AI Certification: An Accreditation Perspective

Susanne Kuch[1]
Deutsche Akkreditierungsstelle (DAkkS), Stabsbereich Akkreditierungsgovernance, Forschung und Innovation, Berlin, Germany

Raoul Kirmes
Deutsche Akkreditierungsstelle (DAkkS), Stabsbereich Akkreditierungsgovernance, Forschung und Innovation, Berlin, Germany

Abstract

AI regulations worldwide set new requirements for AI systems, leading to thriving efforts to develop testing tools, metrics and procedures to prove their fulfillment. While such tools are still under research and development, this paper argues that the procedures to perform conformity assessment, especially certification, are largely in place. It provides an overview of how AI product certifications work based on international standards (ISO/IEC 17000 series) and what elements are missing from an accreditation perspective. The goal of this paper is to establish a common understanding of how conformity assessment in general and certification in particular work regarding AI systems.

2012 ACM Subject Classification Computing methodologies → Artificial intelligence; General and reference → Computing standards, RFCs and guidelines

Keywords and phrases certification, conformity assessment, market entry, accreditation, artificial intelligence, standard

Digital Object Identifier 10.4230/OASIcs.SAIA.2024.14

Category Practitioner Track

Acknowledgements Special thanks are due to Mattis Jacobs for his support with this paper, and to Dominic Deuber and Svenja Reisinger for their feedback.

1 Introduction

Artificial intelligence (AI) applications are increasingly relevant in every business sector. Some AI systems pose a risk to important commodities worthy of protection, such as health, environment or fundamental rights. This rationale drove the European Union's approach in developing Regulation (EU) 2024/1689 (AI Act) [9]. Although the EU AI Act may be the most comprehensive, other governments like the US or China have provisions with different approaches and regulatory depth [11, 10].

Where regulations exist, compliance must be ensured without establishing new trade barriers. Therefore, the World Trade Organization (WTO) demands for processes of conformity assessment and accreditation[2] [12]. With the accreditation of a conformity assessment body (CAB), it is ensured that its results can be recognized by other WTO member states as equivalent to their own national CAB results. In this way, accreditation supports international trade since it helps companies avoid duplicating costly conformity assessment procedures in other WTO member countries for companies. In order to enable this system, accreditation

[1] Corresponding author

[2] Accreditation is a third-party attestation with a conformity assessment body (CAB) as object of conformity. The purpose is to ensure a CAB's formal demonstration of its competence, impartiality and consistent operation in performing specific conformity assessment activities. The authority of an accreditation body can be derived from governments. In Europe, accreditation bodies are government-authorized bodies according to Regulation (EC) No 765/2008.

bodies are tasked with assessing the competence, impartiality, and independence of conformity assessment bodies in the sense that these CABs perform their work reliably and in a comparable as well as reproducible manner. Thus, all the procedures and methods applied and used by CABs to perform conformity assessment need to be objective, reproducible and comparable. Therefore, one purpose of this paper is to outline how the established system of conformity assessment works in order to ensure comparability and reproducibility and how this can be applied in the context of AI systems – in an international context as well as with reference to the AI Act. The other is to clarify the distinction of conformity assessment and AI system "testing" and "validation".

This paper introduces the methodology of the international conformity assessment system in section 2 which is based on the internationally accepted standards series of ISO/IEC 17000, the so-called ISO CASCO toolbox. It will be used to subsequently focus in section 3 on the conformity assessment activity of certification. Here, it will be outlined how certifications with focus on AI systems (mainly seen as products) should be developed in general based on the internationally existent system. It will also briefly explain the distinction between conformity assessment and AI system "testing" and "validation" with respect to the development cycle. This will be followed by the specifications of conformity assessment for the European context in accordance with the EU AI Act in section 4. The paper concludes by outlining some paradigmatic gaps at the technical level of AI systems and by directing to areas where further scientific and standardization work is needed from an accreditation point of view.

