

Designing an AI System Certification Scheme

**An Attempt at Joint Certification via
Integrating Risk Control Measures and
Organizational Management**

Table Of Contents

1. The Growing Need for AI System Certification	2
2. Limitations of Existing Certification Approaches and the Need for Joint Certification..	3
2.1 Evaluation of Performance and Specification Conformity	3
2.2 Evaluation of Risk Control Measures	4
2.3 Evaluation of the Provider's Organization.....	5
2.4 AI System Evaluation: The Joint Certification Approach.....	6
3. Method of Joint Certification for AI Systems	7
3.1 Structure of ISO/IEC 42001: AI Management System	7
3.2 Joint Certification Framework for AI Systems	9
3.3 Similar Initiatives in Fields Other Than ISO	11
3.3.1 EU AI Act.....	12
3.3.2 NIST AI Risk Management Framework (NIST AI RMF)	12
3.3.3 Japan's Government AI Procurement Guidelines.....	13
3.3.4 ISMAP (Information System Security Management and Assessment Program for Government Information Systems)	14
4. Advantages of the Joint Certification Approach.....	15
4.1 Flexible and Resilient Response to Unpredictable Risks.....	15
4.2 Procedural Efficiency through Reuse of Management System Certification	15
4.3 Scalability through Connection with Certification in Individual Domains	16
4.4 Ensuring Conformity with AI-Related Laws and Regulations	18
5. Conclusion: The Future of Joint Certification	20

This white paper was prepared by Smart Governance Ltd. based on the results of joint research conducted with Microsoft Corporation.

1. The Growing Need for AI System Certification

Artificial Intelligence (AI) is rapidly permeating all aspects of society. This influx spans from internal organizational-level utilization, such as research and document creation, to regulated domains like financial services, autonomous driving, and medical technologies.

While AI systems significantly enhance our productivity and quality of life, the scope, characteristics, and quantity of risks posed by AI also increase in tandem. These risks are diverse and complex, thus posing a newfound challenge for governance. They encompass a wide array of potential impacts, including: harms concerning human life due to malfunctions in autonomous vehicles or medical devices; threats to individual rights and to democracy; psychological harm; the amplification of bias leading to discrimination and inequality; and even threats to national security CBRN (Chemical, Biological, Radiological, and Nuclear) systems if AI vulnerabilities allow bad actors to take control.

Such risks from AI systems present a new gap in trust and accountability mechanisms, leading to a newfound demand for the introduction of certification schemes for AI systems. In standard context, certification refers to a “provision by an independent body of written assurance (a certificate) that the product, process, service, or system in question meets specific requirements. Certification is also known as third-party conformity assessment.”¹

The certification mechanism is crucial especially in high-risk areas, as sole reliance on companies' voluntary risk control may be insufficient for ensuring high levels of assurance. Regulators recognize these gaps; for example, the EU AI Act mandates providers of high-risk AI systems to conduct conformity assessments and, in some instances, requires third-party certification.² In addition, in many countries, discussions are underway to revise or establish certification mechanisms for AI-embedded products in heavily regulated areas, such as automobiles and medical devices. As AI continues to impact life, business, and society, the expectation for certification of AI systems by independent, expert third parties is increasingly necessary.³

¹ ISO, Conformity Assessment (<https://www.iso.org/conformity-assessment.html>)

² EU AI Act, Article 43: Conformity Assessment (<https://artificialintelligenceact.eu/article/43/>)

³ AISI, National Status Report of AI Safety in Japan 2024 (https://aisi.go.jp/assets/pdf/j-aisi_report_2024_ja.pdf). Additionally, although not directly related to certification, the "New International Standards Strategy" published by the Intellectual Property Strategy Headquarters of the Cabinet Secretariat on June 3, 2025, pages 24 (<https://www.kantei.go.jp/jp/singi/titeki2/chitekizaisan2025/pdf/kokusaisenryaku.pdf>), highlights AI as a key area within Japan's strategic fields in standards.

2. Limitations of Existing Certification Approaches and the Need for Joint Certification

Several challenges arise when considering possible mechanisms for certifying AI systems, as multiple perspectives exist on the evaluation criteria and scope. The figure below introduces three common existing approaches to certification:

- 1) Certification of the product's conformity to intended performance or specification.
- 2) Assessment of the product's risk control measures and actions.
- 3) Evaluation of the organization that provides and manages the product or service.

[Figure 1 - Classification of Evaluation Approaches]

(2.1) Product Evaluation Approach	(2.2) Product Risk Control Approach	(3) Organization-level Approach
[Evaluation Criteria] Whether the product or service complies with specific requirements such as certain performance standards or technical specifications. [Main Evaluation Items] Physical safety of the product/service (e.g., heat resistance, insulation) and performance testing methods. [Strengths & Weaknesses] Strengths: Effective when the risks are known and solutions for those risks are established. Weaknesses: Difficult to apply to highly complex or dynamically changing systems. [Example Standards] IEC 62368-1 (Information and Communication Technology Equipment) IEC 60335 (Household Electrical Appliances), etc.	[Evaluation Criteria] Whether control measures have been implemented to reduce risk arising from the product or service. [Main Evaluation Items] Identification and assessment of risks for each product/service, selection, implementation, and documentation of countermeasures, etc. [Strengths & Weaknesses] Strengths: Providers with knowledge of the AI system can make flexible decisions on risk mitigation. Weaknesses: Difficult for external parties to assess the adequacy of countermeasures, and in fast-evolving fields, the control measures may become outdated quickly. [Example Standards] ISO/IEC 27001 Annex A ISO/IEC 42001 Annexes A, B, etc.	[Evaluation Criteria] Whether the organization providing the product or service has a system in place to continuously execute and improve risk management. [Main Evaluation Items] Implementation of the PDCA (Plan-Do-Check-Act) cycle within the organization, etc. [Strengths & Weaknesses] Strengths: The organization ensures control measure appropriateness and is less prone to obsolescence, adapting well to changing circumstances. Weaknesses: For AI systems with broad or autonomous risk contexts, the organization alone may not be able to guarantee reliability. [Example Standards] ISO/IEC 9001 ISO/IEC 27001 Main Body ISO/IEC 42001 Main Body

Below, we analyze the objects, methods, and challenges associated with certification when applying these three different approaches to AI systems.

