Assessment of Trustworthiness of an Industrial AI System using ALTAI – A Feasibility Study

Shukun Tokas[1,*], Aida Omerovic[1] and Ragnhild Halvorsrud[1]

[1]SINTEF Digital, Norway

Abstract

This paper reports an exploratory feasibility study of applying the EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) to an off-the-shelf AI-powered product recommender integrated into an industrial AI-enabled service. ALTAI is a voluntary self-assessment checklist, developed to guide users through structured questions and recommended practices aligned with the key requirements for trustworthy AI. Three researchers with complementary expertise and the (industry) case provider each conducted independent ALTAI self-assessments. The objective was to evaluate whether ALTAI can be applied for the purpose of trustworthiness assessment of a procured, off-the-shelf AI component under realistic industrial information constraints. The soundness of ALTAI itself is outside the scope. Our experiences are reported on, in terms of both the process and the resulting assessment outputs. The process facilitated a shared understanding of the trustworthiness requirements and helped uncover trustworthiness concerns. The outputs, which include trustworthiness scores and improvement recommendations, revealed substantial variation across the four assessments. The variation was partly caused by differences in expertise, interpretation, assumptions and limited access to detailed technical and organisational documentation. Overall, ALTAI proved feasible for structuring reflection and discussion about the trustworthiness of the case under consideration. The paper shares our experience of applying ALTAI, and derives recommendations both for practitioners using the framework and for further improvement of the tool.

Keywords

AI, ALTAI, Trustworthiness, Assessment, Feasibility study

1. Introduction

AI is penetrating more and more into critical domains such as energy, banking, transportation, healthcare, and e-commerce, where AI-powered services interact closely with humans and influence organisational and societal processes. As dependence on complex and sometimes unpredictable technological ecosystems grows, ensuring that these systems are trustworthy becomes increasingly important. Trustworthiness, however, is not only a property of an algorithm or model. It also depends on how an AI component is integrated, governed, monitored, and experienced within a broader socio-technical system. Assessing the trustworthiness of AI-enabled services is therefore both necessary and inherently challenging. In many cases, especially when AI systems or AI-enabled components are procured as commercial off-the-shelf products or third-party services rather than developed in-house, deployers face an information asymmetry. They are responsible for integrating, monitoring, and governing the system in a concrete use context, but they may not have access to detailed technical documentation about model development, training data, evaluation procedures, limitations, risk controls, or design decisions.

Several frameworks and guidelines have been proposed to support the development and assessment of trustworthy AI, including standards, regulatory initiatives, and ethical guidelines. One widely used practical instrument is the EU Assessment List for Trustworthy Artificial Intelligence (ALTAI) [1], developed by the independent High-Level Expert Group on Artificial Intelligence (HLEG-AI) appointed by the European Commission [2]. In 2019, HLEG-AI published its *Ethics Guidelines for Trustworthy*

TRUST-AI: The Second European Workshop on Trustworthy AI. Organized as part of the International Joint Conference on Artificial Intelligence - IJCAI/ECAI 2026. August 2026, Bremen, Germany.

*Corresponding author.

Shukun.Tokas@sintef.no (S. Tokas); Aida.Omerovic@sintef.no (A. Omerovic); Ragnhild.Halvorsrud@sintef.no (R. Halvorsrud)

0000-0001-9893-6613 (S. Tokas); 0000-0001-8868-597X (A. Omerovic); 0000-0002-3774-4287 (R. Halvorsrud)

AI [3]. In the guidelines, "HLEG-AI developed seven principles for AI which are intended to help ensure that AI is trustworthy and ethically sound" [4]. The seven principles include: *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*.

These requirements are also reflected in the European AI Act [4]. ALTAI is a voluntary self-assessment checklist that particularly applies to AI systems that directly interact with users, and it is primarily addressed to developers (providers) and deployers of AI systems, whether self-developed or acquired from third parties [2]. The tool guides assessors through structured questions and produces visual assessment results and mitigation-oriented recommendations. For each requirement, ALTAI provides introductory guidance and relevant definitions in the Glossary, and the online tool [5, 1] contains additional explanatory notes for many of the questions.

