Marketplace

AI Certification

AI MARKETPLACE

AI Certification

Publisher:
Prof. Dr.-Ing. Roman Dumitrescu
Heinz Nixdorf Institute, Paderborn University, Germany

Authors:
Ashwin P. S. Venkatesh (Heinz Nixforf Institute)
Dr. Christian Soltenborn (Heinz Nixdorf Institute)

Contents

1 Introduction

The use of Artificial Intelligence (AI) is key to the advancement of technology across domains. However, the use of AI comes with its challenges, especially in the context of trustworthiness and ethical implications. Therefore, the certification of AI development processes will promote the adoption of AI systems by guaranteeing a certain standard of safety, security, and reliability.

Certification can be classified into two categories: 1) certification of an organization and its processes, and 2) certification of the software as a product produced by these organizations. In both cases, the goal of certification is increased trust in the organization and its products for the end customer. Certification of some well-established standards can be acquired by dedicated certifying institutions, for instance, the ISO 9001 standard. This will guarantee that the organization meets the quality and regulatory requirements as defined by the ISO standard.

While certification standards exist for traditional software systems, the certification of Artificial Intelligence systems is still an open research topic. To this end, the European Commission has taken the initiative and has recommended a regulatory framework and published certification concepts for AI systems. Since then, there have been public discussions and publications that provide constructive criticism of the recommendations of the EU Commission. The central recommendation is to uphold the ethical principles of the EU and at the same time to avoid over-regulation and hinder innovation.

The AI Marketplace (AIM) plays an important role as an enabler by bridging the gap between AI service providers and users. Therefore, establishing mutual trust between the participants by establishing standard certification schemes becomes crucial. Potential users of AI would use the AIM to find suitable partners for their use cases. On one hand, AI users need to be able to trust the companies listed on the platform. On the other hand, the AI service providers would want to increase their credibility and trustworthiness on the platform. To serve these requirements, the AIM has proposed a three-level certification scheme for AI service providers that are based on the latest developments and publications in the trustworthy AI development domain.

There are already efforts and publications towards trustworthy AI development. However, in a recently conducted a survey with the potential users of AI on our platform, we found out that 75% of the 36 participants are not aware of any certification standard or organization that is AI-specific. To this end, this document also aims at raising awareness regarding AI certification.

The rest of this document is structured as follows: chapter 2 analyzes the relevant certification standards and the scientific literature on trustworthy AI. Chapter 3 presents the AI certification schemes of AIM. Finally, chapter 4 concludes the document.

AI certification ensures safety, security, and reliability, promoting trust and adoption of AI systems.

A survey found 75% of potential users are unaware of AI-specific certifications.

2 Existing Certification Schemes for AI-based Systems

While traditional software systems have existed for a few decades now, AI-based systems have seen rapid development in the past decade. However, the law and regulations around the AI topic have not caught up to the speed of development of these systems. Recently there has been growing interest in establishing standards for the trustworthy development of AI systems in Europe. In this section, we discuss existing and ongoing efforts towards certification for AI systems from three organizations that are the most prominent and relevant in the German context: the EU commission (see section 2.1), KI.NRW (see section 2.2), and KI-Bundesverband (see section 2.3). Section 2.4 concludes the chapter with a discussion of the most relevant scientific literature on the trustworthy development of AI systems.

The EU Commission promotes trustworthy AI development by publishing guidelines and recommendations.

2.1 European Commission

The European Commission is committed to promoting the development of AI in the European context. To this end, the commission has established a high-level expert group on AI for creating a guideline for the trustworthy development of AI systems and to make policy recommendations. The high-level group has published the “Ethics Guidelines for Trustworthy AI” [Hig19] document that proposes a framework for trustworthy AI development. This was followed by the publication of “Assessment List for Trustworthy AI” [Hig20] which is intended for self-evaluation of compliance. The EU Commission also released the proposed regulation on AI, the so-called EU Artificial Intelligence Act.

Ethics Guidelines for Trustworthy AI

The ethics guidelines proposed by the high-level commission intend to promote the development of trustworthy AI by creating a framework for achieving lawful, ethical, and robust AI. The following statements summarize the core ideas of the paper:

- AI should be developed and deployed in accordance with respect for human autonomy, prevention of harm, fairness, and explicability.
- Special attention must be given to situations involving children, people with disabilities, and vulnerable groups.
- Adequate measures must be taken proportional to the magnitude of the risks.
- AI systems must meet seven key requirements for trustworthy AI:

1. Human Agency and Oversight
2. Technical Robustness and Safety
3. Privacy and Data Governance
4. Transparency
5. Diversity, Non-discrimination and Fairness
6. Societal and Environmental well-being
7. Accountability.

Assessment List for Trustworthy AI

The high-level working group on AI published an assessment list for trustworthy AI-based systems through extensive interviews with companies. The assessment list is intended for flexible use by organizations to investigate potential risks generated by their implementation of an AI system. All seven requirements from the ethics guidelines document are further elaborated with concrete questions. The assessment list is intended to be reviewed with a multidisciplinary team of experts from data scientists, and compliance officers to the management. The assessment list is also made available as an interactive online version[1]. Figure 1 shows an excerpt of the questionnaire for the requirement "human agency and oversight".