2.1 Evaluation of Performance and Specification Conformity

The first approach evaluates whether an AI system adheres to predefined performance or technical specifications. This approach has been commonly applied to industrial tangible products. For example, under Japan's Electrical Appliance and Material Safety Act, specified electrical appliances, such as electric massagers, are required to acquire the PSE mark by undergoing certification to ensure they meet safety performance standards regarding heat resistance, insulation, and other relevant factors.⁴

⁴ METI, Electrical Appliances and Materials Safety Act
(<https://www.meti.go.jp/english/policy/economy/consumer/pse/index.html>)

Likewise, with AI systems, we can assess their intended design by evaluating elements such as their training data or algorithmic architecture, or by testing their performance against the required criteria.

Note: This approach is sometimes referred to as ‘product certification,’ where a product is verified to meet predefined standards. However, to avoid confusion with the risk control measures approach outlined in Section 2.2, this paper will use the term ‘specification conformity’ to describe this approach.

This approach of evaluating conformity to predefined requirements is practical when the product undergoes minimal changes from the test phase to market launch. It is also effective when risks are known and methods for addressing them are already established; typically, for traditional tangible products. However, for continuously updated, adaptive systems, especially technologies like AI that possess an advanced level of complexity, this approach may not be practical. AI systems have a dynamic nature, where behavior changes depending on the training data and operating environment, making it difficult to assess risk based only on a single temporal snapshot.

In addition, since the decision-making process of AI is a highly complex black box, one may not be able to quantify unknown impacts in advance. Furthermore, it is challenging to define and quantify qualitative value judgments, such as fairness and ethical appropriateness, as objective and quantitative evaluation indicators. Therefore, rigidly demanding specific performance or specifications in AI system conformity assessment carries the risk of strongly inhibiting the development and utilization of AI systems.

Notably, in some high-risk AI systems that involve hardware, such as autonomous driving vehicles and surgical robots, the demand for legal protection through performance and specification conformity assessment becomes stronger. In such cases, it is plausible that sector-specific laws may require particular performance and specifications, or that standards may present models for these performance and specifications. Nevertheless, since these are fundamentally established for individual sectors, it would likely be difficult or impractical to establish a generalized, common performance and specification standard applicable to AI systems.

2.2 Evaluation of Risk Control Measures

An approach could also involve evaluating whether proper risk control measures are in place to address the anticipated risks associated with these systems. This involves identifying and assessing the risks that each respective AI system may pose, and determining whether appropriate countermeasures have been implemented to mitigate them.

Similar methods have been generally adopted for system-based risks, such as cybersecurity and privacy-oriented scenarios. For example, Annex A of ISO/IEC 27001, a standard for information security management systems, provides control measures for a system that can

be put in place to mitigate risks identified as part of risk assessment activities..⁵In one scenario, an organization may identify that their systems contain highly confidential data at risk of tampering, and choose to implement control A.8.24 from Annex A on cryptography, in which “rules for the effective use of cryptography, including cryptographic key management, shall be defined and implemented.”

A key advantage of this approach is that it allows providers (including those who develop or offer AI systems as a service) to determine the most effective risk mitigation strategies, as they are better informed about the nature of individual AI systems and their associated risks.

Furthermore, unlike the uniform criteria used in Approach 2.1, the flexibility this approach gives organizations in determining how to handle risks makes it particularly useful for certifying systems like AI, where the risk landscape is both extensive and constantly evolving.

However, a challenge remains in how a third party or certification body can assess whether the selected risk control measures are compliant or satisfactory. The primary issue lies in an information gap, where the certification body often has less information than the AI system provider, leading to an imbalance in available data.

In addition, in the case of AI systems, where both technology and social acceptance are constantly evolving, an inherent pacing issue arises. If a set of control measures is deemed appropriate at the time of certification, the rapid pace of AI development means that these measures may quickly become outdated. Therefore, unless the organization developing or operating the AI system has a continuous process in place to evaluate and update control measures, it would be difficult to conclude that the organization’s risk response is sufficient for constantly evolving AI systems. Indeed, control measures themselves have historically never been assessed in isolation. Rather, their adequacy is evaluated within the broader context of the management system clauses, which we will address next.

2.3 Evaluation of the Provider’s Organization

Another common type of conformity assessment is a management system certification, such as ISO 9001 and ISO/IEC 27001, which, although it includes system requirements in Annex A, is still performed as a management system certification rather than a product certification. .