Despite broader considerations and extensive information, applying ALTAI in practice raises several challenges: *(i)* Trustworthiness assessments are highly use case and context-dependent, and *(ii)* AI systems today are often developed, procured, and deployed in diverse ways. Some organisations build AI systems in-house, while others integrate open-source components, third-party APIs, or off-the-shelf AI services into existing platforms. This diversity complicates assessment. Moreover, several ALTAI questions require technical, organisational, legal, and domain-specific knowledge that is unlikely to be held by a single assessor or role. Because trustworthiness of AI-enabled technologies is highly context-dependent, we argue that a major weakness of ALTAI is its limited context awareness and lack of decision support for trade-off analysis between requirements and mitigations of gaps. One method known for bringing human perspectives and context into complex processes is customer journey mapping, which is widely used also in non-commercial fields [6]. Anchored in the end-user, a journey is structured into granular touchpoints involving multiple actors, systems, and communication channels, while preserving the context and intent of the human actor. Journey methods can therefore serve as a contextual lens for trustworthiness assessment by making human goals and intentions explicit.

In this paper, we report on a feasibility study in which ALTAI was applied to an anonymised industrial case: an AI-powered product recommender embedded in a consumer-to-consumer circular-economy platform in Norway. To define the assessment scope, we used customer journey descriptions to identify the relevant AI-enabled touchpoint and situate the recommender within the broader service context. Three researchers with complementary expertise independently conducted ALTAI assessments, and the case provider conducted a separate self-assessment. We then gathered experiences from the process undergone and compared the resulting radar diagrams and recommendation outputs.

The research questions which this study has been driven by, include:

1. To what degree is ALTAI feasible and appropriate as a framework for assessment of trustworthiness of a re-commerce system?
2. Which competences and roles should the analysis team members represent, in order to conduct an ALTAI based trustworthiness assessment?
3. How reliable are the findings from the trustworthiness assessment of an e-commerce platform for the broader target group of the critical domains such as energy, banking, transportation and health?

The purpose of this study is not to provide a definitive trustworthiness score for ProdRecAI. Rather, we examine the feasibility of applying ALTAI on an embedded AI component in a real-world industrial setting. Based on our exploration, we derive recommendations for practitioners using ALTAI and for future improvement of the tool.

2. Method

This study was conducted as an exploratory feasibility study. The method consisted of three main steps: 1) defining the assessment scope, 2) conducting ALTAI assessments, and 3) consolidating the experiences and the resulting assessment outputs.

Case context and assessment scope: For confidentiality reasons, we anonymize the industry case provider in this article. The case provider operates a Norwegian C2C digital platform where users can buy and sell used goods. The platform supports buyers in searching for, purchasing, and receiving products through a digital service ecosystem involving sellers, the platform operator, and logistics partners. The buyer journey shown in Figure 1 starts with product search and selection, followed by a purchase request sent to the seller through the platform. Payment is initiated at this stage, while the platform simultaneously coordinates communication with a logistics partner regarding package pickup and delivery. The buyer is notified when the seller accepts the purchase request, after which the remaining journey concerns shipment tracking and delivery.

