Continuous Assessment-Driven Requirement Elicitation for Trustworthy AI Systems

Anna Fedorova[1] (✉), Wolfgang Stefani[1], Christoph Heitz[1], and Ricardo Chavarriaga[1]

Zurich University of Applied Science, Centre for Artificial Intelligence, Winterthur, Switzerland fedo@zhaw.ch

Abstract. As AI systems increasingly impact critical decisions, ethical principles should be integrated throughout the entire development lifecycle. This is particularly relevant when these technologies influence high-stakes decisions and critical infrastructure. Regulatory efforts emphasize the need to adopt an ethics-by-design approach, i.e., embedding ethics from the earliest stages of development. However, most assessment frameworks are intended for use after development is complete, which conflicts with this approach. Relying solely on ex-post evaluations risks overlooking ethical issues that should be addressed earlier. This highlights the need for assessment methods suitable for early and continuous evaluation throughout the system lifecycle. We propose an assessment approach that differentiates between ethical concerns and risks that are: (1) within the current scope, (2) relevant at later stages, and (3) not relevant. This differentiation enables stage-appropriate, effective application of ethics-by-design principles. We demonstrate this in a practical use case, where we adapt the EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) to elicit positive functional and non-functional requirements during the design phase of an AI-assisted decision-making system for railway management. We show how the method elicits a list of Technology Readiness Levels (TRL) sensitive requirements. We also provide suggestions to improve trustworthiness assessment tools suitable for early-stage evaluation.

Keywords: trustworthy AI · ethical frameworks · ethics-by-design

1 Introduction

As artificial intelligence (AI) systems become increasingly embedded in high-stakes and safety-critical applications, ensuring their ethical and trustworthy design has become a main concern. Trustworthiness assessment tools have emerged as practical instruments to guide developers and stakeholders in aligning AI systems with ethical principles (c.f., [1], [2]). Reviews of existing frameworks can be found in [3], [4].

One such tool is the Assessment List for Trustworthy Artificial Intelligence (ALTAI), developed by the European Commission as a framework for *ex-post*

self-assessment of AI systems' trustworthiness [5]. Although ALTAI was originally intended for use after a system's development, it is increasingly applied in *ex-ante* contexts [6], [7].

When applied during early development stages - such as during proof-of-concept or low Technology Readiness Level (TRL) phases - the use of assessment tools designed for ex-post assessments may fail to capture risks that only manifest in later stages. This can lead to incomplete or misleading assessments and may prevent the effective integration of ethical principles across the full AI lifecycle. The European Commission, among others, emphasizes the importance of *ethics-by-design* [8], which involves embedding ethical considerations from the earliest stages of development, rather than addressing them ex-post, i.e., once the system is fully implemented.

We argue that self-assessment tools must be designed to be applied throughout the entire design cycle, instead of focusing on ex-post assessments. Consequently, these assessment tools must be designed to explicitly acknowledge the project scope and intended TRL, as well as support the anticipation of risks that may manifest at later stages of the cycle or higher TRL and provide guidance that is appropriate for that stage. By doing so, they should provide guidance that is appropriate for the current stage of the lifecycle, and guide potential actions to be taken at later stages. For this purpose, they must differentiate between elements and characteristics that are (1) relevant and actionable at the current stage, (2) likely to emerge in future phases, which should be anticipated, and (3) currently out of scope.

We illustrate this by identifying limitations on the use of ALTAI for early-stage assessments and propose a methodology to support stage-appropriate analysis of the trustworthiness of AI systems. This paper presents a methodology for translating trustworthy assessment questions into positive functional and non-functional requirements suitable for early stages of the development lifecycle, thus enabling *Trustworthiness-by-design*. We illustrate its use in a use case of AI-assisted decision-making in railway scheduling.