For example, ISO 9001 certification does not directly require that individual products meet specific quality standards or whether risk control measures are applied to particular products. Instead, it certifies that the organization has established and operates a system to manage processes—from product design and manufacturing to inspection—and is committed to continuous improvement. The objective of certification is to enable “organizations of all sizes

⁵ ISO/IEC 27001 Annex A.8 Technological controls 24. Cryptography (<https://www.iso.org/standard/27001>)

and sectors to improve their performance, meet customer expectations and demonstrate their commitment to quality."⁶

This method aims to ensure the provider guarantees the reliability of the product and the effectiveness of risk control measures through the organization's internal mechanisms, rather than having a third party directly perform the product assessment. This method is expected to bridge the information gap between the AI system provider and the external certification body, which was a challenge in Approach 2.2. Additionally, while specific control measures may become obsolete due to changes in technology and society, management system certification — which assesses whether "an organization has a system in place to evaluate and improve control measures as needed"— offers the advantage of being more adaptable.⁷

In the context of AI systems, additional complexities and nuances must be considered. Because of the dynamic nature of AI systems, a management system approach can be critical. Organizations must be able to demonstrate that they have a robust system in place of human management to assess risks and impacts, evaluate processes, and update as necessary. However, there are limits to an approach that relies solely on the organization or an internal team to mitigate risks. For example, in the United States, the approval process for software medical devices was restructured to simplify or eliminate the individual product approval process for companies. This approach emphasized the management requirements of the organization manufacturing the device, rather than focusing on the approval of individual medical devices. It ultimately did not function effectively due to factors such as the complexity of software development processes, a lack of clear criteria for organizational evaluation, and the need to certify device-specific elements, including clinical performance and cybersecurity.⁸ This highlights the limitations of an organization-led approach.

2.4 AI System Evaluation: The Joint Certification Approach

When examining AI systems, where risk levels are diverse and system content changes dynamically, each of the certification approaches—evaluation of product specification (2.1), product risk control measures (2.2), and organizational management (2.3)—has apparent limitations.

Among these, applying the product specification approach (2.1) is often challenging in rapidly changing areas, such as AI algorithms. Therefore, as a basic framework for AI system certification, examining both the control measures adopted for each product (2.2) and the organizational management that evaluates, implements, and improves them (2.3) is considered to lead to more substantive risk management for AI systems and, as a result, promote trust.⁹

⁶ ISO 9001:2015 (<https://www.iso.org/standard/62085.html>)

⁷ ISO 14001:2015 (<https://www.iso.org/standard/60857.html>)

⁸ FDA, The Software Precertification (Pre-Cert) Pilot Program: Tailored Total Product Lifecycle Approaches and Key Findings (<https://www.fda.gov/media/161815/download?attachment>)

⁹ In addition to these certifications, there may be cases where separate certification is required for hardware products (e.g., self-driving cars, AI medical devices, etc.), especially when certain performance and specifications are required.

Such an approach, evaluating both control measures and the organization, can be referred to as joint certification.¹⁰ This approach relies on the basis that a system's trustworthiness is guaranteed only when an organization or process with a reliable management system consistently enforces the proper control measures, striking a balance between innovation, control, and responsibility.

Introducing such an approach with AI would require the following elements:

- 1) Evaluation criteria for AI systems' risk control measures on the basis of product certification, but for a digital, non-tangible product (i.e., the AI system).
- 2) Evaluation criteria for the organization providing AI systems.
- 3) A mechanism for integratively evaluating and granting certification, either issued as two separate certifications at one time, or a single certification with two objects of conformity.

3. Method of Joint Certification for AI Systems

In the following passages, we will examine how these elements can be fulfilled by referencing existing standards and frameworks.

3.1 Structure of ISO/IEC 42001: AI Management System

ISO/IEC 42001 is widely recognized currently as the only certification-based international standard that focuses on responsible AI systems.¹¹ The framework consists of 10 main body clauses and four annexes.

The main body has a structure similar to management system standards such as ISO 9001 and ISO/IEC 27001, following a standard high-level structure of Plan-Do-Check-Act (PDCA), a cycle framework designed to promote continuous improvement through iterative planning, execution, evaluation, and refinement of processes.¹²

⁹ In addition to these certifications, there may be cases where separate certification is required for hardware products (e.g., self-driving cars, AI medical devices, etc.), especially when certain performance and specifications are required.

¹⁰ Joint certification has already been implemented for some physical products apart from AI products. For example, ISO/IEC 9001 and Energy Star are combined for air conditioning units.

¹¹ ISO/IEC 42001: 2023 (A.5.3/B.5.3) (<https://www.iso.org/standard/42001>)

¹² PECB, The Plan-Do-Check-Act (PDCA)

(<https://pecb.com/article/the-plan-do-check-act-pdca-cycle-a-guide-to-continuous-improvement>)

Annex A of ISO/IEC 42001 comprehensively presents control objectives, controls, and implementation guidance for responding to risks associated with the design and operation of AI systems. For instance, it outlines control measures, such as impact assessments of their AI systems, and documents the findings. Proceeding, Annex B provides detailed guidance on implementation. As guidance for the control measures in the AI system impact assessment process, it specifies key elements to implement, including individuals' legal positions, well-being, fundamental human rights, and broader societal implications.¹³ Thus, Annexes A and B of ISO/IEC 42001 serve as key elements of the standard, Annex A providing normative controls that an organization must consult to determine whether risks are addressed appropriately, and Annex B providing informational guidance for how to adequately implement the chosen controls.

In addition, Annex C highlights potential organizational objectives and identifies discrepancies within risk management, while Annex D presents specific use cases across various industrial sectors. These annexes are informative and serve solely as reference materials, without normative value or as part of the standard's requirements.

