



BSI Standards Publication

Data governance and quality for AI within the European context

National foreword

This Published Document is the UK implementation of CEN/CLC/TR 18115:2024.

The UK participation in its preparation was entrusted to Technical Committee ART/1, Artificial Intelligence.

A list of organizations represented on this committee can be obtained on request to its committee manager.

Contractual and legal considerations

This publication has been prepared in good faith, however no representation, warranty, assurance or undertaking (express or implied) is or will be made, and no responsibility or liability is or will be accepted by BSI in relation to the adequacy, accuracy, completeness or reasonableness of this publication. All and any such responsibility and liability is expressly disclaimed to the full extent permitted by the law.

This publication is provided as is, and is to be used at the recipient's own risk.

The recipient is advised to consider seeking professional guidance with respect to its use of this publication.

This publication is not intended to constitute a contract. Users are responsible for its correct application.

This publication is not to be regarded as a British Standard.

© The British Standards Institution 2024
Published by BSI Standards Limited 2024

ISBN 978 0 539 32048 0

ICS 35.240.01

Compliance with a Published Document cannot confer immunity from legal obligations.

This Published Document was published under the authority of the Standards Policy and Strategy Committee on 30 November 2024.

Amendments/corrigenda issued since publication

Date	Text affected
------	---------------

ICS 35.240.01

English version

Data governance and quality for AI within the European context

Gouvernance et qualité des données pour l'IA dans le
contexte européen

Datenmanagement und -qualität für KI im
europäischen Kontext

This Technical Report was approved by CEN on 30 September 2024. It has been drawn up by the Technical Committee CEN/CLC/JTC 21.

CEN and CENELEC members are the national standards bodies and national electrotechnical committees of Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Republic of North Macedonia, Romania, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Türkiye and United Kingdom.



CEN-CENELEC Management Centre:
Rue de la Science 23, B-1040 Brussels

Contents	Page
European foreword.....	4
Introduction	5
1 Scope	8
2 Normative references	8
3 Terms and definitions.....	8
3.1 General	8
3.2 Data governance	10
3.3 Data quality.....	12
4 Abbreviations.....	14
5 JRC research and data-related standards on AI.....	15
5.1 General	15
5.2 Research: Data quality requirements for inclusive, non-biased and trustworthy AI.....	16
5.3 Data-related standards on AI for data governance and data quality.....	18
5.3.1 General	18
5.3.2 A short description of the standards mentioned in Figure 4 (taken from www.iso.org).	19
6 Data governance	24
7 Data quality.....	35
8 Elements for data, data sets, information for testing and evaluation	45
9 Data governance and data quality for large European contexts	50
9.1 General	50
9.2 Italian government: Strategy program on Artificial Intelligence.....	50
9.3 Italian agency application of data quality model for public administrations	51
9.4 Spanish experience on data Governance: Data Office	52
9.5 European governance relating to the Directive on inclusivity and accessibility	53
10 General considerations on innovative technology: Ethics, Governance, AI Act.....	54
11 Potential challenges.....	56
11.1 General.....	56
11.2 Stakeholders' engagement.....	56
11.3 Contextualization	56
11.4 Critical infrastructures.....	57
11.5 Ethics and regulatory challenges.....	57
11.6 Interoperability	58
11.7 Big volume of data.....	59
12 Best practices from organizations, industries and research activities.....	59
12.1 General	59

12.2	AI in healthcare: the MES-CoBraD approach	59
12.3	Overview of industries that stand out for their approach to data governance	60
	Bibliography.....	62

Figures

Figure 1	— Connections of Legislations, Standards, Guidelines & Monitoring specifications.....	7
Figure 2	— Active organizations mentioned in JRC.....	17
Figure 3	— Standards and Technical reports mentioned in JRC	18
Figure 4	— Clusters of standards, TS, TR data-related	18
Figure 5	— Example of relationships among quality aspects of ISO/IEC 5259-2, ISO/IEC 25059 and AI Act , eliciting new requirements to be harmonized	23
Figure 6	— European legal references and ISO standards for AI on data quality (or complementary)	26
Figure 7	— Data governance framework.....	27
Figure 8	— Data governance flow at European level.....	30
Figure 9	— Data managing integration and synthesis of experiences	32
Figure 10	— Data Governance summary	34
Figure 11	— Data Quality Measures and Data Life Cycle Model	36
Figure 12	— Relationship among quality models, characteristics, QM, QME, property, target entity	38
Figure 13	— Data life cycle framework	44
Figure 14	— Example of conceptual perspective visualization of data testing and evaluation...47	
Figure 15	— Visualization of elements for governance of data, data sets, testing.....47	
Figure 16	— Example of ontological contextual schema of elements resulting in the conference online held in October 2020 with 100 speakers [13].....	49

Tables

Table 1	— Main documents considered for data governance framework.....	24
Table 2	— Type of governance and multi-level point of view	28
Table 3	— Characteristics of the data quality model adapted from ISO/IEC 25012	37
Table 4	— Characteristics of data quality models from ISO 8000, ISO/IEC 25012 and ISO/IEC 5259-2	41

European foreword

This document (CEN/CLC/TR 18115:2024) has been prepared by Technical Committee CEN/JTC 21 “Artificial Intelligence”, the secretariat of which is held by DS.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. CEN shall not be held responsible for identifying any or all such patent rights.

