ISSN: 2508-7894
KJAI website: http://acoms.kisti.re.kr/kjai
doi: http://dx.doi.org/10.24225/kjai.2023.11.3.1

Trends of Artificial Intelligence Product Certification Programs*

Yejin SHIN[1], Joon Ho KWAK[2], KyoungWoo CHO[3], JaeYoung HWANG[4], Sung-Min WOO[5]

Received: April 14, 2023. Revised: August 02, 2023. Accepted: September 02, 2023.

Abstract

With recent advancements in artificial intelligence (AI) technology, more products based on AI are being launched and used. However, using AI safely requires an awareness of the potential risks it can pose. These concerns must be evaluated by experts and users must be informed of the results. In response to this need, many countries have implemented certification programs for products based on AI. In this study, we analyze several trends and differences in AI product certification programs across several countries and emphasize the importance of such programs in ensuring the safety and trustworthiness of products that include AI. To this end, we examine four international AI product certification programs and suggest methods for improving and promoting these programs. The certification programs target AI products produced for specific purposes such as autonomous intelligence systems and facial recognition technology, or extend a conventional software quality certification based on the ISO/IEC 25000 standard. The results of our analysis show that companies aim to strategically differentiate their products in the market by ensuring the quality and trustworthiness of AI technologies. Additionally, we propose methods to improve and promote the certification programs based on the results. These findings provide new knowledge and insights that contribute to the development of AI-based product certification programs.

Keywords: AI, Artificial Intelligence, AI products, Certification Programs

Major Classification Code: Artificial Intelligence

1. Introduction

Artificial intelligence (AI) technology has been continuously evolving with its adoption in various industries. AI is gradually being used in everyday applications that were previously considered unimaginable, such as diagnosing diseases and serving customers. AI algorithms based on data-driven deep neural networks can exhibit excellent performance approaching that of human judgment in several fields and are rapidly replacing most classical machine learning algorithms. However, the process of using AI involves certain inherent risks and potential side effects (Shin, 2022). The perception that machines are replacing human labor can raise questions about the credibility of AI technologies that relate to fairness and transparency. In addition, most AI algorithms are black-box models, and explaining how their results were derived to developers and consumers can be challenging. This is an important issue in high-risk fields where safety is a top priority. Therefore, the need to inform consumers regarding

* This work was supported by the Korean MSIT (Ministry of Science and ICT) as Establishing the foundation of AI Trustworthiness(TTA) & This work was supported by Education and Research promotion program of KOREATECH in 2023.

1 First Author. Principal Researcher, AI Trustworthiness Verification Team, TTA, Republic of Korea. Email: yepp1252@tta.or.kr
2 Second Author. Team Manager, AI Trustworthiness Verification Team, TTA, Republic of Korea. Email: pentarous@tta.or.kr
3 Third Author. Principal Researcher, AI Trustworthiness Verification Team, TTA, Republic of Korea. Email: pinokio622@tta.or.kr
4 Fourth Author. Principal Researcher, AI Trustworthiness Verification Team, TTA, Republic of Korea, Email: jyhwang@tta.or.kr
5 Corresponding Author. Assistant Professor, School of Electrical, Electronics and Communication Engineering, KOREATECH, Republic of Korea, Email: innosm@koreatech.ac.kr

the safety and trustworthiness of AI-based products has been established in the relevant literature. Safety and trustworthiness should be considered critical thresholds for adopting AI in practical applications (Byun, 2019).

Many guidelines to evaluate the quality and trustworthiness of AI technologies have been released by institutions in various countries along with the results of associated research (Cihon et al., 2021; Tao et al., 2019; Machlev et al., 2021; Breck et al., 2017; CDEI, 2021; AI HLEG, 2020). Moreover, public- and private-sector AI product certification programs have been developed in recent years to document that these guidelines have been adopted and implemented (Genovesi & Julia, 2022). However, these certification programs are not uniform and operate with different testing criteria and standards. Hence, considerable further development remains necessary before these approaches can become well-known and be adopted throughout the industry. In this study, we examine AI-based product testing and certification trends used in different countries, compare several certification programs, and propose a method to improve such certification programs based on the results of our analysis.

2. Artificial Intelligence Product Certification Program Used in Different Countries

2.1. IEEE: CertifAIEd

The Institute of Electrical and Electronics Engineers (IEEE) has established CertifAIEd (IEEE, 2021) to assess the ethical impacts of AI systems through an Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS). Products, services, and technologies based on autonomous and intelligent systems (AIS) can be certified by CertifAIEd. The certification confirms that an AIS operates in a predictable and consistent manner and fosters trust with users. Four certification factors are considered, including transparency, accountability, algorithmic bias, and privacy. The proof-of-concept for the certification framework used in CertifAIEd was completed by assessing an e-mail classification system provided by Wiener Stadtwerke, an Austrian company, in 2021.

