JRC 139430

European Commission

SCIENCE FOR POLICY BRIEF

# Harmonised Standards for the European AI Act

## HIGHLIGHTS

→ The European Union adopted the AI Act in August 2024, and the provisions for high-risk AI systems will start to apply after a transition period of 2 or 3 years[1].

→ European harmonised standards for the AI Act, provided they are published in the Official Journal of the EU, will grant a legal presumption of conformity to AI systems developed in accordance with them.

→ European standardisation organisations, led by CEN and CENELEC, are in the process of drafting the necessary AI standards, following a request from the European Commission.

→ This brief discusses some of the key characteristics expected from upcoming standards that would support the implementation of the AI Act.

## INTRODUCTION AND STATE OF PLAY

### The AI Act

The European Union AI Act [1], the first-ever legal framework on Artificial Intelligence (AI), entered into force on August 1st, 2024. The AI Act is part of a wider package of policy measures in the EU to support the development of trustworthy AI while strengthening its uptake, investment and innovation in the EU.

Among the key aims of the AI act are to ensure that AI systems respect the safety, health and fundamental rights of individuals, and to address the risks of very powerful AI models. The AI Act provides a uniform approach to address these issues across the EU.

This document is concerned with the essential requirements laid down in the AI Act for high-risk AI systems, and with the role of technical standards in defining how to meet them in practice.

After a transition period of 2 or 3 years[1], i.e. starting August 2026, high-risk AI systems will have to comply with requirements related to risk management, data quality and governance, logging and traceability, technical documentation, transparency, human oversight, accuracy, robustness, and cybersecurity.

Compliance with these requirements will be ensured through the establishment of a quality management system and through conformity assessment before placement on the market.

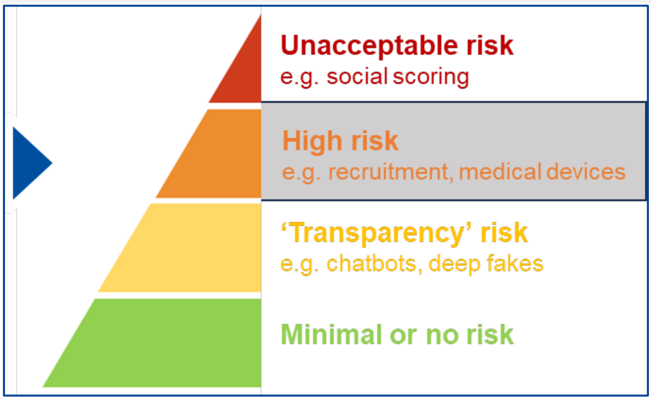
[1] A transition period of 3 years is defined for systems embedded in products already subject to third-party conformity assessment identified in Annex I to the Regulation, and 2 years for other high-risk systems identified in Annex III to the Regulation

**THE ROLE OF STANDARDS**

> "Standards are important instruments to support the implementation of Union policies and legislation and to ensure a high level of protection of safety and fundamental rights for all persons in the Union. Standards can also support the establishment of equal conditions of competition and a level playing field for the design and development of AI systems, in particular for small and medium-sized enterprises that develop AI solutions." [2]

### High-risk AI systems

The AI Act defines various categories of AI system according to their risks. Most AI systems and applications pose minimal or no risks. Certain AI systems are subjected to transparency obligations, e.g. when they interact with natural persons or pose risks of impersonation or deception. High-risk AI systems are limited to those that could have a significant harmful impact on the health, safety or fundamental rights of individuals. For these, the AI Act defines a clear set of requirements, as extensively discussed in this document. Finally, the Regulation prohibits certain AI practices that pose unacceptable risks.



## Standardisation for the AI Act

The AI Act defines the essential requirements that high-risk AI systems must satisfy in order to guarantee their safety. In line with other pieces of product legislation, technical standards define concrete approaches that can be adopted to meet these requirements in practice. These are harmonised standards, developed by European Standardisation Organisations upon request from the European Commission to support compliance with EU legislation. After being assessed and published in the Official Journal of the EU, standards will confer providers of high-risk AI systems with presumption of conformity with the relevant legal obligations.