Any feedback and questions on this document should be directed to the users’ national standards body. A complete listing of these bodies can be found on the CEN website.

Introduction

This document aims to provide an overview of the relevant regulations in the European context and connected international standards, paying particular attention to data governance and data quality topics. Relevant regulations considered are:

- “Council of Europe” Ad hoc Committee on AI (CAI) that produced “Recommendation CM/Rec (2020) of the Committee of Ministers to member States on the human rights impact of algorithmic systems” and the deliverable “possible elements of a legal framework on Artificial Intelligence, based on the Council of Europe’s standards on human rights, democracy and the rules of law” (2021) [1].
- “European strategy for data” (2020), which is essential to govern new technologies and create business opportunities.
- “Artificial Intelligence Act” (2024), which aims to ensure that AI systems placed on the market and used in the EU are safe and respect fundamental rights. Attention is given specifically to:
 - Article 10 “Data and data governance” describing the quality criteria specifying aspects of training, validation and testing of data sets.
 - Article 15 “Accuracy, robustness, and cybersecurity” describing essential quality characteristics that can be extended to a general data quality model; consistency between terms and definitions is a common goal of this document, as well as of future TS and EN standards.
 - Articles where standard quality characteristics are mentioned (see Figure 5).
- “Data Governance Act” (2022) providing a framework aiming:
 - to increase trust in data sharing across areas;
 - to develop common European data spaces in strategic domains (e.g. health, environment, energy, agriculture, mobility, finance, manufacturing, public administration);
 - to strengthen mechanism to increase data availability and overcome technical obstacles to the reuse of data.
- “Data Act” (2023): key elements include the reinforced data portability and data sharing, rules governing the processing data shared, model contracts, access and use data held by private companies, data and cloud interoperability, databases containing data from IoT, restriction on data sharing.
- “Open data Directive” (EU 2019/1024): provides common rules for a European market for government-held data, including the re-use of public sector information.

In addition, Regulation (EU) 2016/679 of the European Parliament and the Council on the protection of natural persons about the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC – GDPR, it is also considered in this document. The General Data Protection Regulation – GDPR, entered into force on May 2016, creates a harmonized set of rules applicable to processing of all European personal data. The objective of GDPR is to ensure that personal data enjoys a high standard of protection everywhere in the EU, increasing legal certainty for both individuals and organizations proceeding data, and offering a higher degree of protection for individuals and their fundamental rights. According to ISO/IEC 22989 types of organizations are e.g. commercial enterprises, governments agencies, not-for-profit organizations. The objective of GDPR is to provide a

consistent and high level of protection of natural persons regarding the processing of personal data and the free movement of such data and to remove the obstacles to the flow of personal data within the Union. In addition, GDPR ensures a common level of protection of the rights and freedoms of natural persons concerning the processing of such data all over the Member States, increasing legal certainty for both individuals and organizations proceeding with data and offering a higher degree of protection for individuals and their fundamental rights.

GDPR takes into consideration also the activity of processing personal data by Artificial Intelligence systems (see processing reported in 3.2.10), as we will see explaining characteristics of data quality containing specific requirements on this topic strongly related to some principles of GDPR and as can also be seen in some documents of the Council of Europe COE [1].

Another important aspect of quality underlined in this document it is related to accessibility for disabled users. In this case also we will describe the concepts explaining characteristics of data quality the value of accessibility, and understandability of data. The accessibility quality characteristic related to a European legislative regulation is a good example of governance of data that are obtained with a global vision by monitoring the activities in progress in each Country. A similar approach of governance, global and local, can be extended in the future to the large applications of AI, developing specific EN Standards or Technical specification.

Finally, some considerations on ethics are reported to reinforce some aspects related to data use.

The European Commission and the Member States put forward a 'Coordinated Plan on Artificial Intelligence' - COM (2018) 795 - with the stated goal of maximizing AI investments impact both at European and national levels and strengthening synergies and cooperation among Member States. To this end, Member States were strongly encouraged to develop their own national AI strategies (e.g. with Guidelines and monitoring specifications) to achieve these aims, in conformance with laws.