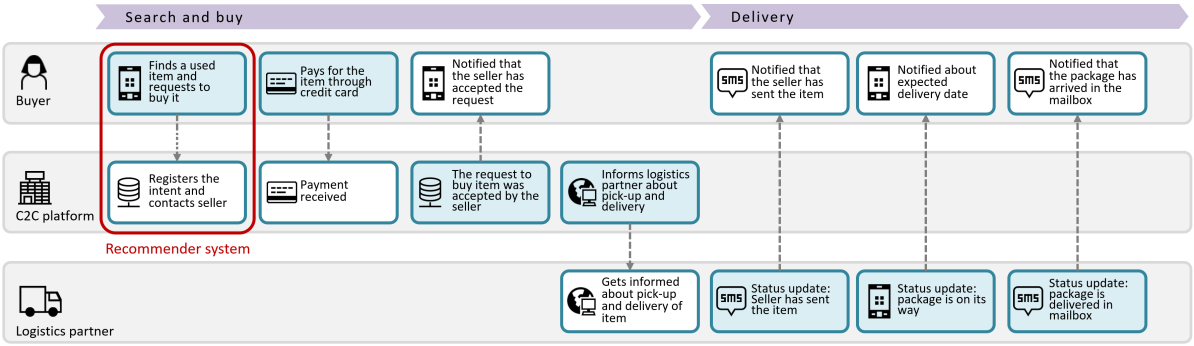

Figure 1: Buyer journey showing the main actors and touchpoints, used to scope the system for assessment.

To contextualize the assessment, we identified the touchpoints where AI functionality supported the user experience. This helped us narrow the assessment scope from the overall platform to a specific AI-enabled component: the product recommender. We refer to the scoped component as ProductRecommenderAI, ProdRecAI, (highlighted in Figure 1).

The system under assessment and information availability: ProdRecAI is an off-the-shelf AI-powered component integrated into the platform. It generates personalised product recommendations based on user interactions with the service, such as interests, browsing behaviour, previous purchases, and demographic characteristics. More broadly, the case provider uses AI across several functionalities, including search, seller recommendations on product pages, and image-to-text support in the manual product listing flow.

Detailed technical and organisational documentation of ProdRecAI was, however, not available to the external analysts. In particular, they did not have access to a complete architecture description, model documentation, training-data documentation, performance evaluations, bias or fairness analyses, security test results, or detailed governance documentation. Nor was this merely a restriction imposed on the external analysts: because the component is procured, the case provider's own senior representatives could not answer several questions about its internals either; as the provider's technical lead put it, "we have not developed a foundation model, we have bought standard off-the-shelf products". *This situation is not unique to our case. It reflects a recognised challenge in AI governance.* The NIST AI Risk Management Framework highlights that organisations deploying third-party AI systems often depend on the transparency provided by the supplier and may have only limited access to information about model development, design decisions, evaluation procedures, limitations, and risk controls (Section 3.4, Section 5.1-GOVERN 6) [7]. Third-party providers may provide limited documentations for reasons of intellectual property protection, security, commercial confidentiality, or system complexity. Such asymmetry makes trustworthiness assessment inherently more difficult. Given these constraints on the ideal conditions of information availability, we deliberately did not attempt to obtain the vendor's proprietary documentation: the purpose of the study was to examine the feasibility of applying ALTAI under the information availability conditions that deploying organisations realistically face.

Assessment team: The assessments involved three analysts, all authors of this paper, with

complementary expertise in human–computer interaction, AI/ML, and data protection. The analysts conducted independent ALTAI self-assessments based on information provided by the case provider and their own inspection of the service. In addition, the case provider conducted a separate assessment. This assessment was completed jointly by two employees: one in a senior management role and one in a senior technical role.

Independent assessments using ALTAI tool: We used the publicly available web-based ALTAI tool (https://altai.insight-centre.org/) [5]. First, the three analysts jointly reviewed the ALTAI framework, its seven pillars of trustworthiness, and the accompanying documentation. To establish a shared understanding of the ALTAI tool, the three analysts first conducted independent pilot assessments of the *Human Agency and Oversight* pillar. They documented interpretation challenges, ambiguities, and information gaps through structured comments, which were subsequently used to support calibration and discussion. After this calibration step, each analyst independently completed a full ALTAI assessment of ProdRecAI. The case provider completed an independent ALTAI self-assessment. This resulted in four assessments in total: three from the (external) analysts and one from the case provider.