Given the key role played by standards, the European Commission has engaged with European Standardisation Organisations since the proposal for the AI Regulation was initially presented in April 2021. In May 2023, well in advance of the adoption of the AI Act by the European Parliament and the Council, the Commission adopted a formal standardisation request [2], which was accepted by CEN-CENELEC.

Standards build on the consensus of a substantial number of stakeholders: the standardisation request explicitly demands measures to facilitate the participation of representatives from all sectors and types of organisations besides large industry players, such as small and medium-sized enterprises and societal stakeholders. Considering this, creating a technical standard from the ground up is a process that can take a significant amount of time. On the other hand, European standardisation for the AI Act can make use of international standardisation activities from ISO and IEC, which can be adopted and recognised in the European context.

At the time of publication of this brief, shortly after adoption of the final legal text, drafting of many of the requested standards is underway, and an update to the standardisation request is in preparation. Despite this, the process to develop standards has been slower than anticipated by standardisation stakeholders. Reaching consensus on new work items required for the AI Act and their scope has often proven challenging, testing the limits of decision-making processes in standardisation committees and leading to delays. As consensus on fundamental topics starts to emerge thanks to participants' efforts, steady progress is required to complete drafting on time.

## CHARACTERISTICS OF AI STANDARDS

We present a series of characteristics that harmonised standards for the AI Act are expected to display, based on the analysis of the final legal text and the standardisation request.

Standardisation deliverables requested by the European Commission cover 10 concrete aspects of AI.

These are presented in Table 1, together with various key considerations and priorities related to their content, which standardisers may consider to support alignment with the legal text. In addition, harmonised standards for the AI Act should display a broad set of essential quality attributes, independently of the technical area of AI trustworthiness that they address [3]. This section presents some of these key attributes.

### Tailored to the objectives of the AI Act

Standards must specifically address and prioritise the risks that AI could pose to the health, safety and fundamental rights of individuals. However, existing international standardisation efforts tend to focus on protecting the objectives of organisations using AI [4]. There are fundamental differences between managing risks to organisational objectives and addressing possible risks of AI systems to individuals. The latter should be the focus of standards supporting the implementation of the AI Act.

### Oriented to AI systems and products

Standards for the AI Act should complement the organisation-centric view of existing international documents with a system and product-centric view, to ensure that the techniques and processes defined in them systematically address the risks of products and services using AI. Standards should cover all the phases of the product lifecycle, from initial inception of the AI system, when risks can already start to be identified and assessed, up to post-market placement stages, i.e. those in which AI systems are monitored while in operation, and possibly updated and modified.

### Sufficiently prescriptive and clear

In order to support compliance with the Regulation, standards must define requirements that AI systems must meet. On one side, documents that contain only guidance and recommendations cannot be used for presumption of conformity. On the other hand, standards containing an excessive amount of requirements would be highly counterproductive, especially if these are overly abstract or open.

Requirements in standards should define in clear and explicit terms the criteria and priorities that AI providers must observe when implementing them, as well as when assessing compliance. In essence, standardisers should strive to capture in precise terms the processes, techniques and methods needed to make AI systems trustworthy in a verifiable manner, ensuring they address all identified risks in line with the Regulation, while being mindful of the implementation burden.

### Applicable across sectors and systems

Standards should define, to the extent possible, horizontal requirements, i.e. requirements that are applicable to various types of AI systems across sectors. These can be complemented, when necessary, with requirements that apply to specific sectors or to specific types of systems, such as computer vision systems or natural language processing systems. In any case, certain degree of flexibility will be required when applying these requirements to specific AI systems in concrete operational environments. It should be reasonably clear for AI providers, when presented with horizontal requirements, how to identify suitable ways to apply them to their AI products in light of their intended use and the identified risks. Therefore, standards are also expected to provide the necessary guidance to support their application in specific contexts.