1.1 The Assessment List for Trustworthy AI (ALTAI)

The High-Level Expert Group on Artificial Intelligence, appointed by the European Commission, developed Ethics Guidelines for Trustworthy Artificial Intelligence [9]. These guidelines are supported by the Assessment List for Trustworthy AI (ALTAI), a comprehensive self-assessment tool based on this framework for achieving Trustworthy AI (TAI). ALTAI is structured around seven key requirements: **Human Agency and Oversight**: Ensuring AI systems support human decision-making and autonomy, with adequate human oversight mechanisms. **Technical Robustness and Safety**: Addressing resilience, security, accuracy, reliability, and fall-back plans. **Privacy and Data Governance**: Ensuring data protection and governance align with GDPR. **Transparency**: Emphasizing traceability, explainability, and communication. **Diversity, Non-discrimination, and Fairness**: Avoiding bias, ensuring accessibility, and involving stakeholders. **Societal and Environmental Well-being**: Considering

the impact on society, environment, democracy, and the working environment, including changes in working arrangements and skills of people interacting with an AI system. **Accountability**: Ensuring auditability and risk management. To avoid confusion with the technical requirements of the use cases, we will refer to the ALTAI requirements further in this paper as dimensions of trustworthiness.

The ALTAI-based assessment requires going through a set of yes/no questions on each of the dimensions and their subsections. The ALTAI questionnaire is available as a checklist and an online interactive tool [5]. The tool is intended for flexible use, organizations are expected to adapt it to their specific needs and sectors. Its guidelines specifically mention that the assessors might need additional questions to complement the assessment depending on the application. It encourages a multidisciplinary approach, involving various stakeholders such as AI designers, data scientists, legal experts, and management.

2 Requirement elicitation based on ALTAI

We extend the ALTAI tool to enable assessments applicable at different stages of the lifecycle. The extended methodology allows not only to assess the trustworthiness dimensions, but also elicitation of functional and non-functional requirements explicitly related to these dimensions.

We followed an iterative approach, starting with an initial description of the use case. The use case description must be defined by a multidisciplinary set of stakeholders to include domain-specific knowledge from relevant areas. It describes the AI systems' tasks and provides an overview of the necessary functions and tasks of the system.

As a starting point, we use a use case description following standards IEC PAS 62559 [10] and ISO/IEC 24030:2024 [11]. The detailed description of the use case methodology is out of scope for this paper and can be found in [12]. Alternative methods can be used to define the use case and its context for subsequent requirement elicitation.

The use case is then analyzed following the ALTAI structure to identify relevant trustworthiness dimensions that need to be addressed at the current or later lifecycle stages (See Fig. 1).

Next, based on the description of the use cases, the group of stakeholders identify requirements of the functions based on the business needs, without explicitly identifying any specific technologies or products. The same document can also cover 'non-functional' requirements, such as constraints, performance, security, and data interactions with other applications or systems [10].

The next stage of the process is to perform the trustworthiness assessment based on the current use case description using ALTAI. This allows to identify new functional and non-functional requirements in an iterative process by analyzing responses to the ALTAI questionnaire. The process starts with the definition of the scope of analysis, including the stage of the life-cycle and intended TRL. For each of these questions, stakeholders must determine whether it is "Relevant for the scope", "Relevant for the full system" or "Not relevant".

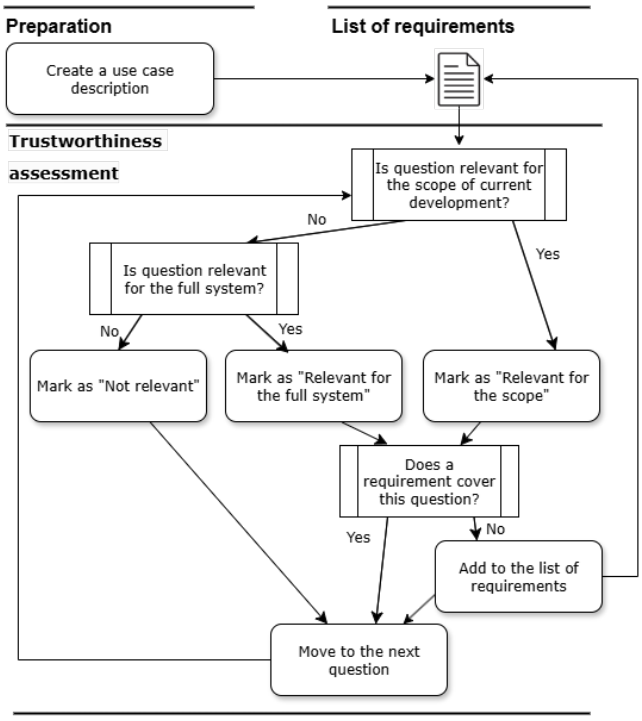

Fig. 1. Methodology for eliciting requirements based on trustworthy assessment.