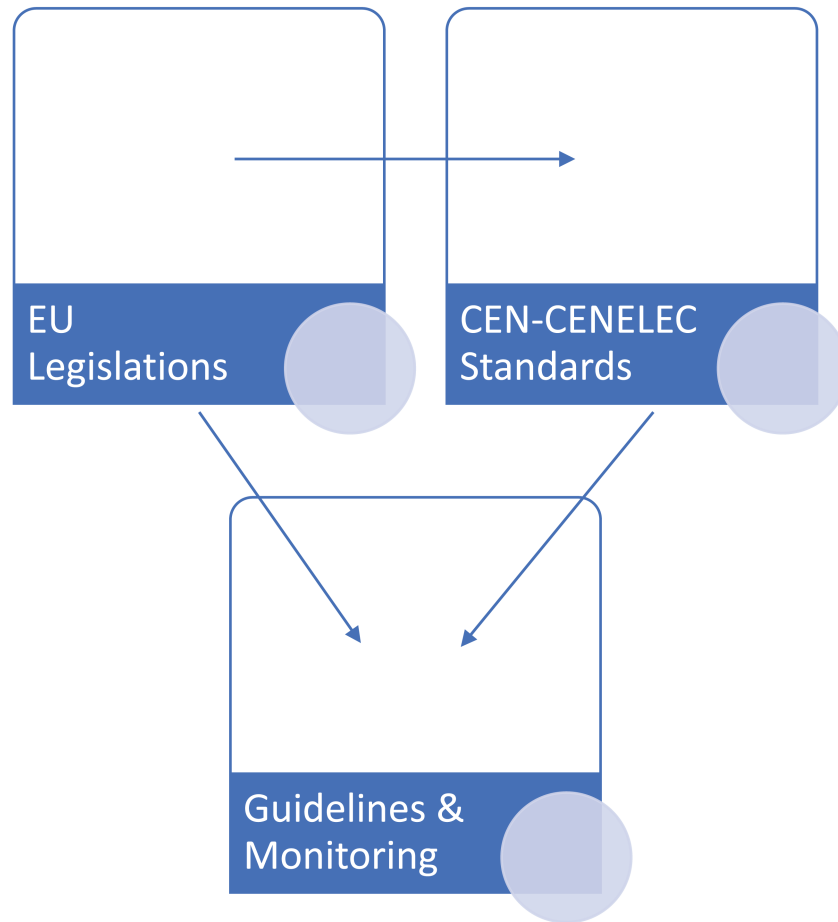


Figure 1 — Connections of Legislations, Standards, Guidelines & Monitoring specifications

EU AI Act and CEN-CENELEC JTC21 are harmonizing legislations and Standards. Guidelines & Monitoring can be developed by Member States / Companies: examples are quoted in Clause 9 and 12 of this TR. Following these perspectives, the goal of this document is promoting a complement to the overview of a common terminology and language on Artificial Intelligence to facilitate innovation, communications, coordination, planning and agreements between European countries, national visions, enterprises, projects and products realization oriented to quality and mitigating risks. For innovation management the approach taken in the ISO 56000 family can be considered. For social motivation and responsibility, ISO 26000 can contribute to sustain the inclusiveness and ethics principles.

1 Scope

This document provides an overview on AI-related standards, with a focus on data and data life cycles, to organizations, agencies, enterprises, developers, universities, researchers, focus groups, users, and other stakeholders that are experiencing this era of digital transformation.

It describes links among the many international standards and regulations published or under development, with the aim of promoting a common language, a greater culture of quality, giving an information framework.

It addresses the following areas:

- data governance;
- data quality;
- elements for data, data sets properties to provide unbiased evaluation and information for testing.

2 Normative references

There are no normative references in this document.

NOTE For the application of this document: users and stakeholders can apply the standards listed depending on their context of use and in compliance with the laws.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

- ISO Online browsing platform: available at <https://www.iso.org/obp/>
- IEC Electropedia: available at <https://www.electropedia.org/>

Note 1 to entry: Terms and definitions have been divided into General, Data Governance and Data quality.

3.1 General

3.1.1

Artificial Intelligence

AI

research and development of mechanisms and applications of AI systems

Note 1 to entry: Research and development can take place across any number of fields such as computer science, data science, humanities, mathematics, and natural sciences

[SOURCE: ISO/IEC 22989:2022¹]

¹ ISO/IEC 22989:2022/AMD1 is under development.

3.1.2**AI system**

engineered system that generates outputs such as content, forecasts, recommendations, or decisions for a given set of human-defined objectives

Note 1 to entry: The engineered system can use various techniques and approaches related to artificial intelligence to develop a model to represent data, knowledge, processes, etc. which can be used to conduct tasks.

