Technical Specification

ISO/IEC TS 42119-2

First edition
2025-11

Artificial intelligence — Testing of AI —

Part 2: Overview of testing AI systems

Intelligence artificielle — Test des IA —

Partie 2: Vue d'ensemble du test de systèmes d'IA

Reference number
ISO/IEC TS 42119-2:2025(en)

© ISO/IEC 2025



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2025

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

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

Published in Switzerland

Contents

Page

Foreword ... **iv**
Introduction ... **v**
1 Scope ... **1**
2 Normative references ... **1**
3 Terms and definitions ... **1**
4 Abbreviated terms ... **8**
5 Introduction to AI systems and software testing ... **8**
5.1 General ... 8
5.2 AI system life cycle ... 8
5.3 AI system functional view ... 9
5.4 Risk-based testing ... 10
5.5 Test processes ... 11
5.5.1 General ... 11
5.5.2 Test processes in the context of the AI system life cycle ... 11
5.6 Test documentation ... 12
5.7 Testing stakeholders ... 12
6 Identifying risks in AI systems ... **12**
7 Test approaches for testing AI systems ... **13**
7.1 Introduction to test approaches for AI systems ... 13
7.2 Test levels ... 13
7.3 Test types ... 14
7.3.1 Introduction ... 14
7.3.2 Common test types ... 14
7.3.3 Specialist data quality test types ... 15
7.3.4 Specialist AI model test types ... 18
7.3.5 Static testing of knowledge engineering systems ... 20
7.4 Test design techniques and measures ... 20
7.4.1 Introduction ... 20
7.4.2 Common test design techniques ... 21
7.4.3 Common test coverage measures ... 22
7.4.4 Specialist test coverage measures ... 23
Annex A (informative) **Introduction to software testing** ... **25**
Annex B (informative) **Characteristics of AI systems** ... **28**
Annex C (informative) **Example risk assessment** ... **29**
Bibliography ... **33**

© ISO/IEC 2025 – All rights reserved

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 jointly by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittees SC 7 *Software and systems engineering* and SC 42, *Artificial intelligence*.

A list of all parts in the ISO/IEC 42119 series can be found on the ISO and IEC websites.

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.

© ISO/IEC 2025 – All rights reserved

Introduction

This document facilitates understanding of how ISO/IEC/IEEE 29119-1, ISO/IEC/IEEE 29119-2, ISO/IEC/IEEE 29119-3, ISO/IEC/IEEE 29119-4 and ISO/IEC 20246 apply to the testing of AI systems.

The purpose of ISO/IEC/IEEE 29119 (all parts) is to define an internationally agreed set of standards for software testing that can be used by any organization when performing any form of software testing.

ISO/IEC/IEEE 29119-1 introduces software testing concepts, which can be applied to any AI system.

ISO/IEC/IEEE 29119-2 comprises test process descriptions that define the software test processes at the organizational level, test management level and dynamic test levels. It supports dynamic testing, functional and non-functional testing, manual and automated testing and scripted and unscripted testing, and can be utilized for the testing of any software-based system, including AI systems.

ISO/IEC/IEEE 29119-3 defines software test documentation. The requirements specified for templates and examples of test documentation defined in ISO/IEC/IEEE 29119-3 can be met in the test documentation for any AI system.

ISO/IEC/IEEE 29119-4 defines test design techniques, which can be utilized for the testing of AI systems and their components.

ISO/IEC 20246 defines processes and templates for work product reviews, including inspections, walkthroughs and technical reviews.

This document explains how ISO/IEC/IEEE 29119-2 can be adopted for the testing of AI systems and their components and how the test documentation templates defined in ISO/IEC/IEEE 29119-3 can be implemented when testing AI systems and their components. This document also explains how ISO/IEC 20246 can be adopted for the review of AI systems and related documentation. This document is structured as follows:

— Clauses 1 to 4 define the scope, normative references, terms and definitions and abbreviated terms;

— Clause 5 defines concepts of AI system architectures, the AI system life cycle and testing processes and documentation;

— Clause 6 explains how risk is identified for AI systems;

— Clause 7 defines test approaches suitable for testing AI systems and components;

— Annexes A to C provide supporting details and examples.

