INTERNATIONAL STANDARD

ISO/IEC 25059

First edition
2023-06

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

Reference number
ISO/IEC 25059:2023(E)

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

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 document 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 or www.iec.ch/members_experts/refdocs).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and https://patents.iec.ch. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

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. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by Joint 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 and www.iec.ch/national-committees.

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 International Standards on SQuaRE, see Annex A for further information.

AI systems require additional properties and characteristics of systems to be considered, and stakeholders have varied needs. AI systems have different properties and characteristics. For example, AI systems can:

— replace human decision-making;

— be based on noisy, or incomplete data;

— be probabilistic;

— adapt during operation.

According to ISO/IEC TR 24028,[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.

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

ISO/IEC TR 24028 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.[3]

This document outlines an application-specific AI system extension to the SQuaRE quality model specified in ISO/IEC 25010.

AI systems perform tasks. One or more tasks can be defined for an AI system. Quality requirements can be specified for the evaluation of task fulfilment.

The quality model is considered from two perspectives, product quality as described in Clause 5 and quality in use in Clause 6. The relevance of these terms is explained, and links to other standardization deliverables (e.g. the ISO/IEC 24029 series[4][5]) are highlighted.

ISO/IEC 25012:2008[6] 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.[7]

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

1 Scope

This document outlines a quality model for AI systems and is an application-specific extension to the standards on SQuaRE. 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 22989:2022, *Information technology — Artificial intelligence — Artificial intelligence concepts and terminology*

ISO/IEC 23053:2022, *Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 22989:2022, ISO/IEC 23053:2022 and the following 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/

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/IEEE 15939:2017, 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 measure of software quality, external measure of software quality or software quality in use measure are described in the quality model in ISO/IEC 25010.

[SOURCE: ISO/IEC 25040:2011, 4.61]

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 — change reduction to treatment.]

3.1.5
transparency
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.

3.2 Product quality

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

3.2.2
functional adaptability
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 using machine learning models, do not usually provide functional correctness in all observed circumstances.

[SOURCE: ISO/IEC 25010:2011, 4.2.1.2, modified — Note to entry added.]

3.2.4
intervenability
degree to which an operator can intervene in an AI system's functioning in a timely manner to prevent harm or hazard

3.2.5
robustness
degree to which an AI system can maintain its level of functional correctness under any circumstances

3.3 Quality in use

3.3.1
societal and ethical risk mitigation
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.