Note 2 to entry: AI systems are designed to operate with varying levels of automation.

[SOURCE: ISO/IEC 22989:2022²]

3.1.3**element**

smaller part of an architecture

EXAMPLES records, fields, format, metadata, images, etc.

[SOURCE: ISO/IEC 25024:2015; in ISO/IEC 25024:2015, 4.19, the term is used with reference to the architecture of data and to computer program domain such as data model or data dictionary.]

3.1.4**framework**

reusable design (models or code) that can be refined (specialized) and extended to provide some portion of the overall functionality of many applications

[SOURCE: IEEE 1320.2-1998 (R2004)]

3.1.7**life cycle**

evolution of a system, product, service, project or other human-made entity, from conception through retirement

[SOURCE: ISO/IEC 22989:2022; ISO/IEC/IEEE 15288:2023]

3.1.8**measure**

variable to which a value is assigned as the result of measurement

Note 1 to entry: the term measure is used to refer collectively to base measures, derived measures and indicators.

[SOURCE: ISO/IEC 25024:2015, 4.26, ISO/IEC 25010:2011, 4.4.5, ISO/IEC/IEEE 15939:2017]

3.1.9**measurement**

set of operations having the object of determining a value of a measure

[SOURCE: ISO/IEC 25024:2015; ISO 3951-5:2006]

² ISO/IEC 22989:2022/AMD1 is under development.

3.1.10

metric

defined measurement method and measurement scale

[SOURCE: ISO/IEC 14102:2008]

3.1.11

process

set of interrelated or interacting activities which transforms inputs into outputs

[SOURCE: ISO/IEC/IEEE 12207:2017]

3.1.12

product

result of a process

[SOURCE: ISO/IEC/IEEE 12207:2017]

3.1.13

property

property to quantify

property of a target entity that is related to a quality measure element and which can be quantified by a measurement method

[SOURCE: ISO/IEC 25021:2012, Figure 5, reported in Figure 12 of this document]

3.1.14

quality model

defined set of characteristics and of relationships between them, which provides a framework for specifying quality requirements and evaluating quality

[SOURCE: ISO/IEC 25000:2014]

3.1.15

system

combination of interacting elements organized to achieve one or more stated purposes

[SOURCE: ISO/IEC 25000:2014]

3.2 Data governance

3.2.1

corporate governance

system by which corporations are directed and controlled

[SOURCE: ISO/IEC 38500:2024]

3.2.2

data governance

execution and enforcement of authority over the definition, production, and usage of data related assets

[SOURCE: IEEE 7005:2021]

3.2.3**data governance framework**

strategy, policies, decision-making structures and accountabilities, through which the organization's governance arrangements operate on data

[SOURCE: ISO/IEC TR 38502:2017, modified – the data are specified]

3.2.4**governance**

process for establishing and enforcing strategic goals and objectives, organizational policies, and performance parameters

[SOURCE: Software Extension to the PMBOK (R) Guide Fifth Edition) ISO/IEC/IEEE 21840:2019]

3.2.5**governing body**

person or group of people who are accountable for the performance and conformance of the organization

[SOURCE: ISO/IEC 5259-5³; ISO/IEC 38500:2024]

3.2.6**management**

system of controls and processes required to achieve the strategic objectives set by the organization's governing body

[SOURCE: ISO/IEC/IEEE 21840:2019]

3.2.7**strategy**

organization's overall plan of development, describing the effective use of resources in support of the organization in its future activities. It involves setting objectives and proposing initiatives for action

[SOURCE: ISO/IEC/IEEE 24765]

3.2.8**process**

predetermined course of events that occur during the execution of all or part of a program

[SOURCE: ISO/IEC 2382:2015]

3.2.9**personal data**

<processing of personal data> any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person

[SOURCE: Regulation (EU) 2016/679 (GDPR) [28], Article 4 (1)]

³ Under preparation. Current stage: ISO/IEC FDIS 5259-5:2024.

3.2.10

processing of personal data

operation or set of operations which is performed on personal data or on sets of personal, whether or not by automated means, such as collection, recording, organization, structuring, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction

[SOURCE: Regulation (EU) 2016/679 (GDPR) [28], Article 4 (2)]

3.2.11

product

result of a process

[SOURCE: ISO/IEC/IEEE 12207:2017, 3.1.36; ISO/IEC/IEEE 24748-1:2024, 3.34]

3.3 Data quality

3.3.1

analytics

composite concept consisting of data acquisition, validation, processing, including quantification, visualization and interpretation

[SOURCE: ISO/IEC 5259-1:2024, modified; ISO/IEC 20546:2019, modified]

3.3.2

big data

extensive datasets, primarily in the data characteristics of volume, variety, velocity, and/or variability, that require a scalable technology for efficient storage, manipulation, management and analysis

[SOURCE: ISO/IEC 20546:2019]

3.3.3

data

reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing

Note 1 to entry: Data can be processed by humans or by automatic means.