This distinction is our extension of the method described in [7], and it allows differentiation during the assessment between the elements or characteristics of the system, which are not considered in the full use case, those that must be considered at the current lifecycle stage or scope, and those relevant for a fully implemented system.

In an iterative process, for each question, the first step is to assess whether is relevant for the scope of analysis (i.e., lifecycle stage and TRL). In the affirmative case, it is marked as "Relevant for the scope". If it is not relevant for the scope, the group should assess its relevance for the full scope of the use case. If the question becomes relevant at a later phase or for systems with higher TRL, it is marked as "Relevant for the full system". This process is repeated for all assessment questions.

The next step focuses on eliciting the technical and non-technical requirements. First, it must be checked if existing requirements already address the relevant questions. In this case, no new requirement will be generated. Otherwise, a new requirement is defined. After this, the use case documentation is updated to explicitly link each requirement to the relevant trustworthiness questions.

Since the ALTAI questionnaire was designed for ex-post analysis, when new requirements are generated, its questions were rephrased in the form of positive requirements. This methodology allows us to keep account of the necessary functionality and ensure that the AI system is aligned with the trustworthiness dimensions. Its application throughout the lifecycle supports the requirement for a continuous risk management included in the EU AI Act [13].

3 Application to a use case

We tested the methodology in a use case of AI-based assisted decision making for railway network control. Since it corresponds to an application in critical infrastructure, it has higher requirements for trustworthiness. This work was performed as part of the EU-funded project AI4REALNET. Below, we present a brief description of the use case. The full description is available in [12].

Business problem: The demand for railway network capacity in terms of efficiency, network capacity, traffic density and traffic management resilience is growing rapidly. Simultaneously, environmental awareness results in changing mobility regulations and the need for novel dispatching technologies.

Objective: AI-based solution must enhance dispatchers' capabilities by assisting them in rescheduling trains or automating decision-making processes and providing support for human decision-making in complex operation scenarios. The AI system should detect deviations in real time, assess their impact on the overall schedule, and predict issues that may arise in the future.

The AI4REALNET project aims at developing proof of concept (POC) systems with a maximum TRL of 5 (i.e., limited functionality compared to the full system). Tests will be performed in a simulated environment focusing on the technical feasibility.

3.1 Requirement elicitation

The use case description was formulated by the domain experts, industry representatives and relevant stakeholders. Next, they performed the ALTAI-based assessments. The questionnaire was shared as a collaborative document, which enabled workshop participants, stakeholders, developers, and other contributors to provide answers to individual questions. A designated domain expert managed the process, ensuring effective coordination. The collected input was carefully evaluated to assess the relevance of each question to the use case. The assessment results were documented in a report including questions, decisions and supporting arguments. This document summarizes the outcome in a table, which included a full list of the ALTAI questions and corresponding columns: **Relevance** - relevancy of the question for the UC (Relevant for the system - R full, relevant for scope - R scope, Not relevant - NR); **Consideration** – a summary of comments that stakeholders had on the topic, including rationale for the decision; **Measure** – if the question is relevant, list of requirements that address it. An excerpt from this table is presented in Table 1.

Table 1. Excerpt from the table with the Trustworthiness Assessment.

