ISSN 1831-9424

JRC Technical Report

Analysis of the preliminary AI standardisation work plan in support of the AI Act

Soler Garrido, Josep
Fano Yela, Delia
Panigutti, Cecilia
Junklewitz, Henrik
Hamon, Ronan
Evas, Tatjana
André, Antoine-Alexandre
Scalzo, Salvatore

2023

Joint
Research
Centre

EUR 31518 EN

This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The contents of this publication do not necessarily reflect the position or opinion of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the European Union concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Contact information
Name: Josep Soler Garrido
Address: Edificio EXPO, Avda Inca Garcilaso sn, 41092 Seville, Spain
Email: josep.soler-garrido@ec.europa.eu

EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC132833

EUR 31518 EN

PDF ISBN 978-92-68-03924-3 ISSN 1831-9424 doi:10.2760/5847 KJ-NA-31-518-EN-N

Luxembourg: Publications Office of the European Union, 2023

How to cite this report: 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*, Publications Office of the European Union, Luxembourg, 2023, doi:10.2760/5847, JRC132833.

Contents

Abstract

This report provides an analysis of the standardisation roadmap in support of the AI Act (AIA). The analysis covers standards considered by CEN-CENELEC Joint Technical Committee (JTC) 21 on Artificial Intelligence (AI) in January 2023, evaluating their coverage of the requirements laid out in the legal text. We found that the international standards currently considered already partially cover the AIA requirements for trustworthy AI defined in the regulation. Furthermore, many of the identified remaining gaps are already planned to be addressed by dedicated European standardisation. In order to contribute to the debate on the AI standards and support the work of standardisers, this document presents an independent expert-based analysis and recommendation, by highlighting areas deserving further attention of standardisers, and pointing, when possible, to additional relevant standards or directly providing possible additions to the scope of future European standards in support of the AI Act.

Authors

Josep Soler Garrido

Delia Fano Yela

Cecilia Panigutti

Henrik Junklewitz

Ronan Hamon

Tatjana Evas

Antoine-Alexandre André

Salvatore Scalzo

Executive Summary

We present an expert analysis of standards considered by CEN-CENELEC JTC 21 for adoption in support of the Artificial Intelligence Act. In scope are standards in the list presented in the JTC21 plenary meeting in January 2023, which includes international ISO/IEC standards as well as new CEN-CENELEC work items to be developed at European level. This preliminary list of standards has been analysed from the lens of the deliverables requested in the draft standardisation request from the European Commission.

Risk Management

In the current work plan, risk management coverage is provided mainly through ISO/IEC 23894 on AI risk management. However, the alignment of this standard with AI Act needs is rather limited given its broad and unspecific focus on organizational risks, and limited presence of risks to fundamental rights, health and safety considered in the AI Act proposal. In addition, it is a high level, non-prescriptive standard, providing guidance instead of concrete requirements. In light of this, a new European standardisation deliverable on risk, as currently proposed by CEN-CENELEC JTC 21 is of particular importance, and a promising candidate to address some of the above shortcomings. This upcoming work, the Check List for AI Risk Management (CLAIRM), is expected to provide a more granular set of technical requirements related to managing AI risks, with specific risk sources, harms and countermeasures. This should take the form of a prescriptive European Norm with a practical nature, setting concrete risk management requirements and enabling AI providers to select and implement risk treatment measures. It is also expected to contribute to avoiding excessive fragmentation in terms of risk management standards needed for the AI Act. Within this context, various standards with risk-related content have been actively discussed in standardisation meetings, including ISO/IEC 23894, 42001 and 31000, but none of them fully addresses the risks considered in the AI Act. Therefore, new standardisation work undertaken at JTC21 level represents an opportunity to address European needs regarding AI risk management in a single reference, preventing a situation whereby AI providers need to adopt an excessive number of standards to achieve a compliant risk management system. This central European reference can be complemented in specific aspects with pre-existing international work, for example covering process-oriented risk management considerations, or detailed coverage of crucial technical aspects, such as in the area of identification and treatment of unwanted bias.

Data Governance and Data Quality

These aspects are covered mainly through the ISO/IEC 5259 series on data quality for analytics and machine learning. In particular, part 3 provides comprehensive coverage of AIA requirements in terms of data governance. Regarding data quality, a comprehensive catalogue of quality attributes listed in part 2 of this series include those that are most relevant in the context of the AI Act. However, additional implementation requirements are needed to ensure proper selection and prioritisation of quality attributes in line with AI Act risks. In its current form, this series of standards has a broad scope, with data quality broadly defined as data meeting the organization's requirements. In this sense, the data quality attributes most related to AI Act requirements are only superficially covered. Sharpening the focus of this standard to better align it with AI Act legal requirements would also bring the additional benefit to reduce its implementation overhead for AI providers, as full compliance appears to require a comprehensive list of work products to be produced. Specifications such as ISO/IEC TS 12791 and IEEE 7003 covering aspects of unwanted bias, and in particular their clauses on data bias, are particularly complementary in this case, given their technical orientation and their potential to more sharply focus on the risks considered in the AI Act.

Record Keeping

Logging and record keeping are covered only to a limited degree in ISO/IEC 42001, as one of the optional controls to consider for the implementation of risk treatment options. Therefore, it is expected that further requirements on this topic will be provided in new work item, such as the CEN-CENELEC JTC21 "AI trustworthiness characterization", which in its outline already includes record keeping and traceability as trustworthiness characteristics. New specific work being proposed by JTC21 Working Group 3 is particularly relevant and encouraging. These planned ENs and any other new work items proposed in response to the AI Act standardisation request should aim to provide detailed coverage of logging and traceability mechanisms for situations that could result in risks, supporting oversight of AI systems, but also pay special attention to conformity assessment and post-market monitoring aspects. Among the standards reviewed by the time of publishing this report, one of the few mature sources with relevant content in this aspect is the IEEE 7001 on transparency, and especially the requirements provided to cover needs of incident investigator profiles, which

appear to be aligned with AIA priorities. This calls on the need to explore modalities for possibly integrating appropriate elements of IEEE standards into future European standardisation deliverables.

Transparency

Transparency in the AI Act, with a focus on provision of information to users, is covered to a certain extent in ISO/IEC 42001, which provides specification for an AI management system. Relevant controls listed include "System documentation and information for users" and "Understandability and accessibility of provided information". Additional, more detailed coverage is expected in new European work item(s) as well, notably the "AI trustworthiness characterization" work item. Its outline already includes transparency as a concern. It is important to ensure that this standard provides technical detail and potentially templates for the documentation of AI systems and datasets for user stakeholders with various profiles and levels of expertise. Other relevant sources to consider are emerging community practices for documentations of AI systems, models and datasets coming from industry and academia, which provide a good basis for future templates containing many relevant technical information elements for users of high-risk AI systems

Human Oversight

This is an aspect expected to be covered mostly in new work items, such as JTC 21 "AI trustworthiness characterization" work, which in its outline mentions three different aspects of oversight, including transparency, monitorability, explainability/interpretability, and intervenability/controllability. The range of considerations is encouraging, as it is crucial that this document considers a wide range of approaches and measures for human oversight, e.g. organisational, technical measures (redundancy, controllability), training measures, and proper design of human-machine interfaces, among others. It is also important that the risk of automation bias is properly addressed in harmonised standards. Limiting standardisation focus to technical approaches such as explainable AI techniques is not a suitable option, especially given that many of these approaches are still in research stage and are not yet technically robust. Indeed, standards are expected to rely on consolidated methods, and set realistic and achievable human oversight goals based on generally accepted techniques.

Accuracy and Robustness

In terms of international standards already considered, partial coverage of robustness is provided by the ISO/IEC 24029 series. Part 1 is a technical report, which could be a useful reference with guidance on robustness metrics for supervised classification/regression models using statistical and empirical approaches. Part 2 of this series is a technical specification describing formal methods for applications where these approaches are practically feasible. However, as is the case with other requirements, substantial additional coverage of accuracy and robustness should be provided in the JTC21 work item on "AI trustworthiness characterization". Its outline already describes a general framework for coverage of most trustworthiness requirements and their interdependencies, including criteria and observables to measure robustness. It is recommended that standardisers deliver clear requirements for AI providers, in order to ensure the appropriate selection and justification of accuracy and robustness criteria, metrics and thresholds. In addition, the specific challenges to verification and validation testing of accuracy and robustness for machine learning systems should be addressed. These elements should also include a discussion of accuracy and robustness for AI systems beyond classification and supervised machine learning. Additionally, in its current scope, this document covers mostly testing and measurement aspects. This needs complementation with guidance on design methods and practices for accuracy and robustness.

Cybersecurity

The current work plan considers adoption of ISO/IEC 27001, a widely adopted high-level standard specifying how to set up an information security management system. This is a generic document with a focus on organisational aspects, and is generally applicable to providers of AI products. In this regard, it provides good coverage of classical cybersecurity concerns, especially when used in conjunction with ISO/IEC 27002, which provides a list of security controls. However, the ISO/IEC 27000 series does not provide any meaningful coverage of AI-specific cybersecurity threats, AI-specific attacks or AI-specific security controls. These are not currently considered in the outline of CEN-CENELEC JTC21 new work items. AI cybersecurity risks should be covered in new European work on risk management, as part of a planned checklist of risks and risk mitigations. In addition, specifications covering threats, detection and mitigation measures, and security controls for AI-specific cybersecurity risks are required. On these, the upcoming ISO/IEC 27090 is a promising specification to follow.