4 Abbreviated terms

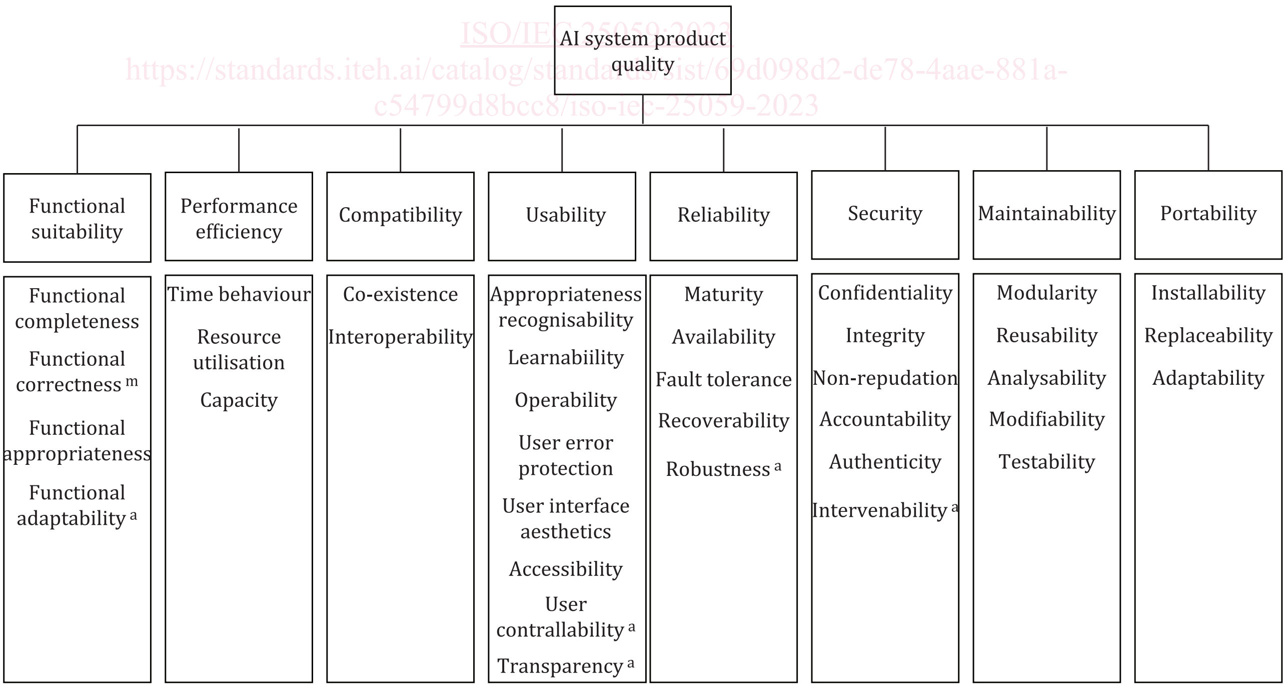
AI artificial intelligence

ML machine learning

5 Product quality model

5.1 General

An AI system product quality model is detailed in Figure 1. The model is based on a modified version of a general system model provided in ISO/IEC 25010. New and modified sub-characteristics are identified using a lettered footnote. Some of the sub-characteristics have different meanings or contexts as compared to the ISO/IEC 25010 model. The modifications, additions and differences are described in this clause. The unmodified original characteristics are part of the AI system product model and shall be interpreted in accordance with ISO/IEC 25010.



[a] New sub-characteristics.

[m] Modified sub-characteristics.

Figure 1 — AI system product quality model

Each of these modified or new sub-characteristics are listed in the remainder of this clause.

5.2 User 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 in a timely manner. Enhanced controllability is helpful if unexpected behaviour cannot be completely avoided and that can lead to negative consequences.

User controllability is related to controllability, which is described in ISO/IEC 22989:2022, 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, production data and the results of previous actions taken by the system. The concept of functional adaptability subsumes that of continuous learning, as defined in ISO/IEC 22989:2022, 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 cause additional unhelpful outcomes to become more likely based on the system's previous choices. Weightings of a decision path with relatively high uncertainty, reinforced based on previous AI system decisions, can result in higher likelihood of unintended negative outcomes. In this fashion, functional adaptability can reinforce negative human cognitive biases.

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. The AI system product quality model amends the description since AI systems, and particularly 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 can be found in ISO/IEC TS 4213.[11]

Additionally, there can be a trade-off between characteristics such as performance efficiency,[12] robustness[13] and functional correctness.

Annex C provides further information about why functional correctness is preferred to other terms such as the more general performance to describe the correctness of the model.

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 (see Annex C for discussion on the term performance) under any circumstances including:

— the presence of unseen, biased, adversarial or invalid data inputs;

— external interference;

— 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 TR 5469:—[1].[14]

Robustness is discussed in ISO/IEC TR 24028:2020, 10.7,[2] and methods for assessment are described in ISO/IEC TR 24029-1[4] and defined in ISO/IEC 24029-2.[5]

5.6 Transparency

Transparency is sub-characteristic of usability in the product quality model and a sub-characteristic of satisfaction in the quality in use model.

It relates to the degree to which appropriate information about the AI system is communicated to stakeholders.

Transparency of AI systems can help potential users of AI systems to choose a system to fit their requirements, improving stakeholders' knowledge about the applicability and the limitations of an AI system, and assisting with the explainability of AI systems.

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

Transparent AI 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 AI systems can be built of well-documented subcomponents whose interfaces are explicitly described. The transparency of AI systems eases investigations of system malfunctions.

A system with low transparency has internal workings which are difficult to inspect externally. 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]

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 enable operation and control, which emphasizes the importance of an AI system's user interface. Compared to operability, intervenability is more fundamental from a quality perspective and is intended to prevent an AI system from doing harm or hazard.

Intervenability is related to controllability, which is described in ISO/IEC 22989:2022, 5.15.5.

1) Under preparation. Stage at the time of publication ISO/IEC CD TR 5469:2023.