Each assessment resulted in a radar diagram summarizing pillar scores and a list of recommendations (to mitigate risks). We then compared the four assessments to examine convergence and divergence across assessors. This comparative analysis formed the basis for assessing the feasibility, usefulness, and limitations of ALTAI for evaluating trustworthiness in real-world AI systems.

3. Results and discussion

The ALTAI tool produced two outputs for each assessment: (i) a *radar diagram*, which visualises the assessment results across the seven dimensions of trustworthiness, and (ii) a set of *recommendations* aimed at improving weaker areas. Figure 2 presents a consolidated representation of the four radar diagrams generated from the independent ALTAI assessments of ProdRecAI.

Analyst 1´s profile is compact, with all scores concentrated near the centre: *transparency* and *diversity, non-discrimination and fairness* score zero, *privacy and data governance* is moderate, and the remaining dimensions are in the low-to-mid range, suggesting either substantial uncertainty or a perception of low

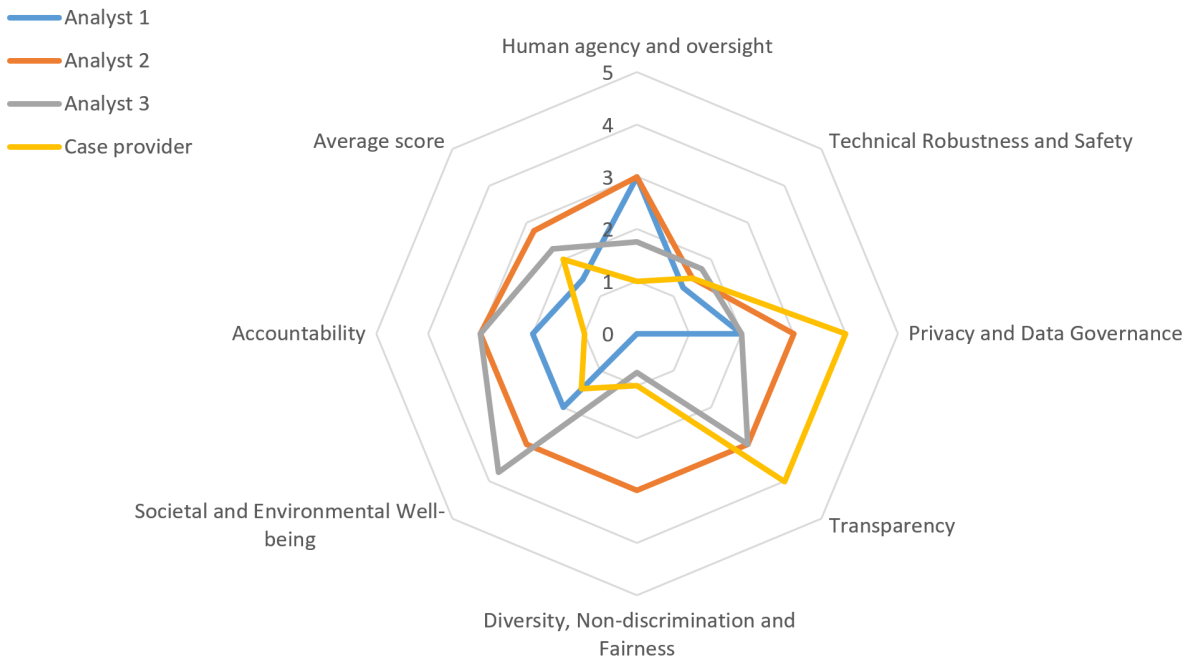

Figure 2: Comparison of ALTAI scores from the three analysts and the case provider shown in a radar chart.

trustworthiness. Analyst 2 produced the broadest profile, with most pillars in the mid-range around 3.0 but *technical robustness and safety* notably lower at around 1.5. Analyst 3´s profile is uneven: *societal and environmental* well-being is highest (approximately 3.5–4.0), *transparency* and *accountability* around 3.0, while *diversity, non-discrimination and fairness* falls below 1.0. The case provider´s profile is also uneven: *privacy and data governance* and *transparency* reach approximately 4.0, while the remaining dimensions cluster around 1.0–1.5. This suggests that the assessor viewed the system as relatively stronger in data handling and transparency-related practices, but weaker in robustness, fairness, societal impact, and accountability mechanisms.