The aim of the ISO/IEC 42119 series is to provide requirements and guidance on the testing of AI components and systems.

Other parts of the ISO/IEC 42119 series include:

— ISO/IEC TS 42119-3 describes approaches and provides guidance on processes for the verification and validation analysis of AI systems;

— ISO/IEC TS 42119-7 provides technology-agnostic guidance for conducting red teaming assessments on AI systems;

— ISO/IEC TS 42119-8 provides definitions, concepts, requirements and guidance related to assessing prompt-based text-to-text AI systems that utilize generative AI.

© ISO/IEC 2025 – All rights reserved

Artificial intelligence — Testing of AI —

Part 2:
Overview of testing AI systems

1 Scope

This document provides requirements and guidance on the application of the ISO/IEC/IEEE 29119 series to the testing of AI systems. This document follows a risk-based approach and uses risks associated with AI systems, and their development and maintenance, to identify suitable test practices, approaches and techniques applicable to AI systems and their components. When the test practices, approaches and techniques are already specified in the ISO/IEC/IEEE 29119 series, this document provides additional detail and describes their application in the context of AI systems.

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/IEEE 29119-2, *Software and systems engineering — Software testing — Part 2: Test processes*

ISO/IEC/IEEE 29119-3, *Software and systems engineering — Software testing — Part 3: Test documentation*

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/

3.1
artificial intelligence
AI
<discipline> research and development of mechanisms and applications of *AI systems* (3.3)

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 22989:2022, 3.1.3]

3.2
AI model
machine-readable representation of *knowledge* (3.15)

Note 1 to entry: An *ML model* (3.19) and the knowledge captured from experts as rules in an expert system are both forms of AI model.

© ISO/IEC 2025 – All rights reserved

3.3
AI system
artificial intelligence 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* (3.1) to develop a model to represent data, *knowledge* (3.15), 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 22989:2022, 3.1.4]

3.4
bias
<artificial intelligence> systematic difference in treatment of certain objects, people or groups in comparison to others

Note 1 to entry: Treatment is any kind of action, including perception, observation, representation, *prediction* (3.24) or decision.

[SOURCE: ISO/IEC TR 24027:2021, 3.3.2, modified –Oxford comma removed in definition and note to entry.]

3.5
classification model
<machine learning> *machine learning model* (3.19) whose expected output for a given input is one or more classes

[SOURCE: ISO/IEC 23053:2022, 3.1.1]

3.6
concept drift
phenomenon where the statistical properties of *input data* (3.14) change over time, leading to decreased *model* (3.20) performance

3.7
data quality
characteristic of data that the data meet the organization's data requirements for a specified context

[SOURCE: ISO/IEC 5259-1:2024, 3.5]

3.8
data quality characteristic
category of data quality attributes that has a bearing on *data quality* (3.7)

[SOURCE: ISO/IEC 25012:2008, 4.4, modified —"bears" replaced by "has a bearing".]

3.9
dataset
collection of data with a shared format

EXAMPLE 1 Micro-blogging posts from June 2020 associated with hashtags #rugby and #football.

EXAMPLE 2 Macro photographs of flowers in 256x256 pixels.

Note 1 to entry: Datasets can be used for validating or *testing* (3.39) an *AI model* (3.2). In a *machine learning* (3.18) context, datasets can also be used to train a machine learning algorithm.

[SOURCE: ISO/IEC 22989:2022, 3.2.5]

© ISO/IEC 2025 – All rights reserved

3.10
expert system
AI system (3.3) that accumulates, combines and encapsulates *knowledge* (3.15) provided by a human expert or experts in a specific domain to infer solutions to problems

[SOURCE: ISO/IEC 22989:2022, 3.1.13]

3.11
explainability
<artificial intelligence> property of an *AI system* (3.3) to express important factors influencing the AI system results in a way that humans can understand

Note 1 to entry: It is intended to answer the question "Why?" without actually attempting to argue that the course of action that was taken was necessarily optimal.