An assessment list for trustworthy AI systems was published after extensive company interviews.

Are end-users or subjects informed that they are interacting with an AI system?

Could the AI system affect human autonomy by generating overreliance by end-users?

Could the AI system affect human autonomy by interfering with the end-user's decision-making process in any other unintended and undesirable way?

Did you put in place any procedure to avoid that the AI system inadvertently affects human autonomy?

Figure 1: Excerpt of the questionnaire for the requirement "human agency and oversight" [Hig20]

[1] https://altai.insight-centre.org/

EU Artificial Intelligence Act

On 21 April 2021, the European Commission released the draft for the EU Artificial Intelligence Act[2]. After deciding on proposed amendments, the act is likely to be passed into law in 2023. The EU Artificial Intelligence Act addresses all AI applications that are developed using machine learning, Bayesian approaches, or logic and knowledge-based approaches, and that produce output like recommendations, content, predictions, etc.

The EU Artificial Intelligence Act defines 4 risk levels for AI applications. Depending on the risk level, different obligations and restrictions apply (see Figure 2):

The EU Artificial Intelligence Act, expected to pass in 2023, defines four risk levels for AI applications with corresponding obligations.

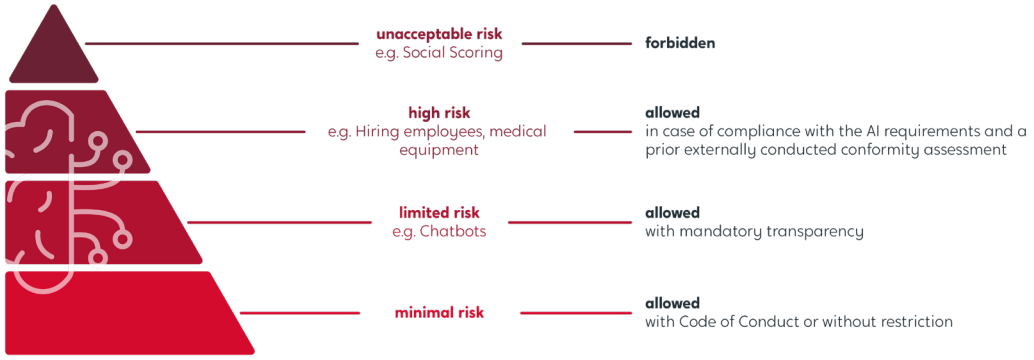

Figure 2: AI risk levels proposed by EU Commission[2]

- **Unacceptable risk** AI applications are prohibited. These applications include subliminal techniques, manipulation, social scoring, and biometrics.
- **High risk** AI applications are subject to strict obligations before they can be put on the market. Such AI applications deal with critical infrastructure, educational or vocational training, safety components of products, employment, management of workers and access to self-employment, essential private and public services, law enforcement, migration, asylum and border control management, and administration of justice and democratic processes.
- **Limited risk** AI applications need to fulfill specific transparency obligations. This category includes AI applications such as chatbots, emotion recognition and biometric categorization systems, and systems generating deepfake or synthetic content.
- **Minimal risk** AI applications are allowed without restrictions. Spam filters or AI-enabled video games represent examples for minimal risk AI applications.

[2] https://artificialintelligenceact.eu/the-act/

2.2 KI.NRW – Künstliche Intelligenz Nordrhein-Westfalen

KI.NRW[3] in association with Fraunhofer IAIS is working towards establishing certification schemes for AI systems. To this end, KI.NRW has published a white paper discussing the AI certification topic under the title “Trustworthy Use of Artificial Intelligence” [Fra19]. The whitepaper is followed by another publication with a detailed catalog for the inspection of AI systems [Fra21].

Whitepaper: Trustworthy Use of Artificial Intelligence

The whitepaper is meant to be a starting point for the interdisciplinary development of AI certification schemes between IT, law, philosophy, and ethics. The agenda was to create material for neutral auditors to be able to cross-check AI applications for trustworthiness from an ethical and legal perspective. The publication describes six audit areas specific to AI as shown in Figure 2.

The KI.NRW and Fraunhofer IAIS whitepaper outlines audit areas focusing on safety, ethics, and legal compliance.