### Aligned with the state of the art

The current pace of advancement of AI is unprecedented, and many techniques move from research labs to products in very short timeframes. Consequently, many modern AI systems, such as those based on generative AI, are not addressed by existing international standards. The European Commission standardisation request focuses on high-risk AI systems and excludes obligations for general-purpose AI models defined in the Regulation. However, the standards requested should be applicable to all high-risk AI systems, including those integrating general purpose AI models as components. Therefore, standardisers should fully consider state-of-the-art AI techniques and modern AI system architectures when defining requirements for high-risk AI systems.

### Cohesive and complementary

The various areas of standardisation covered by the request of the European Commission cannot be considered in isolation. Standards for the AI Act should be cohesive and complementary with one another, ensuring that the many aspects of AI trustworthiness are covered in the resulting documents with a clear logical structure and organization that facilitates their adoption, and that they explicitly capture the various interdependencies and trade-offs between requirements. It is expected that a small number of standards will capture the core horizontal requirements for compliance with the AI Act, referencing as appropriate other documents for more detailed guidance on the implementation of specific clauses for concrete types of systems. This will require close coordination and constant communication between standardisation working groups.

Table 1: Standardisation deliverables requested by the European Commission

| Risk Management | Data Governance and Quality |
|---|---|
| Standards should specify a risk management system for products and services using AI. The requirements captured by standards should be aimed at identifying and mitigating risks of AI systems on the health, safety and fundamental rights of individuals. This is a novel aspect for AI standardisation, as the orientation of published and ongoing ISO/IEC work takes a very different approach in terms of risk objectives and definitions [3] [4].<br>Harmonised standards for other EU product safety legislation can provide a reference, as standards for the AI Act will also be product-oriented. This is in contrast to existing ISO/IEC work, which often focuses on the organisations using AI. But even if product safety standards are taken as a reference, new elements specific to AI have to be considered. These include technical ones, e.g. related to the software nature and lifecycle of AI, as well as non-technical ones, such as considering fundamental rights in risk assessment and mitigation plans.<br>Risk management requirements defined in standards should provide strong assurance and documented evidence that all the relevant foreseeable risks of the AI system have been identified and addressed. The effectiveness of risk mitigation measures put in place should be demonstrated through testing and evaluation based on suitable processes, metrics and thresholds. Indeed, testing for the purposes of identifying risk management measures and ensuring compliance with legal requirements is a key aspect of Article 9 of the AI Act.<br>In summary, standards do not have to prescribe specific risk treatment measures for every AI system. However, they should define clear and explicit requirements on the processes and outcomes to be achieved, as well as key criteria and priorities that providers of AI systems must observe, e.g. when testing mitigation measures. | Standards should cover both data governance and management aspects, as well as dataset quality aspects for AI Act compliance, as captured in Article 10 of the Regulation.<br>This will require consideration of various aspects not covered in existing ISO/IEC work. First and foremost, standardisation should explicitly take into account the risks identified as part of the risk management process. Indeed, it is expected to require the adoption of data-related measures specifically tailored to these risks, not just aiming to support broader organisational objectives [4]. In addition to the definition of explicit criteria for the selection of data quality metrics and governance processes, standards should define the evidence required to support these choices, i.e. to demonstrate their suitability and effectiveness in complying with legal obligations.<br>At the technical level, standards should address some of the unique aspects of data-driven AI systems, e.g. those based on machine learning. These include the possible effects of unwanted biases in datasets. In this context, the legal text makes strong emphasis on the statistical properties of datasets, e.g. in relation to representativeness, correctness and completeness. Standards should specify the key technical methods and techniques that support these in practice, including criteria and priorities to observe when defining and measuring these properties for specific AI systems.<br>The lifecycle of AI systems is also extensively referenced in the legal text, and standards should cover data governance through the various development and operational phases of AI products. In particular, they should consider the increasing complexity of data provenance in modern AI systems, and the often-unexpected ways in which data quality issues result in downstream risks when AI systems are in operation. |
| **Record Keeping** | **Transparency** |
| Standardisation on record keeping for the AI Act, in line with Article 12 of the Regulation, should define clear requirements to ensure that objectives in the legal text related to tracing and recording of events and information in AI systems are met. These include the identification of situations that may result in risks, and in general, recording of all operation and performance aspects of AI systems needed to monitor compliance with the full set of legal requirements, including after being placed on the market and deployed in operation.<br>Standards must ensure logging and record keeping systems consider all of the relevant events, triggers and information elements that are needed to support these objectives. It may not be possible for horizontal standards to prescribe exhaustive sets of information elements and events to capture by every AI system. Indeed, logging subsystems for individual AI products will | Standards on AI transparency should define all the relevant transparency information required to support compliance with Article 13 of the Regulation and the corresponding obligations and needs of providers and deployers of high-risk AI systems.<br>In practice, this is expected to cover a broad range of information elements, including those needed to support understanding of how the AI system works, its characteristics, capabilities, strengths, limitations and performance, in line with the legal text. Relevant aspects include, in particular, information about any potential circumstances that could give rise to risks, as well as information to support understanding of the outputs of the system, enabling deployers to make informed decisions, such as to oversee the operation of the system and take corrective action if needed. |