Overall, the four independent assessments reveal substantial variation in the perceived trustworthiness of ProdRecAI, although the assessors used the same tool and an agreed scope. A partial point of convergence is the low scoring of *technical robustness and safety* across assessments. In contrast, *transparency* varies from low (Analyst 1) through moderate (Analysts 2 and 3) to high (case provider), underscoring how role and information access shape results. The results therefore indicate that ALTAI radar diagrams should not be interpreted as objective measurements of trustworthiness; rather, they are useful discussion artefacts that make visible where assessors agree, disagree, or lack sufficient information. These differences underscore the importance of (i) access to sufficiently detailed information about the system component(s) under assessment, and (ii) assessing AI systems within their broader socio-technical context, as several dimensions (e.g., societal impact, fairness, and accountability) may not be fully captured when evaluating isolated components.

In addition to the radar diagrams, ALTAI generates mitigation-oriented **recommendations** for each trustworthiness pillar. These recommendations are provided separately for each pillar. To illustrate their nature, selected examples from Analyst 4's (i.e., the case provider) output are presented below.

Technical robustness and safety

Put in place measures to ensure the integrity, robustness and overall security of the AI system against potential attacks over its lifecycle.

Inform users as soon as possible if some new threats are detected.

Privacy and Data Governance

Take measures to consider the impact of the AI system on the right to privacy, the right to physical, mental and/or moral integrity and the right to data protection.

Consider establishing mechanisms that allow flagging issues related to privacy or data protection concerning the AI system.

These examples illustrate that the ALTAI recommendations are relevant and aligned with key trustworthiness principles; however, they are largely generic and broad in scope. In our case, they were useful in supporting reflection and prompting discussions around potential risks, helping to surface issues that may otherwise be overlooked. At the same time, the recommendations require substantial interpretation and contextualisation before they can be translated into concrete actions for ProdRecAI. A key limitation is that the resulting recommendations appear to be primarily oriented towards organisations that develop their own AI systems, without adequately accounting for the diversity of today's AI ecosystem, where these systems may be built in-house, assembled from open-source components, or procured as off-the-shelf services. This distinction is important: an organisation integrating an off-the-shelf AI component faces different responsibilities and mitigation options than one developing a model from scratch, yet the recommendations do not differentiate between these contexts. Another limitation is the lack of traceability between the questionnaire responses and the generated recommendations, which makes it difficult to assess the reasoning behind the recommendations and limits prioritisation of mitigation efforts.

The analysts also noted that several ALTAI questions were difficult to interpret because they are formulated at a very generic level. The questionnaire does not distinguish between stakeholder roles, such as AI architects, software developers, product owners, and domain experts, although these groups possess different knowledge and responsibilities; asking all questions of all participants may reduce both

the precision of, and analysts´ confidence in, the answers. Domain-specific examples accompanying the questions, e.g., from e-commerce, autonomous vehicles, or AI-assisted medical image analysis, could help assessors interpret the intent of the questions more consistently. Similarly, some recommended measures appear insufficiently adapted to different types of AI systems: the recommendation to "deploy a 'stop button' or procedure to safely abort an operation when needed" is meaningful for autonomous systems, but is difficult to interpret in the context of an online shopping platform.