2 The methodology of conformity assessment

According to ISO/IEC 17000, conformity assessment means the demonstration that specified requirements are fulfilled.[3] Conformity assessment must account for all different kinds of *objects of conformity assessment* such as products, processes, organizations, persons, services, data, systems, materials, designs. Hence, a variety of different types of conformity assessment activities exist. These are mainly testing (including calibration and proficiency testing), inspection, validation/verification and certification (of management systems; persons; products, processes and services). While those activities are all different, they follow the same functional approach determined in ISO/IEC 17000, Annex A. In order to ensure the comparability and reproducibility of the results, each activity and the needed procedure to fulfil the functional approach are defined in the respective conformity assessment standard (ISO/IEC 17000 series), with every one of it internationally agreed upon to support the WTO mutual recognition approach of statements of conformity.

Conformity assessment can be performed by three different entities: the organization providing a product or service (first-party), organizations with user interest (second-party) or independent accredited conformity assessment bodies (third-party). The accredited third-party conformity assessment body is the only one that is allowed to conduct certification as special type of conformity assessment activity according to ISO/IEC 17000 [2].

Third-party certification is therefore an important independent evaluation method to demonstrate certain levels of quality, transparency and trust of a product or service, management system or a competent person to other market players and authorities.

[3] Those specified requirements are very often determined in standards and less commonly in technical specifications (in terms of regulation (EU) 1025/2012). Seldom are they directly determined within national regulations.

3 General certification scenarios for AI systems

For AI systems, the following certifications[4] are particularly relevant: management system certifications (ISO/IEC 17021-1) [3] and certifications for products, processes and services (ISO/IEC 17065) [4].

3.1 Management system certification

Organizations may certify their artificial intelligence management system (AIMS) to show other market participants their competency in accordance with ISO/IEC 42001 [6].[5] To obtain an AIMS certificate, the organization needs an accredited conformity assessment body (CAB) according to ISO/IEC 17021-1, which offers ISO/IEC 42001 certifications based on ISO/IEC 42006[6]. With such an AIMS certification an organization demonstrates that it is competent to set-up and effectively run a management system for AI systems. This kind of certification can serve as useful evidence for tender procedures or business-to-business relationships in international trade.

The objective of an AIMS is to support an organization in ensuring that the AI systems that are in the scope of the AIMS are developed or deployed as intended and with the respective applicable (quality) requirements (often specified in specific AI system standards). An AIMS is especially important and crucial if requirements for the AI technology itself are not (widely) agreed upon (and standardized) due to a lack of scientific basis (e.g. regarding benchmarks). This becomes especially evident in case of complex AI systems (black-box models). The reason is that they have a limited predictive reliability. Thus, an AI management systems enables the monitoring of the environment in which a complex (black-box model) is deployed and can restrict the impacts of such a model's limited predictive reliability. An AIMS achieves this by monitoring the intended operating conditions and managing, for example, to some degree the processes for data input and generated output. By this, an AIMS can to a certain extent indirectly fill the gap of the non-existent widely scientifically agreed methods to assess some technical requirements for a complex AI model.

3.2 Product certification

Product certifications are performed by accredited third-party CABs only according to ISO/IEC 17065 and follow specific schemes according to ISO/IEC 17067 [5]. ISO/IEC 17065 allows combining different conformity assessment activities (CAA) to evaluate specific AI systems or components thereof (e.g. data, AI model, software). Those *objects of conformity assessment* are assessed against specified requirements listed in the applied certification scheme.

The CAB has to decide on a case-by-case basis which conformity assessment activities (CAA) are applicable to perform its client's assignment. The following CAA are expected to be used during a certification process according to ISO/IEC 17065: *inspection* and *audit*, in some cases laboratory *testing* (with defined or standardized metrics) of software and hardware components or *validation* of statements (e.g. transparency, reliability, level of explainability). Through these conformity assessment activities, a third-party CAB will assess whether or not the set requirements of the specific scheme that was applied to the respective client are met.