[SOURCE: ISO/IEC 25012:2008, ISO/IEC 2382:2015]

Note 2 to entry: The reinterpretable representation is connected to the data attributes that enable to be read and interpreted by users (see ISO/IEC 25012:2008, 5.3.2.7).

3.3.4

data life cycle

cycle composed of 10 stages, i.e. idea conception, business requirements, data planning, data acquisition, data preparation, building model, system deployment, system operation, data decommissioning, system decommissioning

[SOURCE: ISO/IEC 8183:2023, Clause 5]

3.3.5

data management

disciplined process that plans for acquirers and provides stewardship for business and technical data, consistent with requirements, throughout the data life cycle

[SOURCE: IEEE 7005:2021, 3.1]

3.3.6

data processing

systematic performance of operations upon data

[SOURCE: ISO/IEC 2382:2015]

3.3.7

data provenance record

record of the ultimate derivation and passage of a piece of data (3.3.3) through its various owners or custodian

[SOURCE: ISO 8000-2:2022, 3.8.4]

3.3.8

data quality

degree to which the characteristics of data satisfy stated and implied needs when used under specified conditions

[SOURCE: ISO/IEC 25012:2008]

3.3.9

data quality management

coordinated activities to direct and control an organization with regard to data quality

[SOURCE: ISO 8000-2:2022, 3.8.2]

3.3.10

data quality model

defined set of characteristics which provides a framework for specifying data quality requirements and evaluation data quality

[SOURCE: ISO/IEC 25012:2008]

3.3.11

data file

set of related data records treated as a unit

Note 1 to entry: In ISO/IEC 25024:2015, data set is a synonym of data file.

3.3.12

dataset

collection of data with a shared format and goal-relevant content

[SOURCE: ISO/IEC 22989:2022, 3.2.5, modified in ISO/IEC 5259-2⁴]

⁴ Under preparation. Current stage: ISO/IEC FDIS 5259-2:2024

3.3.13

data strategy

organization's overall plan of development, describing the effective use of data in support of the organization in its future activities

Note 1 to entry: It involves setting a policy, objectives and proposing initiatives for action.

3.3.14

information

knowledge concerning objects, such as facts, events, things, or ideas, including concepts, that within a certain context have a particular meaning

[SOURCE: ISO/IEC 2382:2015, quoted in ISO/IEC 25012:2008, and ISO/IEC 25024:2015]

3.3.15

non-personal data

all data that does not qualify as personal data (3.2.9)

3.3.16

provenance

information on the place and time of origin, derivation or generation of a dataset, proof of the dataset, or a record of past and present ownership of the dataset

[SOURCE: ISO/IEC 11179-33:2023, 3.11]

3.3.17

synthetic data

data that has been generated using a purpose-built mathematical model or algorithm, with the aim of solving a (set of) data science task(s)

[SOURCE: [38]]

3.3.18

quality in use

extent to which the system or product, when it is used in a specific context of use, satisfies, or exceeds stakeholders needs to achieve beneficial goals or outcomes

[SOURCE: ISO/IEC 25019:2023, 3.1.15]

4 Abbreviations

AI	Artificial Intelligence
AWI	Approved Work Item
CD	Committee Draft
CEI	Italian Electrotechnical Committee
CEN	European Committee for Standardization
CENELEC	European Electrotechnical Committee for Standardization
CLC	CENELEC
DGA	Data Governance Act
DIS	Draft International Standard

DLC	Data Life Cycle
EU	European Union
GDPR	General Data Protection Regulation
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
ISO	International Organization for Standardization
LC	Life Cycle
JTC 1	ISO-IEC Joint Technical Committee for Information Technology
JTC 21	CEN-CENELEC Joint Technical Committee for Artificial Intelligence
JRC	Joint Research Centre
ML	Machine learning
SQuaRE	Software Quality Requirements and Evaluation
TR	Technical Report
TS	Technical Specification
UNI	Italian National Unification Body

5 JRC research and data-related standards on AI

5.1 General

The Artificial Intelligence website of the European Commission's Joint Research Centre (JRC) is the official EU website reporting and monitoring of the development, uptake, and impact of Artificial Intelligence in Europe.

It is possible to explore [AI Watch by topic](#) that takes a holistic view of what is impacting AI:

- topics on AI: Enablers, Landscape, Standards, Evolution of Technology, Trustworthy
- tools
- countries: all the European countries
- publications: that can be filtered by a keyword, type, data, area
- collaborations
- data
- events
- news

It offers a dynamic vision according to the evolutions taking place. Below is a static overview of the situation of data and AI standards.

