ISO/IEC ~~TR~~DTR 5469~~:202x(E)~~

ISO/IEC JTC 1/SC 42~~/WG 3~~

Secretariat: ANSI

Date: 2023-09-22

Artificial intelligence — Functional safety and AI systems

~~Technical Report~~

~~Warning for WDs and CDs~~

~~This document is not an ISO International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.~~

~~Recipients of this draft are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.~~

FDIS stage

Contents

Foreword

ISO (the International Organization for Standardization) ~~is a~~ and IEC (the International Electrotechnical Commission) form the specialized system for worldwide ~~federation of national standards~~standardization. National bodies ~~(~~that are members of ISO ~~member bodies). The work~~or IEC participate in the development 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~~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.~~ 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~~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~~ 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

~~Attention is drawn~~ISO and IEC draw attention to the possibility that ~~some of~~ the ~~elements~~implementation of this document may ~~be~~involve the ~~subject~~use of (a) patent~~ rights. ISO~~(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.~~ 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~~) 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~~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~~www.iso.org/members.html and www.iec.ch/national-committees.

Introduction

The use of artificial intelligence (AI) technology in industry has increased significantly in recent years and AI has been demonstrated to deliver benefits in certain applications. However, there is limited information on specification, design, and verification of functionally safe AI systems or on how to apply AI technology for functions that have safety-related effects. For functions realized with AI technology, such as machine learning (ML), it is difficult to explain why they behave in a particular manner and to guarantee their performance. Therefore, whenever AI technology is used in general and especially when it is used to realize safety-related systems, special considerations are likely to arise.

The availability of powerful computational and data storage technologies makes the prospect of large-scale deployment of ML possible. For more and more applications, adopting machine learning (as an AI technology) is enabling the rapid and successful development of functions that detect trends and patterns in data. This makes it possible to induce a function's behaviour from observation and to quickly extract the key parameters that determine its behaviour. Machine learning is also used to identify anomalous behaviour or to converge on an optimal solution within a specific environment. Successful ML applications are found in analysis of, for example, financial data, social networking applications and language recognition, image recognition (particularly face recognition), healthcare management and prognostics, digital assistants, manufacturing robotics, machine health monitoring and automated vehicles.

In addition to ML, other AI technologies are also gaining importance in engineering applications. Applied statistics, probability theory and estimation theory have, for example, enabled significant progress in the field of robotics and perception. As a result, AI technology and AI systems are starting to realize applications that affect safety.

Models play a central role in the implementation of AI technology. The properties of these models are used to demonstrate the compatibility of AI technology and AI systems with functional safety requirements. For instance, where there is an underlying known and understood scientific relationship between the key parameters that determine a function's behaviour, there is likely to be a strong correlation between the observed input data and the output data. This leads to a transparent and sufficiently complete model as the basis for AI technology. In this case, compatibility of the model with functional safety requirements is demonstrated. However, AI technology is often used in cases where physical phenomena are so complex or at such a small scale, or unobservable without influencing the experimental data, that consequently there is no scientific model of the underlying behaviour. In this case, the model of the AI technology is possibly neither transparent nor complete and the compatibility of the model with functional safety requirements is hard to demonstrate.

Machine learning is used to create models and thus to extend the understanding of the world. However, machine-learnt models are only as good as the information used to derive the model. If the training data does not cover important cases, then the derived models are incorrect. As more known instances are observed they are used to reinforce a model, but this biases the relative importance of observations, steering the function away from less frequent, but still real, behaviours. Continuous observation and reinforcement moves the model towards an optimum or it ~~overemphasises~~overemphasizes common data and overlook extreme, but critical, conditions.

In the case of continuous improvement of the model through the use of AI technology, the verification and validation activities in order to demonstrate its safety integrity are undermined as the function behaviour progressively moves away from the rigorously tested, ideally deterministic and repeatable behaviour.

The purpose of this document is to enable the developer of safety-related systems to appropriately apply AI technologies as part of safety functions by fostering awareness of the properties, functional safety risk factors, available functional safety methods and potential constraints of AI technologies. This document also provides information on the challenges and solution concepts related to the functional safety of AI systems.

Clause 5Clause 5 provides an overview of functional safety and its relationship with AI technology and AI systems.

Clause 6Clause 6 describes different classes of AI technology to show potential compliance with existing functional safety International Standards when AI technology forms part of a safety function. Clause 6Clause 6 further introduces different usage levels of AI technology depending on their final impact on the system. Finally, Clause 6Clause 6 also provides a qualitative overview of the relative levels of functional safety risk associated with different combinations of AI technology class and usage level.

Clause 7Clause 7 describes, based on ISO/IEC 22989, a three-stage realization principle for usage of AI technology in safety-related systems, where compliance with existing functional safety International Standards cannot be shown directly.

Clause 8Clause 8 discusses properties and related functional safety risk factors of AI systems and presents challenges that such use raises, as well as properties that are considered when attempting to treat or mitigate them.

Clauses 9, 10 and 11Clauses 9, 10 and 11 show possible solutions to these challenges from the field of verification and validation, control and mitigation measures, processes, and methodologies.

The Annexesannexes provide examples of application of this document and additional details. Annex A providesaddresses how IEC 61508-3 is applied to AI technology elements, and annex BAnnex B provides examples to how to apply three-stage realization principles and define various properties. Annex CAnnex C describes more detailed processes related to clause 9.3. Annex D9.3. Annex D shows the mapping between safety lifecyclelife cycle in ISO/IEC 61508-3 and AI system life cycle in ISO/IEC 5338:—[1].

