



Publicly

This is a preview of ISO/PAS 8800:2024. [Click here to purchase the full version from the ANSI store.](#)

Specification

ISO/PAS 8800

Road vehicles — Safety and artificial intelligence

Véhicules routiers — Sécurité et intelligence artificielle

**First edition
2024-12**

This is a preview of ISO/PAS 8800:2024. Click [here](#) to purchase the full version from the ANSI store.



COPYRIGHT PROTECTED DOCUMENT

© ISO 2024

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

Foreword	vi
Introduction	vii
1 Scope	1
2 Normative references	1
3 Terms and definitions	2
3.1 General AI-related definitions.....	2
3.2 Data-related definitions.....	7
3.3 General safety-related definitions.....	9
3.4 Safety: Root cause-, error-and failure-related definitions.....	11
3.5 Miscellaneous definitions.....	12
4 Abbreviated terms	14
5 Requirements for conformity	15
5.1 Purpose.....	15
5.2 General requirements.....	15
6 AI within the context of road vehicles system safety engineering and basic concepts	16
6.1 Application of the ISO 26262 series for the development of AI systems.....	16
6.2 Interactions with encompassing system-level safety activities.....	17
6.3 Mapping of abstraction layers between the ISO 26262 series, ISO/IEC 22989 and this document.....	20
6.4 Example architecture for an AI system.....	22
6.5 Types of AI models.....	23
6.6 AI technologies of a ML model.....	23
6.7 Error concepts, fault models and causal models.....	24
6.7.1 Cause-and-effect chain.....	24
6.7.2 Root cause classes.....	26
6.7.3 Error classification based on the safety impact.....	27
7 AI safety management	28
7.1 Objectives.....	28
7.2 Prerequisites and supporting information.....	28
7.3 General requirements.....	28
7.4 Reference AI safety life cycle.....	31
7.5 Iterative development paradigms for AI systems.....	33
7.6 Work products.....	34
8 Assurance arguments for AI systems	35
8.1 Objectives.....	35
8.2 Prerequisites and supporting information.....	35
8.3 General requirements.....	36
8.4 AI system-specific considerations in assurance arguments.....	36
8.5 Structuring assurance arguments for AI systems.....	37
8.5.1 Context of the assurance argument.....	37
8.5.2 Categories of evidence.....	38
8.6 The role of quantitative targets and qualitative arguments.....	39
8.7 Evaluation of the assurance argument.....	40
8.8 Work products.....	41
9 Derivation of AI safety requirements	41
9.1 Objectives.....	41
9.2 Prerequisites and supporting information.....	42
9.3 General requirements.....	42
9.4 General workflow for deriving safety requirements.....	43
9.5 Deriving AI safety requirements on supervised machine learning.....	46
9.5.1 The need for refined AI safety requirements.....	46

This is a preview of ISO/PAS 8800:2024. [Click here to purchase the full version from the ANSI store.](#)

	9.5.4	Restricting the occurrence of AI output insufficiencies	50
	9.5.5	Metrics, measurements and threshold design	54
	9.5.6	Considerations for deriving safety requirements	55
	9.6	Work products	56
10		Selection of AI technologies, architectural and development measures	56
	10.1	Objectives	56
	10.2	Prerequisites	56
	10.3	General requirements	56
	10.4	Architecture and development process design or refinement	57
	10.5	Examples of architectural and development measures for AI systems	58
	10.6	Work products	62
11		Data-related considerations	62
	11.1	Objectives	62
	11.2	Prerequisites and supporting information	62
	11.3	General requirements	62
	11.4	Dataset life cycle	63
	11.4.1	Datasets and the AI safety lifecycle	63
	11.4.2	Reference dataset lifecycle	64
	11.4.3	Dataset safety analysis	65
	11.4.4	Dataset requirements development	71
	11.4.5	Dataset design	74
	11.4.6	Dataset implementation	75
	11.4.7	Dataset verification	75
	11.4.8	Dataset validation	76
	11.4.9	Dataset maintenance	77
	11.5	Work products	77
12		Verification and validation of the AI system	78
	12.1	Objectives	78
	12.2	Prerequisites and supporting information	78
	12.3	General requirements	78
	12.4	AI/ML specific challenges to verification and validation	80
	12.5	Verification and validation of the AI system	81
	12.5.1	Scope of verification and validation of the AI system	81
	12.5.2	AI component testing	84
	12.5.3	Methods for testing the AI component	86
	12.5.4	AI system integration and verification	88
	12.5.5	Virtual testing vs physical testing	88
	12.5.6	Evaluation of the safety-related performance of the AI system	89
	12.5.7	AI system safety validation	90
	12.6	Work products	91
13		Safety analysis of AI systems	91
	13.1	Objectives	91
	13.2	Prerequisites and supporting information	92
	13.3	General requirements	92
	13.4	Safety analysis of the AI system	93
	13.4.1	Scope of the AI safety analysis	93
	13.4.2	Safety analysis based on the results of testing	95
	13.4.3	Safety analysis techniques	95
	13.5	Work products	97
14		Measures during operation	97
	14.1	Objectives	97
	14.2	Prerequisites and supporting information	98
	14.3	General requirements	98
	14.4	Planning for operation and continuous assurance	99