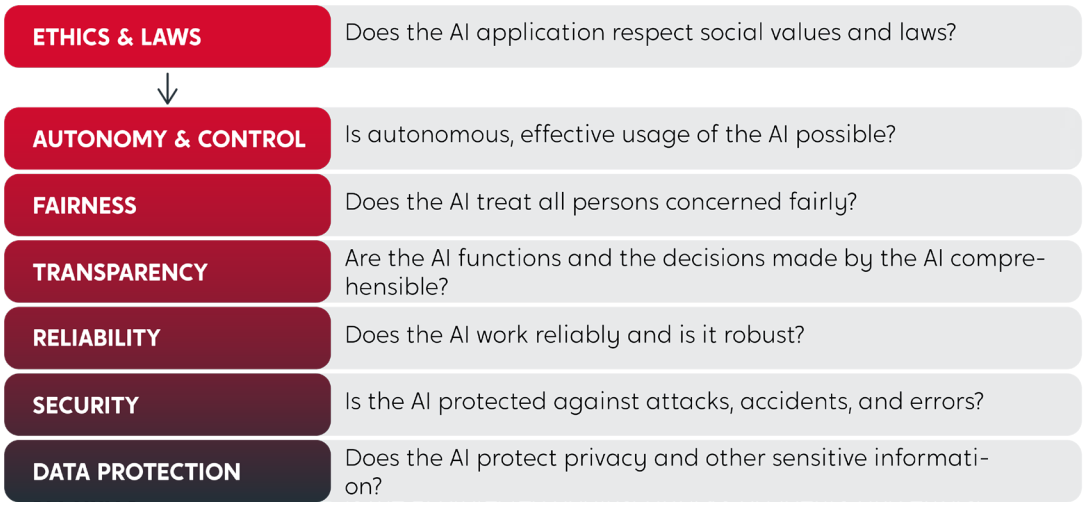

Figure 3: Audit areas from the KI.NRW and Fraunhofer IAIS whitepaper [Fra19], [Fra21]

[3] https://www.ki.nrw/en/ai-certification/

The authors discuss AI-specific challenges in the context of these audit areas and highlight key challenges in the area:

- AI applications should be designed in a way that allows us to verify if they work safely and reliably.
- Applications should be aligned with the ethical and legal framework.
- In addition to technical safeguards, it should be clear under what circumstances the use of AI is ethically and legally acceptable.
- An interdisciplinary team from IT, philosophy, and law is required to resolve the challenges of AI certification.
- Ethics come from historical experiences gained by people and therefore is difficult to implement ethics as a code. Therefore, a new AI development guideline that accounts for a universal human value system is required.

Inspection Catalog for AI Systems

The initial whitepaper by KI.NRW established the direction of creating a standard for trustworthy AI development. As a follow-up, the inspection catalog is a document of over 160 pages detailing the guidelines for developing trustworthy AI systems with high-quality standards. The document is targeted at two groups: 1) as a basis for neutral auditing organizations to evaluate AI systems developed by companies, and 2) as a guideline for developers who are designing and developing AI systems.
The inspection catalog first formally defines a structure and life cycle of an AI appliction. It further builds on the six aspects of trust areas described in the whitepaper (see Figure 1). The catalog proposes a two-phase approach to testing the AI system, with the so-called “Top-down” and “Bottom-up” testing approach. First, the AI risks are identified and analyzed. Based on this, criteria are derived. Second, each risk area is mitigated by taking measures that reduce the risk and fulfill the previously defined criteria.

The detailed inspection catalog provides guidelines for trustworthy and neutral auditing of AI systems.

2.3 KI-Bundesverband

KI-Bundesverband[4] is a German AI association that represents more than 250 innovative companies that focus on the development and application of AI. In the following sections, two whitepapers published by KI-Bundesverband are discussed. The first publication, “KI Guetesiegel” [KI 19], is about defining and adhering to a common understanding of ethical values concerning AI systems. This includes a self-declaration form that represents self-commitment by the organizations that sign them. The second publication, “Position Paper on EU-Regulation of Artificial Intelligence by the German AI Association” [KI 21], gives six key recommendations for the regulation of AI in the European context.

KI Guetesiegel (AI seal of approval)

KI Guetesiegel aims towards establishing a seal of approval for ethical product and service development of AI systems. It offers German companies an opportunity to strengthen compliance with basic quality parameters of trustworthy AI development. The whitepaper proposes four quality criteria as guiding principles for trustworthy AI development: ethics, impartiality, transparency, and security.

- **Ethics:** AI development should maintain basic European values: human dignity, freedom, democracy, equality, and the rule of law.
- **Impartiality:** old stereotypes and prejudices that might already be present in the data should not be reinforced by the AI system. Therefore, trained personnel should use appropriate analysis techniques to detect and mitigate this bias.
- **Transparency:** every step from data pre-processing, feature engineering, model creation, and model evaluation must be well-documented.
- **Security:** an AI system processes data just like classic data processing systems. Therefore, the same requirements for the confidentiality and integrity of data processing apply to AI systems as well.

KI Guetesiegel, a seal of approval, encourages German companies to adhere to ethical, neutral, transparent, and secure AI development.

The AI service provider company will be given a “KI Guetesiegel” upon submitting a signed self-declaration of commitment towards the guidelines published by KI-Bundesverband.

[4] https://ki-verband.de/

Position Paper on EU-Regulation of AI

KI-Bundesverband proposes a nine-point plan for AI regulation in the EU. The recommendations were part of the discussions regarding the European Commission's whitepaper on AI. Authors note that a general regulation on AI is not feasible and therefore ethical aspects of an AI must be evaluated on a case-by-case basis concerning the individual context or use case. Therefore, a framework is proposed for risk evaluation of new use cases to make regulatory restrictions proportional to the risk. Special considerations are taken to make sure that there should be a low barrier to entry. Such that the regulations are not a burden to small enterprises and startups.

The KI-Bundesverband proposes a nine-point plan for EU AI regulation, including a case-by-case evaluation.

2.4 Relevant Scientific Literature

This section presents a summary of recent literature that study novel techniques for testing and establishing trust in machine learning and artificial intelligence software systems.