[4] Certifications of persons according to ISO/IEC 17024 can be important but are not considered here. The focus is on the needed conformity assessment activities for market access that determines the AI system as a "product".

[5] This applies as well for providers of no or low-risk AI systems in Europe.

[6] ISO/IEC 42006 is currently under development and is likely to be published by the beginning of 2025.

Table 1 Overview of AI systems and its components as objects of conformity mapped towards the applicable conformity assessment activities.

Object of conformity assessment (AI system and/or components)	**Conformity assessment activity/activities**
Organization (of the AI provider or user)	*Audit* according to ISO/IEC 17021-1 (for AIMS: ISO/IEC 42006)
Dataset	*Inspection* according to ISO/IEC 17020 *Testing* according to ISO/IEC 17025
AI model(s)	*Inspection* according to ISO/IEC 17020 *Validation* according to ISO/IEC 17029 *Testing* according to ISO/IEC 17025
Software (for user interaction)	*Inspection* according to ISO/IEC 17020 *Testing* according to ISO/IEC 17025
Hardware	*Inspection* according to ISO/IEC 17020 *Testing* according to ISO/IEC 17025
AI system (incl. all components)	*Certification* based on ISO/IEC 17065, including surveillance and monitoring activities

Table 1 shows the AI system and its components as objects of conformity assessment, mapped to the respective conformity assessment activity which can either be used for a certification scheme or as a standalone conformity assessment activity.

3.3 Identifying gaps within conformity assessment

While non-complex (white-box) AI systems (e.g. knowledge-based systems) can be assessed and certified with the various conformity assessment standards mentioned in Table 1 with a certain level of reliability and certainty, complex (black-box model) AI systems (e.g. deep neural networks (DNN)) have to be assessed indirectly by using CAAs. The focus of indirect assessments is on management systems as well as monitoring activities through embedded inspection that form crucial parts of a certification scheme according to ISO/IEC 17067. Due to their abstract design, certification procedures according to ISO/IEC 17065 are capable of handling high levels of complexity[7] which is why no additional certification standard on this level of the ISO/IEC 17000 series is required for AI systems.

What is necessary instead is the development of specific schemes in accordance with ISO/IEC 17067 and the monitoring tools for complex (black-box) AI systems. Furthermore, to support the conformity assessment activities, there is the need to develop and publish the specific technical standards on specific methods that specify testing or inspection procedures for AI systems through academia and standardization organizations. An example for such missing standards would be component testing, e.g. for datasets (bias) or AI models (robustness or security features). From an accreditation point of view, it is important to have objective measurands that determine for example if there is bias in the data (or what threshold of bias is acceptable) in order to assess the dataset in a comparable manner or what threshold defines an AI model as "robust" and how it is determined. Together with a measurand for data bias or for robustness, the specific testing procedures for bias or robustness need to be standardized to ensure comparability and reproducibility.

[7] This is already the case for many different products that mix e.g. software with hardware components in highly regulated areas (e.g. for electronic products in medical area or for construction products, etc.).

In general, accreditation sees a subtle yet recognizable distinction between existent development specific testing and validation methods and procedures as applied by the providers and the methods and procedures used for conformity assessment. The latter need to be internationally aligned through standardization by recognized standardization organizations in order to fulfil the WTO mutual recognition requirement. Also the methods and procedures need to be validated and, if applicable, calibrated such that the tools are applicable to many different objects of conformity (AI systems or the specifically tested components) in order to ensure comparability and reproducibility of the results in an objective manner. Such specifications can only be discussed and agreed upon in standardization and need to be scientifically proven as objective measurand. It may also be necessary to determine whether synthetic datasets should be developed and provided for testing procedures according to ISO/IEC 17025 to serve as objective test or reference datasets. Therefore, high priority must be given to the ongoing work in standardization and in academia to define and develop the needed measurands as well as evaluation methods for AI systems and the components.