Quality Management

The AI Management System described by ISO/IEC 42001 is broadly in line with the AI Act requirement for a Quality Management System. However, this is a high-level standard that is designed to provide as much flexibility and freedom as possible to organizations that apply it, which in its current form limits its usefulness as a means to ensure regulatory compliance. Prominently, the risks considered as well as the controls implemented for risk treatment are all optional, and any link to regulatory compliance is intentionally avoided. As such, European standardisers are encouraged to propose mechanisms to make this standard more concrete, prescriptive and tailored to the European regulation, either when adopting it initially as an EN, or as a harmonised standard for the AI Act at a later stage. It would be particularly relevant to make any necessary controls that directly address AI Act requirements mandatory, and demand stronger justifications for controls both excluded and adopted for AI risk treatment, including the provision of evidence of their effectiveness. In addition, more technical depth and concrete implementation requirements for some of the processes and controls is needed, especially for those not prominently covered in dedicated standards. In particular, coverage of post-market monitoring aspects should be extended in order to include specific mechanisms to monitor for risks and negative impacts to individuals. Requirements in terms of AI system impact assessment by providers of high-risk AI systems, i.e. details on how to assess potential negative impact on individuals, also deserve more prominence in this standard, making them a strict requirement and establishing links to upcoming European standardisation work on risk management. Similarly, further guidance on the management of modifications to the high-risk AI system would be required, with special consideration to those that could affect the AI system's risk profile.

Conformity Assessment

The current work plan contains a proposal for adoption of ISO/IEC 42006, a standard defining requirements for conformity assessment bodies auditing an AI management system based on ISO/IEC 42001. While still at an early stage, this is expected to be a relevant standard, given the importance of the quality management system in assessing conformity with the regulation. In addition to auditing the AI management system, standardisers should consider the need for specifications covering conformity testing of the AI systems themselves. These should cover competencies for assessment of high-risk AI systems, i.e. methodological and procedural steps for auditing them. Conformity assessment bodies would also benefit from standardisation deliverables providing guidance on processes and techniques to determine compliance with concrete trustworthiness requirements, and on the identification of key indications and triggers that justify more in-depth exploration and testing of AI systems during conformity assessment procedures.

Outlook

The analysis presented in this report shows that considerable progress has been made by European standardisers in a short period to prepare a work plan in response to the upcoming standardisation request for the AI Act. The adoption of international work, notably from ISO/IEC, will provide a valuable starting point to cover the various requirements for high-risk AI systems laid down in the AI Act. However, several standardisation gaps remain to be addressed. Furthermore, some of the areas where additional standardisation is required exhibit strong European specificities. For instance, technical specifications are required that specifically target the AI risks that the European regulation considers, i.e. those to fundamental rights, health and safety of individuals. In addition, the specific requirements considered in the regulation cannot be considered in isolation. Given the interdependencies and trade-offs involved, many important technical requirements and guidance for providers of high-risk AI systems are arguably best captured as part of a European Norm that comprehensively covers the various trustworthiness requirements laid down in the AI Act. Crucially, both of these aspects, risk and trustworthiness characteristics, are captured in the preliminary work plan by JTC21 in the form of new European standardisation deliverables.

1 Introduction

The Commission proposal for the AI Act is, at the time of publishing this report, entering the final stages of negotiations by the European Parliament and Council. Once the regulation enters into force, and after a transitional period, high-risk AI systems will have to comply with a set of AI trustworthiness requirements prior to their placing on the market or putting into service in the European Union. Harmonised and European standards developed by European Standardisation Organisations, while being voluntary instruments, will play a key role by defining technical solutions to fulfil those requirements. Furthermore, compliance with harmonised standards would provide a legal presumption of conformity to AI providers. The European Commission has already started the process to adopt a standardisation request providing a formal mandate to European Standardisation Organisations to develop the necessary standards. This request is expected to be formally adopted in the first half of 2023, marking the start of a period of four months during which the standardisation bodies addressed by the request should prepare and submit a work programme for the provision of the standardisation deliverables requested. However, even prior to the formal adoption of the standardisation request, substantial preparatory work has already been undertaken, especially in the context of the CEN-CENELEC JTC21 Special Advisory Group. Accordingly, a preliminary list of international and European standards expected to be part of the work plan has been prepared. In this report, we take this set of standards, as presented in the JTC21 Plenary meeting in January 2023, as a starting point, and provide an initial analysis of their coverage in relation to different requirements for high-risk AI systems in the AI Act proposal. This analysis is mainly intended as an input to the European Standardisation Organizations to support the further development of this work programme in response to the AI Act standardisation request.

Table 1 Standards considered for harmonization by CEN-CENELEC JTC21 WG1, as presented in the plenary meeting on 16/17 January 2023

ISO/IEC 22989	Artificial Intelligence concepts and terminology
ISO/IEC 23053	Framework for Artificial Intelligence (AI) system using Machine Learning
ISO/IEC 42001	AI management system
ISO/IEC 23894	AI Risk Management
ISO/IEC 5259-part 1	Data quality for analytics and machine learning (ML) - Overview, terminology, and examples
ISO/IEC 5259-part 2	Data quality for analytics and machine learning (ML) Data quality measures
ISO/IEC 5259-part 3	Data quality for analytics and machine learning (ML) Data quality management requirements and guidelines
ISO/IEC 5259-part 4	Data quality for analytics and machine learning (ML) Data quality process framework
ISO/IEC 27001:2013	Information security management systems
ISO/IEC 42006	Requirements on bodies performing audit and certification of AI management systems
CEN-CENELEC Risk	AI Risk catalogue and management
CEN-CENELEC Trustworthiness	AI trustworthiness characterisation

2 Coverage analysis

In this report, we individually consider the 10 items considered in Annex I of the draft Standardisation Request to support the AI Act, for which standards and standardisation deliverables are requested:

- risk management system for AI systems,
- governance and quality of datasets used to build AI systems,
- record keeping through logging capabilities by AI systems,
- transparency and information provisions to the users of AI systems,
- human oversight of AI systems,
- accuracy specifications for AI systems,
- robustness specifications for AI systems,
- cybersecurity specifications for AI systems,
- quality management system for providers of AI systems, including post-market monitoring process,
- conformity assessment for AI systems.

For each of these items, we discuss coverage provided by standards identified by CEN-CENELEC in the preliminary work plan. In general, while standardisation deliverables tend to be highly specialised and focussed, there is not a one-to-one mapping between the document reviewed and items in the standardisation request. Indeed, some standards may address multiple requirements at once, while others may just provide foundational elements that support other standards, but not directly address any specific requirement. Examples of the latter are standards such as ISO/IEC 22989 and ISO/IEC 23053, which are essential foundational standards covering terminology. These supporting standards, considered in addition to the 10 standardisation items in the standardisation request, are important documents that should also be identified and provided as part of a complete and consistent set of European standardisation deliverables. However, given this report's focus on specific requirements of the AI Act proposal, supporting standards are not discussed in detail. We can refer the reader to the analysis made by the JRC's in 2022 of these standards, reproduced in the annex of this report. On the other hand, our analysis also looks, when appropriate, beyond standards currently considered in the preliminary work plan. Additional items considered include some direct requests from the SAG group, such as specifications on bias (ISO/IEC 24024 and ISO/IEC 12791). In addition, other technical specifications identified by the JRC are referenced if they are deemed as potentially relevant, providing recommendations to address any gaps identified in the current work plan.

2.1 **Risk Management**

The AI Act acknowledges the socio-economic benefits that AI can bring but also highlights the new risks that come with this technology and defines requirements to address them. Within the AI Act context, in line with the EU product safety legislation approach, risk is understood as the potential to adversely affect individuals or the society. Article 9 of the AI Act requires a risk management system to be established in relation to high-risk AI systems. Such system shall be put in place to manage different types of risks through risk management measures which are identified through testing. Therefore, the three main objects of consideration here are *risk assessment*, *testing* and *risk management*.

The *risks* considered by Article 9 correspond to both inherent and residual risks of the AI system. Foreseeable risks shall be identified and analysed, and risk monitoring should be in place in order to identify and evaluate not foreseeable risks, e.g. using data gathered from a post-market monitoring system; all of them are to be eliminated or mitigated through *risk management measures*. Those risks which cannot be eliminated, shall be mitigated or controlled, and any residual risks after risk treatment should be judged acceptable. Additionally, the risk management measures put in place shall take into consideration the capacity and training of the user and the environment in which the AI is deployed. In order to identify risk management measures, Article 9 requires *testing*, which will also serve as a tool to ensure the consistent AI performance and compliance with Chapter 2 requirements.

This ecosystem of risks, risk management measures and testing, shall be managed in accordance with the requirements of the AI Act through a risk management system to be established by a provider and implemented,

documented, maintained and updated regularly. Such a system shall encompass the entire AI lifecycle, be continuous and be applied in an iterative manner.

Therefore, standardisers should provide technical requirements and guidance on:

- **risk assessment**, including how to identify and analyse foreseeable risks, and estimate and evaluate not foreseeable risks, e.g. through monitoring, and how to analyse and assess residual risks. This should notably include AI system **assessment of impacts** as part of the risk management activities carried out by AI providers, i.e. how to identify, evaluate and address potential negative impacts of the AI system on individuals and society.
- **risk treatment**, i.e. how to eliminate or reduce all types of AI risks in the context of the regulation, including the mitigation and control of those risks which cannot be eliminated
- **testing,** including how to identify risk management measures, and ensure compliance with Chapter 2 before placing the AI in the market, employing suitable metrics and probabilistic thresholds
- **risk management system**, including how to establish, implement, document, maintain and update such a system, and ensure it is continuous, iterative and takes place through the whole AI lifecycle.