In this clause is highlighted part of the Conference and Workshop organized online on 8 June 2022 by Joint Research Centre - JRC, the European Commission's Science and knowledge Service, with the participation of more than 178 persons from 36 countries, out of which 137 from 21 EU member states.

This research is considered for this document a good introduction to the tentative extension of standards on AI concerning data (see 5.2).

5.2 Research: Data quality requirements for inclusive, non-biased and trustworthy AI

The report “*Data quality requirements for inclusive, non-biased and trustworthy AI*” [2] is available at <https://data.europa.eu/doi/10.2760/365479>

Even if the data and data quality perspective is horizontal, parallel sessions was focused in the research on different sectors:

4.1 Education and employment

4.2 Law enforcement and the public sector

4.3 Finance

4.4 AI for media, including social media, content moderation, recommender systems

4.5 Medicine and healthcare

4.6 Industrial automation and robotics

with a lot of detailed information.

For example, in 4.5.2 of the report [2], included in 4.5 (Medicine and healthcare), the challenges addressed are mentioning data set properties and data quality aspects:

- legal compliance
- completeness and correctness of the data
- currentness
- inter and intra-data consistency
- representativeness of data
- balancedness
- avoidance of bias.

To guarantee the reader a harmonized view of the data quality characteristics according to the standards, under development, also requested by the JRC [2], it is useful to consider:

- The data quality inherent characteristics mentioned in 4.5.2 (compliance, completeness, currentness, consistency) are an essential basis of inherent data quality (system and domain independent), that can be completed with accuracy, credibility;
- The data quality for data set is defined as representativeness, balancedness, avoidance of bias.

In Clause 8 of this document, a complete data quality model is defined based on specific international standards.

The JRC report is quoted in this document when appropriate many with precise indications.

“The European Standardisation Organisations CEN and CENELEC recognized the urgent need for AI standardization and launched last year the Joint Technical Committee 21 ‘Artificial Intelligence’ (JTC 21), responsible for the development and adoption of standards for AI, as well as providing guidance to other technical committees concerned with AI.” (see JRC document, 2.4)

As reported in JRC “agreements between European Standardisation Organisations (ESO) and International Standardisation Organisations (ISO), as well as relevant ad-hoc initiatives, can ensure that international standards can be used at European level (also as harmonized standards)”. In terms of international AI standardization ISO/IEC JTC1 SC42 (Joint Technical Committee 1, Subcommittee 42) is the main source, with a considerable history of AI work, and a substantial number of standards published or on development at different stages”. (JRC document - Clause 2.3).

In the JRC document, 4.5.1, the following figure (Figure 2) is shown on the state of art, ongoing standardization activities.

NOTE The mentioned title in the figure “Data governance & quality for AI” was related to the “ad hoc group 5”, before the modification to “Data governance & quality for AI within the European context”. Ad hoc group 5 is now the JTC21 WG3.