4 European specifications and needs

With the AI Act in force, specific requirements need to be fulfilled by European AI operators.

Operators of no- or low-risk AI systems (Art. 3, No. 8, AI Act), only need to fulfil transparency requirements. In contrast, operators of high-risk AI systems must meet all requirements in Chapter III, and demonstrate it by using pre-defined conformity assessment procedures. According to Article 43 AI Act, there are two options: *self-declaration* (first-party) or conformity assessment by a notified body (third-party) (as *product certification*). Yet, where the AI system is part of another product (Annex I, especially Section A), the relevant conformity assessment under those legal acts is required.

All of these options align with the New Legislative Framework (NLF)[8] and reflect procedures outlined in different modules of decision 768/2008/EC. Module A of this decision applies to the self-declaration and module H1[9] to a notified body. For AI systems listed in Annex I, the modules of these legal acts apply with the preferred conformity assessment standard (of the 17000 series) to be used according to EA-2/17 [8]. Often, this is also ISO/IEC 17065.

Looking at the second option outlined in Article 43 of the AI Act, notified bodies need to perform conformity assessment (product certification according to ISO/IEC 17065). Thus, also in the European context, the general approach of product certification as outlined in section 3.2 remains applicable. However, the legal consequences differ. In the EU, only a positive assessment by a notified body for a high-risk AI system (here as result of a product certification), where required, permits affixing the CE-marking. Furthermore, the certification process is outlined in Annex VII AI Act. It includes the assessment of the quality management system (QMS) in accordance with Art. 17 of the AI Act via the conformity assessment activity *audit* following ISO/IEC 17021-1, and the technical documentation in accordance of Art. 11 of the AI Act regarding the high-risk AI system via an *inspection* in order to evaluate if all requirements of Chapter III are comprised by the technical documentation.

[8] The NLF was adopted in 2008 in order to establish a common legal framework for placing goods on the internal market and ensure a high quality of those products placed on the market. The NLF consists of regulation (EC) 765/2008, decision 768/2008/EC and regulation (EU) 2019/1020.

[9] The European Cooperation for Accreditation (EA) declares in its EA Accreditation for Notification (AfN) Project report from July 2024 [7] that the preferred standard for regulation (EU) 2024/1689 (AI Act) is ISO/IEC 17065.

Accordingly, it applies for the European context as well that specific certification schemes according to ISO/IEC 17067 are developed to ensure compliance with the AI Act based on the already existent CAA (ISO/IEC 17000 series). Since many other harmonized regulations may interfere with the AI Act, it may be feasible to develop this within the different economic sectors in order to take the specific sectoral regulations as well as business practices and applications into account. Thus, for a notified body to attest conformity for an operator, these scheme developments are now crucial. For potential EU-specific objective measurands (e.g. for high quality data), it is necessary that first the standardization experts define the technical requirements for the respective objective of conformity assessment (e.g. data), and subsequently determine which kind of conformity assessment activities may need a specification (e.g. specific testing measurand according to ISO/IEC 17025 for "high quality data" or the development of a procedure to provide "reference data" according to ISO 17034 [1]).

5 Conclusion

In conclusion, we can summarize that the general framework for conducting conformity assessment and especially certification is provided by the ISO/IEC 17000 series and it is generally sufficient for complex systems. This paper demonstrated that this also applies for complex (black-box) and non-complex (white-box) AI systems, since both can be assessed based on these conformity assessment activities. Hence, no additional standards specifically targeting the level of the conformity assessment activities (ISO/IEC 17000 series) are required. However, this paper also made clear what is missing. In particular, there is a need to develop sector and technology-specific certification schemes according to ISO/IEC 17067 – both internationally and at the European level. Furthermore, there is still a gap in standardization regarding objective scientifically proven measurands for certain components (e.g. bias in datasets; security features; robustness) and their calibration as well as the applicable reproducible testing methods. Here, there might be even the need to develop European specific standard testing methods. However, this can only be evaluated once the applicable AI system standards to meet the requirements of the AI Act are developed. From an accreditation perspective, there might also be the need to develop objective synthetic test and reference datasets to ensure a high quality of the testing procedures in laboratories and to reliably assess AI model validation procedures. With regard to complex (black-box) AI systems, there is also a lack of scientific basis that determines whether conformity is presumed or not. Thus, more technical work needs to be done here.