[SOURCE: ISO/IEC 22989:2022, 3.5.7]

3.12
feature
<machine learning> measurable property of an object or event with respect to a set of characteristics

Note 1 to entry: Features play a role in training and *prediction* (3.24).

Note 2 to entry: Features provide a machine-readable way to describe the relevant objects. As the algorithm will not go back to the objects or events themselves, feature representations are designed to contain all useful information.

[SOURCE: ISO/IEC 23053:2022, 3.3.3]

3.13
ground truth
<machine learning> value of the target variable for a particular item of labelled *input data* (3.14)

Note 1 to entry: The term ground truth does not imply that the labelled input data consistently corresponds to the real-world value of the target variables.

[SOURCE: ISO/IEC 22989:2022, 3.2.7]

3.14
input data
data for which an *AI system* (3.3) calculates a predicted output or inference

[SOURCE: ISO/IEC 22989:2022, 3.2.9]

3.15
knowledge
<artificial intelligence> abstracted information about objects, events, concepts or rules, their relationships and properties, organized for goal-oriented systematic use

Note 1 to entry: Knowledge in the *AI* (3.1) domain does not imply a cognitive capability, contrary to usage of the term in some other domains. In particular, knowledge does not imply the cognitive act of understanding.

Note 2 to entry: Information can exist in numeric or symbolic form.

Note 3 to entry: Information is data that has been contextualized, so that it is interpretable. Data is created through abstraction or measurement from the world.

[SOURCE: ISO/IEC 22989:2022, 3.1.21]

3.16
label
target variable assigned to a sample

[SOURCE: ISO/IEC 22989:2022, 3.2.10]

© ISO/IEC 2025 – All rights reserved

3.17
life cycle
evolution of a system, product, service, project or other human-made entity, from conception through retirement

[SOURCE: ISO/IEC/IEEE 15288:2015, 4.1.23]

3.18
machine learning
ML
process of optimizing *model* (3.20) parameters through computational techniques, such that the model's behaviour reflects the data or experience

[SOURCE: ISO/IEC 22989:2022, 3.3.5]

3.19
machine learning model
ML model
mathematical construct that generates an inference or *prediction* (3.24) based on *input data* (3.14) or information

EXAMPLE If a univariate linear function ($y = \theta_0 + \theta_1 x$) has been trained using linear regression, the resulting model can be $y = 3 + 7x$.

Note 1 to entry: A machine learning model results from training based on a machine learning algorithm.

[SOURCE: ISO/IEC 22989:2022, 3.3.7, modified –"ML model" added as a preferred term.]

3.20
model
physical, mathematical or otherwise logical representation of a system, entity, phenomenon, process or data

[SOURCE: ISO/IEC 22989:2022, 3.1.23]

3.21
neural network
NN
neural net
artificial neural network
<artificial intelligence> network of one or more layers of *neurons* (3.22) connected by weighted links with adjustable weights, which takes *input data* (3.14) and produces an output

Note 1 to entry: Neural networks are a prominent example of the connectionist approach.

Note 2 to entry: Although the design of neural networks was initially inspired by the functioning of biological neurons, most works on neural networks do not follow that inspiration anymore.

[SOURCE: ISO/IEC 22989:2022, 3.4.8]

3.22
neuron
<artificial intelligence> primitive processing element which takes one or more input values and produces an output value by combining the input values and applying an activation function on the result

Note 1 to entry: Examples of nonlinear activation functions are a threshold function, a sigmoid function and a polynomial function.

[SOURCE: ISO/IEC 22989:2022, 3.4.9]

© ISO/IEC 2025 – All rights reserved

3.23
parameter
model parameter
internal variable of a *model* (3.20) that affects how it computes its outputs

Note 1 to entry: Examples of parameters include the weights in a *neural network* (3.21) and the transition probabilities in a Markov model.

