, IEEE) for daily data management and governance?	ć	ć
79	3.8	Did you establish oversight mechanisms for data collection, storage, processing and use?	ć	ć
80	3.9	How are data labels and annotations created?	ć	ć
81	3.10	Did you put in place processes to ensure the quality and integrity of your data?	ć	ć
82	3.11	Do you track and measure data quality using relevant metrics (e.g., accuracy, completeness)?	ć	ć
83	3.12	Protocols, processes and procedures are in place and implemented to manage and ensure proper data governance ?	ć	ć
84	3.13	Did you assess who can access users' data, and under what circumstances?	ć	ć
85	3.14	Did you ensure that these persons are qualified and understand the details of data protection policy?	ć	ć
86	3.15	Did you ensure an oversight mechanism to log when, where, how, by whom and for what purpose data was accessed?	ć	ć
		Actual Score	ć	ć
		Total Potential Score	ç	ç
		7.0 - Program Management Percentage	**0.00%**	**0.00%**

Scoring Summary per Pillar

Artificial Intelligence Ethics Assessment

Scoring Summary per Pillar

	Criteria	Organization Self Assessment	External Assessment	Target %
		Overall	Overall	
1.0	Human agency and oversight	100.00%	0.00%	100.00%
2.0	Technical robustness and safety	73.91%	0.00%	100.00%
3.0	Privacy and data governance	0.00%	0.00%	100.00%
4.0	Transparency	7.69%	12.50%	100.00%
5.0	Diversity, non-discrimination and fairness	5.88%	5.88%	100.00%
6.0	Societal and environmental well-being	25.00%	0.00%	100.00%
7.0	Accountability	87.50%	0.00%	100.00%
	Total Score (%)	44.21%	2.17%	



References

- Ademola, O. E. (2024). Detailing the Stakeholder Theory of Management in the AI World: A Position Paper on Ethical Decision-Making. *Advances in Multidisciplinary & Scientific Research Journal Publication*. https://doi.org/10.22624/aims/sij/v10n1p1
- Agbese, M., Mohanani, R., Khan, A., & Abrahamsson, P. (2023). Implementing AI Ethics: Making Sense of the Ethical Requirements. *Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering*, 62–71. https://doi.org/10.1145/3593434.3593453
- Aguiar, D. C., Pamplona Salomon, V. A., & Pereira Mello, C. H. (2015). An ISO 9001 Based Approach for the Implementation of Process FMEA in the Brazilian Automotive Industry. *International Journal of Quality & Reliability Management*. https://doi.org/10.1108/ijqrm-09-2013-0150
- Ahmed, A., Mathrani, S., & Jayamaha, N. (2021). An Integrated Lean and ISO 14001 Framework for Environmental Performance: An Assessment of New Zealand Meat Industry. *International Journal of*

Lean Six Sigma. https://doi.org/10.1108/ijlss-05-2021-0100

- *AI Act | Shaping Europe's digital future*. (2024, April 30). https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
- Aref, E. (2024). *Team Dynamics And Conflict Resolution: Integrating Gen AI in Project Based Learning to Support Students' Performance*. ASEE Annual Conference, Portland, Oregon.
- Aref, E., & Fredericks, T. (2021). A New Course Development in Usability Engineering: Hands-On Learning Based on Research Work. *2021 ASEE Virtual Annual Conference Content Access Proceedings*, 36599. https://doi.org/10.18260/1-2--36599
- Barafort, B., Mesquida, A. L., & Mas, A. (2016). *How to Integrate Risk Management in IT Settings Within Management Systems? Comparison and Integration Perspectives From ISO Standards*. https://doi.org/10.1007/978-3-319-38980-6_19
- Ben-Asher, N., & González, C. (2015). Effects of Cyber Security Knowledge on Attack Detection. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2015.01.039
- Chen, W., Liu, C., Xing, F., Peng, G., & Yang, X. (2022). Establishment of a maturity model to assess the development of industrial AI in smart manufacturing. *Journal of Enterprise Information Management*, *35*(3), 701–728. https://doi.org/10.1108/JEIM-10-2020-0397
- Commission, E., Directorate-General for Communications Networks, C., & Technology. (2020). *The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment*. Publications Office. https://doi.org/doi/10.2759/002360
- Diaz-Rodriguez, N., Del Ser, J., Coeckelbergh, M., de Prado, M. L., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the Dots in Trustworthy Artificial Intelligence: From AI Principles, Ethics, and Key Requirements to Responsible AI Systems and Regulation. In *arXiv*. http://dx.doi.org/10.48550/arXiv.2305.02231
- Eberl, J. K., & Drews, P. (2021). Digital Leadership – Mountain or Molehill? A Literature Review. In F. Ahlemann, R. Schütte, & S. Stieglitz (Eds.), *Innovation Through Information Systems* (pp. 223–237). Springer International Publishing.
- *Ethics of Artificial Intelligence | UNESCO*. (2022). https://www.unesco.org/en/artificial-intelligence/recommendation-ethics
- European Commission. (2019, April 8). *Ethics guidelines for trustworthy AI | Shaping Europe's digital future*. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
- Freiman, O. (2023). Making sense of the conceptual nonsense 'trustworthy AI.' *AI and Ethics*, *3*(4), 1351–1360. https://doi.org/10.1007/s43681-022-00241-w