Finally, we collected qualitative feedback from the case provider to better understand the perceived usefulness of ALTAI in their setting, through nine followup questions, of which two are reported here, due to space constraints:

Question 1: To what extent was this analysis useful for evaluating ProdRecAI in terms of quality/functionality? Did this provide insight that will affect future functionality/service?
Answer *"Very little useful. We experienced that the questions were more aimed at AI companies, where one typically develops their own algorithms, or builds the core of the product on top of AI solutions. "*
Question 2: What suggestions for improvement do you have for this method and tool, so that it becomes more appropriate for evaluating the trustworthiness of AI-supported services?
Answer *"We think it was very thorough, and should therefore be comprehensive."*

The case provider perceived limited practical value for evaluating service quality or functionality, partly because several questions appeared to assume an organisation that develops its own AI algorithms. However, the feedback highlights a key strength of ALTAI: its comprehensiveness fosters broad reflection on trustworthiness. This supports our assessment that ALTAI is useful as a structured reflection and discussion tool, facilitating cross-disciplinary discussions among various stakeholders, posing critical questions, and enhancing stakeholders´ knowledge of both the system and its trust-related aspects.

Suitability of ALTAI for different types of AI systems. Our experience suggests that the effectiveness of an ALTAI assessment depends on the technical design, degree of opacity, and intended use of the system under assessment. Three characteristics of ProdRecAI limited ALTAI's effectiveness in our case. First, ProdRecAI is a procured, off-the-shelf component. Consequently, questions that presuppose developer-level knowledge of the system architecture, training data, design decisions, and evaluation procedures could not be answered with confidence on the basis of the information available to analysts. Although ALTAI is addressed to developers and deployers alike [5], many questions implicitly assume the developer's perspective, as also reflected in the case provider's feedback above. Second, the component is largely opaque to the deploying organisation. This was particularly consequential for the *Diversity, Non-discrimination and Fairness* dimension. Recommender systems may be susceptible to popularity bias, feedback loops, and discrimination risks arising from the use of demographic characteristics [8]. However, without access to model-level documentation, training data, or evaluation results, these risks could only be considered at the level of observable system behaviour and the organisational controls surrounding the system. Third, ProdRecAI is an embedded component of a broader service rather than a stand-alone AI system, which makes the system boundary a matter of interpretation; moreover, the service is of moderate criticality, so questions and recommendations aimed at safety-critical systems are difficult to translate to this context. Taken together, our findings suggest that ALTAI is most effective for AI systems that are developed in-house or with full transparency from the supplier, that constitute a clearly bounded system rather than a component embedded in a broader service, and whose intended use is of sufficient criticality that questions on safety, human oversight, and abort mechanisms are meaningful. It is less appropriate — in its current form — for procured, opaque, embedded AI components. For such components, its usefulness is likely to depend on clear scoping, role-specific guidance, supplier involvement, and explicit mechanisms for documenting assumptions and assessment uncertainty.

Subjectivity and value judgments. Completing ALTAI inevitably involves subjective value judgments: assessors must interpret broad questions and weigh incomplete evidence. ALTAI can therefore not be used to obtain uniform and fully objective results. Moreover, the degree of subjectivity also varies across the seven pillars. Questions on *technical robustness and safety* are in principle empirically verifiable given adequate documentation, whereas *societal and environmental well-being* and *fairness*

require normative judgments on which analysts may differ even with complete information about the system; *privacy and data governance* combines verifiable compliance elements with subjective judgments of proportionality and impact. This mirrors risk management scholarship, including risk management under the EU AI Act [4, 9], where risk acceptability is recognised as partly a normative question involving trade-offs between values and interests rather than a purely technical measurement [7, 9]. For ALTAI, this implies that variation across analysts has two distinct sources: differences in information access, which better documentation and scoping can reduce, and differences in value judgments, which cannot be eliminated but can be made explicit and documented.

4. Our recommendations

Based on our feasibility study, we propose the following recommendations for practitioners applying ALTAI in trustworthiness assessments, as well as for further improving the tool.