¹³ These measures do not have to be implemented in their entirety and can be developed and implemented independently according to the characteristics of the organization (ISO/IEC 42001, 6.1.3)

The structure of ISO/IEC 42001 is shown in Figure 2 below:

[Figure 2 - ISO/IEC 42001 Structure]



3.2 Joint Certification Framework for AI Systems

Based on the above analysis, the necessary elements for possible joint certification of AI systems can be considered within the following framework:

- 1) Evaluation Criteria for AI System Control Measures: ISO/IEC 42001 (Annexes A, B)
- 2) Evaluation criteria for the organization providing AI systems: ISO/IEC 42001 Main Body
- 3) A mechanism for integratively evaluating and granting certification (e.g., a conformity assessment scheme): ISO/IEC 42007 High-level framework and guidance for the development of conformity assessment schemes for AI systems (currently under development by ISO/IEC SC42 JWG6)

Regarding element 1), the evaluation criteria for the AI system's control measures are deemed suitable, with ISO/IEC 42001 Annex A/B serving as an effective basis for assessing whether appropriate risk management is being applied to each product. As mentioned prior, for AI systems with uncertain risks and dynamically evolving systems or services, an approach that examines risk management and documentation processes as control measures for each product is usually, if not always, more suitable than rigid evaluation criteria focused on conformity to prescriptive specifications.

With element 2), the evaluation criteria for the organization providing the AI systems or product, the main body of the ISO/IEC 42001 should be applicable. As stated above, even if a list of control measures for a particular AI system is deemed appropriate, it is not guaranteed to remain continuously valid. Therefore, this element critically examines whether the organization developing and operating the AI system has implemented a system for continuously assessing and updating control measures, ensuring ongoing effectiveness in managing risks.

Element 3 is the most critical piece of the joint certification approach. As mentioned, the structure of ISO/IEC 42001 is particularly well-suited to this approach. It allows for a scheme where the management system requirements in the main body are assessed separately from the system-specific controls in its annexes. Crucially, updated ISO/IEC 17067¹⁴ now provide guidance on how to integrate multiple objects of conformity into a single certification scheme. Furthermore, the development of high-level guidance for creating AI-specific product certification schemes (ISO/IEC 42007) is already underway within international standardization organization (ISO/IEC JTC 1/SC 42 and ISO/CASCO). While this future guidance will not be a certification scheme in itself, its concepts will provide a clear and harmonized framework for scheme owners. This will help move the industry forward in creating practical conformity assessment schemes that encompass the joint certification concept.

Certification schemes that aim to evaluate both product-specific requirements and organizational requirements when granting certification also exist outside of the AI realm. Typical examples are Type 5 and Type 6 certification schemes, referenced in ISO/IEC 17067. Specifically, a Type 5 certification scheme not only assesses whether the product itself meets the required standards but also evaluates whether the management system of the manufacturing facility producing the product is operating effectively. Furthermore, after certification, it combines product sampling tests from the market with regular audits of the organization, thereby continuously ensuring conformity in product and organizational aspects. Conversely, a Type 6 certification scheme is primarily used for certifying processes or services, evaluating both the control measures for each product and the management system or direct organization that produces them. However, although the assessment methods within these types of certification schemes examine both the management systems and products, there is still only one object of conformity (the product/process) and only one resulting certification (certification of the product/process).

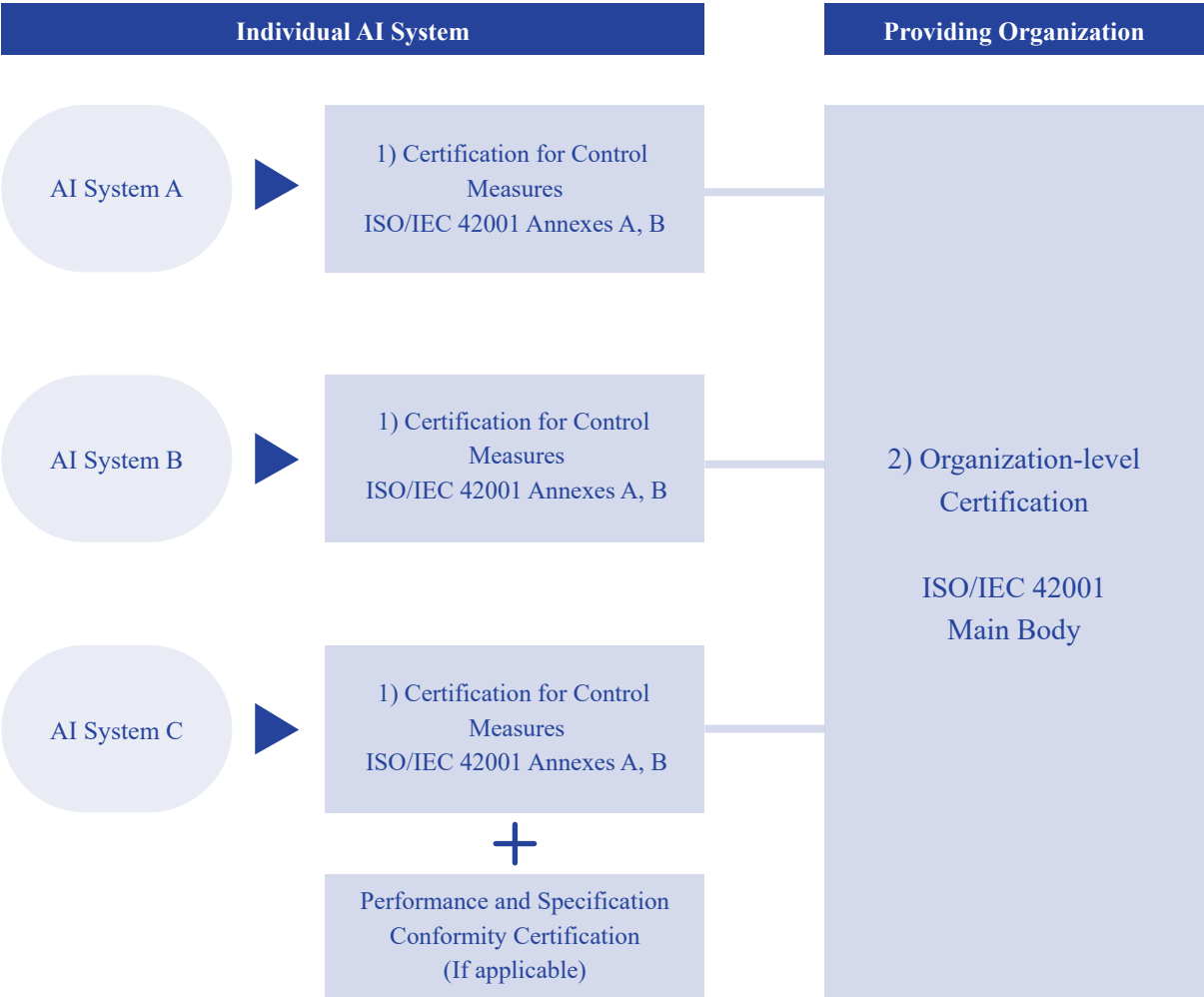
Therefore, if these certification schemes were applied to each AI system, it could require entire recertification of not only the different control measures for individual products or services, but also the organization itself — an entity that, while separate from individual AI systems, is not always directly tied to the specific risks of each new AI system. This is why considering new approaches to multiple objects of conformity such as implementing a two-stage certification process (on the AI system itself and one on the organization) could be