On the other hand, it became evident that in all AI system contexts, AIMS certifications play a vital role for maintaining trust in AI systems and can be considered as even more important than other management systems being used more conventionally as quality assurance measure. The reason is that an AI management system allows to control the context of deployment and to some degree the processes of data input and generated output of an AI system. With regard to complex (black-box) AI systems and the limited predictability of their behaviour, AIMS certifications are thus the only trust anchor available. Consequently, the importance and trustworthiness of the system of accreditation and conformity assessment in general, and in particular of those accredited CABs for these AIMS certifications, should not be underestimated.

References

1 DIN Media. DIN EN ISO 17034:2017-04 Allgemeine Anforderungen an die Kompetenz von Referenzmaterialherstellern (ISO 17034:2016); Deutsche und Englische Fassung EN ISO 17034:2016.

2 DIN Media. DIN EN ISO/IEC 17000:2020-09 - Konformitätsbewertung - Begriffe und allgemeine Grundlagen (ISO/IEC 17000:2020); Dreisprachige Fassung EN ISO/IEC 17000:2020 - ISO/IEC: ISO/IEC 17000:2020.

3 DIN Media. DIN EN ISO/IEC 17021-1:2015-11 - Konformitätsbewertung - Anforderungen an Stellen, die Managementsysteme auditieren und zertifizieren - Teil 1: Anforderungen (ISO/IEC 17021-1:2015); Deutsche und Englische Fassung EN ISO/IEC 17021-1:2015.

4 DIN Media. DIN EN ISO/IEC 17065:2013-01 - Konformitätsbewertung - Anforderungen an Stellen, die Produkte, Prozesse und Dienstleistungen zertifizieren (ISO/IEC 17065:2012); Deutsche und Englische Fassung EN ISO/IEC 17065:2012.

5 DIN Media. DIN EN ISO/IEC 17067:2013-12 - Konformitätsbewertung - Grundlagen der Produktzertifizierung und Leitlinien für Produktzertifizierungsprogramme (ISO/IEC 17067:2013); Deutsche und Englische Fassung EN ISO/IEC 17067:2013.

6 DIN Media. ISO/IEC 42001:2023-12 - Informationstechnik – Künstliche Intelligenz – Managementsystem.

7 European Accreditation. EA Accreditation for Notification (AfN) Project. Report. Last accessed July,30,2024. URL: https://european-accreditation.org/wp-content/uploads/2023/04/AFN-PROJECT-2024.pdf.

8 European Accreditation. EA Document on Accreditation for Notification Purposes. Last accessed July 30,2024. URL: https://european-accreditation.org/publications/ea-2-17-m/.

9 European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, 2016/797/EU and 2020/1828/EU (Artificial Intelligence Act). URL: http://data.europa.eu/eli/reg/2024/1689/oj.

10 The State Council of the People's Republic of China. China moves to support generative AI, regulate applications. Last accessed 2024-07-30. URL: https://english.www.gov.cn/news/202307/13/content_WS64aff5b3c6d0868f4e8ddc01.html.

11 White House of the United States of America. Executive order on the safe, secure, and trustworthy development and use of artificial intelligence. Last accessed 2024-07-30. URL: https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/?utm_source=link.

12 WTO. Agreement on Technical Barriers to Trade. Last accessed 2024-07-26. URL: https://www.wto.org/english/docs_e/legal_e/17-tbt_e.htm.