For practitioners:

1. ALTAI should be considered as a framework that facilitates the assessment process—particularly structured discussion, reflection, and brainstorming—rather than as a tool for deriving a definitive level or score of trustworthiness for the system under analysis.
2. Assessment teams should be diverse and interdisciplinary to ensure the comprehensive competencies needed for the assessment. Where possible, the process should involve developers, deployers, operators, domain experts, and representatives of end users. No single role is likely to possess all the technical, organizational, legal, and contextual knowledge needed to reliably address various ALTAI dimensions.
3. The assessment process should be conducted collaboratively and iteratively, allowing for discussion, calibration, and refinement of responses. Differences in interpretations and assessments across team members reveal uncertainties about the system and ambiguities in the tool itself. These differences should be documented and used to calibrate responses across iterations.
4. Customer journeys and associated modelling languages can support context understanding and facilitate identification of the relevant architectural aspects which should be considered in the evaluation.

For further improvement of the tool:

1. The concept of Trustworthiness of AI systems is in general underspecified as well as broad, implying that the guidance on the requirements and the underlying questions should be further elaborated and more detailed.
2. *Traceability* between the ALTAI questionnaire, requirements, and resulting recommendations should be made more explicit within the tool. Each mitigation recommendation in the output should be linked to the specific questions, responses, and ALTAI pillars that triggered it. This would enable analysts to better understand the rationale behind the recommendations and to prioritise mitigation actions more effectively.
3. ALTAI should provide support for *context-sensitive prioritisation of requirements*. The relative importance of ALTAI pillars may vary depending on the type of AI system and deployment environment. For example, privacy and data governance may be particularly critical in healthcare settings, whereas transparency and human oversight may be more important in high-impact decision-support systems. The importance of the requirements should be provided, e.g. through weights and rationale. Alternatively, there should be support in the tool to specify importance of the requirements and the related criteria (or questions posed).
4. The tool should have support for *explicit expressions of uncertainty*. This would enable respondents (i.e., analysts or users) to indicate their level of confidence when answering individual questions, as well as to document the rationale and assumptions underlying their responses. The tool should

enable the uncertainty expressed to be propagated from questions to the recommendations, thereby reflecting an overall level of confidence in the assessment results. Uncertainty representations should be designed to be intuitive and interpretable for users with diverse backgrounds and levels of expertise.

5. The tool should have support for an *adaptive assessment process* that can be tailored to the specific domain, role, and system configuration or scope. Such adaptability would make the tool more relevant across diverse real-world scenarios, such as including developers, deployers, integrators of off-the-shelf AI components, and organisations assessing AI-enabled service components rather than complete AI systems.
6. The tool should support an early scoping of the target of the analysis, in order to focus the analysis on the relevant trustworthiness requirements and the related factors. For example, analysis of trustworthiness of a foundational model would focus on different aspects of the ALTAI tool compared to analysis of an off-the-shelf AI which constitutes a part of a broader system.

5. Limitations, validity and reliability

The study is based on a single industry case and one type of AI-enabled service. The findings should therefore be interpreted as an exploratory feasibility study rather than as generalizable evidence. Certain ALTAI questions appeared to assume that responding user or organization develops and trains its own AI model, rather than integrating a third-party component into a broader service ecosystem. The assessment was further limited by the response options available in ALTAI tool. The tool provides limited support for *expressing uncertainty*, *marking questions as not relevant*, or *documenting assumptions underlying responses*. In particular, the absence of a *"don't know"* option forced analysts to provide best-guess answers. These limitations likely contributed to variation across assessments. Also, analysts did not have full access to all technical and organizational documentation, which introduced uncertainty into their assessment. We consider incomplete documentation of the component a defining characteristic of the deployment context under study rather than an avoidable flaw; however, it does mean that the study cannot cleanly separate variation caused by information asymmetry from variation caused by ambiguity in ALTAI's questions or by genuine differences in value judgments. Our conclusions are therefore restricted to the feasibility and process value of applying ALTAI under such conditions; we make no claims about the actual trustworthiness of ProdRecAI, nor about ALTAI's performance under conditions of full documentation.