[SOURCE: ISO/IEC 22989:2022, 3.3.8]

3.24
prediction
<artificial intelligence> primary output of an *AI system* (3.3) when provided with *input data* (3.14) or information

Note 1 to entry: *Predictions* (3.24) can be followed by additional outputs, such as recommendations, decisions and actions.

Note 2 to entry: Prediction does not necessarily refer to predicting something in the future.

Note 3 to entry: Predictions can refer to various kinds of data analysis or production applied to new data or historical data (including translating text, creating synthetic images or diagnosing a previous power failure).

[SOURCE: ISO/IEC 22989:2022, 3.1.27]

3.25
production data
data acquired during the operation phase of an *AI system* (3.3), for which a deployed AI system calculates a predicted output or inference

[SOURCE: ISO/IEC 22989:2022, 3.2.12]

3.26
retraining
<machine learning> updating a *trained model* (3.40) by *training* (3.41) with different *training data* (3.42)

[SOURCE: ISO/IEC 22989:2022, 3.3.10]

3.27
risk
combination of the probability of occurrence of a negative event and the severity of that event

[SOURCE: ISO/IEC Guide 51:2014, 3.9, modified — Definition revised.]

3.28
sample
atomic data element processed in quantities by a *machine learning* (3.18) algorithm or an *AI system* (3.3)

[SOURCE: ISO/IEC 22989:2022, 3.2.13]

3.29
stakeholder
any individual, group, or organization that can affect, be affected by or perceive itself to be affected by a decision or activity

[SOURCE: ISO/IEC 38500:2015, 2.24, modified —Comma removed after “be affected by” in the definition.]

3.30
static testing
testing (3.39) in which a test item is examined against a set of quality or other criteria without the test item being executed

EXAMPLE Reviews, static analysis.

© ISO/IEC 2025 – All rights reserved

[SOURCE: ISO/IEC/IEEE 29119-1:2022, 3.78]

3.31
supervised machine learning
machine learning (3.18) that makes only use of labelled data during *training* (3.41)

[SOURCE: ISO/IEC 22989:2022, 3.3.12]

3.32
task
<artificial intelligence> action required to achieve a specific goal

Note 1 to entry: Actions can be physical or cognitive. For instance, computing or creation of *predictions* (3.24), translations, synthetic data or artefacts or navigating through a physical space.

EXAMPLE Classification, regression, ranking, clustering and dimensionality reduction.

[SOURCE: ISO/IEC 22989:2022, 3.1.35]

3.33
test approach
high-level test implementation choice, typically made as part of the test strategy design activity

EXAMPLE 1 The use of model-based *testing* (3.39) for the functional system testing.

EXAMPLE 2 Typical choices made as test approaches are *test level* (3.36), *test type* (3.38), *test design technique* (3.35), *test practice* (3.37) and the form of *static testing* (3.30) to be used.

[SOURCE: ISO/IEC/IEEE 29119-1:2022, 3.83]

3.34
test data
<machine learning> data used to assess the performance of a final *model* (3.20)

Note 1 to entry: Test data is disjoint from *training data* (3.42) and *validation data* (3.44).

[SOURCE: ISO/IEC 22989:2022, 3.2.14]

3.35
test design technique
test technique
procedure used to create or select a test model, identify test coverage items, and derive corresponding test cases

EXAMPLE Equivalence partitioning, boundary value analysis, decision table testing, branch testing.

Note 1 to entry: The test design technique is typically used to achieve a required level of test coverage.

Note 2 to entry: Some *test practices* (3.37), such as exploratory testing or model-based testing are sometimes referred to as “test techniques”. Following the definition in the ISO/IEC/IEEE 29119 series, they are not test design techniques as they are not themselves providing a way to create test cases, but instead use test design techniques to achieve that.

[SOURCE: ISO/IEC/IEEE 29119-1:2022, 3.94]

3.36
test level
one of a sequence of test stages, each of which is typically associated with the achievement of particular objectives and used to treat particular *risks* (3.27)

EXAMPLE The following are common test levels, listed sequentially: unit/component testing, integration testing, system testing, system integration testing, acceptance testing.