In the current CEN-CENELEC JTC21 proposal for a new work item "AI risk catalogue and risk management", some of the above elements are to be covered within the presented scope. The proposal builds upon forthcoming international standards, such as ISO/IEC 23894 and 42001, focusing on providing further detail on the analysis of risks and possible countermeasures.

ISO/IEC DIS 23894 builds on the generic ISO/IEC 31000 guidelines on risk management, providing additional information, when necessary, on AI-specific requirements. Both standards share structure, and the content of 31000 is covered by 23894, which expands its requirements in relation to AI systems. Even though the attributes of the risks management system described by the ISO/IEC documents provide a valuable guidance for Article 9 requirements, the implementation specifications on those processes are limited and not prescriptive. More importantly, in their current form, this guidance is not suitable for the AI Act due to a fundamental difference between AI Act legal provisions and ISO/IEC standards on what is considered to be a risk.

In the ISO/IEC documents, the definition of a risk focuses on broad organisational objectives, including the objectives of the AI system itself. Accordingly, when ISO/IEC refers to risk, it means a positive or negative outcome on the AI system itself like, for example, the system's performance. The AI Act understands risk as the negative impact that an AI system could have on individuals and societies. Therefore, the AI Act puts individuals at the centre of attention when defining risk. Unfortunately, this mismatch in focus, renders the implementation guidelines provided by ISO/IEC 23894 unsuitable to cover AI Act requirements in their current form. Objectives and priorities to manage the risks an organisation developing an AI system faces, as addressed by ISO/IEC standards, are in many cases not directly transferrable to manage the risks AI systems pose to health, safety and fundamental rights of individuals. In other words, while closely interlinked, the methods needed to make sure a product works as intended are not necessarily the same as the ones needed to ensure such product is causing no harms to individuals or societies.

In light of this, the new JTC21 work on risk should develop a different approach, aiming in particular to capture European Union understanding of risks to health and safety as well as EU values and fundamental rights. Where appropriate, it can take existing ISO/IEC documents such as ISO/IEC 23894 as a guidance.

Similarly, new JTC21 work on risk is also expected to build upon ISO/IEC 42001 draft standard for AI management system, where risk management is a component of it. ISO/IEC 42001 specifies the processes to follow to manage AI systems as the main body of text, followed by more specific implementation details for controls in the two normative annexes. Section 6 on planning provides a high level description of the processes to be considered to assess and treat risks, and is therefore one of the most relevant sections for the AI Act.

However, in its current form, many aspects of risk management in the ISO/IEC 42001 draft standard are only optional considerations and, like ISO/IEC 23894, primarily focus on risks for organisations using AI systems. Despite this, and even though we encounter the same mismatch in terms of risk definition, ISO/IEC 42001 does to some extent and in an optional manner take into account the potential consequences to individuals and societies in its guidelines to risk assessment (6.1.2) and AI system impact assessment (6.1.4). As further discussed in section 2.8 on Quality Management, stronger requirements and additional implementation guidance is required on these important aspects, and this is an area where JTC21 work items on risk could complement international standardisation.

Some concrete areas that new JTC21 work on risk could cover are the following.

- **More depth in risk assessment**. Provide a clear and operational overview of risk sources, as well as methods for estimating and evaluating foreseeable and not foreseeable risks, e.g. through monitoring. Methodologies and criteria for assessment of residual risks should also be considered.
- **Concrete guidance on risk treatment**. Concrete methods to eliminate or reduce risks should be provided, including technical specifications covering mitigation and control techniques not currently covered in ISO/IEC standards. This is expected to be in the scope of the proposed JTC21 work item on risk, including a checklist and catalogue for AI risk management.
- **Appropriate guidelines for the risk management system**. Provide a risk management system tailored to the needs of the AI Act. This should be focused on the AI product development lifecycle, i.e. focused on the development of AI products, and in line with the risks considered in the AI Act proposal, ensuring coverage of the whole AI lifecycle.
- **Testing**. Provide, when not covered by requirement-specific standards, general requirements and guidance on how to test risk management measures, and how to perform testing to ensure consistent performance for its intended purpose in compliance with AI Act requirements.

To achieve the above, we discuss some additional standardisation resources, not in the preliminary work plan, which could be taken into consideration.

Regarding risk treatment, the ISO/IEC standards on bias, such as ISO/IEC 12791, already provide relevant requirements and guidance to identify, analyse, estimate, evaluate, mitigate, control, eliminate and reduce certain AI risks. Initial versions of this ISO/IEC work on bias are, however, limited in scope, as they cover unwanted bias in classification and regression machine learning tasks. Despite this, it is a prescriptive document and still under active development, with the potential to be expanded and aligned to AI Act needs in terms of risks and AI systems in scope. An additional perspective on AI bias is offered by the IEEE 7003 standard, which defines minimum criteria to ensure these systems don't cause unintended, unjustified nor unacceptable bias. Even though only the risks related to bias are considered by these standards, they still serve as a useful reference, providing technical specifications for risk management, as unwanted bias is one of the major risk sources in AI systems.

A new document in early stages of development, ISO/IEC 42005 aims to provide further guidelines for AI impact assessment by providers of high-risk AI systems, with the advantage that it does not display the objective mismatch of the aforementioned ISO/IEC ones. In the future, this document may complement 42001 in the assessment of risks and impacts of the AI system to users and individuals, providing specific methods, and should be followed closely.

Finally, none of the standards considered so far include substantial guidance of testing from a risk management perspective, as referred to in the AI Act. In any case, testing aspects specific to the individual trustworthiness requirements are expected to be covered by requirement-specific standards. Nevertheless, general considerations for risk-oriented testing of AI systems, as well as testing any relevant AI risk mitigation measures, should be in the scope of new JTC21 work on risk.

2.2 Data Governance and Data Quality

In the context of Article 10, standards are expected to cover data governance and data quality aspects. In terms of data governance, important considerations are:

- coverage of dataset design choices, data collection, data preparation and processing steps throughout the lifecycle, including any annotation and labelling steps,
- relevant assumptions, e.g. with regard to what the data used in AI systems represents,
- assessment of availability, quantity and suitability of data,
- notably, examination of unwanted biases in data,
- any data gaps and shortcomings.

Regarding data quality, attention should be paid to quality attributes that are relevant to the risk-oriented nature of the regulation, notably their relevance, representativeness, correctness and completeness, as well as their statistical properties, considering the specific settings where the AI systems are to be deployed.

The ISO/IEC 5259 series of standards is the main reference in the current work plan to address data quality and data governance. An aim of this series is to provide tools and methods to assess and improve the quality of data used for analytics and machine learning, defining a full data quality framework.

In terms of data governance, part 3 is the most relevant element of this series for the AI Act, as it defines quality management requirements and guidelines. Clauses 6 and 7, in particular, define specific data quality management requirements and recommendations, covering the entire lifecycle. This is complemented with part 4, which provides guidance and good practices for achieving data quality for machine-learning systems, considering specific approaches, such as supervised, unsupervised, semi-supervised, reinforcement learning.

This matches well with the requirements for data governance in the AI Act. However, there are two important considerations to make:

- **Complexity.** The complexity for adoption of this entire data governance framework appears to be high, as for each stage it demands a comprehensive list of work products. It would be important for standardisers to delineate which specific parts of this standard and work products are relevant for compliance with AI Act requirements[1], in a way which is proportionate, effective and aligned with the objectives of the AI Act.
- **European specificities.** The standard does not go into detail on a number of relevant issues important form the point of view of the AI Act.. For instance, unwanted data bias is just an item captured in a long list, without further elaboration on how it should be identified, measured and addressed, as these are highly relevant to the specific risks to individuals addressed in the AI Act.

In terms of data quality, part 2 of this series provides a comprehensive catalogue of data quality attributes. However, an important consideration concerns the definition of data quality, which in this series is defined from the perspective of data meeting organizational requirements. This is a broad and generic definition, which, for adoption in the EU context, would require alignment with data quality priorities in the AI Act, such as mitigating any potential risks to individuals. This is a similar concern as in the case of risk, pointing to specificities arising from the European AI regulation. In this sense, upcoming European standards are expected to adopt a view of data quality that reflects the potential role of data in the emergence of risks to individuals, and the use of best data quality and governance practices to mitigate them. Naturally, many of the data quality characteristics defined in part 2 of the 5259 series are relevant to this end. However, further specification is required to guide the selection and assessment of the right data quality attributes in line with European priorities as captured in the AI Act.

Based on the above, this standard would ideally be complemented with other specifications that are more focused on the specific areas of concerns of the AI Act, such as the impact of AI systems trained with data on individuals, specifying data quality requirements from this point of view. For example, documents covering bias, such as ISO/IEC 24027, ISO/IEC 12791 and IEEE 7003 appear to be very complementary. The latter, for instance, provides a comprehensive description of bias sources in data, with practical guidance for developers like the consideration of proxies for protected attributes, assessment of data from external sources, consideration of pre-processing in bias mitigations, and general dataset design considerations. Further considerations on data quality, including from specific sectors, can be found in the recent JRC report (Balahur-Dobrescu, et al., 2022).