The study has limited *external validity*, as it is based on a single service component within one industry case, and the high variability across assessments makes it difficult to generalise the findings to other contexts. The results also indicate limited *construct validity*, since several ALTAI questions were broad, abstract, and open to interpretation; consequently, analysts may not have assessed the underlying trustworthiness constructs in the same way. While ALTAI is intended to be applied by an interdisciplinary analysis group, many of its questions are specific and technically detailed, presupposing detailed technical insight to the system — that was not available to any of the analysts in our study. The variation across the four radar diagrams accordingly indicates *low inter-rater reliability*. Under these information conditions this is an expected and informative outcome about ALTAI's sensitivity to information access and interpretation, rather than only a deficiency of the assessment.

The reliability of these findings for critical domains such as energy, banking, transportation, and health must be considered low: neither the scores nor the specific concerns transfer. What can be expected to transfer are the process-level insights: the competence requirements, information prerequisites, and the value of ALTAI as a structured reflection instrument.

6. Conclusion

This study examined the feasibility and practical value of applying ALTAI for assessing the trustworthiness of an AI-enabled service in an industrial setting. The findings indicate that ALTAI is feasible to

apply and effective for structuring reflection, discussion, and identification of trustworthiness-related concerns. However, the framework should not be treated as an objective measure of trustworthiness. Its main value lies in supporting a structured assessment process that helps organisations uncover assumptions, uncertainties, knowledge gaps, and differences in interpretation. The study also highlights the importance of iterative calibration and careful documentation of assessment rationale. Although the graphs that resulted from the analysis had too much variation, the analysts found the process itself useful, as it facilitates understanding of trustworthiness and the degree to which its criteria are fulfilled. Overall, ALTAI shows strong potential to support organisations in building and maintaining trustworthy AI systems. To strengthen ALTAI's practical applicability, the tool should provide better support for context-sensitive assessment, traceability, uncertainty expression, and adaptation to different roles, domains, and system boundaries. Such enhancements are essential to strengthen its robustness and usability in real-world settings. Future work should include additional case studies across domains, system types, and provisioning models.

Acknowledgments

This research was supported by *(i)* SINTEF project "CAT: Context-aware Trustworthiness Management of Human-Centric AI", funded by the Basic Funding through the Research Council of Norway; and *(ii)* "Privacy@Edge: Privacy-aware Edge and Data Subjects (338909)", funded by Research Council of Norway.

Declaration on Generative AI

During the preparation of this work, the author(s) used Microsoft 365 Copilot (for work) and ChatGPT-5.5 in order to: Grammar and spelling check, remove repeated sentences, improve sentence formulation and clarity. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] High-Level Expert Group on Artificial Intelligence, The assessment list for trustworthy artificial intelligence (ALTAI) for self assessment (Comprehensive Guide), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68342, 2020.

[2] High-Level Expert Group on Artificial Intelligence, the European AI Alliance, High-level expert group on artificial intelligence, https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai, 2020.

[3] High-Level Expert Group on Artificial Intelligence, Ethics Guidelines For Trustworthy AI, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai, 2019.

[4] European Parliament, Council of the European Union, EU Artificial Intelligence Act, Official Journal of the European Union L (2024) 1–116. URL: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.

[5] High-Level Expert Group on Artificial Intelligence, The Assessment List for Trustworthy Artificial Intelligence (ALTAI), https://altai.insight-centre.org/, 2020.

[6] Y. Tueanrat, S. Papagiannidis, E. Alamanos, Going on a journey: A review of the customer journey literature, Journal of business research 125 (2021) 336–353.

[7] National Institute of Standards and Technology (NIST), Artificial intelligence risk management framework (AI RMF 1.0) (2023) 100–1. URL: https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf.

[8] J. Chen, H. Dong, X. Wang, F. Feng, M. Wang, X. He, Bias and debias in recommender system: A survey and future directions, ACM Transactions on Information Systems 41 (2023) 1–39.

[9] J. Schuett, Risk management in the Artificial Intelligence Act, European Journal of Risk Regulation 15 (2024) 367–385. doi:10.1017/err.2023.1.