DRAFT INTERNATIONAL STANDARD

ISO/IEC DIS 25059

ISO/IEC JTC **1**/SC **42**

Secretariat: **ANSI**

Voting begins on:
2022-07-12

Voting terminates on:
2022-10-04

Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems

ICS: 35.080

This document is circulated as received from the committee secretariat.

Reference number
ISO/IEC DIS 25059:2022(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 42, *Artificial intelligence*.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

High-quality software products and computer systems are crucial to stakeholders. Quality models, quality requirements, quality measurement, and quality evaluation are standardized within the SQuaRE series of international standards (ISO/IEC 25000 [1] to ISO/IEC 25099).

AI systems require additional properties and characteristics of systems to be considered, and stakeholders have varied needs. This is because the AI system can be used for decision-making tasks, can be based on noisy and incomplete data, can give probabilistic predictions in some cases, can learn from data and can adapt during operation. Also relevant is the increased automation that occurs in such systems.

According to ISO/IEC TR 24028:2020 [2], trustworthiness has been understood and treated as both an ongoing organizational process as well as a non-functional requirement specifying emergent properties of a system — that is, a set of inherent characteristics with their attributes — within the context of quality of use as indicated in ISO/IEC 25010:2011 [3].

ISO/IEC TR 24028:2020 discusses the applicability to AI systems of the ISO/IEC 250xx – SQuaRE series that have been developed for conventional software. According to ISO/IEC TR 24028:2020, the SQuaRE series does not sufficiently address the data-driven unpredictable nature of AI systems. While considering the existing body of work, ISO/IEC TR 24028:2020 identifies the need for developing new standards for AI systems that can go beyond the characteristics and requirements of conventional software development.

ISO/IEC TR 24028:2020 contains a related discussion on different approaches to testing and evaluation of AI systems. It states that for testing of an AI system, modified versions of existing software and hardware verification and validation techniques are needed. It identifies several conceptual differences between many AI systems and conventional systems and concludes that "the ability of the [AI] system to achieve the planned and desired result … may not always be measurable by conventional approaches to software testing". Testing of AI systems is addressed in ISO/IEC TR 29119-11:2020 [4].

This document outlines an application-specific AI system extension to the SQuaRE series quality model specified in ISO/IEC 25010:2011 [3]. This document is meant to be used in conjunction with the SQuaRE series.

AI systems perform tasks. One or more tasks can be defined for an AI system. For the evaluation of task fulfillment it is necessary to specify quality requirements with evaluation measures.

In Clause 3 this document defines terminology that can be used to define quality requirements for AI systems. In Clause 5 and 6 the relevance of these terms is explained, and links to other standardization deliverables (e.g. the ISO/IEC 24029 series [5][6]) are highlighted.

ISO/IEC 25012:2008 [7] contains a model for data quality that is complementary to the model defined in this document. ISO/IEC 25012:2008 is being extended for AI systems by the ISO/IEC 5259 series of standards [8].

1 Scope

This document outlines a quality model for AI systems and is an application-specific extension to the SQuaRE series. The characteristics and sub-characteristics detailed in the model provide consistent terminology for specifying, measuring and evaluating AI system quality. The characteristics and sub-characteristics detailed in the model also provide a set of quality characteristics against which stated quality requirements can be compared for completeness.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 25010:2011, *Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models*

ISO/IEC 22989:—[1], *Information technology — Artificial intelligence — Artificial intelligence concepts and terminology*

ISO/IEC 23053:—[2], *Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)*

ISO/IEC 24029-2:—[3], *Artificial intelligence (AI) — Assessment of the robustness of neural networks — Part 2: Methodology for the use of formal methods*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 22989:—, ISO/IEC 23053:—, and the following apply.

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

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1 General

3.1.1
measure, noun
variable to which a value is assigned as the result of measurement

Note 1 to entry: The term “measures” is used to refer collectively to base measures, derived measures, and indicators.