FactSheets: Increasing Trust in AI Services through Supplier's Declaration of Conformity

The FactSheets are inspired by the practice in other industries to use standardized supplier declarations of conformity. Although these declarations may not have legally binding, they can still be used to enhance trust between the end user and the supplier. The authors suggest a comprehensive list of declarations related to AI services and further provide two concrete examples [ABH+19].
FactSheets are intended to be a voluntary declaration and propose possible extensions for the future. A FactSheet is based on the supplier's declaration of conformity (SDoC). An SDoC is a written assurance with evidence of conformity to a specific requirement. It is used in many industries to create trust between suppliers and clients.

Some of the key elements typically included in a FactSheets are:

- **Purpose:** overview of the intended uses of the service.
- **Domains and applications:** information about the application domain and how the service will be used
- **Basic performance:** information about the assessment of the service performance. For instance, the dataset that was tested on, testing methodology,

FactSheets, a voluntary supplier declaration of conformity, enhance the trust between the end user and the supplier.

performance metrics such as accuracy, error rates, etc.

- **Security:** details about how the service could be attacked the steps taken to mitigate them.

Model Cards for Model Reporting

"Model Cards" is a research paper from Google that recommends that each AI model can be accompanied by documentation detailing their characteristics. The purpose of Model Cards is to improve transparency and accountability in the development and deployment of machine learning models. It is noted that there is no standard documentation scheme to communicate the characteristics of a trained model. The authors also provide two examples to showcase their concept. Google has released a website[5] with two reference model cards as an initial step in establishing this as a standard [MWZ+19].

Some of the key elements typically included in a Model Card are:

- **Model details:** information about the model version, date, license, architecture, hyperparameters, etc.
- **Intended use case:** describes the problem the model is designed to solve, including the intended users and potential out-of-scope use cases.
- **Performance metrics:** information about the model's accuracy and other performance metrics such as accuracy, precision, recall, F1 score, AUC, etc. Could also include comparison of the performance with other relevant models and benchmarks.
- **Training data:** information about the data used to train the model. Providing this may not be possible at all times, but minimal information that can be made public should be provided here. For instance, size, diversity, and distribution.
- **Ethical considerations:** potential ethical considerations should be documented, such as fairness, privacy, transparency, and accountability.
- **Caveats and recommendations:** information about potential limitations or biases in the model, such as underrepresentation of certain groups in the training data. It could also contain known limitations of the model architecture, training, or evaluation processes.

Model Cards, a concept from Google, aim to improve AI transparency by detailing characteristics of AI models.

[5] https://modelcards.withgoogle.com/about

Datasheets for Datasets

Like the previous two publications, this work from Google and Microsoft is inspired by datasheets in the electronics industry, where each component is accompanied by its characteristics and recommendations. The authors recommend a similar datasheet for machine learning datasets to enable better communication and transparency between dataset creators and users [GMV+21].
The intention is to encourage dataset creators to take caution during the creation and maintenance of datasets. Thereby, documenting any underlying assumptions and the potential implications of its use. On one hand, the dataset creator is more cautious. The dataset consumer, on the other hand, can ensure that they have the information that is needed to make full use of the dataset.

Google and Microsoft propose datasheets for machine learning datasets, enhancing transparency and communication between creators and users.

Some of the key elements typically included in a Datasheet are:

- **Motivation:** lists the reasons why the dataset was created, and the problem or question it addresses.
- **Composition:** description of the data sources, including the diversity of the dataset, as well as any potential biases or limitations in the data.
- **Collection process:** description of how the data was collected. For instance, whether people were involved in the process, or if any ethical review process was conducted.
- **Preprocessing:** description of any preprocessing steps applied to the data, such as normalization or removal of instances.
- **Uses:** description of how the data should and should not be used.
- **Distribution:** information on how the dataset is made available to users, including any privacy or access restrictions.
- **Maintenance:** description of how the dataset will be maintained and who will be responsible for the support and maintenance of the dataset.

Further Related Work

The following are some of the recent publications in the trustworthy AI development research area:

Machine Learning Testing: Survey, Landscapes, and Horizons: This paper provides a comprehensive survey of ML testing research by reviewing 138 scientific papers regarding testing approaches for ML systems [ZHM+22].

Testing and Quality Validation for AI Software-perspectives, Issues, and Practices: This paper discusses testing and quality validation techniques for AI software features [TGW19].

Recent publications in trustworthy AI development research cover ML testing, data quality, software certification, and development processes.

The Dataset Nutrition Label: A Framework to Drive Higher Data Quality Standards: The authors propose a diagnostic framework that intends to provide a standardized view of the core components of a dataset [HHN+18]. The intention is similar to model cards and datasheets as discussed earlier.

On Testing Machine Learning Programs: In this paper, authors review existing testing practices for ML solutions found in the literature and identify gaps in them, followed by recommendations for future research directions [BK20].

Software Certification: Methods and Tools (Dagstuhl Seminar 13051): In this edition of the Dagstuhl Seminar, experts from both academia and industry met and published discussions on the challenges and best practices of certification technologies [CHH+13].

The Five Laws of SE for AI: In this paper, the authors discuss five software engineering principles that should be applied to AI software to improve the development processes [Men20].

3 AI Marketplace Certification Schemes

AI is an emerging field. Naturally, it has novel challenges when it comes to certification of its functionality. To create trust, a software system needs to be designed and implemented in such a way that the desired functionality can be tested and verified. However, given the black-box nature of many AI techniques, such deterministic tests are challenging to implement.