Dimension: 3 - Privacy and Data Governance		
Question: Is your AI system being trained, or was it developed, by using or processing personal data (including special categories of personal data)?		
Relevance	**Consideration**	**Measure**
NR	Simulation environments are used for training.	None
Dimension: 5 - Diversity, Non-discrimination and Fairness		
Question: Did you establish a strategy or a set of procedures to avoid creating or reinforcing unfair bias in the AI system, both regarding the use of input data as well as for the algorithm design?		
Relevance	**Consideration**	**Measure**
R full	Bias avoidance must be considered during development and monitored after implementation.	Introduce Fairness requirements (Fa-1 and Fa-2) to ensure that the system fairly distributes unavoidable delays throughout the system and does not unfairly favor specific Railway Undertaking Operating Managers (RUOMs)
Dimension: 7 - Accountability		
Question: Did you establish a process to discuss and continuously monitor and assess the AI system's adherence to this Assessment List for Trustworthy AI (ALTAI)?		
Relevance	**Consideration**	**Measure**
R scope	Monitoring processes are out-of-scope for the POC. They must be considered for the full scope system.	No measure for this scope

This structure is similar to [7]. The introduced novelty is the distinction between "relevant for full system" and "relevant for scope". It allows keeping track of functionalities planned for different versions of the system and evaluating their TRL according to the maturity and level of detail of requirements.

The same stakeholder group, which defined the use case description, then analyzed existing requirements and identified gaps, proposing new requirements where necessary. All pertinent requirements were incorporated into the official use case documentation. This systematic evaluation and documentation of ethical considerations serve to justify the decisions made within the AI4REALNET project. An excerpt from this list is shown in Table 2.

The final outcome of the process is (i) an assessment Table, where each relevant question is assigned a trustworthiness-related requirement, and (ii) an updated list of Use Case requirements.

3.2 Results of the assessment

To evaluate the role of trustworthy dimensions for different scopes, we show the proportion of questions marked as relevant for the full scope and for the intended

Table 2. Example of elicited requirements. The full list is available in [14].

ALTAI dimension	Subdimension	AI4REALNET requirements
Diversity, Non-discrimination and Fairness	Avoidance of Unfair Bias	The system should not unfairly favor specific RUOMs. Rescheduling in railway operations must impact the RUOMs fairly. Measures should be put in place to ensure that these constraints are observed.
Technical robustness and safety	General safety	Identify the system's levels of exposure to such threats, both in terms of quantity and duration.
Accountability	Risk management	Put in place by design mechanism in case of applications that can adversely affect individuals in terms not only of hazard but also exposure and vulnerability.

scope (POC) in Figure 2. Notice that we don't report the spider diagrams produced by the online ALTAI assessment tool, as there is no information on how exactly the information from responses is used to define the visualization.

We can see that assessment of the POC AI system yields a relatively small number of ALTAI questions as relevant. Since the systems will be developed and tested in a simulated environment, trustworthiness characteristics, such as security, are marked as not relevant for this scope as they will depend on other elements of the whole system. Likewise, in this scope, safety in extreme situations will not be tested as the simulated data only includes the most common scenarios and tests for these cases will be planned for later stages or systems with higher TRL. Overall, the most relevant dimensions for this scope are Human Agency and Oversight, Transparency, Societal and Environmental Well-being.

When analyzing the trustworthiness of the full scope system, the relevance of all dimensions increases. 4 dimensions have over 80% relevance: Accountability, Technical Robustness and Safety, Privacy and Data Governance, Transparency.

In the following paragraphs, we summarize the responses and considerations gathered from the stakeholders for each of the ethical dimensions making clear the difference in the functionalities of POC and full use case.

Human Agency and Oversight. The AI system's interaction with end-users influences their autonomy and decision-making processes, introducing the risk of overdependence. Targeted training to ensure users understand the system's functionality and limitations can mitigate this risk. Although the AI does not simulate human-like interaction, it may still encourage compulsive usage patterns. In railway operations, varying oversight models - such as Human-in-the-Loop and Human-in-Command - require well-defined protocols to allow safe human intervention when the AI is in control. Oversight mechanisms should also detect undesirable outcomes and enable the management of the system's adaptive learning behavior. These mechanisms must already be tested in the POC to estimate the impact on the operators. These test should inform decisions about full-scale deployment.