[1] Under preparation. Stage at the time of publication: ISO/IEC FDIS 5338:2023.

Artificial intelligence — Functional safety and AI systems

1 Scope

This document describes the properties, related risk factors, available methods and processes relating to:

— ~~Use~~use of AI inside a safety related function to realize the functionality;

— ~~Use~~use of non-AI safety related functions to ensure safety for an AI controlled equipment;

— ~~Use~~use of AI systems to design and develop safety related functions.

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 22989:2022, *Information technology — Artificial intelligence — Artificial intelligence concepts and terminology*

3 Terms and definitions

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

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

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

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

3.1
safety
freedom from *risk* ~~(3.3)~~(3.3) which is not tolerable

[SOURCE: IEC 61508-4~~, ed. 2.0 (~~:2010~~)~~, 3.1.~~1~~11]

3.2
functional safety
part of the overall *safety* ~~(3.1)~~(3.1) relating to the EUC (Equipment Under Control) and the EUC control system that depends on the correct functioning of the E/E/PE (Electrical/Electronic/Programmable Electronic) safety-related systems and other risk reduction measures

[SOURCE: IEC 61508-4~~, ed. 2.0 (~~:2010~~)~~, 3.1.12]

3.3
risk
functional safety risk
<functional safety> combination of the probability of occurrence of *harm* ~~(3.5)~~(3.5) and the severity of that *harm* ~~(3.5)~~(3.5)

Note 1 to entry: For more discussion on this concept, see Annex A of IEC 61508-5.

[SOURCE: IEC 61508-4~~, ed. 2.0 (~~:2010~~-04)~~,, 3.1.6, ~~added <~~modified — Added < functional safety~~>~~ > domain]

3.4
risk
organizational risk
<organizational> effect of uncertainty on objectives

Note 1 to entry: An effect is a deviation from the expected. It can be positive, negative or both and can address, create or result in opportunities and threats.

Note 2 to entry: Objectives can have different aspects and categories and can be applied at different levels.

Note 3 to entry: Risk is usually expressed in terms of risk sources, potential events, their consequences and their likelihood.

Note 4 to entry: This is the core definition of risk. As risks are specifically focused on *harm* ~~(3.5)~~(3.5) a discipline specific definition of *risk* ~~(3.3)~~(3.3) is used in this document in addition to the core risk definition.

[SOURCE: ISO 31000:2018, 3.1, ~~added~~modified — Added <organizational> domain and ~~added~~ Note 4 to entry]

3.5
harm
physical injury or damage to the health of people, or damage to property or the environment

[SOURCE: IEC 61508-4~~, ed. 2.0 (~~:2010~~)~~,, 3.1.1]

3.6
hazard
potential source of *harm* ~~(3.5)~~(3.5)

[SOURCE: IEC 61508-4~~, ed. 2.0 (~~:2010~~)~~,, 3.1.2]

3.7
hazardous event
event that may result in *harm* ~~(3.5)~~(3.5)

[SOURCE: IEC 61508-4~~, ed. 2.0 (~~:2010~~)~~,, 3.1.4]

3.8
system
~~combination~~arrangement of ~~interacting~~parts or elements ~~organized to achieve one or more~~that together exhibit a stated ~~purposes~~behaviour or meaning that the individual constituents do not

[SOURCE: ISO/IEC/IEEE 15288:~~2015, 4.1~~2023, 3.46, ~~removed~~modified — Removed the ~~four~~three Notes to entry]

3.9
systematic failure
failure, related in a deterministic way to a certain cause, which can only be eliminated by a modification of the design or of the manufacturing process, operational procedures, documentation or other relevant factors

[SOURCE: IEC 61508-4~~, ed. 2.0 (~~:2010~~)~~,, 3.6.6]

3.10
safety-related system
designated system that both

— implements the required safety functions necessary to achieve or maintain a safe state for the EUC; and

— is intended to achieve, on its own or with other E/E/PE safety-related systems and other risk reduction measures, the necessary safety integrity for the required safety functions

[SOURCE: IEC 61508-4, ed. 2.0 (:2010), 3.4.1]

3.11
safety function
function to be implemented by an E/E/PE safety-related system or other risk reduction measures, that is intended to achieve or maintain a safe state for the EUC, in respect of a specific *hazardous event* (3.7)(3.7)

[SOURCE: IEC 61508-4, ed. 2.0 (:2010), 3.5.1]

3.12
equipment under control
EUC
equipment, machinery, apparatus or plant used for manufacturing, process, transportation, medical or other activities

Note 1 to entry: The EUC control system is separate and distinct from the EUC.

[SOURCE: IEC 61508-4, ed. 2.0 (:2010), 3.2.1]

3.13
programmable electronic
PE
based on computer technology which can be comprised of hardware, software and of input and/or output units

Note 1 to entry: This term covers microelectronic devices based on one or more central processing units (CPUs) together with associated memories, etc.

EXAMPLE The following are all programmable electronic devices:

— microprocessors;

— micro-controllers;

— programmable controllers;

— application specific integrated circuits (ASICs);

— programmable logic controllers (PLCs);

— other computer-based devices (e.g. smart sensors, transmitters, actuators).

[SOURCE: IEC 61508-4, ed. 2.0 (:2010), 3.2.12]