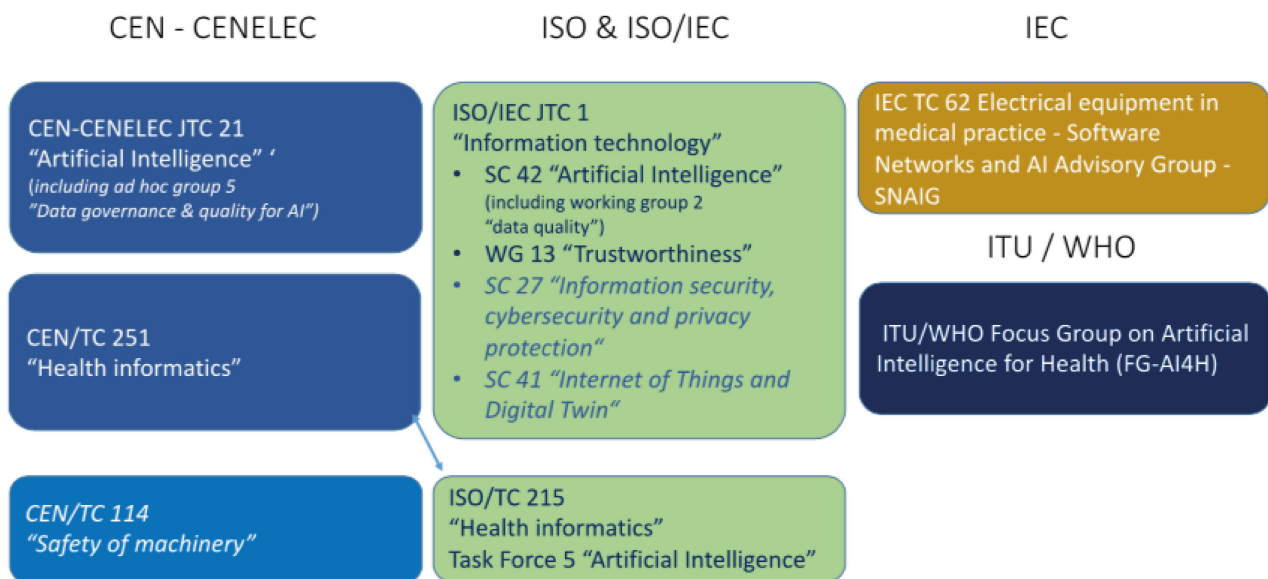


Figure 2 — Active organizations mentioned in JRC

The section ISO/IEC JTC1 has been extended in Figure 2, considering relevant existing international standards published or under development.

The current international standards or documents in the field of AI are summarized below, listing relevant documents developed by subcommittees of ISO/IEC JTC 1 on Information Technology:

- SC42 for Artificial Intelligence;
- SC40 for IT Governance;
- SC7 for testing in Software Engineering field.

Before to proceed collecting standards the following feature is described also in 4.5.1 of the JRC document, giving an idea, as reported in Figure 3, of initial perspective to be extended, with particular attention to data quality.

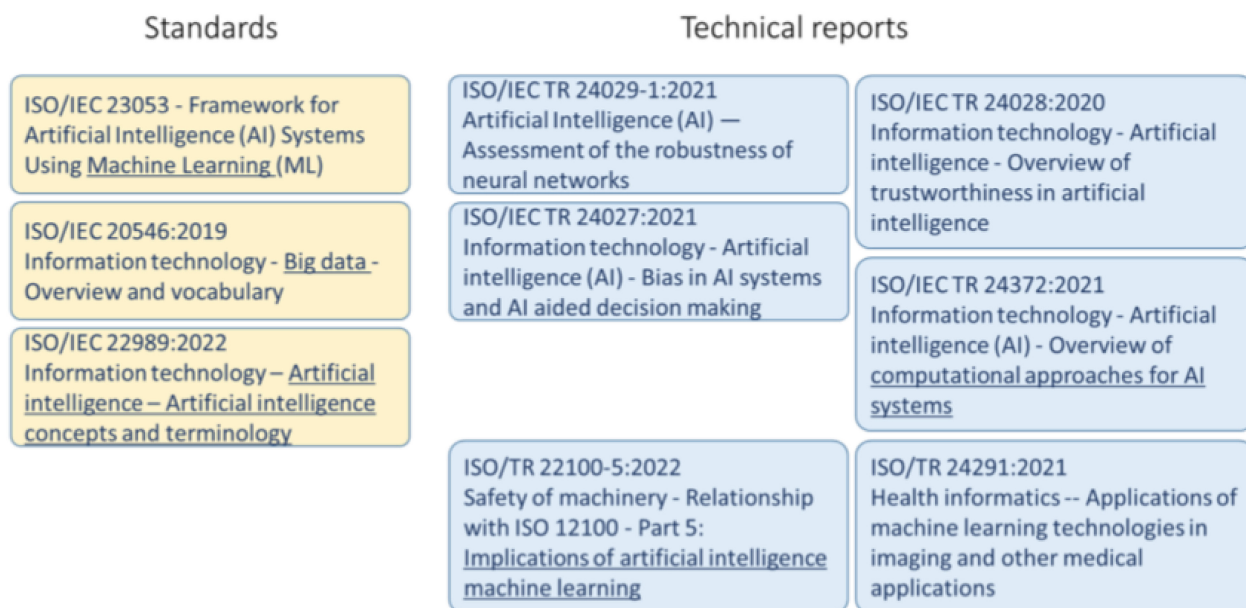


Figure 3 — Standards and Technical reports mentioned in JRC

Relevant ISO/IEC publications (standards and TRs) concern horizontal aspects of AI (e.g. robustness, bias, machine learning = ML) as well as health specific aspects (i.e. ML applications for imaging and other medical applications).

In the following there are core published standards or TRs (including DIS - Draft International Standards) or under development (AWI - WD - CD - Committee Draft stage). Because of the fast evolution of information, it is recommended to verify the stage of each standard at www.iso.org.

5.3 Data-related standards on AI for data governance and data quality

5.3.1 General

Figure 4 shows clusters of documents developed by different commissions and working groups.