2.3 **Record Keeping**

Record-keeping specifications should cover aspects related to logging during operation of high-risk AI systems, enabling traceability. In particular, they are expected to be relevant to situations that may result in risks or lead to substantial modification. These specifications are also expected to specifically consider how to facilitate post-market monitoring, as well as potentially conformity assessment.

International standards already considered in the work plan partially cover logging and record-keeping aspects. ISO/IEC 42001, defining an AI management system, covers these aspects in the list of controls to consider for the implementation of risk treatment options. In particular, it states that automatic record keeping of event logs should be enabled while the AI system is in operation when necessary. In this sense, implementation guidance in Annex B is in line with AI Act needs but lacks detail and is not strongly focused on the identification of risks or post-market monitoring aspects. Indeed, it refers to these aspects only superficially, stating that logging

[1] Notwithstanding the usefulness of a data standard considering needs arising from various relevant legislations. Data processors are expected to observe multiple regulatory requirements, and having these covered by a reduced number of standards would be highly beneficial.

should enable assessment of the impact of the system to relevant interested parties, and monitoring of undesirable performance.

Therefore, newly developed standards in response to the AI Act standardisation request must provide most of necessary coverage. In particular, further coverage is anticipated in the CEN-CENELEC NWIP on AI trustworthiness characterization. Indeed, record keeping and traceability are included in the list of trustworthiness characteristics, as assurance concerns, with a set of criteria and observables expected to be developed as part of this standardisation deliverable. In addition, a dedicated PWI on logging is being prepared within JTC21 WG3, and it is expected to cover relevant aspects of record management for monitoring and auditing AI systems. It is important to ensure that these documents demand –and provide guidance on the implementation of– concrete and practical measures and techniques for effective record keeping, in a way which is proportionate to the risks and context of use of specific AI systems. They should be broadly applicable, allowing AI providers to define events, metrics and information to be recorded that are tailored to specific AI systems and their associated risks, and which support AI system oversight, conformity assessment and post-market monitoring needs.

Existing international standards can potentially provide relevant content for record-keeping requirements. In particular, certain parts of IEEE 7001 on transparency could be a useful reference. Specifications in this document for incident investigator stakeholders are particularly relevant in the context of Article 12 of the AI Act. The clauses in this standard cover practical implementation aspects, such as the use of standard formats, recording of timestamped data, and coverage of inputs, outputs and decisions. Similarly, lifecycle coverage is comprehensive, and it establishes the need to facilitate investigation of incidents, with a focus on supporting inspections in the case of failures of autonomous systems that result in harms. Notably, it demands traceability of internal processes in AI systems leading to incidents and provides practical implementation guidance to enable reconstruction of incidents, including cases where neural networks are used. It also prescribes the storage of relevant internal information such as decision-making logic and the availability of adequate auditing tools. These are all important aspects to capture in technical specifications on record-keeping for the AI Act.

2.4 **Transparency**

At its core, Article 13 is about documentation and provision of information to users of for high-risk AI systems. It states that these systems should be designed and developed so that their operation is *transparent*, in order to enable users to understand and use the system appropriately. The article also states that high-risk AI systems must come with instructions for use, which must be in a digital format that is concise, complete, correct and clear, and easily accessible and comprehensible to the users. The instruction must include the identity and contact details of the provider, the performance specifications of the AI system including the level of accuracy, robustness, cybersecurity, and any known or foreseeable risks to health and safety or fundamental rights. Furthermore, Article 13 requires providing accessible and understandable information about the data used to train the AI system. Also, the instructions must include information about the human oversight measures (see recommendations for Article 14) and the expected lifetime of the system and any necessary maintenance measures.

Considering the standards in the preliminary work plan, the draft standard ISO/IEC 42001 on AI management covers some of these information elements within its normative Annex B. In particular, section B.8.2 on *system documentation and information for users*, with minor modifications, such as the inclusion of data specifications and information about cybersecurity in the user documentation, in conjunction with section B.3 on *understandably and accessibility of provided* information, could provide reasonable coverage of Article 13. However, these sections are part of an annex presenting controls which are only optional for organisations to adopt, establishing only the obligation to provide a justification for inclusion or exclusion. In order to comply with the AI Act, controls B.8.2 and 8.3 would need to be extended in terms of implementation guidance, and either become mandatory or be subject to stricter justifications by organisations.

Some aspects of transparency of AI systems are also discussed in ISO/IEC DIS 25059:2022(E) "Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems". ISO/IEC 25059 considers transparency of an AI system to be related to the amount of information available on the system and the way it is communicated to relevant stakeholders. The standard also mentions that "*highly transparent and modular systems can be built of well-documented subcomponents whose interfaces are explicitly described*", which is in line with the idea of transparency of Article 13. However, this does not directly specify neither how to craft a good documentation for the subcomponents of the AI system nor how to properly describe its interfaces.

The reviewed version of JTC21 proposed PWI on Trustworthiness mentions transparency as one of the trustworthiness characteristics, but at this early stage does not go into much detail on how it will be addressed. An initial insight on how transparency could be approached by this PWI is given by the positioning of the standard with respect to other trustworthiness projects. In relation to transparency, ISO/IEC TS 12792 "Information technology — Artificial intelligence — Transparency taxonomy of AI systems" is mentioned.

ISO/IEC TS 12792 is a working draft for a transparency taxonomy of AI systems. It aims to produce a foundational standard with a taxonomy that can be used in other standards, or to improve communication on transparency issues between stakeholders. The definition of AI Transparency provided in the standard (which is that of ISO/IEC 22989 - 5.15.8) aligns with that of Article 13. Such definition of transparency of an AI system focuses on providing appropriate information about the system to stakeholders, including the goals, limitations, design choices, assumptions, features, models, algorithms, training methods, and quality assurance processes used. Additionally, it also covers providing information about the data used to produce the system and its protection, as well as the purpose, and how the system was built and deployed. This standard, while not a prescriptive document, could potentially provide a very valuable foundation to later define concrete transparency requirements, e.g. through a common understanding of transparency mechanisms and relevant technical information elements to be considered by AI providers, including practical guidance for the provision of information to users and other stakeholders. As such, it is important to ensure that the standard provides comprehensive coverage of all the key transparency elements mentioned in the AI Act, including both technical aspects, e.g. those linked to the design, validation and operation of AI systems, as well as non-technical considerations, e.g. related to the context where AI systems are used, and their performance and potential risks to individuals and society.

In conclusion, harmonised standards concerned with transparency, and in particular the relevant sections of the upcoming JTC21 work on trustworthiness, should specify the process for documenting in a clear and understandable way both the AI model and its training/testing/validation data. Future harmonised standards on AI transparency could also define templates for reporting information on AI systems and describing their transparency level. In this regard, the ISO/IEC TS 12792 working draft already refers to non-normative transparency forms tailored to concrete stakeholders, mentioning concrete formats. In line with this standard's objectives, standardizers can draw inspiration from widely used frameworks such as *Datasheets for Datasets* (Gebru, et al., 2021) and *Model Cards for Model Reporting* (Mitchell, et al., 2019) complementing them with all the information required by Article 13. An analysis of some of these community initiatives from the lens of the AI Act has been recently published (Hupont, Micheli, Delipetrev, Gómez, & Soler Garrido, 2023).

2.5 **Human Oversight**

Article 14 lays out requirements for human oversight of high-risk AI systems. It states that these systems should be designed and developed in a way that allows for effective oversight by natural persons during use, including through the use of appropriate human-machine interface tools. Human oversight aims at preventing or minimizing risks to health, safety or fundamental rights that could emerge when a high-risk AI system is used. Human oversight should be ensured through measures identified and built into the AI system by the provider or, if appropriate, to be implemented by the user. These measures should allow individuals to fully understand the capabilities and limitations of the AI system, detect and address any issues, prevent automation bias, interpret the system's output, override or reverse the system's output and intervene on the system's operation. Additionally, for certain types of high-risk AI systems, actions or decisions based on the system's output should be verified and confirmed by at least two natural persons.

From the set of standards in the preliminary work plan, ISO/IEC 42001 provides some coverage of Article 14 concerns, emphasizing the requirement for human oversight measures. However, it does not provide any technical specification for the implementation of human oversight measures. These will be covered by upcoming JTC21 standardisation. Specifically, the PWI on Trustworthiness mentions four aspects of Human Oversight: Transparency, Monitorability, Explainability/Interpretability, and Intervenability/Controllability. At this early stage, the PWI does not go into further details of these. However, an insight on how they might be addressed comes from the positioning of this standard with respect to other trustworthiness projects and standards. In particular, for the Oversight trustworthiness characteristics, three standards are mentioned: ISO/IEC TS 12792, ISO/IEC TS 8200, and ISO/IEC TS 6254, each relating to one or more of the four aspects listed. Concerning Transparency, an analysis of how ISO/IEC TS 12792 relates to Article 13 on Transparency is available in the

previous subsection. In addition to those considerations, we note that ISO/IEC 12792 correctly acknowledges the importance of transparency in enabling other AI properties (such as human oversight).