[SOURCE: ISO/IEC 15939:2007, 3.15]

3.1.2
measure, verb
make a measurement

[SOURCE: ISO/IEC 25010:2011, 4.4.6]

3.1.3
software quality measure
measure of internal software quality, external software quality or software quality in use

Note 1 to entry: Internal software quality, external software quality and software quality in use are described in the quality model in ISO/IEC 9126-1

[1] Under preparation. Stage at the time of publication ISO/IEC FDIS 22989:2022.
[2] Under preparation. Stage at the time of publication ISO/IEC FDIS 23053:2022.
[3] Under preparation. Stage at the time of publication ISO/IEC CD 24029-2:2021.

[ISO/IEC 25030:2007, A.55]

3.1.4
risk treatment measure
protective measure
action or means to eliminate hazards or reduce risks

[SOURCE: ISO/IEC Guide 51:2014, 3.13], modified to change reduction to treatment in order to align with ISO/IEC 23894:—[4]]

3.2 Product quality

3.2.1
user controllability
<quality model> degree to which a user can appropriately intervene in an AI system's functioning in a timely manner

3.2.2
functional adaptability
<quality model> degree to which an AI system can accurately acquire information from data, or the result of previous actions, and use that information in future predictions

3.2.3
functional correctness
degree to which a product or system provides the correct results with the needed degree of precision

Note 1 to entry: AI systems, and particularly those machine learning methods, do not usually provide functional correctness in all observed circumstances. Therefore, it is necessary to measure the correctness and incorrectness carefully.

[SOURCE ISO/IEC 25010:2011, 4.2.1.2, added note]

3.2.4
robustness
<quality model> degree to which an AI system can maintain its level of performance under any circumstances

[SOURCE ISO/IEC 22989:—, modified to be the degree to which the system has the property, 3.5.12]

3.3 Quality in use

3.3.1
societal and ethical risk mitigation
<quality model> degree to which an AI system mitigates potential risk to society

Note 1 to entry: Societal and ethical risk mitigation includes accountability, fairness, transparency and explainability, professional responsibility, promotion of human value, privacy, human control of technology, community involvement and development, respect for the rule of law, respect for international norms of behaviour and labour practices.

3.3.2
transparency
<quality model> degree to which appropriate information about the AI system is communicated to relevant stakeholders

Note 1 to entry: Appropriate information for AI system transparency can include aspects such as features, components, procedures, measures, design goals, design choices and assumptions.

[SOURCE ISO/IEC 22989:—, modified to be the degree to which the system has the property, 3.5.14]

[4] Under preparation. Stage at the time of publication ISO/IEC DIS 23894:2022.

3.3.3
intervenability
<quality model> degree to which an operator can intervene in an AI system's functioning in a timely manner to prevent harm or hazard

4 Abbreviations

AI artificial intelligence
ML machine learning

5 Product quality model

5.1 General

An AI system product quality model is detailed in Figure 1 below. The model is based on a modified version of a general system model provided in ISO/IEC 25010:2011. New and modified sub-characteristics are identified using an asterisk. Some of the sub-characteristics have different meanings or contexts as compared to the ISO/IEC 25010:2011 model. The modifications, additions and differences are described in this Clause. The original characteristics shall be interpreted as defined in ISO/IEC 25010:2011.