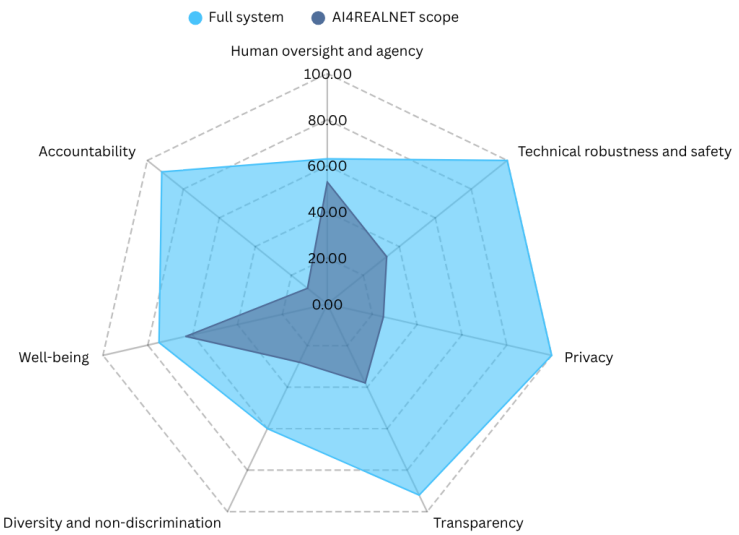

Fig. 2. Comparison of the number of relevant questions in the trustworthiness assessment for full scope and the intended scope (POC).

Technical Robustness and Safety. While the project addresses some robustness and safety dimensions, further detail is needed during full system deployment (e.g., safety mechanisms like collision avoidance are managed by a separate component). Resilience against adversarial threats is considered within the POC, while aspects like certification and long-term security strategies are beyond the project's scope. Ensuring fault tolerance and ongoing human involvement in technical evaluations is critical. System performance must be closely monitored, and uncertainty in AI outputs should trigger human review and control transfer mechanisms. Continuous learning functions must be documented and interpretable to preserve both system reliability and human authority.

Privacy and Data Governance. Since the system does not process personal data, immediate privacy issues are limited. However, as the system evolves, future stages should include mechanisms to address potential privacy risks. While current data governance complies with regulatory expectations, specific data privacy requirements are deemed unnecessary for the POC due to the non-sensitive nature of the data in use.

Transparency. Ensuring traceability through record-keeping of system decisions - including inputs and outputs - is fundamental for transparency. The explainability of AI behavior is prioritized to help operators grasp its objectives, decision logic, and learning patterns. Effective communication between the AI and human users is essential to minimize misuse and establish trust. System outputs are clearly distinguishable from human actions, with users kept informed about the system's capabilities and limitations at all times.

Diversity, Non-discrimination, and Fairness. The AI system must avoid introducing or reinforcing biases during its development, particularly in ensuring equitable outcomes such as neutral delay distribution across railway operators. Inclusive stakeholder engagement, including public and professional workshops, ensures that development aligns with societal expectations and user needs.

Societal and Environmental Well-being. The system can potentially bring environmental benefits by improving operational efficiency. Its deployment will also influence workforce roles and skill requirements, which must be considered in the design. Collaboration with end-users and human factors specialists is encouraged to address these impacts. While the system will require new skills, the development of related training programs falls outside the POC scope.

Accountability. System accountability is supported through exhaustive documentation and logging practices, which are key for retrospective analysis and performance tracking. Although comprehensive risk management planning is not part of POC scope, existing logging infrastructure provides a strong basis for future ethical oversight and accountability mechanisms.

4 Recommendations for improving the trustworthy assessment tools

Create questionnaires tailored to different development stages of the AI lifecycle. Conducting a trustworthiness assessment from the early stages and throughout the entire life cycle of development is beneficial for envisioning measures to ensure trustworthiness-by-design. However, ALTAI was developed as a post-hoc assessment, and the formulation of questions reflects it.

Our approach to continuous trustworthiness assessment is compatible with relevant regulatory requirements. The EU AI act, Art 9. states that risk management "shall be understood as a continuous iterative process planned and run throughout the entire lifecycle of a high-risk AI system, requiring regular systematic review and updating" [13]. Likewise, the General Data Protection Regulation (GDPR), Art. 35, requires that Data Protection Impact Assessments must be conducted "prior to processing" and continuous regular assessments are required during the lifecycle as means to address any potential risks [15].