Concerning monitorability and intervenability/controllability, the PWI on Trustworthiness correctly recognises them as linked to Article 14. Indeed, according to the article, the human overseeing the system must be able to override or reverse the output of the system and intervene on its operation. Intervenability and Controllability are two of the dimensions of quality to be taken into consideration when assessing an AI system in the ISO/IEC DIS 25059 "Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems". However, ISO/IEC 25059 is not a prescriptive document, and does not detail how to achieve these goals. Controllability is discussed in more prescriptive terms in ISO/IEC TS 8200 "Information technology — Artificial intelligence — Controllability of automated artificial intelligence systems". This standard covers relevant areas of AI system control such as state observability and state transition, control transfer process and cost, reaction to uncertainty during control transfer, and verification and validation approaches. While this standard is useful to define the technical prerequisites necessary to enable control over certain AI systems, it is an early draft with a very specific focus, and as such it is not expected to cover the broad set of approaches available for human oversight and control of AI systems, as outlined below.

Finally, concerning explainability and interpretability, the PWI mentions ISO/IEC TS 6254 "Information technology — Artificial intelligence — Objectives and approaches for explainability of ML models and AI systems". A previous JRC report already provides preliminary considerations on this standard (Soler Garrido, et al., 2023). ISO/IEC 6254 describes approaches and methods that can be used to achieve the explainability objectives of different stakeholders concerning the AI system's behaviour output and results. It identifies several characteristics of explainability (explanation needs, form, approaches, and technical constraints) and uses them to categorise many existing eXplainable AI (XAI) approaches. In this regard, it is important to note the tools listed in ISO/IEC TS 6254 are not necessary to ensure Human Oversight as described in Article 14. Importantly, the AI Act does not explicitly call for specific solutions such as XAI or interpretable/transparent-by-design models. Furthermore, it should be noted that many of these technical solutions currently suffer from a number of limitations, e.g., lack of reliability and robustness, lack of a standard framework for explanations evaluation, and an underexplored relationship between AI explanations and automation bias. For this reason, harmonised standards should be broad and seek alternative solutions that are tried-and-tested and effective. An upcoming study (Panigutti, et al., 2023) provides some examples of technical and non-technical measures for human oversight, such as:

- **Training of the human overseeing the system**. The provider of the AI system should specify the type of training necessary to ensure human oversight and prevent automation bias. This should include at least training on the characteristics, capabilities and limitations of the AI system, its performance on different subgroups of individuals. To increase its effectiveness, training could include interactive training materials and scenarios-based training. Scenarios should include "edge cases" or "failure scenarios" in which the AI system is not performing as expected but the failure is not immediately obvious. Furthermore, the human overseeing the system shall be made aware of the eventuality of automation bias (tendency to over-rely on automation) and a proper training on the topic shall be carried out before the individual is assigned to the oversight role. Harmonised standards should provide guidance on the type of training that the human overseeing the system should have before operating the system.
- **Appropriate design of human-AI interfaces.** The AI provider should design the human-AI interface in a way that makes it easy for a trained professional to understand the system's output. For example, the design of the interface should take into consideration cognitive ergonomics principles and be customised to accommodate different operators' needs, skills and training. For example, the interface should use simple and consistent terminology and format. It is also important that the interface to explicit signal whenever the AI system is uncertain (e.g., confidence scores) or unable to provide an answer (e.g., when a model receives input data that is significantly different from the data it was trained on). Furthermore, to prevent automation bias, it is important to design the human-AI interface in a way that encourages cognitive involvement in the task. This can be achieved by, for example, avoiding simple "yes" or "no" outputs and instead providing a range of outputs along with their probabilities, or by presenting an output that requires understanding and processing by a human before any decision is made. Harmonized standards should provide guidance on how to design the human-AI interface of high-risk AI systems in this way.
- **Organisational measures**. The AI system provider should specify which organisational measures are the most appropriate for the system in use. Organisational measures for human oversight might include

implementing redundancy and avoiding single points of oversight failure, as well as proper assignment of roles and responsibilities for the individuals overseeing the system. An example for cybersecurity would be the adoption of ISO/IEC 27001 on an information security system, which defines a range of security roles that employees of an organisation (e.g. the provider) should fill in order to comply with the standard and ensure a resilient information security management. ISO/IEC 42001 also includes considerations related to the assignment of roles and responsibilities to ensure that the AI management system conforms to the requirements, however it does not specify the precise roles and responsibilities that should be assigned. Harmonised standards should provide guidance on organisational measures to ensure that the humans overseeing the system are aware of and accountable for their roles in monitoring and intervening in the AI system as needed.

- **Ensure controllability and intervenability.** AI providers should provide sufficient mechanisms so that the human overseeing the system can easily understand and control it. For example, by designing the human-machine interface to frame and limit the foreseeable misuses of the system. In this context, a good practice to follow could be the co-design of the user interface with stakeholders is (e.g., focus groups, mock-up co-design, user testing, and usability tests in operational settings). Harmonized standards should provide guidance on identifying the proper mechanisms for controllability and intervenability.

2.6 **Accuracy and Robustness**

The first two angles reflected in the technical requirements listed in Article 15 are accuracy and robustness, which can be considered classical technical elements to ensure that a system fulfils its design goals. They are also discussed in more detail in recital 50, and are covered in two sections in the draft standardisation request. It should be noted that the term accuracy is not used in the narrow sense of statistical accuracy, as may be the case in the more technical literature. In our context, and as clarified in the standardisation request, accuracy refers to the capability of the AI system to perform the task for which it has been designed. In this sense, accuracy standards are expected to support –and guide the selection of– a broad range of possible metrics for evaluating the performance of AI systems.

The main elements that can be drawn from these sources are:

- High risk AI systems should be designed and developed in such a way that they achieve, in light of their intended purpose, an appropriate level of accuracy and robustness, and that they perform consistently in those respects throughout their lifecycle.
- Providers should declare relevant accuracy and robustness metrics and levels in the accompanying instructions of use
- Concerning robustness, high-risk AI systems shall be resilient as regards to errors, faults or inconsistencies that occur within the system or the environment in which the system operates, in particular due to their interaction with natural persons or other systems.
- The robustness of high-risk AI systems may be achieved through technical redundancy solutions, including backup or fail-safe plans.

These considerations extend to AI systems which continue to learn after being placed on the market or put into service, notably in respect to feedback loops.

Coverage in existing ISO/IEC standards of the technical aspects of these requirements on accuracy and robustness is limited. Both are handled on a high level in overview standards such as ISO/IEC 22989 on concepts and terminology, or ISO/IEC 24028 on an overview of trustworthiness characteristics. Most notably, robustness is covered by the ISO/IEC 24029 series. ISO/IEC 24029-1 on robustness assessment for neural networks gives an overview of metrics and measures currently available to assess the robustness of neural networks, including available statistical and empirical measures of robustness. It is a relevant document for the AI Act, but it presents some limitations in terms of technical coverage. An important one is that it appears to be mostly applicable to classical AI applications, such as classification and regression, excluding more recent approaches such as generative models, or more recent classes of large, general purpose AI systems. In addition, adversarial robustness is not prominently covered in this report, making the document less applicable for AI cybersecurity requirements. Finally, it is a technical report, and therefore not a prescriptive document. Nevertheless, it provides solid guidance, emphasising that it is currently limited to statistical and empirical approaches, which are impacting robustness testing of AI systems compared to conventional software. Considering AI Act requirements, this specification should be complemented with guidance on how to set acceptable thresholds for the different statistical and empirical robustness methods, considering the context of use and risks of

specific AI systems. These can later be complemented with specific thresholds, although this is likely to be a subject of vertical standardisation as it heavily depends on the particular application considered.

Part 2 of this series, ISO/IEC 24029-2, is a technical specification in development that covers formal robustness verification methods. Despite the narrow scope of the topic, formal verification techniques are relevant for the AI Act. They are the only set of techniques that can in theory provide strict guarantees in terms of robustness of an AI system, and establish strong links to techniques used in classical software testing. However, in their current state of development, these techniques have limitations in terms of their applicability, in particular regarding their scalability to larger neural network architectures.

The newly planned JTC 21 work item on trustworthiness is particularly relevant to address some of the standardisation needs outlined. According to the initial proposal reviewed, both accuracy and robustness are in scope. The elements already captured on robustness are particularly promising. The set of robustness criteria, including stability, sensitivity, relevance and reachability, seems appropriate. Even if it is not a breakdown of robustness seen in the conventional Machine Learning literature, many of the elements covered are well-known features of (AI) system robustness. In general, the scope of the new JTC21 work item makes it a more general reference than similar ISO/IEC work, and it certainly has the potential to cover more technical ground than, for example, 24029-1, in addition to being a more prescriptive document setting out concrete technical requirements.

There are other ongoing standardisation work items at the international stage covering validation, verification and testing aspects, which could in the future be useful references for accuracy and robustness. Two relevant documents are ISO/IEC AWI TS 29119-11 on the testing of AI systems and ISO/IEC AWI TS 17847 on verification and validation analysis of AI systems. In a more advanced stage is ISO/IEC DIS 25059 on a quality software model for AI systems. They could contribute to the definition of suitable accuracy and robustness metrics and testing methods. In general terms, standards currently in an advanced development stage provide limited coverage for either accuracy or robustness. The new JTC21 work item on trustworthiness is therefore a promising development, and we can already identify important technical elements that should be considered in its development.

The discussion on technical limits and possible approaches for measuring accuracy and robustness of AI systems is expected to go beyond statistical and empirical measures for supervised machine learning in either classification or regression applications. Important additional considerations not only include verification methods and theoretical limits, but also accuracy and robustness metrics for generative, unsupervised and large-scale models, which are so far not covered in the reviewed documents and which, while being part of a constantly evolving state-of-the-art, should be considered by standards. Similarly important are methods and approaches to define thresholds and acceptable levels for accuracy and robustness according to the context of use and risks posed by AI systems.