AI systems pose a challenge in terms of legal and ethical frameworks. Establishing clear guidelines for the development of AI is a continuously evolving topic. Researchers, politicians, and experts in human aspects of AI are debating this issue over the last few years. To this end, organizations such as KI.NRW[6], European Commission, and Fraunhofer IAIS[7] have interdisciplinary experts studying this topic and publishing their results for public debate. Therefore, certification schemes designed to certify AI software should consider the latest guidelines and recommendations from various discussions around this topic.

AI service providers are companies offering AI software products in the marketplace. The certification of service providers is essential to establish mutual trust between users of AI and AI service providers. Existing industry-standard certification schemes are a good indicator that a particular organization has established standard working procedures. However, there is no well-established certification scheme that yet specifically targets trustworthy AI development. Moreover, obtaining an industry-standard certification is not feasible for small-scale companies and startups. Therefore, certification schemes should include schemes with a low barrier to entry. Furthermore, as suggested by recent publications in the trustworthy AI community, certification schemes should consider the severity of the use case the AI service is targeting and proportionately regulate the requirements.

The AI Marketplace offers three certification schemes aiming to increase trust between AI providers and AI users.

Based on these recommendations, the AI Marketplace offers three certification schemes to increase mutual trust between AI software providers and AI users in the platform. These schemes are inspired and adopted based on existing certification schemes and scientific literature. Specifically, AI Marketplace promotes the adoption of guidelines and best practices for trustworthy AI development published by KI.NRW and the EU Commission. The three AI certification schemes are named:

1. AI company certification
2. Extended AI company certification
3. AI application certification

The level of trust increases from the AI company certification to the AI application certification. However, the effort for certification increases as well.

AI providers must go through the pro-

[6] https://www.ki.nrw/en/
[7] https://www.iais.fraunhofer.de/

cesses defined to obtain an AI badge for the respective AI certification level. The badge will then be displayed on the platform along with the offerings. Subsequently, the three certification levels and the process steps for obtaining them are described.

3.1 AI Company Certification

The AI company certification is primarily about raising awareness in the community about existing efforts toward establishing best practices for trustworthy AI development. The certification scheme is designed as a low-barrier-to-entry for AI solution providers. Therefore, based on “Self-declaration”, i.e., the provider is responsible for adhering to the guidelines of trustworthy AI software development published by KI.NRW. The AI company certification addresses the EU low and limited risk levels.

The AI company certification raises awareness and requires self-declaration to adhere to trustworthy AI guidelines.

Figure 4 shows the process steps for obtaining the AI company certification:

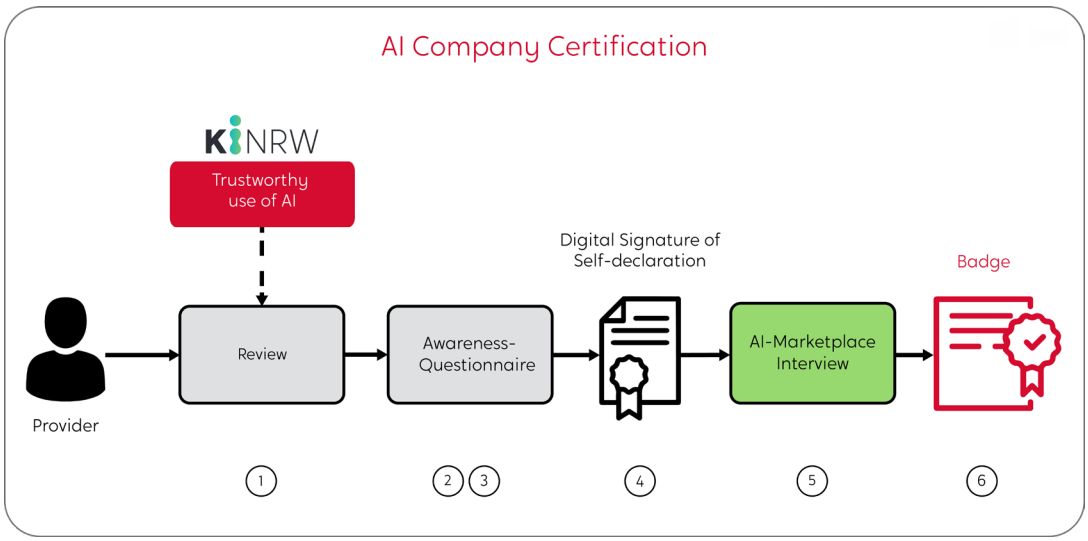

Figure 4: Process for AI company certification

1. **Review:** The applicant reviews the whitepaper on the trustworthy use of AI from KI.NRW.
2. **Awareness Questionnaire:** The applicant answers the questionnaire (see attachment) as reassurance.
3. **Self-declaration:** The applicant submits a self-declaration form stating that the applicant is aware of the guidelines on trustworthy AI development and would strive to adhere to them.

4. **Digital Signature:** The form needs to be submitted with a digital signature by the applicant using keys validated by a mutually trusted certification authority.
5. **AI Marketplace Interview:** The digitally signed self-declaration and the awareness questionnaire are submitted to AI Marketplace. The documents will be approved by the platform after internal checks and a one-on-one meeting with the applicant.
6. **Badge:** After approval, an AI company badge will be displayed in the profile of the applicant on AI Marketplace.