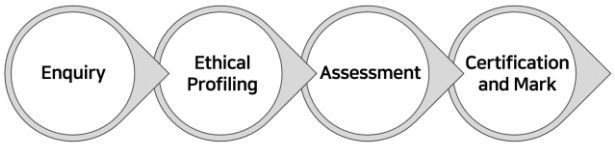

Figure 1: Certification process of IEEE CertifiAIEd

The certification process proceeds through several steps as illustrated in Fig. 1., which include an initial inquiry, ethical profiling, conducting an assessment, and finally issuing a certificate and mark. In the inquiry step, the project scope of the assessment and certification is determined through an agreement between the IEEE Authorized Assessor and the certification applicant. Subsequently, the impacts of the results of the operation of the product are assessed and relevant risks are identified based on ethical profiling. In the assessment step, the IEEE Authorized Assessor provides the assessment criteria and the certification applicant submits evidence that the product meets each criterion. The assessor compiles a Case for Ethics document, which establishes that the product meets the assessment criteria based on each piece of evidence. The Case for Ethics document is then delivered to the IEEE Authorized Assessor and used to write an assessment report that includes the results of the review and suggestions for further improvement and to determine whether certification should be granted. Once the process establishes that certification should be granted for a given product, a certificate, and the IEEE CertifAIEd™ mark are issued.

2.2. WEF: Certification Framework for Responsible Use of Facial Recognition

The World Economic Forum (WEF) has also established a certification framework (Louradour, 2020) along with regulations for facial recognition technology. Facial recognition technology has resulted in racial discrimination and violations of civil liberties guaranteed by law. Hence, the WEF has emphasized specific guidelines to ensure the trustworthiness and safety of AI applications through a robust governance framework and has announced an associated certification framework. This certification targets any public- or private-sector organizations that apply facial recognition technology for flow management and is applied to systems in the design stage or operating facial recognition systems.

The certification factors include the risks associated with a facial recognition system and issues with bias and discrimination, privacy, performance, and accessibility for vulnerable people. These factors are self-assessed through a questionnaire. When the applicant organization is ready, reviewers visit the site and conduct an audit of the system to be certified.

The results of the audit are reported in five categories, including Major Non-conformance, Minor Non-conformance, Sensitive Points, Strengths, and Notes. Major Non-conformance is the only category that affects the issuance of a certificate. Problems reported as instances of Major Non-conformance must be resolved before the certificate is issued as they are subject to corrective action. Complete definitions of the five categories are provided in Table 1.

Table 1: Categories in WEF certification

Category	Definition
Major non-conformance	Non-fulfilment of a requirement calling into question the operation, efficiency, or improvement of the facial recognition management system. Major non-compliance must be addressed through corrective action before certification can be issued.
Minor non-conformance	Failure to meet a specified requirement that does not in itself compromise the effectiveness or improvement of the facial recognition management system. Minor non-compliance should be addressed through corrective action but does not by itself prevent the issuance of a certification.
Sensitive Point	A latent risk of non-compliance. Evidence of compliance with the requirements of the certification framework has been obtained, but the organization must modify its practices to eliminate this latent risk.
Strength	Practice that exceeds the usual level of performance observed in response to the certification requirements.
Note	Observations about the system's compliance with the requirements of the audit framework.

Upon the completion of the audit, an audit report is sent to the applicant organization, which must then submit a plan to take corrective action for further improvements. Afterward, a decision is made on whether to grant the certification based on the audit report and the reviewers' recommendations. If a certificate is granted, this certification is issued for one year from the completion date of the corrective action for any Major Non-conformance.

2.3. LNE: AI Certification

The AI certification provided by the French *Laboratoire national de métrologie et d'essais* (LNE) (LNE, 2021) was developed to define standard requirements for the design, development, evaluation, and maintenance of all types of AI functionalities using machine learning. Therefore, this certification targets all business sectors in which AI is used.

The certification items are defined in terms of four processes as listed in Table 2, including design, development, evaluation, and management. To obtain the certification, developers and suppliers of AI solutions must meet the requirements of the AI process certification standard: which include "design, development, evaluation, and maintenance in operational conditions". In the design procedure, the specifications are validated alongside both normative and regulatory requirements to confirm that they reflect the requirements. In the development stage, the learning quality of any databases is verified to check whether the specifications are adequately adhered to in the functionality of the AI technologies. The system's conformity to the specifications defined before deployment is examined in the evaluation stage, and the definitions of evaluation protocols and metrics are validated. Finally, in the maintenance stage, the organization checks whether the defined specifications are maintained throughout the operation of the system to be certified and whether it complies with the relevant regulations.