Another important element is how to conduct conformity testing, including verification, validation and auditing of accuracy and robustness for machine learning models, connecting to the change of paradigm that the statistical nature of AI brings, compared for classical software testing methodologies. Examples include the test oracle problem, or missing test coverage criteria for complex neural network models. This is further discussed in the context of conformity assessment in section 2.9. Finally, addressing the accuracy, robustness and validation testing challenges of continuously learning systems (including “feedback loops” in the AI Act description, Article 15) is crucial, as techniques may be used that allow AI models to keep learning after deployment in a self-regulated manner.

In summary, some important considerations for the future development of accuracy and robustness standards include:

- **Metrics and thresholds.** The question of how to select and justify metrics and their thresholds is crucial, and standards should set clear demands for providers of high risk AI systems in line with the accuracy and robustness requirements in Article 15 of the AI Act.
- **Design considerations.** Regarding the scope of new JTC21 work on trustworthiness, the question arises whether it focuses mostly on testing and measuring of these trustworthiness characteristics, or also include guidance on design methods and practices for accuracy and robustness of AI models. These are also in scope of Article 15 of the AI Act and should be covered in standards supporting it.
- **Coverage of recent AI models and architectures.** Standardisers should also consider how the current ISO/IEC technical reports and specifications on measuring robustness, such as ISO/IEC 24029-1 and

ISO/IEC 24029-2, could be complemented in order to cover unsupervised, generative and large models, including language models.

Some preliminary standardization work items initiated within JTC21, e.g. on the accuracy of NLP systems under Working Group 3, and others, are particularly encouraging as they appear to specifically target some of these gaps. These new activities will be covered by future analysis.

2.7 **Cybersecurity**

The third angle reflected in the technical requirements listed in Article 15 is cybersecurity. It is also discussed in more detail in recital 51, which in turn has informed, to a significant extent, the proposed formulation for cybersecurity in the draft standardisation request. The main elements drawn from these sources are:

- High risks AI systems should be ensured and designed to be resilient against attempts to alter their use, behaviour and performance; and to compromise their security properties by malicious third parties exploiting the AI systems' vulnerabilities.
- Organisational and technical solutions shall be implemented to address these goals.
- A cybersecurity risk assessment shall be done for high regulatory-risk AI systems. Note the two different levels of risk: cybersecurity risk vs. regulatory risk from the AI Act.
- Technical solutions shall be appropriate to the relevant circumstances and risks.

Organisational and technical solutions shall thus include, where appropriate, measures to prevent and control cyberattacks against AI assets such as data, models, other digital assets or the underlying ICT infrastructure. Adversarial examples and data poisoning are specifically mentioned, but it should be assumed that the generally mentioned cyberattacks could include other AI-specific attacks targeting AI assets such as AI-backdoors, model inversion or membership attacks, which also touch upon data protection and privacy, and connect thus heavily with Article 10.

In general, we find that coverage of cybersecurity in standards in the current work plan is limited, and cybersecurity is not prominently covered in the scope of JTC21 new work items. In particular, new work items defined so far provide no coverage of AI-specific cybersecurity topics, such as handling and threat modelling adversarial attacks, measuring adversarial robustness or hardening of AI models.

The main reference considered in the current list is ISO/IEC 27001. This is a high-level standard defining how to set up an information security management system. It is useful and well aligned with the AI Act, in that it provides a risk-based approach for cybersecurity, and it is widely adopted. It focuses on organisational aspects, which are in principle applicable to AI products in so far as they are software products. Many classical cybersecurity requirements, threats, vulnerabilities and security controls are applicable to AI systems. More details on this and other cybersecurity standards can be found in a recent ENISA report (Bezombes, Brunessaux, & Cadzow, 2023). It should be noted that the 27001 standard is only useful together with ISO 27002, which lists the possible security controls to be implemented. However, ISO/IEC 27001/2 cannot provide full coverage when considering AI-specific threats, as AI is not explicitly addressed in these documents. Therefore, an important question is how to cover AI-specific cybersecurity concerns in standardisation.

Regarding the new CEN-CENELEC work item(s) on risk, currently in definition, its scope so far does not contain specific elements on cybersecurity risks, but those could be added to the list of risk checks to be developed. Similarly, regarding the JTC21 work item on trustworthiness being planned, cybersecurity is prominently missing from the list of trustworthiness characteristics. Therefore, current coverage of AI cybersecurity in existing standards or in the scope of JTC21 work packages is limited. A range of elements are not currently addressed and would need to be either included in standards under development or covered in new standardisation work at international or European level, for example from JTC21 or other relevant committees such as JTC13.

Notable areas requiring dedicated coverage include threat modelling of AI-specific cyberattacks and their management, how to measure adversarial robustness of AI models against such attacks, increasing the resilience of AI models against cyberattacks and the development of AI-specific security controls. Some of these topics touch upon questions of active research, requiring standardisers to identify techniques that are generally accepted and mature for technical specifications. Additional important aspects that need coverage are taxonomies and terminology at the intersection of AI and cybersecurity, AI cybersecurity risk assessment considerations, including the adaptation of classical security controls to cover security needs of AI assets, and new supply chain security risks for AI systems connected to dataset provenance and the proliferation of pre-trained models, AI libraries and foundational AI services. Most of these are also covered in classical cybersecurity

and standards such as the ISO 27000 series, but are in need of tailoring in order to handle the challenges of modern AI models.

In summary, concrete considerations and open points going forward include.

- **Clarify the role of the ISO/IEC 27000 series.** The effectiveness of ISO/IEC 27001 depends on other standards such as ISO/IEC 27002 with the list of security controls to be implemented.
- **Consider the interplay of management standards.** In particular, the integration of ISO/IEC 27001 with other management standards in the work plan. This includes ISO/IEC 42001 and new European standardisation work to be undertaken on the area of AI risk management.
- **Increase coverage of AI-specific cybersecurity aspects**. Currently considered standards don't focus on AI specific cybersecurity issues or mitigation measures. A plan should be in place to complement standards currently in the work plan with AI cybersecurity considerations.
- **Consider ongoing international standardisation on AI cybersecurity.** Connected to the previous point, standards under development, notably ISO/IEC 27090, have the potential to provide relevant AI cybersecurity guidance. European contributions should ensure alignment with AI Act standardisation needs, in particular to include coverage of mitigation measures and security controls against cyberattacks that may contribute to the specific AI risks considered in the AI act.

2.8 Quality Management

Article 17 of the AI Act requires a quality management system to be put in place and be fully documented. Such a quality management system shall be proportionate to the organisation's size and shall take into account the *AI ecosystem* by providing a strategy for regulatory compliance, technical specifications, communication handling with national authorities and by establishing an accountability framework. The quality system shall additionally encompass *techniques and procedures* as well as *systems* for a wide range of related activities. Specifically, compliance with Article 17 requires coverage of:

- How to take into account the high-risk AI ecosystem by defining how to set up:
    - Strategy for regulatory compliance
    - Technical specifications to be applied, including standards
    - Communication handling with national authorities
    - An accountability framework
- Techniques and procedures for:
    - AI design, design verification and development.
    - AI examination, testing and validation
    - Data management
    - Quality control and assurance
    - Reporting of serious incidents and of malfunctioning
    - Record keeping of all relevant documentation and information
    - Conformity assessment
    - Management of AI modifications
- Systems for:
    - Risk management
    - Post-marketing monitoring
    - Resource management

Naturally, a single standard cannot provide the necessary technical specification to be compliant with all these requirements. Indeed, one would expect the implementation details on the risk management system to be covered by a dedicated standard and the standard covering the overall quality management system

requirements referencing it when specifying the need for a risk management system to be put in place. Similarly, one would expect dedicated standards on data management and conformity assessment to provide additional technical specifications.

Nonetheless, a standard establishing a quality system considering all the aforementioned aspects –integrating them into a systematic quality management system that is documented– is needed to ensure compliance with the AIA, enabling conformity assessment as well as quality monitoring post market placement.

ISO/IEC 42001 draft standard for AI management provides considerable overlap with Article 17 needs. The most relevant content in the context of Article 17 can be found in the normative Annex B, containing implementations guidelines for AI controls. However, the aim of this standard is to provide as much freedom as possible to organisations when selecting the necessary controls, and this approach is not fully compatible with regulatory requirements such as those laid down in the AI Act. In short, under the current ISO/IEC 42001 specification, organisations are free to judge which of the controls are most relevant and hence free to choose which controls to adopt, and the standard does not provide comprehensive requirements on the set of acceptable justifications for the choice of controls, nor establishes a link to regulatory requirements.

Indeed, the need and form of justifications come as a soft requirement in the reviewed version of ISO/IEC 42001. Since most of its coverage of Article 17 is provided by the normative Annex B, the choice of controls and the criteria for selecting and implementing them is of paramount importance, as regulatory compliance would ultimately rely on it. JTC21 should take this point into consideration when adopting ISO/IEC 42001 as a European Norm, or in the harmonisation process, in order to ensure the necessary controls for regulatory compliance are mandatory, and that the justification provided for both inclusion and exclusion of controls is in line with the AI Act, ensuring that the text is more specific on what constitutes an acceptable justification.

Regarding the coverage of ISO/IEC 42001 of the requirements laid out in Article 17, there is a wide content overlap, yet the standard does not always include enough technical implementation details. Most of the implementation detail for the controls is in the form of guidance, without any requirements. Therefore, this standard needs to be complemented with other specifications that further detail methodologies and implementation requirements, for example covering AI design considerations, quality control and assurance, testing or data management.