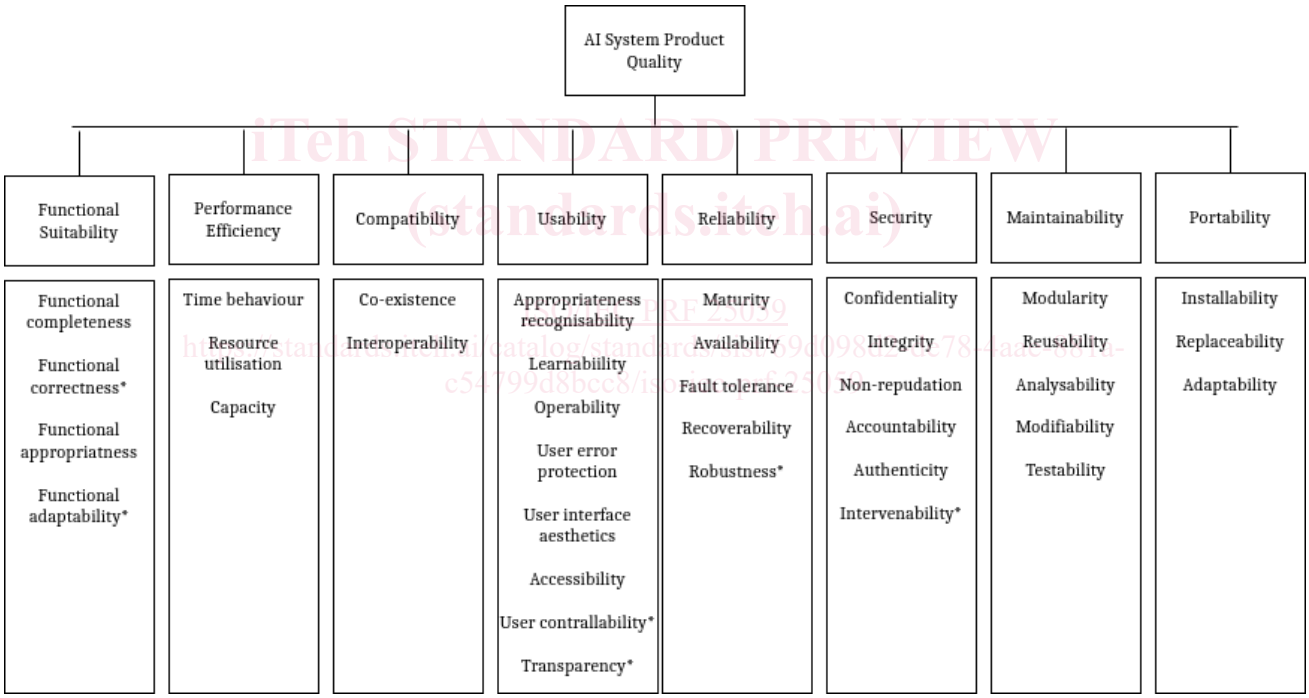

Figure 1 — AI system product quality model

Each of these modified or added sub-characteristics is listed in the remainder of this Clause.

5.2 Controllability

User controllability is a new sub-characteristic of usability. User controllability is a property of an AI system such that a human or another external agent can intervene in its functioning. Enhanced controllability is helpful if unexpected behaviour cannot be completely avoided and that would lead to negative consequences.

User controllability is related to controllability, which is described in ISO/IEC 22989:—, 5.12.

5.3 Functional adaptability

Functional adaptability is a new sub-characteristic of functional suitability. Functional adaptability of an AI system is the ability of the system to adapt itself to a changing dynamic environment it is deployed in. AI systems can learn from new training data, operational input data and the results of previous actions

taken by the system. The concept of functional adaptability is broader than that of continuous learning systems, as defined in ISO/IEC 22989:—, 5.11.9.2.

Continuous learning is not a mandatory requirement for functional adaptability. For example, a system that switches classification models based on events in its environment can also be considered functionally adaptive.

Functional adaptability in AI systems is unlike other quality characteristics as there are system specific consequences that cannot be interpreted using a straight-line linear scale (e.g. bad to good). Generally, higher functional adaptability can result in improvements for the outcomes enacted by AI systems.

For some systems, high functional adaptability can modify the AI system based on added information that can improve the overall output of the system. In other systems, high functional adaptability can cause additional unhelpful outcomes to become more likely based on the system's previous choices (i.e. weightings of a decision path with relatively high uncertainty that is reinforced based on decisions previously selected by the AI system) providing a higher likelihoods of unintended negative outcomes (e.g. reinforcing a negative human cognitive bias).

While conventional algorithms usually produce the same result for the same set of inputs, AI systems, due to continuous learning, can exhibit different behaviour and therefore can produce different results.