3.2 Extended AI Company Certification

The extended AI company certification scheme is aimed at organizations that have already obtained one or more industry-standard AI certifications. For obtaining such certifications, elaborated quality checks have already taken place. Therefore, this adds more credibility to the self-declaration from the provider. The AI certification scheme addresses the EU low and limited risk levels.

Figure 4 shows the process steps for obtaining the extended AI company certification. Process steps 1)-4) are identical to the AI company certification:

The extended AI company certification is aimed at organizations with generic industry-standard certifications, providing more credibility.

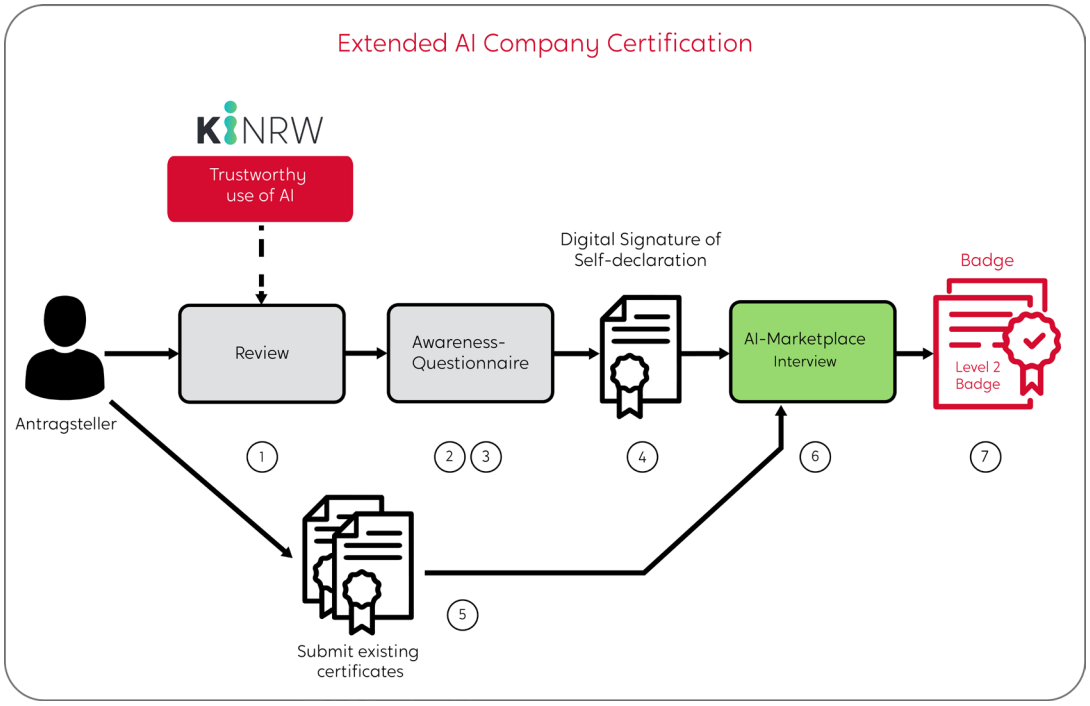

Figure 5: Process for extended AI company certification

1. **Review:** The applicant reviews the whitepaper on the trustworthy use of AI from KI.NRW.
2. **Awareness Questionnaire:** The applicant answers the questionnaire (see attachment) as reassurance.
3. **Self-declaration:** The applicant submits a self-declaration form stating that the applicant is aware of the guidelines on trustworthy AI development and would strive to adhere to them.
4. **Digital Signature:** The form needs to be submitted with a digital signature by the applicant using keys validated by a mutually trusted certification authority.
5. **Existing Certificates:** The applicant submits existing AI certifications from the following list of accepted certifications:
   - ISO 9001 [ISO9001]
   - ISO 27001 [ISO/IEC27001]
   - ISO/IEC DIS 27070 [ISO/IEC27070]
   - ISO/IEC 33001 [ISO/IEC33001]
   - ISO/IEC 90003 [ISO/IEC90003]
   - BSI C5 [Bun20]
   - Other certifications can also be considered based on merit.
6. **AI Marketplace Interview:** The digitally signed self-declaration, the awareness questionnaire, and the AI certificates are submitted to AI Marketplace. The documents will be approved by the platform after internal checks and a one-on-one meeting with the
7. applicant.

   Badge: After approval, an AI company badge and the further approved AI certificates will be displayed in the profile of the provider on AI Marketplace.

3.3 AI Application Certification (Coming Soon)

The AI application certification scheme is the most elaborate. It involves a manual audit by an expert. This scheme is aimed at use cases that are critical and therefore must be scrutinized individually. In this way, the EU high risk level is addressed. Figure 7 shows the process steps for obtaining the AI application certification:

The AI application certification scheme involves manual audits by experts for critical AI use cases.