require consideration of their intended use, their risks, and possibly concrete practical trade-offs, e.g. related to the volume of data that the corresponding AI systems generate and their resources. However, standards can define a minimum set of generally applicable information elements, including generic data points (e.g. timestamps, versions, user actions and decisions) as well as AI-specific ones, e.g. related to machine learning models.

Besides these essential information elements, standards should set clear requirements on how to establish a logging plan for AI systems, and how to implement, test and document it, including the key criteria that AI system providers have to consider. This includes, for example criteria to define which events to log. Application of these requirements must guarantee that logging solutions are practical and cover all of the above-mentioned objectives of the AI Act.

Existing standardisation work at international level could serve as a solid foundation, especially if the published version the upcoming ISO/IEC AI transparency taxonomy sufficiently covers all the transparency elements required by the AI Act, including those related to the risks of AI to individuals and society. Provided that this is the case, additional standardisation work could set its focus on defining concrete requirements on top of this foundation, clearly defining the scope and structure of transparency artefacts for compliance, and the objectives and criteria that providers are expected to meet when producing them.

These requirements should collectively ensure that transparency information ultimately produced for a high-risk AI system is comprehensive, meaningful, accessible and understandable for its intended audience, in line with the needs of Article 13 of the AI Act.

### Human Oversight

Standardisation for the AI Act is expected to define clear requirements that support providers of high-risk AI systems in selecting, implementing and verifying the effectiveness of human oversight measures, in line with Article 14 of the Regulation.

There is a broad range of possible oversight measures to ensure that AI systems stay within intended operational constraints, and that natural persons are able to control and override its outputs if necessary. These may range from highly technical measures to enhance understanding of system decisions, to user interfaces that increase monitoring and interaction capabilities of operators, to various types of training measures, to name just a few.

Providers of high-risk AI systems should be able to translate the requirements defined in standards into concrete oversight measures from the wide range of available options, considering the intended use of their specific AI systems and the risks identified. Indeed, it may not be possible to anticipate and exhaustively prescribe oversight measures for every AI system. However, standards should define a set of clear requirements on how the selection of human oversight measures has to be carried out, and how these have to be implemented and tested.

The application of these requirements should lead to verifiable outcomes regarding the oversight of AI systems, and should clearly define the parameters and criteria to consider when testing the effectiveness of human oversight measures in preventing and minimising risks posed by the AI system, involving natural persons as required.

### Accuracy

Standards on accuracy for the AI Act in line with Article 15 of the Regulation are expected to define requirements that support providers of high-risk AI systems in the selection of relevant and effective accuracy metrics and thresholds. In addition, standards should define the processes, methods and techniques to adopt in order to measure accuracy reliably, and to report it following best practices.

In a similar manner as for other requirements, it may not be possible for AI standards to prescribe accuracy metrics and thresholds for every high-risk AI system, or define in full detail how to measure them. Despite this, a number of standards on AI accuracy focusing on some types of systems, such as NLP or computer vision, are currently under development, and these will provide specific requirements and guidance for some applications.