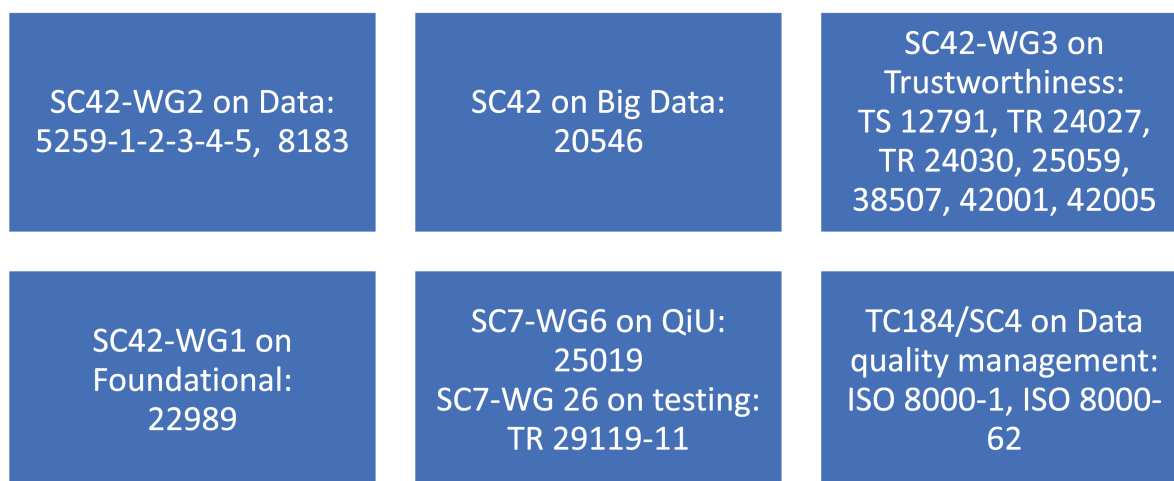


Figure 4 — Clusters of standards, TS, TR data-related

5.3.2 A short description of the standards mentioned in Figure 4 (taken from www.iso.org)

5.3.2.1 ISO/IEC 5259-1 Artificial intelligence – Data quality for analytics and machine learning (ML) – Part 1: Overview, terminology, and examples

This document provides the means for understanding and associating the individual documents of the ISO/IEC 5259 series and is the foundation for conceptual understanding of data quality for analytics and machine learning.

5.3.2.2 ISO/IEC 5259-2⁵ Artificial intelligence – Data quality for analytics and machine learning (ML) – Part 2: Data quality measures

This document provides a data quality model containing measures, and a guidance on reporting data quality in the context of analytics and machine learning (ML). This document builds on ISO 8000 series, ISO/IEC 25012, and ISO/IEC 25024. The aim of this documents is to enable organizations to achieve their data quality objectives and requirements. ISO/IEC 5259-2⁶ “Data quality measures” satisfy the needs related to the quality of “individual data” (data item) and “datasets” (group of data) for representativeness, control of bias and so on. In the Annex are reported information on measurement, uml model, overview and categories of quality characteristics, synthetic data.

5.3.2.3 ISO/IEC 5259-3 Artificial intelligence – Data quality for analytics and machine learning (ML) – Part 3: Data quality management requirements and guidelines

This document specifies requirements and provides guidance for establishing, implementing, maintaining, and continually improving the quality for data used in the areas of analytics and machine learning. This document does not define a detailed process, methods or metrics. Rather, it defines the requirements and guidance for a quality management process along with a reference process and methods that can be tailored to meet the requirements in this document.

5.3.2.4 ISO/IEC 5259-4 Artificial intelligence – Data quality for analytics and machine learning (ML) – Part 4: Data quality process framework

This document provides general common organizational approaches, regardless of type, size or nature of the applying organization, to ensure data quality for training and evaluation in analytics and machine learning. It is applicable to training and evaluation data that comes from different sources, including data acquisition and data composition, data preparation, data labelling, evaluation, and data use.

5.3.2.5 ISO/IEC 5259-5⁷ Artificial intelligence – Data quality for analytics and machine learning (ML) – Part 5: Data quality Governance

This document provides a data quality governance framework for analytics and machine learning to enable governing bodies of organizations to direct and oversee the implementation and operation of data quality measures, management, and related processes with adequate controls throughout the data life cycle. This document can be applied to any analytics and machine learning. This document does not define specific management requirements or process requirements specified in ISO/IEC 5259-3 and ISO/IEC 5259-4, respectively.

5.3.2.6 ISO/IEC 8183:2023 Information Technology – Artificial intelligence – Data lifecycle framework

This document provides an overarching data life cycle framework that is instantiable for any AI system from data ideation to decommission. This document is applicable to the data processing throughout the

⁵ Under preparation. Current stage: ISO/IEC FDIS 5259-2:2024.

⁶ Under preparation. Current stage: ISO/IEC FDIS 5259-2:2024.

⁷ Under preparation. Current stage: ISO/IEC FDIS 5259-5:2024.