In the certification procedure, first, the application document is reviewed, and then an initial audit is conducted. Feedback is then required on any non-conforming factors in the audit results. Afterward, the official certification committee audits the report, and a certificate can be issued based on the results of the audit according to their decision. Following the certification, the audit is performed every year for three years.

Table 2: Requirements LNE certification

Certified procedures	Definition	Examples of audited procedures
Design	Transforming an expression of requirements into functional specifications	Specification and inclusion of normative and regulatory requirements
Development	Translating the specifications into an evaluation-ready version of the AI functionality	Learning Quality of Databases
Evaluation	Verification of the conformity of the system to the specifications defined before its deployment	Definition of evaluation protocols and metrics in every evaluation tool to allow reporting on the effectiveness of these intelligent systems.
Maintaining operational conditions	Ensuring compliance of the AI functionality with the defined specifications after its deployment and throughout its operation	All features necessary to maintain operating conditions. AI systems can evolve throughout their life with performance degradations

2.4. Korean Standards Association: AI+ Certification

The Korean Standards Association (KSA) provides AI+ Certification (KSA, 2023) as a certification program for AI product quality. The certification targets any software (SW), services, or information and communication technology products that use AI, such as manufacturing solutions and smart homes.

The AI+ Certification procedure is divided into certification audits and tests as shown in Table 3. The

certification audit involves an on-site audit of the quality management system subject to certification based on the ISO 9001 standard. The certification tests examine the performance of AI elements and the product quality of software components. The AI performance tests measure accuracy, precision, recall, specificity, and the mean absolute error for any models that are used. The software product quality tests apply the ISO/IEC 25023 software quality standard and include tests for factors such as functional conformity, performance efficiency, compatibility, and usability. The certification procedure is as follows. When an applicant organization applies for certification, the KSA accepts the application, performs certification audits, and issues a certificate. The testing agency WISESTONE performs the certification tests and issues an official test report.

Table 3: Two types of AI+ certification

Category	Target	Criteria
Certification Audit	Quality management system	1) KS Q ISO 9001/ISO 9001 2) AI+ additional requirement
Certification Test	Quality of AI products	1) AI performance test 2) ISO/IEC 25023 International standard for system & SW measurement 3) ISO/IEC 25021 International standard for package SW quality requirements and tests

3. Comparative Analysis and Improvements of Certification Programs

In the comparisons among the four certification programs examined in Section 2, we found relatively few commonalities between the programs such as key test items or certification procedures. As a result of examining the characteristics of each of these certification programs, we found that the WEF limits the targets of their certification to systems that apply facial recognition technology; this differs from the other programs, which target all AI products. In contrast, LNE defines certification items in terms of AI processes; this differentiates it from the other programs that establish certification items based on characteristics of systems using AI such as transparency, privacy, and performance. Unlike the other programs, the IEEE program includes test items for AI ethics in the certification. The KSA evaluates products based on the ISO/IEC 25000 series, a software quality evaluation model, and has added certain AI performance metrics as certification factors. It may be observed that most AI product certification programs are extensions of conventional software quality certifications, and the test items and scope are not formalized but rather determined through agreement with the test client. There are no guidelines for evaluating the quality, safety, and trustworthiness of AI products. As AI products proliferate, private certification programs will continue to emerge; if each program uses different test items or standards, this ecosystem may lead to several confusions in terms of industrial practice. Hence, there is a need for a method to evaluate whether any given certification program satisfies the relevant policies or certification requirements for safe and reliable AI; we highlight the need to establish a foundational standard for this purpose. If an international standard is established through agreements on the evaluation of the quality and trustworthiness of AI products, mutual recognition of different certification systems implemented by various countries and institutions would also be expected.

4. Conclusion and Suggestions

Interest in the safe use of AI is continuing to increase globally with the rapid advancement of AI technology and the widespread adoption of related services. Accordingly, companies have recognized that ensuring AI quality and trustworthiness is a strategic differentiator in the market and are investing certain efforts in implementing such assurances. Due to the characteristics of AI technology and its wide variety of applications and uses, the quality, safety, and trustworthiness of AI systems can be considered questionable—that is, whether any given implemented technology suffices to fulfill its intended function and whether it involves a risk of incidents, among other concerns. In response, studies have been conducted to evaluate the quality and trustworthiness of AI technologies, and guidelines have been released at the national and institutional levels. Furthermore, product certification programs are under development in the public and private sectors to examine and validate the quality of AI products.