- Goirand, M. (2023). Mapping the Flow of Knowledge as Guidance for Ethics Implementation in Medical AI: A Qualitative Study. *Plos One*. https://doi.org/10.1371/journal.pone.0288448
- Golpayegani, D., Pandit, H. J., & Lewis, D. (2023). Comparison and Analysis of 3 Key AI Documents: EU's Proposed AI Act, Assessment List for Trustworthy AI (ALTAI), and ISO/IEC 42001 AI Management System. In L. Longo & R. O'Reilly (Eds.), *Artificial Intelligence and Cognitive Science* (Vol. 1662, pp. 189–200). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-26438-2_15
- Gutierrez, C. I., Aguirre, A., Uuk, R., Boine, C. C., & Franklin, M. (2023). A Proposal for a Definition of General Purpose Artificial Intelligence Systems. *Digital Society*, 2(3), 36. https://doi.org/10.1007/s44206-023-00068-w
- Huriye, A. Z. (2023). The Ethics of Artificial Intelligence: Examining the Ethical Considerations Surrounding the Development and Use of AI. *American Journal of Technology*. https://doi.org/10.58425/ajt.v2i1.142
- Lewis, D., Hogan, L., Filip, D., & Wall, P. J. (2020). Global Challenges in the Standardization of Ethics for Trustworthy AI. *Journal of ICT Standardization*, 123–150. https://doi.org/10.13052/jicts2245-800X.823
- Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., Yi, J., & Zhou, B. (2023). Trustworthy AI: From Principles to Practices. *ACM Computing Surveys*, *55*(9), 1–46. https://doi.org/10.1145/3555803
- Li, N. (2023). Ethical Considerations in Artificial Intelligence: A Comprehensive Disccusion From the Perspective of Computer Vision. *SHS Web of Conferences*. https://doi.org/10.1051/shsconf/202317904024
- Minkkinen, M., Niukkanen, A., & Mäntymäki, M. (2024). What about investors? ESG analyses as tools for ethics-based AI auditing. *AI & SOCIETY*, *39*(1), 329–343. https://doi.org/10.1007/s00146-022-01415-0
- Prikshat, V., Patel, P., Varma, A., & Ishizaka, A. (2022). A Multi-Stakeholder Ethical Framework for AI-augmented HRM. *International Journal of Manpower*. https://doi.org/10.1108/ijm-03-2021-0118
- Singh, A., & Pandey, J. (2024). Artificial intelligence adoption in extended HR ecosystems: Enablers and barriers. An abductive case research. *Frontiers in Psychology*, *14*, 1339782. https://doi.org/10.3389/fpsyg.2023.1339782
- Vianello, A., Laine, S., & Tuomi, E. (2023). Improving Trustworthiness of AI Solutions: A Qualitative Approach to Support Ethically-Grounded AI Design. *International Journal of Human–Computer Interaction*, *39*(7), 1405–1422. https://doi.org/10.1080/10447318.2022.2095478

- Vyhmeister, E., & Castane, G. G. (2024). TAI-PRM: Trustworthy AI—project risk management framework towards Industry 5.0. *AI and Ethics*. https://doi.org/10.1007/s43681-023-00417-y
- Wibowo, E. J., & Ramli, K. (2022). Impact of Implementation of Information Security Risk Management and Security Controls on Cyber Security Maturity (A Case Study at Data Management Applications of XYZ Institute). *Jurnal Sistem Informasi*. https://doi.org/10.21609/jsi.v18i2.1146
- Yusuff Marican, M. N., Razak, S. A., Selamat, A., & Othman, S. H. (2023). Cyber Security Maturity Assessment Framework for Technology Startups: A Systematic Literature Review. *Ieee Access*. https://doi.org/10.1109/access.2022.3229766
- Zografaki, E., Magoutas, A., & Chountalas, P. (2017). Focus on Human Resources as a Mediating Factor Between ISO 9001 Certification and Continuous Quality Improvement. *International Journal of Decision Sciences Risk and Management*. https://doi.org/10.1504/ijdsrm.2017.084010

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.