¹⁴ ISO/IEC 17067: 2013 (<https://www.iso.org/standard/55087.html>)

beneficial for both the certification body and the acquiring party. As illustrated in Figure 3, the proposed joint certification scheme allows the organizational component—which aligns with the main body of ISO/IEC 42001—to be certified once, with that certification being applicable across multiple AI systems. Thus, when the same organization provides different AI systems, only the control measures component of each new AI system would need to be certified, rather than a recertification audit every time.

The above content is illustrated via Figure 3 below:

[Figure 3 - Joint Certification for AI Systems]



3.3 Similar Initiatives in Fields Other Than ISO

Mechanisms that promote trust via evaluating control measures and the organizational structure supporting them are also adopted in areas beyond international standards bodies. Below, we explore examples outside the ISO context.

3.3.1 EU AI Act

The EU AI Act is a product safety regulation for AI systems that came into force in August 2024.¹⁵ The legislation requires, among others, conformity assessment for high-risk AI systems, which demonstrates whether a product, service, or management system meets the requirements stipulated in laws and standards. In the EU AI Act, the specific core requirements for this conformity assessment are: 1) requirements related to product-specific control measures concerning AI system risk management and record-keeping, and 2) requirements concerning the organization or management system that supports them.

More specifically, as part of the requirements for “control measures” for AI systems, providers of high-risk AI systems must establish and maintain a risk management system throughout the system's lifecycle, ensure data governance, prepare technical documentation, perform recordkeeping, provide information to users, and implement human oversight measures.¹⁶ These aspects are similar to measures in ISO/IEC 42001 Annexes A and B. In addition, the requirements regarding the “organization” of high-risk AI system providers mandate that they implement a Quality Management System (QMS) and establish a framework to ensure ongoing conformity.¹⁷ This QMS requirement closely aligns with the content of the existing management standard, specifically, the main body of ISO/IEC 42001.

Therefore, the EU AI Act requires evaluation of both control measures and organizational management to achieve the safety of high-risk AI systems. Because the EU AI Act operates within a product certification framework, requirements for conformity are based upon ISO/IEC 17065 and management system certification alone is insufficient to demonstrate that requirements are being met. However, because the joint certification approach proposes to assess and certify both the management system components (clauses 4-10, per ISO/IEC 17021) and the system requirements (Annex A, per ISO/IEC 17065), a joint certification based conformity assessment scheme could be designed specifically for the EU AI Act using ISO/IEC 42001 as a framework.

3.3.2 NIST AI Risk Management Framework (NIST AI RMF)

The AI Risk Management Framework (AI RMF), published by the U.S. National Institute of Standards and Technology (NIST), guides organizations to manage risks arising from AI systems as a voluntary practice.¹⁸ This framework also integratively addresses specific control measures and the organizational governance that supports them.

¹⁵ EU AI Act, (<https://artificialintelligenceact.eu/>)

¹⁶ Articles 9-14 of the EU AI Act

¹⁷ Article 17 of the EU AI Act

¹⁸ NIST, [Artificial Intelligence Risk Management Framework (AI RMF 1.0)] (<https://www.nist.gov/itl/ai-risk-management-framework>)

The NIST AI RMF consists of four primary functions: Govern, Map, Measure, and Manage. Among these, "Govern" requires establishing a culture, policies, and accountability structure to address AI risks at an organizational level, serving as the foundation for the other functions. This aspect corresponds to the organizational requirements discussed above. Meanwhile, "Map," "Measure," and "Manage" promote a series of specific activities for identifying, analyzing, and implementing countermeasures for risks throughout an AI system's lifecycle, aligning with the control measures for individual systems' schemes.

Therefore, NIST's AI RMF requires the practice of risk control measures based on the establishment of an organizational-level governance system. Here, too, an approach that integratively evaluates both control measures and organizational management is adopted to ensure system safety and reliability, demonstrating a similar risk-based and continually improving system. However, the NIST AI RMF is not a standard in and of itself, and cannot be certified against.

3.3.3 Japan's Government AI Procurement Guidelines

Japan's *Guideline for Governments' Procurements and Utilizations of Generative AI for the sake of Evolution and Innovation of Public Administration*,¹⁹ hereinafter referred to as the "Government AI Procurement Guidelines," was formulated to promote the utilization of generative AI by government agencies and manage use-case risks. These guidelines adopt an approach that integratively evaluates control measures and organizational structures when government agencies procure AI systems.