In summary, there are important gaps that deserve particular attention by JTC21 when adopting ISO/IEC 42001 or in the harmonisation process for Article 17 of the AI Act:

- **Stronger justification**: justification shall be provided for both inclusion and exclusion of controls. Further specification on what constitutes an acceptable justification is needed. In this sense, it is possible that some controls will have to be made mandatory for compliance with the AI Act.
- **Post-marketing monitoring system coverage**: more detailed requirements on what to monitor for during operation tailored to AIA concerns and risks is needed, as ISO/IEC 42001 currently only demands a minimum of system and performance monitoring, repairs, updates and support. Current ISO/IEC 42001 considerations should be complemented with specific mechanisms to monitor the potential negative impact of the operation of the AI system on individuals and society. Suitable measures are also required in order to identify risks missed during initial risk assessment and early AI lifecycle stages, and if relevant, to identify risks in the context of continuous learning systems.
- **Risk management system**: the assessment of the impact of the AI system on individuals should be part of the risk assessment and gain more prominence. Further additional details on its implementation as part of the risk management process should be provided. Additionally, the risk assessment shall explicitly consider impact to individuals in terms of their health, safety or fundamental rights to be in line with the AI Act. ISO/IEC 42005 could become a useful reference in this context.
- **Management of AI modifications**: Specify in more detail the actions required to manage modifications to the AI system, in particular specifying requirements for the continuous assessment and management of risks whenever required due to significant changes to the AI system.
- **Documentation needs**: ISO/IEC 42001 demands multiple items of documentation to be produced, and more clarity and consolidation of documentation needs is required to simplify the application of this standard.

2.9 **Conformity Assessment**

The Conformity Assessment process in the context of the AI Act (AIA) involves verifying that the requirements for high-risk AI systems, outlined in Title III, Chapter 2 of the AIA, have been met. This can be done, depending on the purpose of the AI system, either by the provider of the AI system, following the procedure described in Annex VI, or with the involvement of a third-party Conformity Assessment Body (CAB) or a Notified Body, following the procedure described in Annex VII. The conformity assessment process involves assessing the Quality Management System (QMS) against the requirements in Article 17, examining the technical documentation (Article 11 and Annex IV of the AIA) to determine compliance with the AIA requirements and verifying that both the design and development process of the AI system and its post-market monitoring (as described in Article 61) are consistent with the technical documentation. In the case of a third-party conformity assessment, Annex VII specifies that the Notified Body can request further evidence or testing from the provider, and, if deemed necessary, it can conduct its own tests requesting access to the training and testing datasets and source code of the AI. According to the AIA, the conformity assessment should be carried out prior to the placement on the market or putting into service of the high-risk AI system. However, a new assessment should also be conducted whenever the system is substantially modified or if the system's intended purpose changes.

ISO/IEC JTC 1/SC 42 is currently working on a standard that specify how to carry out conformity assessment for AI management systems according to ISO/IEC 42001, as well as competencies needed for auditors: the ISO/IEC 42006 "Requirements for bodies providing audit and certification of artificial intelligence management systems". While this standard is in a very early stage of development, it could be a relevant reference for harmonisation, considering that ISO/IEC 42001 provides some of the elements required for the Quality Management System defined in the AI Act.

In addition to ISO/IEC 42006, CEN-CENELEC JTC 21 is currently working on a technical report on conformity assessment, providing a landscape that includes the ISO/IEC 17000 series (the ISO CASCO toolbox) and discussing candidates for harmonisation. In this regard, while useful, the ISO CASCO toolbox provides a generic framework for accreditation schemes that needs complementation with additional requirements in order to be actionable for high-risk AI systems.

Furthermore, in addition to auditing the AI management system, standardisers should also consider the need for specifications covering conformity testing of the AI systems themselves. Additional standardisation activities could focus on two key aspects: (i) tools and procedures to perform conformity assessment of AI systems effectively, and (ii) competencies of CABs.

Tools and procedures. Annexes VI and VII provide high-level guidance on conformity assessment procedures that can be further specified at the technical level in standards. For example, these could specify the conditions under which a CAB requires additional testing from the provider of an AI system, or chooses to carry out its own tests on the AI system, i.e., conditions for more in-depth auditing of AI systems during conformity assessment. Standards could also specify the tools, methodologies and metrics that CABs should use when performing such tests, as well as auditing report templates, and the best practices data access and testing in these scenarios.

Competencies of CABs. Harmonised standards should also cover the specific AI competencies that CABs need for assessment of high-risk AI systems. CABs should have AI-specific expertise to be able to assess whether a system has been adequately tested or to perform the tests themselves. This includes knowledge of AI system design and development process, as well as the socio-technical considerations involved in the use of AI systems. The CAB should also be able to identify and evaluate any potential risks to health, safety, or fundamental rights that may arise when the system is used in accordance with its intended purpose or under conditions of foreseeable misuse.

The following are peculiarities of AI systems that ESOs should envisage to take into consideration within standards for Conformity Assessment:

- **Underspecification**. Conformity assessment of AI systems is often challenging due to underspecification, i.e., lack of clear specifications for the system. In the context of AI systems, this can happen for two reasons: (i) the inherent data-driven nature of AI (the optimal model behaviour is automatically learned from data rather than being specified by developers) (Black, et al., 2022), and (ii) as a consequence of current design and documentation practices not being thorough and solid (D'Amour, et al., 2020). In both cases, underspecification can result in unpredictable behaviour in real-world settings, as the AI pipeline does not fully specify the system's expected behaviour. Harmonised standards should focus on identifying tools and

methodologies to appropriately check the conformance of inherently underspecified systems (first scenario) and identify system's design and documentation aspects related to underspecification that are relevant to conformity checks (second scenario). Finally, harmonised standards should define the relevant competencies required by CABs to ensure they can effectively perform conformity assessment of such underspecified systems.

- **Systems complexity**. Many AI systems, particularly those based on deep learning, and more recent general purpose and generative AI models, are highly complex and operate as black boxes, meaning it is difficult to understand how they work internally. This problem is exacerbated by the fact that production environments might be characterized by many of these black box models integrated with other systems and processes (Black, et al., 2022; Raji, et al., 2020). Complexity and opacity make it challenging to assess whether such systems meet the requirements of the AIA. CABs should have suitable competencies to perform such an evaluation.
- **Data complexity**. Data for training and testing AI systems usually come from various sources. Such a mix might make it difficult to isolate each source's specific contribution to the AI model's final output (Raji, et al., 2020). Furthermore, the wide range of possible input for AI systems can make testing the system in all possible input scenarios challenging (Black, et al., 2022). CABs should have AI-specific competencies to consider the interactions between different data sources and the potential for a wide range of input.
- **Fast and iterative development**. AI development often takes place in a fast and iterative manner, posing challenges to conformity assessment processes, making it difficult to keep up with their evolution and to achieve a comprehensive assessment (Raji, et al., 2020). The iterative nature of AI development also means AI systems are continually modified and updated, making it difficult to apply conformity assessment procedures consistently. Harmonised standards should identify tools and procedures to ensure that conformity assessment is able to determine whether an AI system continues to meet relevant AIA requirements over time.
- **Concept/Data drift**. Another related problem is that of data drift, i.e., when the data used to train the AI model becomes outdated for real-world use (Black, et al., 2022). This might happen for a variety of reasons, including unexpected real-world events changing the data generation process and creating a mismatch between the relationship between input and output learned by the AI model and the real-world relationship. This phenomenon might hinder validity of conformity checks as the AI system might start behaving unexpectedly and potentially cause harm. Harmonised standards need to identify suitable methods for assessing the effectiveness of post-marketing monitoring mechanisms (Mökander, Axente, Casolari, & Floridi, 2022).
- **Socio-technical systems**. To effectively assess conformity with some of the AIA requirements (e.g., Articles 9 and 10), CABs should have knowledge of socio-technical systems, which includes the social and cultural context in which the AI system will be used, as well as the technical aspects of the system itself. For example, CABs should have an understanding of the potential risks and impact of AI systems on society. CABs should also have knowledge of the statistical properties that the data sets should have in order to be suitable for training and testing of each specific high-risk AI system, as well as any element that needs to be considered when testing AI systems for use in specific geographical, behavioural, or functional settings. Harmonised standards should provide guidance on the knowledge and expertise that CABs need to properly assess the conformity of AI systems in socio-technical contexts.
- **Complex human-machine interactions**. CABs should have knowledge of how to assess the sufficiency and appropriateness of human oversight in preventing automation bias and enabling correct use and intervention. This might include evaluating the design and development of the AI system to ensure that it includes appropriate human-machine interface tools and can be effectively overseen by natural persons during use. Harmonised standards should specify the knowledge and expertise that CABs should possess to assess the sufficiency and appropriateness of human oversight in preventing automation bias and enabling correct use and intervention.