5.4 Functional correctness

Functional correctness exists in ISO/IEC 25010:2011. This model amends the description since AI systems, and particularly those probabilistic ML methods, do not usually provide functional correctness because a certain error rate is expected in their outputs. Therefore, it is necessary to measure correctness and incorrectness carefully. Numerous measurements exist for these purposes in the context of ML methods and examples of these as applicable to a classification model with accompanying relevant examples can be found in ISO/IEC TS 4213:—[5] [12].

Additionally, there can be a trade-off between characteristics such as performance efficiency [13], robustness [14] and functional correctness.

5.5 Robustness

Robustness is a new sub-characteristic of reliability. It is used to describe the ability of a system to maintain its level of functional correctness under any circumstances including:

- the presence of unseen, biased, adversarial or invalid data inputs;
- external interference;
- harsh environmental conditions encompassing generalization, resilience, reliability;
- attributes related to the proper operation of the system as intended by its developers.

The proper operation of a system is important for the security of the system and safety of its stakeholders in a given environment or context. Information about functional safety in the context of AI systems can be found in ISO/IEC 5469:—[6] [15].

Robustness is discussed in ISO/IEC TR 24028:2020, 10.7 [2], and methods for assessment are described in ISO/IEC TR 24029-1:2021 [5] and defined in ISO/IEC 24029-2:—.

5.6 Transparency

Transparency is sub-characteristic of usability. It relates to the availability of data regarding AI system processes and how amenable that data is to external inspection. Transparency of AI systems can help potential users of AI systems to choose a system fitting their requirements, improving relevant stakeholders' knowledge about the applicability and the limitations of an AI system as well as assisting with the explainability of AI systems.

Transparency relates to the availability of information about an AI system and the way this information is communicated to relevant stakeholders in accordance with their objectives and knowhow. The

[5] Under preparation. Stage at the time of publication ISO/IEC DTS 4213.2:2022.
[6] Under preparation. Stage at the time of publication ISO/IEC AWI 5469:2020.

transparency information can include a description of an AI system functionality, the system's decomposition, interfaces, ML model(s) used, training data, verification and validation data, performance benchmarks, logs and the management practices of an organization responsible for the system.

Transparent systems document, log or display their internal processes using introspection tools and data files. The flow of data can be trackable at each step, with applied decisions, exceptions and rules documented. Log output can track processes in the pipeline as they permute data, as well as system level calls. Errors are logged explicitly, particularly in transform steps. Highly transparent and modular systems can be built of well-documented subcomponents whose interfaces are explicitly described.

A system with low transparency has internal workings which are difficult to inspect externally. Transparency of AI systems eases investigations of system malfunctions. Unavailability of detailed processing records can impair testability and societal and ethical impact assessment and risk treatment.

Ultimately, transparency of AI systems contributes to establishing of trust, accountability and communication among stakeholders. Some aspects of transparency are discussed in ISO/IEC TR 24028:2020, 10.2 [2].

Transparency is also a sub-characteristic of satisfaction in the Quality in Use model. See 6.3 for further information.

5.7 Intervenability

The extent of intervenability can be determined depending on the scenarios where the AI system can be used. The key to intervenability is to enable state observation and transition from an unsafe state to a safe state. Operability is the degree to which an AI system has attributes that make it easy to operate and control, which emphasizes the importance of an AI system's interface. Compared to operability, intervenability is more fundamental from a quality perspective and intended to prevent an AI system from doing harm or hazard.

Intervenability is related to controllability, which is described in ISO/IEC 22989:—, 5.15.5

6 Quality in use model

6.1 General

An AI system quality in use model is detailed in Figure 2. The model is based on a modified version of a general quality in use model provided in ISO/IEC 25010:2011. New and modified sub-characteristics are identified using an asterisk. Some of the sub-characteristics have different meanings or contexts as compared to the ISO/IEC 25010:2011 model. The modifications, additions and differences are described in this Clause. The original characteristics shall be interpreted as defined in ISO/IEC 25010:2011.