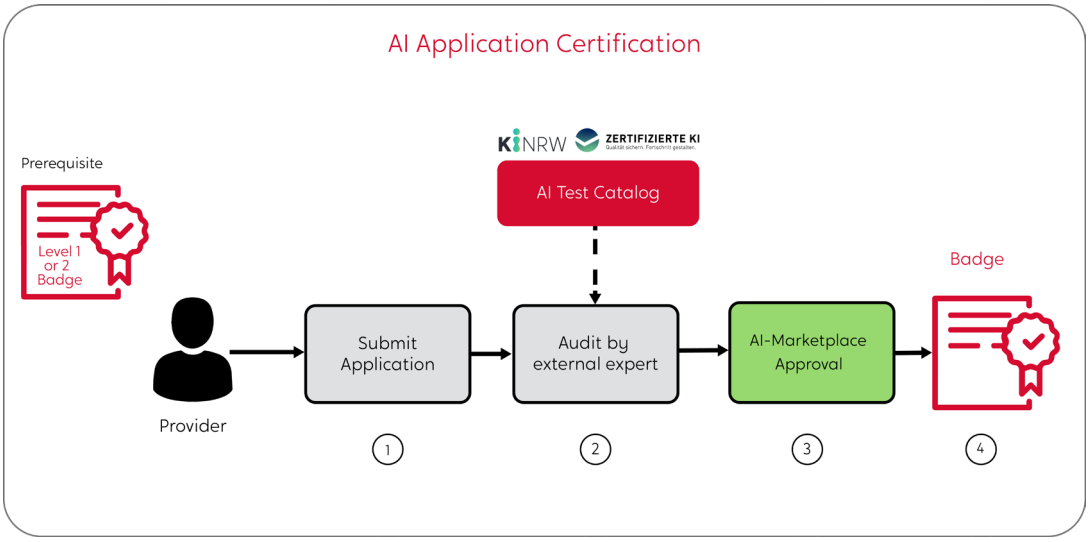

Figure 6: Process for AI application certification

Prerequisite: The applicant has already acquired an AI company badge.

1. **Submit Application:** The applicant provides the executable application, sample dataset, documentation, etc., of the AI application to the auditor upon proper authorization and non-disclosure agreements.
2. **Audit:** The application is audited by experts based on published guidelines.
3. **Approval:** The inputs from the auditor are checked and approved by the AI Marketplace team.
4. **Badge:** After approval, an AI application badge will be displayed on the platform for the specific AI application.

4 Conclusion and Outlook

In this paper, we briefly discussed the core ideas of existing efforts towards certification of trustworthy AI development. We further presented the three certification schemes proposed by the AI Marketplace. These schemes are designed based on the recommendations of the EU Artificial Intelligence Act of the EU Commission and KI.NRW. To this end, the AI company certification scheme has a low barrier to entry while the AI application certification scheme has more rigorous requirements. This is intended to serve different use cases based on the criticality of the respective AI systems.

Going forward, we intend to introduce further automated tests to the requirements to make the overall process more objective and transparent. Furthermore, legal developments, like the adoption of the AI Artificial Intelligence Act, need to be monitored closely. This way, it can be ensured, that the continuously updated AI certification schemes of the AI marketplace consider the latest developments in the field of AI certification.

Bibliography

[Asp22-ol] Arnold, M.; Bellamy, R. K. E.; Hind, M.; Houde, S.; Mehta, S.; Mojsilovic, A.; Nair, R.; Ramamurthy, K. N.; Olteanu, A.; Piorkowski, D.; Reimer, D.; Richards, J.; Tsay, J.; Varshney, K. R.: FactSheets: Increasing trust in AI services through supplier's declarations of conformity. IBM Journal of Research and Develop-ment, (63)4/5, 2019, 6:1-6:13

[BK20] Braiek, H. B.; Khomh, F.: On testing machine learning programs. Journal of Systems and Software, (164), 2020, S. 110542

[Bun20] Cloud Computing Compliance Criteria Catalogue – C5:2020, 2020

[CHH+13] Cofer, D.; Hatcliff, J.; Huhn, M.; Lawford, M.: Software Certification: Methods and Tools (Dagstuhl Seminar 13051). 38 pages, 2013

[Fra19] Thrustworthy Use of Artificial Intelligence. Sankt Augustin, 2019

[Fra21] Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz – KI-Prüfkatalog. Sankt Augustin, 2021

[GMV+21] Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; III, H. D.; Crawford, K.: Datasheets for datasets. Communications of the ACM, (64)12, 2021, S. 86–92

[HHN+18] Holland, S.; Hosny, A.; Newman, S.; Joseph, J.; Chmielinski, K.: The Dataset Nutri-tion Label: A Framework To Drive Higher Data Quality Standards. arXiv, 2018

[Hig19] High-level Expert Group on Artificial Intelligence: Ethics Guidelines for Trustworthy AI, 2019

[Hig20] High-level Expert Group on Artificial Intelligence: Assessment List for Trustworthy Artificial Intelligence (ALTAI) – for self-assessment, 2020

[ISO/IEC27001] Information security, cybersecurity and privacy protection - Information security management systems - Requirements, 2022

[ISO/IEC27070] Information technology - Security techniques - Requirements for establishing virtualized roots of trust, 2021

[ISO/IEC33001] Information technology - Process assessment - Concepts and terminology, 2015

[ISO/IEC90003] Software engineering - Guidelines for the application of ISO 9001:2000 to computer software, 2004