3 Conclusions

In this technical report, we present a detailed analysis of the preliminary list of AI standards in the work plan by CEN-CENELEC JTC21 as of January 2023. This work plan combines the adoption of international standards from ISO/IEC with the development of new work items by JTC 21, in order to cover European specificities. Our analysis shows that this approach fits AI Act standardisation needs very well. On one side, international standards already provide partial coverage of most of the requirements for high-risk AI systems defined in the legal text. On the other hand, there are important elements where international work is not fully aligned with the requirements of the AI Act—in particular in terms of risks considered— or where coverage is insufficient. Indeed, new standardisation deliverables in support of the AI Act is expected to have a clear focus on addressing potential risks posed by AI systems to the fundamental rights, health and safety of individuals in line with EU values In addition, they are expected to be prescriptive —i.e. preferably in the form of European Norms to be later harmonised— and closely tailored to the essential requirements defined in the legal text. Standardisation in the European context is best placed to develop these additional specifications on time for the application of the AI Act.

References

Balahur-Dobrescu, A., Jenet, A., Hupont Torres, I., Charisi, V., Ganesh, A., Griesinger, C., . . . Tolan, S. (2022). *Data quality requirements for inclusive, non-biased and trustworthy AI.* Publications Office of the European Union.

Bezombes, P., Brunessaux, S., & Cadzow, S. (2023). *Cybersecurity of AI and Standardisation.* ENISA.

Black, R., Davenport, J., Olszewska, J., Rößler, J., Smith, A. L., & Wright, J. (2022). *Artificial Intelligence and Software Testing: Building systems you can trust.* BCS Press.

D'Amour, A., Heller, K., Moldovan, D., Adlam, B., Alipanahi, B., Beutel, A., . . . Hormozdiari, F. (2020). Underspecification presents challenges for credibility in modern machine learning. *Journal of Machine Learning Research*.

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM, 64*(12), 86-92.

Hupont, I., Micheli, M., Delipetrev, B., Gómez, E., & Soler Garrido, J. (2023). *Documenting high-risk AI: a European regulatory perspective.* IEEE Computer.

Mitchell, M. W., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I., & Gebru. (2019). Model cards for model reporting. *Proceedings of the conference on fairness, accountability, and transparency*, (pp. 220--229).

Mökander, J., Axente, M., Casolari, F., & Floridi, L. (2022). Conformity assessments and post-market monitoring: A guide to the role of auditing in the proposed European AI regulation. *Minds and Machines*, 241-268.

Panigutti, C., Hamon, R., Hupont Torres, I., Fernández Llorca, D., Fano Yela, D., Junklewitz, H., . . . Gómez, E. (2023). The role of explainable AI in the context of the AI Act. *ACM FAccT (accepted for publication).*

Raji, I., Smart, A., White, R., Mitchell, M., Gebru, T., Hutchinson, B., . . . Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *Proceedings of the 2020 conference on fairness, accountability, and transparency*, (pp. 33-44).

Soler Garrido, J., Tolan, S., Hupont Torres, I., David, F. L., Charisi, V., Gomez Gutierrez, E., . . . Panigutti, C. (2023). *AI Watch: Artificial Intelligence Standardisation Landscape Update.* Publications Office of the European Union.

List of abbreviations and definitions

AI	Artificial Intelligence
AIA	AI Act
CAB	Conformity Assessment Body
CEN	European Committee for Standardization
CENELEC	European Committee for Electrotechnical Standardization
EN	European Norm
hEN	Harmonised European Norm
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronic Engineers
ISO	International Organization for Standardization
JTC	Joint Technical Committee
NWIP	New Work Item Proposal
PWI	Preliminary Work Item
QMS	Quality Management System
SAG	Special Advisory Group
WG	Working Group
XAI	eXplainable AI

List of tables

Annexes

Annex 1. Comments on ISO/IEC 22989 and ISO/IEC 23053

These are relevant and highly informative foundational and horizontal standards that establish a common terminology and concepts. They can serve as a basis for other standards that cover specific technical aspects of AI systems, including those addressing concrete requirements for high-risk AI systems under the European AI regulation proposal. From the point of view of the European Commission, besides ensuring that the standards introduce the necessary terminology for future European and harmonised standards in support of the regulation, a priority is that they capture the main AI-related terms and concepts used in the AI Act, ensuring that the respective definitions are compatible and aligned with their use in the legal text. In some cases, this may imply capturing the various meanings that certain terms can have depending on whether they are used in a technical document, such as a technical specification, or a legal one, such as the regulation proposal or a standardisation request. These are some concrete remarks related to specific terms used in the AI regulation proposal and also reflected in these standards:

- The definition of Artificial Intelligence in the European regulation is currently subject of discussion by the co-legislators, and is therefore not possible to make conclusive comments. However, the current definition of AI system in ISO/IEC 22989 appears to be generally well aligned with that in the Commission proposal.
- ISO/IEC 22989 provides a broad and generally accepted definition of risk, which could potentially be complemented with the more specific meaning of risk in the context of the European AI regulation, which focuses on the potential negative impacts of certain AI-systems, and notably, for high-risk AI systems, the negative impacts to fundamental rights, health and safety of individuals.
- Similarly, the main definition of bias under ISO/IEC 22989 – a systematic difference in treatment of certain objects, people or groups in comparison to others – is a neutral one. The standard further clarifies that other meanings are possible, including the one commonly found in social contexts, referring to the notion that certain differences in treatment are unfair, but in the standard this is instead associated with the term unfairness. It should be highlighted that this latter meaning appear to be better aligned with the use of the term bias in the legal text, where it is linked to discrimination and other unfair differences in treatment.
- The definition currently used of performance as a measurable result appears broad and not specific to the AI field. Complementing this definition with one for AI system performance in line with the use in the legal text, i.e. the ability of an AI system to achieve its intended purpose, which can include quantitative and qualitative measures, would be beneficial.
- ISO/IEC 22989 provides a comprehensive breakdown of stakeholders, differentiating, for example between AI provider and producer. It should be noted that the legal text uses definitions closely linked to product legislation, where the figure of the provider takes the responsibility for placing the product on the market, regardless of the natural or legal person who designed or developed it. Some of the definitions in ISO/IEC under provider (e.g. platform provider) may not technically be providers under EU product legislation.
- The concept of human oversight of AI systems could be explicitly defined and linked to related terms in the standard covering specific aspects, such as controllability, explainability or external supervision. Similarly, human agency and oversight could be added to the definition of trustworthiness as one of its characteristics.
- Some other terms may be used both in legal documentation with a broader meaning as in technical standards. An example of this would be the term accuracy, which in the AI Act refers to the capability of the AI system to perform the task for which it has been designed, and in ISO/IEC 23053 is given a narrower definition corresponding to statistical accuracy, one of several possible metrics for evaluating the performance of AI system.
- Terminology for some of the requirements in the AI Act may not be sufficiently covered. For instance, there is no mentioning of cybersecurity or AI cybersecurity in the standards documents. They also largely omit any elements from adversarial machine learning including defining the field itself, or the nature of adversarial examples, the possible threat of backdoors, membership inference attacks or model theft etc. This appears to be a gap in terminology with respect to the concepts and terms used in the AI Act.

It should be noted that the above list of relevant terms used in the AI regulation proposal and reflected in these foundational standards is not exhaustive, and it is possible that new relevant terms and concepts will be added to the legal text in during the negotiation process between co-legislators (one example could be the concept of

general-purpose AI systems). In the future, it may be beneficial to also cover these concepts in the foundational standards when adopted in the European context.

Furthermore, in some cases, ISO/IEC 22989 and ISO/IEC 23053 extend their scope beyond establishing terminology and describing concepts, covering aspects that may be related to technical approaches available for AI providers to satisfy certain requirements. These may include risk management or data management approaches, techniques for AI robustness, metrics for evaluation of AI systems. In some cases, concrete standards are mentioned as well, e.g. ISO/IEC 23894 in the discussion on risk management or ISO/IEC 24027 for treatment of unwanted bias. While this content is in general useful, informative, and generally in line with the requirements of the European AI regulation, it is important to ensure that these foundational standards do not pre-empt or contradict the content of standards specific to concrete technical requirements, including those still to be developed in the European context.

Finally, it should be explicitly noted that these standards do not intend to be exhaustive in the AI terminology and concepts introduced, or even in the application areas of AI and machine learning mentioned, as these are constantly evolving. For example, in the case of ISO/IEC 23053, the content appears to focus on basic and better-understood ML approaches for classification and regression. Future, requirement-specific standards, e.g. those covering performance metrics, may need to consider systems performing other types of tasks, e.g. detection, recommendation or content generation tasks which may be relevant in the context of high-risk applications.

Most of the comments above are about complementing and extending the definitions and concepts covered the ISO standards, adding remarks that capture, for example, differences between the AI regulation and technical specifications in the use of certain terminology. Ideally, these remarks would be considered –including any necessary modifications to the standards– when adopting them in the European context, e.g. by including a mapping between the definitions of key terms in the context of the European AI regulation and the more technical definitions provided in ISO/IEC 22989 and ISO/IEC 23053. Alternatively, European Standardisation Organizations may choose to address these comments in the course of the development of new technical specifications.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online (european-union.europa.eu/contact-eu/meet-us_en).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

— by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

— at the following standard number: +32 22999696,

— via the following form: european-union.europa.eu/contact-eu/write-us_en.

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (european-union.europa.eu).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (european-union.europa.eu/contact-eu/meet-us_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (eur-lex.europa.eu).

Open data from the EU

The portal data.europa.eu provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

Science for policy

The Joint Research Centre (JRC) provides independent, evidence-based knowledge and science, supporting EU policies to positively impact society

EU Science Hub
joint-research-centre.ec.europa.eu

@EU_ScienceHub
EU Science Hub - Joint Research Centre
EU Science, Research and Innovation
EU Science Hub
@eu_science

Publications Office
of the European Union