However, the primary objective of harmonised standards should be to define a layer of generally applicable requirements that ensure that the selected accuracy metrics and thresholds are demonstrably appropriate and effective in addressing the objectives of the Regulation.

These requirements should be clear for providers of high-risk AI systems, explicitly defining the criteria and priorities to follow when implementing methods to measure accuracy, when choosing between various options regarding metrics, thresholds and benchmarks, and when documenting the necessary evidence at the right level of granularity, in order to certify that their systems are compliant with Article 15 of the AI Act.

### Robustness

Standardisation on robustness should define requirements related to the resilience of high-risk AI systems when deployed, including when facing errors, faults or inconsistencies in the environment of operation, as captured in Article 15 of the Regulation.

### Cybersecurity

Standards on cybersecurity should define technical and organisational measures to achieve a level of cybersecurity that is appropriate to the risks of AI systems. Given the software nature of AI, some controls in existing standards will be applicable, such as those in the ISO/IEC 27000 family. These may be most relevant for the security of the ICT

Providers of AI systems have at their disposal a broad range of technical and organisational measures that can support robustness and prevent harmful or undesirable behaviours. Some techniques for robustness in certain types of AI systems start to be covered by ISO/IEC standardisation, and harmonised standards for the AI Act can build on these. However, it may not be possible for standards to provide an exhaustive catalogue techniques for robustness, or prescribe how it should be measured for every AI system.

Primarily, standards to support the AI Act should aim to define an essential set of horizontal requirements that ensure the robustness of high-risk AI systems, including requirements on the criteria and priorities to guide the selection of robustness measures, and the technical conditions and environments that should be established to effectively measure, monitor and report robustness for various types of high-risk AI systems.

In other words, standards are expected to support AI providers to define robustness metrics in line with the intended use and risks of specific systems, and to implement effective testing protocols, explicitly defining the conditions under which robustness assessment has to be carried out in line with Article 15 of the AI Act.

infrastructure underlying AI systems. However, AI-specific vulnerabilities, such as data poisoning, model poisoning, model evasion and confidentiality attacks, among others, pose new challenges that will require specific coverage in standards in order for these to fully cover legal requirements in Article 15 of the Regulation.

Ongoing standardisation work starts to capture, mostly in the form of guidance, aspects related to AI-specific threats. However, new threats and countermeasures constantly emerge. In light of this, a main objective of new standardisation on AI cybersecurity should be to define essential requirements for the implementation of a security risk assessment and mitigation plan for high-risk AI systems. Even if concrete security measures cannot be mandated in advance for every AI system, standards can define specific security objectives to achieve, and how these should be verified through testing. These objectives are expected to be defined primarily at the system level [5], especially when mitigation measures for component-level vulnerabilities, e.g. those linked to machine learning models or datasets, cannot be expected to be perfectly effective.

Standardisation on cybersecurity and other requirements, e.g. data quality and robustness, should be tightly coordinated, and mutually reference each other as appropriate.

## Quality Management

Standardisation deliverables on quality management for AI should specify how providers of high-risk AI systems have to establish and maintain an effective quality management system that ensures compliance with the Regulation and supports conformity assessment, including through a robust post-market monitoring system.

Similarly as for risk management, quality management system (QMS) standards for existing product safety legislation are a useful reference, supporting compatibility with existing processes in some sectors. However, additional considerations are required to cover the specificities of AI products, whether embedded in physical products or in the form of software services.

New standardization on QMS should adopt a targeted focus on the specific risks addressed by the Regulation, as well as a product-centric view. In addition, it should cover the full AI product lifecycle and all the relevant aspects of compliance defined in Article 17 of the legal text, e.g. those related to planning, resources, product requirements, techniques for AI system design, development and risk management.

Existing international work, such as the ISO/IEC 42001 AI management system standard, while not aligned in objectives and approach with the AI Act [3], contain some relevant clauses at the technical and organizational levels. These could be referenced, as appropriate, by new standardization in quality management for AI, while ensuring that its focus remains on the specific risks and objectives captured in the legal text.