From an organizational standpoint, these guidelines require businesses to establish AI governance structures based on cutting-edge information, develop incident response procedures, and provide necessary human resources (HR) training. This concept is similar to the organizational management requirements discussed so far. Additionally, these guidelines outline detailed control measures that businesses must adhere to for individual AI systems.²⁰ They cover a wide range of specific requirements, including proper data handling, control of harmful information output, security assurance, attribution of intellectual property rights, and the obligation for businesses to cooperate in the event of a risk.

Ultimately, these guidelines require businesses to establish AI governance systems and implement detailed control measures through checklists for individual systems. As a result, both control measures and organizational management are evaluated integratively to ensure system safety and reliability, reflecting an approach similar to ISO/IEC 42001, and a joint certification approach to certification could provide government procurement systems with the assurance necessary for public safety.

¹⁹ Digital Agency, The Guideline for Japanese Governments' Procurements and Utilizations of Generative AI for the sake of Evolution and Innovation of Public Administration

(https://www.digital.go.jp/assets/contents/node/basic_page/field_ref_resources/e2a06143-ed29-4f1d-9c31-0f06fca67afc/6e45a64f/20250527_resources_standard_guidelines_guideline_04.pdf)

²⁰ The 'Procurement Check Sheet and the 'Contract Check Sheet,' seen in the Digital Agency's Government AI Procurement Guidelines form the core of this.

3.3.4 ISMAP (Information System Security Management and Assessment Program for Government Information Systems)

ISMAP (Information System Security Management and Assessment Program) refers to Japan's security assessment program for government information systems, which began in June 2020 with the primary objective of establishing unified security standards for government information systems.²¹ The evaluation process involves a third-party ISMAP Steering Committee assessing whether a given service meets the required security standards. These requirements are outlined in the ISMAP standards,²² which are divided into the following chapters:

- Governance Standards (Chapter 3): These standards require cloud service providers' management to oversee and guide organizational information security activities.
- Management Standards (Chapter 4): These standards require cloud service providers' managers to establish, implement, operate, monitor, maintain, and continually improve their information security management.
- Control Measures Standards (Chapter 5): These standards provide operational personnel of cloud service providers with a set of control measures to select from, based on risk response policies.

The ISMAP Management Standards, which partly cite ISO/IEC 27001, include similar organizational requirements that have been previously discussed above. Furthermore, the ISMAP Control Measures Standards include requirements for control measures, such as "Policies for information security shall be defined, approved by management, issued, and communicated to employees and relevant external parties" (Chapter 5, 1.1). While ISMAP is primarily an evaluation standard for cloud service products and systems, it also assesses control measures and organizational management to ensure system security, adopting an approach similar to joint certification.

The ISMAP Management Standards, which partly cite ISO/IEC 27001, include organizational requirements, while the Control Measures Standards define technical requirements. By evaluating both, ISMAP integratively assesses organizational management and system-specific controls to ensure security.

However, this differs from the joint certification model proposed in this paper. In ISMAP, both management and controls are assessed as components of a single object of conformity—the cloud service—resulting in a single certification for that service. In contrast, the key distinction of joint certification is that it treats the controls as a separate "product" certification, distinct from the management system certification, yet combines them within

²¹ NISC, Digital Agency, Ministry of Internal Affairs and Communications, Ministry of Economy, Trade and Industry 'Overview of the Security Evaluation System for Government Information Systems (ISMAP)' page 8 (https://www.ismap.go.jp/csm?id=kb_article_view&syparm_article=KB0010005&sys_kb_id=c7b753a583e86a10aa68c6a8beaad3c1&spa=1)

²² ISMAP Steering Committee 'ISMAP Management Standards' (https://www.ismap.go.jp/csm?id=kb_article_view&syparm_article=KB0010028&sys_kb_id=757ef62ac3755250076ededb05013195&spa=1)

one overarching scheme. This structural difference is what allows for the reuse and scalability central to the joint certification concept.

4. Advantages of the Joint Certification Approach

In this chapter, we will gain a better understanding of the advantages that the joint certification approach offers.

4.1 Flexible and Resilient Response to Unpredictable Risks

As discussed in the previous sections, joint certification is an approach that combines certification mechanisms for control measures on a product level and organizational-level management to certify AI systems. Its key strength lies in overcoming the limitations of certifying ISO/IEC 42001 as only a management system (even though there are system level controls in Annex A), or each in isolation, as described in Section 2.

While certifying individual risk control measures allows providers, who have an information advantage in their own AI systems, to determine the content and methods of risk response flexibly, it also has a key drawback: external evaluation of the sufficiency of those measures becomes difficult. Additionally, the content of these control measures may become outdated by the time the certification is granted.

In contrast, certification for the organization can bridge the information gap between the system provider and the external certification body by encouraging the organization to implement responsible control measures, thus being more adaptable to changing circumstances than certification for specific control measures. However, AI systems have a broad risk context and may behave autonomously, making it challenging to ensure trust in an AI system solely by certifying the organization.

Joint certification offers a flexible and resilient approach to promote trust, by incorporating the best aspects of these certification schemes.

4.2 Procedural Efficiency through Reuse of Management System Certification

A practical benefit of joint certification is to apply a single AI management system certification obtained at the "organizational" level for multiple AI systems provided by the same organization. In contrast, traditional product certification typically requires separate certifications for each AI system, including at the organizational level, which can result in significant costs and time delays.