ALTAI-like questionnaires or checklists shall be adapted to become actionable instruments to adopt these requirements into the development procedure. For this purpose, the assessment should be designed so as to address specific challenges at each stage of the lifecycle [16]. These assessments should be conducted as early as possible and repeated after any significant change the system undergoes. The identification and mitigation of risks secured in this way should also be consistent with established standards (e.g., ISO/IEC 23894 [17]).

Distinguish between non-relevant and out-of-scope functionality. As we experienced in AI4REALNET project, some features may not be planned for the current development scope but are foreseen for later implementations. If the trustworthiness assessment is performed before the AI system is finalized this functionality will not be covered by ALTAI. Another factor that influences

the consistency of the assessment is variations in the understanding of the final product among different groups of stakeholders. It is hard to generate a robust result if the participants do not agree on the final product. Collaborative work during the assessment process contributes to reaching a uniform understanding of the current scope and the full system, while integrating the different needs and visions among all stakeholder groups.

Design tailored versions of the assessment tools focused on specific domains/application types. To increase the effectiveness and efficiency of the trustworthiness assessment, the questionnaire can be tailored to address the specificities and risks connected to the domain of the Use Case. For example, the assessment of the dimension of Human Accountability and Oversight in ALTAI is directed more to commercial social applications. The current questions do not sufficiently cover the different types of interactions between the user and AI system in the Use Case. It shows how ALTAI does not address all the complexities of applications of AI-assisted decision making in critical infrastructure.

However, other characteristics might be more crucial to assess in other types of applications. The functions or properties not mentioned in the ALTAI questions are not evaluated, disregarding their importance for the AI system and yielding misleading assessments.

We call for further development of ALTAI-like assessment tools to tailor them to the specific needs of different domains and improve the coverage of risks relevant to different applications. This development should include AI and domain-specific experts.

5 Conclusions

As [18] reflects, AI development does not yet have methods to transfer abstract principles into practice in a real-world context comparable to other disciplines, e.g., medicine, especially when it comes to such ethical concepts as responsibility and trustworthiness. The process of transposition of ethical dimensions into practical requirements resonates with the translation of values into design requirements, as described by [19]. The translation of general norms into more specific design requirements demands domain knowledge, is context-dependent and can take place outside specific design processes. Our approach offers a way to perform a structural analysis of ethical dimensions –based on the EU Ethical Guidelines for Trustworthy AI– identify relevancy to the intended AI solutions and translate the analysis into concrete requirements for the development and operational use.

The methodology addresses another challenge mentioned by [18]: "normative practical requirements must be embedded in development processes and functionally implanted in design requirements." The framework has been tested on real world use cases, but including a multistakeholder group of practitioners, AI experts, domain experts and industrial partners.

There is a risk that "ethical principles having been written into the business and use-case, but coded out by the time a system gets to deployment" [20]. Our

approach mitigates this risk by allowing continuous evaluation of the consistency of requirements with the trustworthiness dimensions.

This continuous process allows to update the understanding of the system and its requirements along the AI lifecycle and keep up with the growing TRL. We argue that trustworthiness assessment tools like ALTAI must explicitly acknowledge the intended TRL and project scope - whether focused on a proof of concept, pilot, or full deployment. They should differentiate between ethical concerns and risks that: (1) are within the current scope, (2) may arise at later stages, and (3) are currently not relevant. This clarity enables more effective and stage-appropriate application of trustworthiness-by-design principles.

6 Acknowledgements

AI4REALNET has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101119527 and from the Swiss State Secretariat for Education, Research and Innovation (SERI). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union and SERI. Neither the European Union nor the granting authority can be held responsible for them.

References

[1] M. Poretschkin, A. Schmitz, M. Akila, *et al.*, "Leitfaden zur gestaltung vertrauenswürdiger Künstlicher Intelligenz (KI-Prüfkatalog)," 2021.