## Conformity Assessment

Standards should define the procedures and processes required to assess conformity of high-risk AI systems with the Regulation prior to their placement on the market or being put into service. They should also define criteria for assessing the competence of persons tasked with the conformity assessment activities, whether these are based on self-assessment by the providers or are carried out by external third-party organisations.

Existing standardisation work, such as the ISO CASCO toolbox, provides a basis with generic principles and guidance for conformity assessment, which new standards for the AI Act can leverage as appropriate. Similarly, the specific requirements that AI products should meet, and that will be checked during conformity testing, are being defined and captured in other standards, such as those discussed in this brief.

Therefore, standardisation work for AI conformity assessment can be highly focused and precise, defining in practical terms how conformity assessment procedures, processes and frameworks should be applied and adapted for AI, and in particular high-risk AI systems, in consideration of the legal requirements in the AI Act.

In this context, alignment between standards for conformity assessment and the various high-risk AI system requirements is essential. For example, assessment of the quality management system plays a key role in conformity assessment for the AI Act. Coordination between parallel standardisation work items should ensure that the resulting standards are complementary and fit for purpose.

## DISCUSSION & OUTLOOK

We discuss the state of play of technical standardisation in support of the EU AI Act and provide important observations regarding the content and qualities of upcoming standards. This document is intended to support the alignment of these standards with the needs of the European AI Regulation. After the formal adoption of the AI Act, drafting of future harmonised standards is expected to move into its final stages. It is crucial that standardisation deliverables are available and published well before August 2026, when the obligations for high-risk AI systems become applicable, in order to give sufficient preparation time for providers of high-risk AI systems.

The main committee tasked with creating AI standards for the European Union, the Joint Technical Committee (JTC) 21 of CEN-CENELEC, has recently published an overview of 37 standardisation activities in support of the AI Act. These include adopted international standards as well as home-grown European norms, which are necessary to cover many critical aspects of AI trustworthiness where international standardisation is not fully aligned with the objectives of the AI Act.

The large number of standardisation activities considered highlights the ambition and complexity of the task. However, a subset of those standards – mostly in the form of home-grown documents – are expected to define the bulk of horizontal requirements, such as those highlighted in this brief, and will reference other standards when appropriate, either in a normative or an informative manner.

Considering the short time frame for application of high-risk AI system provisions, drafting of these standards must progress in a fast and steady manner in the next few months. Standardisation stakeholders, from committee chairs and working group convenors to experts making technical contributions, will play a key role in delivering the remaining technical contributions required and minimising non-productive delays.

In this context, it is essential that the level of consensus achieved by technical experts is sufficiently deep. Requirements captured in standards should be clear, precise and actionable, reducing the effort and uncertainty associated with regulatory compliance, especially for small and medium-sized enterprises developing innovative AI solutions.

The European Commission, through the newly established AI Office and the Joint Research Centre, continues to support this crucial effort.

## REFERENCES

[1] 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, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). Regulation-EU-2024/1689

[2] Commission Implementing Decision on a standardisation request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in support of Union policy on artificial intelligence. Register of Commission Documents - C(2023)3215

[3] European AI Office webinar on the Risk management logic of the AI Act and related standards - Streaming Service of the European Commission

[4] Soler Garrido, J., Fano Yela, D., Panigutti, C., Junklewitz, H., Hamon, R., Evas, T., André, A. and Scalzo, S., Analysis of the preliminary AI standardisation work plan in support of the AI Act. JRC132833

[5] Junklewitz, H., Hamon, R., André, A., Evas, T., Soler Garrido, J. and Sanchez Martin, J.I., Cybersecurity of Artificial Intelligence in the AI Act. JRC134461

## DISCLAIMER

This policy brief is authored by Josep Soler Garrido, Sarah de Nigris, Elias Bassani, Ignacio Sánchez, Tatjana Evas, Antoine-Alexandre André and Thierry Boulangé.

The information and views expressed in it are purely those of the authors and do not necessarily reflect an official position of the European Commission.

## COPYRIGHT

## CONTACT INFORMATION

josep.soler-garrido@ec.europa.eu