With joint certification, the assessment of overall management controls (i.e., clauses 4-10) could be conducted at the organizational level, rather than on an AI system-by-AI system basis. This means that once an organizational-level certification is acquired, only the individual risk control measures of each specific AI system may require evaluation at the "product" level. This benefit is expected to reduce certification costs significantly and the time required to obtain certification, thereby promoting streamlined development,

deployment, and market rollout of AI systems. This benefit is vital for organizations developing and operating multiple AI systems, as it enables the entire certification process to become more efficient, particularly from a continuous perspective. Reusing management system certification also contributes to alleviating the burden on certification bodies and helps address the issue of strained resources.

4.3 Scalability through Connection with Certification in Individual Domains

The joint certification approach for AI systems can also be connected to existing management system certifications. For example, the international standard ISO 13485 (Medical devices — Quality management systems — Requirements for regulatory purposes) is referenced in numerous global regulatory frameworks for medical devices.²³ ISO 13485 and ISO/IEC 42001 share a similar basic structure, both requiring an organizational system aligned with the PDCA cycle. However, ISO/IEC 42001 includes additional requirements not found in ISO 13485, taking into account the characteristics of AI systems.

According to the whitepaper issued by Microsoft and EY,²⁴ the comparison of ISO 13485 and ISO/IEC 42001 is illustrated in Figure 4 below:

[Figure 4 - Comparison of ISO 13485 and ISO/IEC 42001]

Metric	What is similar?	What is different in ISO/IEC 42001?
Planning Phase		
Context of the Organization	- Establish and determine the scope of QMS/AI management systems. - Monitor and measure the defined processes. - Apply a risk-based approach to the control of the appropriate processes.	- Consider external and internal issues before determining the scope. - Understand the needs and expectations of interested parties. - Plan processes for continual improvement of AI management systems. - Personal identifiable information handling.
Leadership and Commitment	- Ensure quality, AI policy and objectives are established. - Communicate the importance of meeting requirements. - Ensure the availability of resources. - Ensure AI management system/QMS achieve the intended results.	- AI management system integration: ISO/IEC 42001 further adds integration into organization processes and emphasizes alignment with other organizational policies (Annex B.2.3).
Detailed Planning of Implementation and Changes	- Establish objectives at relevant functions and levels. - Determine the plan to achieve. - Determine the need for changes to the AI management system and QMS system. This shall be conducted in a planned manner.	- Risk management in planning (AI risk assessment, AI risk treatment and AI system impact assessment) before establishing objectives and an implementation plan.
Support and Resources	- Determine and provide human resource competence and training. - Ensure awareness of documentation requirements and control.	- Specific resources for AI systems include data resources, tooling resources, and system and computing resources.

²³ ISO 13485:2016 (<https://www.iso.org/standard/59752.html>)

²⁴ Microsoft and EY, “How can the medical device industry seize the upside of changing AI regulatory requirements? - Proposal for a joint-certification approach” Pages 9—10

Implement and Operate		
Implementation of Controls	- Plan processes that meet organizational requirements. - Control processes and create contingency plans and document risk management processes, as appropriate.	- Focus on AI risk assessment, treatment, and system impact assessment. - Provide potential controls to assess AI system impact in Annexes.
Check Performance		
Measure, Monitor, Analyze, and Evaluate	- Conduct internal audits at planned intervals. - Conduct management review at planned intervals. In ISO/IEC 42001, this is part of the performance evaluation phase, while it is covered in the planning phase in ISO 13485.	- Detect AI system impact and risk assessments at planned intervals.
Sustain Improvements		
Sustain and Act	- There is a need for improvement. - Take corrective action for non-conformities. - Maintain and document investigative records.	- It needs continual improvement.

For example, if a medical device that meets ISO 13485 management system requirements is equipped with AI functionality, clearing the additional requirements for the AI functionality (the right column of Figure 4) could enable it to obtain ISO/IEC 42001 certification simultaneously. This approach would substantially minimize both the time and cost, in contrast to the process of obtaining ISO/IEC 42001 certification independently.

Furthermore, in the financial sector, the EU AI Act suggests the possibility of reusing existing quality management systems and monitoring systems based on EU directives on the prudential supervision of credit institutions.²⁵ Likewise, if parts of organizational management and control measures needed in various fields can be easily transitioned for AI system certification, joint certification can gain further scalability, thereby reducing certification and audit costs even further.

²⁵ Directive 2013/36/EU of the European Parliament and of the Council of 26 June 2013, (<https://eur-lex.europa.eu/eli/dir/2013/36/oj/eng>)

4.4 Ensuring Conformity with AI-Related Laws and Regulations

The implementation of joint certification may enable higher conformity levels with numerous AI-related laws and regulations worldwide. While multiple legal and regulatory discussions are ongoing globally regarding the development and use of AI systems, the EU AI Act is recognized as the most comprehensive regulation among them. Joint certification can play a crucial role in ensuring alignment with these evolving legal frameworks.

As stated in Section 3.3.3, the AI Act adopts an approach that evaluates both control measures and the organization for providers of high-risk AI systems, requiring conformity assessment for the former. The Act mandates third-party conformity assessment for biometric identification systems where harmonized European standards have not yet been established, as well as for AI systems integrated into products that require certification under other EU product safety regulations. For other high-risk AI systems, self-conformity assessment is also mandatory. The criteria for conducting these conformity assessments are to be stipulated in harmonized European standards. However, currently, harmonized European standards have not yet been established.

Yet, a comparison of the items in the AI Act and ISO/IEC 42001, as shown in Figure 5, reveals that many of the requirements in Article 17 of the AI Act overlap with the main body of ISO/IEC 42001. Additionally, the requirements related to control measures in the AI Act are either similar to or identical with those in ISO/IEC 42001 Annexes A and B. Consequently, by establishing adequate control measures and an organizational structure to obtain joint certification, providers of high-risk AI systems can meet many of the obligations outlined in the EU AI Act.