[2] R. V. Zicari, J. Brodersen, J. Brusseau, *et al.*, "Z-inspection®: A process to assess trustworthy AI," *IEEE Transactions on Technology and Society*, vol. 2, no. 2, pp. 83–97, 2021.

[3] N. Kemmerzell, A. Schreiner, H. Khalid, M. Schalk, and L. Bordoli, "Towards a better understanding of evaluating trustworthiness in AI systems," *ACM Computing Surveys*, vol. 57, no. 9, pp. 1–38, 2025.

[4] A. Awadid, K. Amokrane-Ferka, H. Sohier, *et al.*, "AI systems trustworthiness assessment: State of the art," in *Workshop on Model-based System Engineering and AI, 12th International Conference on Model-Based Software and Systems Engineering (Modelsward)*, 2024.

[5] "Interactive version of the Assessment List for Trustworthy AI." (), [Online]. Available: `https://altai.insight-centre.org/` (visited on 06/05/2025).

[6] M. Borg, J. Bronson, L. Christensson, *et al.*, "Exploring the assessment list for trustworthy AI in the context of advanced driver-assistance systems," in *2021 IEEE/ACM 2nd International Workshop on Ethics in Software Engineering Research and Practice (SEthics)*, IEEE, 2021, pp. 5–12.

[7] T. Stefani, F. Deligiannaki, C. Berro, *et al.*, "Applying the assessment list for trustworthy artificial intelligence on the development of AI supported air traffic controller operations," in *2023 IEEE/AIAA 42nd Digital Avionics Systems Conference (DASC)*, IEEE, 2023, pp. 1–9.

[8] European Commission. "Ethics by design and ethics of use approaches for artificial intelligence." (2021), [Online]. Available: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf (visited on 06/06/2025).

[9] European Commission. "High-level Expert Group on Artificial Intelligence." (2018), [Online]. Available: https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai (visited on 06/05/2025).

[10] J. Trefke, J. M. González, and C. Dänekas, "Iec/pas 62559-based use case management for smart grids," in *Standardization in Smart Grids: Introduction to IT-Related Methodologies, Architectures and Standards*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 39–57, ISBN: 978-3-642-34916-4. DOI: 10.1007/978-3-642-34916-4_3.

[11] "ISO/IEC TR 24030:2024," International Organization for Standardization, Standard, Mar. 2023.

[12] AI4REALNET. "D1.1 – Framework and Use Cases." (2024), [Online]. Available: https://ai4realnet.eu/deliverables/ (visited on 06/05/2025).

[13] Council of European Union, "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence," 2024.

[14] AI4REALNET. "AI4REALNET - D1.1 – Supplementary Material – ALTAI Summary." (2024), (visited on 06/05/2025).

[15] European Parliament and Council of the European Union. "Regulation (EU) 2016/679 of the European Parliament and of the Council." (May 4, 2016), [Online]. Available: https://data.europa.eu/eli/reg/2016/679/oj.

[16] Y. Billeter, P. Denzel, R. Chavarriaga, *et al.*, "MLOps as enabler of trustworthy AI," in *11th IEEE Swiss Conference on Data Science (SDS), Zurich, Switzerland, 30-31 May 2024*, 2024. DOI: 10.21256/zhaw-30443.

[17] "Information technology — Artificial intelligence — Guidance on risk management," International Organization for Standardization, Standard, Mar. 2023.

[18] B. Mittelstadt, "Principles alone cannot guarantee ethical AI," *Nature machine intelligence*, vol. 1, no. 11, pp. 501–507, 2019.

[19] I. van de Poel, "Translating values into design requirements," in *Philosophy and Engineering: Reflections on Practice, Principles and Process*, D. P. Michelfelder, N. McCarthy, and D. E. Goldberg, Eds. Dordrecht: Springer Netherlands, 2013, pp. 253–266, ISBN: 978-94-007-7762-0.

[20] J. Morley, L. Floridi, L. Kinsey, and A. Elhalal, "From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices," *Science and engineering ethics*, vol. 26, no. 4, pp. 2141–2168, 2020.