Note 1 to entry: It is not always necessary for a test item to be tested at all test levels, but the sequence of test levels generally stays the same.

© ISO/IEC 2025 – All rights reserved

Note 2 to entry: Typical test level objectives can include consideration of basic functionality for unit/component testing, interaction between integrated components for integration testing, acceptability to end users for acceptance testing.

[SOURCE: ISO/IEC/IEEE 29119-1:2022, 3.108]

3.37
test practice
conceptual framework that can be applied to the organizational test process, the test management process, and/or the dynamic test process to facilitate *testing* (3.39)

[SOURCE: ISO/IEC/IEEE 29119-1:2022, 3.119]

3.38
test type
testing (3.39) that is focused on specific quality characteristics

EXAMPLE Security testing, functional testing, usability testing, and performance testing.

Note 1 to entry: A test type can be performed at a single *test level* (3.36) or across several test levels (e.g. performance testing performed at a unit test level and at a system test level).

[SOURCE: ISO/IEC/IEEE 29119-1:2022, 3.130]

3.39
testing
set of activities conducted to facilitate discovery or evaluation of properties of one or more test items

[SOURCE: ISO/IEC/IEEE 29119-1:2022, 3.131)

3.40
trained model
<machine learning> result of *model* (3.20)*training* (3.41)

[SOURCE: ISO/IEC 22989:2022, 3.3.14]

3.41
training
model training
<machine learning> process to determine or to improve the *parameters* (3.23) of a *machine learning model* (3.19), based on a *machine learning* (3.18) algorithm, by using *training data* (3.42)

[SOURCE: ISO/IEC 22989:2022, 3.3.15]

3.42
training data
<artificial intelligence> data used to train a *machine learning model* (3.19)

[SOURCE: ISO/IEC 22989:2022, 3.3.16]

3.43
validation
confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled

[SOURCE: ISO/IEC 25000:2014, 4.41]

© ISO/IEC 2025 – All rights reserved

3.44
validation data
development data
<machine learning> data used to compare the performance of different candidate *ML models* (3.19)

Note 1 to entry: Validation data is disjoint from *test data* (3.34) and generally also from *training data* (3.42). However, in cases where there is insufficient data for a three-way training, validation and test set split, the data is divided into only two sets – a test set and a training or validation set. Cross-validation or bootstrapping are common methods for then generating separate training and validation sets from the training or validation set.

Note 2 to entry: Validation data can be used to tune *hyperparameters* (3.46) or to validate some algorithmic choices, up to the effect of including a given rule in an expert system.

[SOURCE: ISO/IEC 22989:2022, 3.2.15, modified —"ML" added before “models” in the definition]

3.45
verification
confirmation, through the provision of objective evidence, that specified requirements have been fulfilled

[SOURCE: ISO/IEC 25000:2014, 4.43]

3.46
hyperparameter
characteristic of a *machine learning* (3.18) algorithm (3.3.6) that affects its learning process

Note 1 to entry: Hyperparameters are selected prior to training and can be used in processes to help estimate model parameters.

Note 2 to entry: Examples of hyperparameters include the number of network layers, width of each layer, type of activation function, optimization method, learning rate for neural networks; the choice of kernel function in a support vector machine; number of leaves or depth of a tree; the K for K-means clustering; the maximum number of iterations of the expectation maximization algorithm; the number of Gaussians in a Gaussian mixture.

[SOURCE: ISO/IEC 22989:2022, 3.3.4]

4 Abbreviated terms

AI	artificial intelligence
API	application programming interface
MCDC	modified condition/decision coverage
ML	machine learning
RBT	risk-based testing

5 Introduction to AI systems and software testing

5.1 General

This clause provides a high-level introduction to AI systems and software testing.

The testing of AI systems and their components shall be performed in accordance with ISO/IEC/IEEE 29119-2.

Annex A provides an introduction to software testing.

5.2 AI system life cycle

Figure 1 shows an example of the AI systems life cycle from ISO/IEC 22989 and its relationship with software testing.

© ISO/IEC 2025 – All rights reserved