In this study, we have analyzed four AI product certification programs provided by the international community and suggested some implications and methods to improve such programs based on the results of our analysis. Our results show that the certification programs target AI products with specific purposes such as autonomous intelligence systems and facial recognition technology, or extend a conventional software quality certification based on the ISO/IEC 25000 standard. Notably, because test items are primarily determined through agreement with the test client rather than by using a formalized scope and list of items, we have confirmed that the current certification programs do not reflect the characteristics of AI technologies sufficiently or are currently in an initial stage at which the uniformity of test

items and standards between certification programs has not yet been established.

As the adoption of AI products and services continues to accelerate, demand for services with assured trustworthiness will increase accordingly. However, the proliferation of certification programs with different test items or standards may confuse many industries. Therefore, international agreements should establish a foundational standard for methods to evaluate the quality and trustworthiness of AI products. The widespread adoption of a common method of evaluating trustworthiness may be expected to prevent confusion among companies and organizations developing AI products and enable the development of services and products based on the standard. Furthermore, the demand for certification is expected to increase as consumers seek assurance as to the ethical basis and trustworthiness of AI products. Such certification may be expected to help consumers identify any risks associated with AI products and services that they use.

References

Breck, E., Cai, S., Nielsen, E., Salib, M., & Sculley, D. (2017). The ML Test Score: A Rubric for ML Production Readiness and Technical Debt Reduction. *in 2017 IEEE International Conference on Big Data* (pp. 1123-1132), December 11-14, Boston, MA, USA, http://dx.doi.org/10.1109/BigData.2017.8258038

Byun, S. (2019). A Study on the Ethics Certification Program Based on the Morality Types of AI Robots. *Journal of Ethics*, *1*(126), 73-90. http://dx.doi.org/10.15801/je.1.126.201909.73

Centre for Data Ethics and Innovation (CDEI) (2021). *Independent report*. The roadmap to an effective AI assurance ecosystem. Retrieved April 13, 2023 (actual access date), from https://www.gov.uk/government/publications/the-roadmap-to-an-effective-ai-assurance-ecosystem/the-roadmap-to-an-effective-ai-assurance-ecosystem-extended-version.

Cihon, P., Kleinaltenkamp, M. J., Schuett, J., & Baum, S. D. (2021). AI Certification: Advancing Ethical Practice by Reducing Information Asymmetries. *IEEE Transactions on Technology and Society*, *2*(4), 200-209. https://doi.org/10.48550/arXiv.2105.10356

Genovesi, S., & Julia, M. M. (2022). Acknowledging Sustainability in the Framework of Ethical Certification for AI. *Sustainability*, *14*(7), 4157. https://doi.org/10.3390/su14074157

High-Level Expert Group on AI (AI HLEG) (2020). The Assessment List for Trustworthy Artificial Intelligence (ALTAI). *European Commission*.

IEEE (2021). *IEEE CertifAIEd™*. Retrieved April 13, 2023 from https://engagestandards.ieee.org/ ieeecertifaied.html.

Korean Standards Association (KSA) (2023), *AI+ Certification*. Retrieved April 13, 2023 from https://www.ksa.or.kr/ksa_kr/6962/subview.do.

Laboratoire national de métrologie et d'essais (LNE) (2021). *Certification of processes for AI*. Retrieved April 13, 2023 (actual access date), from https://www.lne.fr/en/service/certification/certification-processes-ai.

Louradour, S., Madzou, L., & Mella, J. (2020). A Framework for Responsible Limits on Facial Recognition. *World Economic Forum*, White Paper.

Machlev, R., Perl, M., Belikov, J., Levy, K. Y., & Levron, Y. (2021). Measuring Explainability and Trustworthiness of Power Quality Disturbances Classifiers Using XAI-Explainable Artificial Intelligence. *IEEE Transactions on Industrial Informatics*, *18*(8), 5127-5137. http://dx.doi.org/10.1109/TII.2021.3126111

Shin, Y. (2022). 2022 Trustworthy AI Development Guidebook. *TTA Journal*, *201*, 21-27.

Tao, C., Gao, J., & Wang, T. (2019). Testing and Quality Validation for AI Software–Perspectives, Issues, and Practices. *IEEE Access*, *7*, 120164-120175. http://dx.doi.org/10.1109/ACCESS.2019.2937107