A comparison of the AI Act and ISO/IEC 42001 is shown in Figure 5 below:

[Figure 5 - Comparison of the EU AI Act and ISO/IEC 42001]

◆ Organization-level Management Systems

ISO/IEC 42001 (Main Body)	EU AI Act
Articles 4–10	Article 17: Quality Management System

◆ Control Measures

ISO/IEC 42001 Annex A	EU AI Act
A.5.2 AI system impact assessment process	Article 9: Risk Management System
A.7.2 Data for development and enhancement of AI systems	Article 10: Data and Data Governance
A.7.3 Data acquisition	
A.7.4 Quality of data for AI systems	
A.6.2.7 AI system technical documentation	Article 11: Technical Documentation
A.6.2.8 AI system recording of event logs	Article 12: Record Keeping
A.8.2 System documentation and information for users	Article 13: Transparency and Provision of Information to Users
A.9.3 Objectives for responsible use of AI systems	Article 14: Human Oversight

Furthermore, beyond cross-sectoral regulations like the EU AI Act, the development of AI-related regulations and certification schemes continues to progress worldwide across various domains, including automobiles, medical devices, and financial systems. Thereby, considering the characteristics of AI systems and the advantages and disadvantages of existing certification schemes discussed in Section 2, the fundamental framework for final authorization or certification is likely to adopt the form of joint certification, as explored in this white paper. Additionally, as mentioned in Section 3, many documents, such as the US NIST AI RMF and Japan's Government AI Procurement Guidelines, adopt a similar structure.

Ultimately, a joint certification mechanism can ensure consistency in AI-related regulations across various countries, reduce conformity costs for businesses entering global markets, and lower the costs for governments in enforcing laws, thereby resulting in more efficient and effective regulation.

5. Conclusion: The Future of Joint Certification

As AI systems become deeply integrated into society, there is a new push for forward-thinking and adaptive certification frameworks. In this white paper, we have analyzed existing certification schemes, identified the limitations of current approaches, and discussed the necessity, concrete methods, challenges, and benefits of joint certification.

In AI system certification, the traditional approach of evaluating product specification conformity for physical products is not always practical due to the dynamic nature and complexity of AI systems. We demonstrated that “joint certification”, which combines the strengths of evaluating control measures for AI system risks and assessing the organization's management systems that provide AI systems, could lead to more substantive trust and confidence in AI systems. Joint certification is based on the idea that the trustworthiness of AI systems is ensured only when an AI system provider with a reliable management system continuously and proactively implements appropriate control measures throughout the AI system's lifecycle. ISO/IEC 42001's Annexes A and B serve as concrete evaluation criteria for guiding control measures in AI systems, while the main body of ISO/IEC 42001 provides evaluation criteria for organizations that develop and/or implement AI systems. Moreover, ISO/IEC 42007, currently under discussion at ISO, should also guidance on how to evaluate both integratively, via an appropriately designed certification scheme, based on the content of ISO/IEC 42001.

Joint certification is flexible in responding to unpredictable risks and is adaptable to change. It offers benefits such as reusing a single organizational certification for multiple AI products, combining AI certification for individual domains like medicine or finance, and potentially replacing some legal intricacies. This approach can significantly reduce the time, cost, and effort required to obtain certification and its associated audits in many situations; thus, joint certification can be a highly significant tool in promoting trustworthiness and practicality to AI systems.

Yet, challenges exist when introducing joint certification. Firstly, it is challenging for a third party to assess whether the content of the control measures selected and implemented by the AI system provider is suitable. Consequently, it is challenging to determine how to effectively combine countermeasures to manage risks for diverse AI systems, especially given their various use cases. Determining the socially acceptable level for various AI risks is also difficult, as even frameworks like the AI Act or the NIST's AI RMF do not specify acceptable thresholds. Furthermore, for general-purpose and multimodal AI, there are innumerable risk scenarios, and multiple entities are expected to perform fine-tuning, which poses further complexities. Given this, even for businesses familiar with AI system risks, it could be extremely difficult to determine what risks are anticipated and to what extent countermeasures should be taken to mitigate them. Thus, with organizational certification, more concrete discussions are required on what kind of organization can continuously design, operate, and improve control measures, and how risk countermeasures should be coordinated among different organizations.



To maximize the effectiveness of joint certification in addressing these challenges, public and private stakeholders must share practical use cases and discuss the nature of control measures and potential organizations for each scenario, as well as explore inter-organizational collaboration. By consistently applying these inclusive processes, more AI system providers can implement optimal control measures and organizational arrangements at minimal cost, and users can place greater trust in AI systems. Such a mechanism aligns with Japan's digital policy promoting 'Agile Governance' and the philosophy of the Act on Promotion of AI formed in May 2025. Joint certification, constantly updated based on the latest best practices, will be a powerful tool for ensuring the trustworthiness of AI systems and accelerating the social implementation of innovation.

Designing an AI System Certification Scheme: An Attempt at Joint Certification via Integrating Risk Control Measures and Organizational Management

Date of Publication: December 2025

Published by: Smart Governance Ltd.
<https://smart-governance.co.jp/>

【Authors】

Hiroki Habuka
Haruki Matsui
Yuchang Cheng (External Contributor)
Takafumi Ochiai
Matteo Posarelli
Mizuho Miyahara

【Advisor】

Kimberly Lucy Director, GRC Standards (Microsoft Corporation)

Copyright © 2025 Smart Governance Ltd.
This white paper was prepared by Smart Governance Ltd. based on the results of joint research conducted with Microsoft Corporation.

All copyrights and related rights to this document belong to Smart Governance Ltd. Unauthorized reproduction, duplication, modification, or distribution and provision to third parties are strictly prohibited.