[ISO9001] Quality Management Systems - Requirements

[KI 19] KI Gütesiegel, 2019

[KI 21] Position Paper on EU-Regulation of Artificial Intelligence by the German AI Association, 2021

[Men20] Menzies, T.: The Five Laws of SE for AI. IEEE Software, (37)1, 2020, S. 81–85

[MWZ+19] Mitchell, M.; Wu, S.; Zaldivar, A.; Barnes, P.; Vasserman, L.; Hutchinson, B.; Spitzer, E.; Raji, I. D.; Gebru, T.: Model Cards for Model Reporting: Proceedings of the Conference on Fairness, Accountability, and Transparency. FAT* ,19: Conference on Fairness, Accountability, and Transparency, 29-31.1.2019, Atlanta GA USA, ACM, Atlanta, GA, USA, 2019, S. 220–229

[TGW19] Tao, C.; Gao, J.; Wang, T.: Testing and Quality Validation for AI Software–perspectives, Issues, and Practices. IEEE Access, (7), 2019, S. 120164–120175

[ZHM+22] Zhang, J. M.; Harman, M.; Ma, L.; Liu, Y.: Machine Learning Testing: Survey, Landscapes and Horizons. IEEE Transactions on Software Engineering, (48)1, 2022, S. 1–36

Attachment: Awareness Questionnaire of the AI Marketplace

Trustworthy AI Awareness Questionnaire

The following statements are summarized from the white paper published by the Fraunhofer IAIS titled ***"Trustworthy Use of Artificial Intelligence"***.

Ticking the following check boxes indicate that you have read and understood relevant sections from the publication.

1.	I understand that the **"Audit Areas"** described in the white paper contains the following six areas:
☐	Autonomy and control
☐	Fairness
☐	Transparency
☐	Reliability
☐	Security
☐	Data protection
2.	I understand that, the following statements are true in the context of **"Autonomy and control"**
☐	Autonomy of Individuals and social groups may not be disproportionately harmed by AI.
☐	Users must be able to withdraw their consent to use an AI application at any time.
☐	The users must have autonomy and control over the AI system's objectives.
☐	It must be possible to entirely turn off the AI application.
3.	I understand that, the following statements are true in the context of **"Fairness"**
☐	Individuals may not be treated unfairly because they belong to a marginalized or discriminated group.
☐	Since AI systems learn from historical data, caution should be exercised to avoid perpetuating any prejudices that may be present in the society.
☐	Fairness measurement should be established clearly first and then be made quantifiable.
4.	I understand that, the following statements are true in the context of **"Transparency"**
☐	It should be made clear that the communication is made with an AI system.
☐	Clear statements about the purpose of the application and potential risks.
☐	Transparency is about the traceability is occurring with an AI application.
☐	Sufficient documentation and explanation should be available such that AI experts can understand the technical details of the system.

Figure A-1: Awareness Questionnaire of the AI Marketplace (page 1/2)

5.	I understand that, the following statements are true in the context of **"Reliability"**
☐	Reliability refers to various aspects of AI: Correctness of output, Model uncertainties, Robustness to harmful inputs, etc.
☐	Quantitative measurements of reliability should be specified as precisely as possible and be formalized in order to ensure that data used for training sufficiently covers the expected inputs to the AI system.
☐	Functionality should be checked at regularly at appropriate intervals.
☐	There must be a back-up plan in case of reliability failure.

6.	I understand that, the following statements are true in the context of **"Security"**
☐	Security of AI systems are at least as high as any other IT systems.

7.	I understand that, the following statements are true in the context of **"Data protection"**
☐	Must ensure relevant data Protection-law regulations, for instance, the General Data Protection Regulation (GDPR) and Germany's Federal Data Protection Acts are observed.
☐	There is risk that a trained model could contain references to a person without actually containing personal data.
☐	AI applications may access personal data only with the consent of the owner.
☐	Measures must be taken to make data anonymous or protect against the potential for re-identification.

8.	Additionally, I'm aware of the following publications in the context of **"Trustworthy AI"**
☐	Reliability refers to various aspects of AI: Correctness of output, Model uncertainties, Robustness to harmful inputs, etc.
☐	Quantitative measurements of reliability should be specified as precisely as possible and be formalized in order to ensure that data used for training sufficiently covers the expected inputs to the AI system.
☐	Functionality should be checked at regularly at appropriate intervals.

Figure A-2: Awareness Questionnaire of the AI Marketplace (page 1/2)

Imprint

Year of publication:
2023

Place of publication:
Paderborn

Publisher:
Prof. Dr.-Ing. Roman Dumitrescu
Heinz Nixdorf Institut, Universität Paderborn

Authors:
Ashwin P. S. Venkatesh (Heinz Nixdorf Institut)
Dr. Christian Soltenborn (Heinz Nixdorf Institut)

Design:
Lena Heller

HEINZ NIXDORF INSTITUT
UNIVERSITÄT PADERBORN

Heinz Nixdorf Institut
Universität Paderborn
Fürstenallee 11
33102 Paderborn
Telefon +49 (0) 5251 | 60 62 67
Telefax +49 (0) 5251 | 60 62 68
www.hni.uni-paderborn.de

Supported by:

Federal Ministry
for Economic Affairs
and Climate Action

on the basis of a decision
by the German Bundestag