This is a preview of ISO/PAS 8800:2024. [Click here to purchase the full version from the ANSI store.](#)

14.5	Continual, periodic re-evaluation of the assurance argument.....	100
14.6	Measures to assure safety of the AI system during operation.....	101
14.6.1	General.....	101
14.6.2	Technical safety measures.....	101
14.6.3	Safe operation guidance and misuse prevention in the field.....	102
14.7	Field data collection.....	103
14.8	Evaluation and continuous development.....	104
14.8.1	Field risk evaluation.....	104
14.8.2	Countermeasures addressing field risk.....	105
14.8.3	AI re-training, re-validation, re-approval and re-deployment.....	105
14.9	Work products.....	106
15	Confidence in use of AI development frameworks and software tools used for AI model development.....	106
15.1	Objectives.....	106
15.2	Prerequisites and supporting information.....	107
15.3	General requirements.....	107
15.4	Confidence in the use of AI development frameworks.....	107
15.5	Confidence in the use of tools used to support the AI-safety lifecycle.....	109
15.6	Principles for data-driven AI model training and evaluation.....	110
15.7	Work products.....	110
	Annex A (informative) Overview and workflow of this document.....	111
	Annex B (informative) Example assurance argument structure for an AI system.....	116
	Annex C (informative) ISO 26262 gap analysis for ML.....	130
	Annex D (informative) Detailed considerations on safety-related properties of AI systems.....	137
	Annex E (informative) STAMP/STPA example.....	139
	Annex F (informative) Identification of software units within NN-based systems.....	144
	Annex G (informative) Architectural and development measures for AI systems.....	147
	Annex H (informative) Typical performance metrics for machine learning.....	162
	Bibliography.....	167

This is a preview of ISO/PAS 8800:2024. [Click here to purchase the full version from the ANSI store.](#)

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 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).

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes 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 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. ISO 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.

This document was prepared by Technical Committee ISO/TC 22, *Road vehicles*, Subcommittee SC 32, *Electrical and electronic components and general system aspects*.

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.

This is a preview of ISO/PAS 8800:2024. [Click here to purchase the full version from the ANSI store.](#)

The purpose of this document is to provide industry-specific guidance on the use of AI systems in safety-related functions. It is not restricted to specific AI methods or specific vehicle functions.

This document defines a framework for managing AI safety that tailors or extends existing approaches currently defined in the ISO 26262 series and in ISO 21448.

Functional safety-related risks associated with malfunctioning behaviour of an AI system are addressed by tailoring or extending relevant clauses from ISO 26262-series.

The risks related to functional insufficiencies in the AI system are addressed by extending the concepts and guidance provided by ISO 21448. A causal model for understanding the sources of functional insufficiencies in the AI system is proposed. The model is used to derive a set of safety requirements on the AI system as well as a set of risk reduction measures.

NOTE 1 ISO 21448 is applicable to intended functionalities where proper situational awareness is essential to safety and where such situational awareness is derived from sensors and processing algorithms, especially functionalities of emergency intervention systems and systems with ISO/SAE PAS 22736 levels 1 to 5 for driving automation. It is therefore possible that systems utilize AI technologies that do not fall within the scope of ISO 21448.

EXAMPLE 1 ISO 21448 does not apply to the development of an engine control unit that uses AI to optimize its performance whereas this document does.

This document recognizes that due to the wide range of applications of AI and associated safety requirements, as well as the rapidly evolving state-of-the-art, it is not possible to provide detailed requirements on the process or product characteristics required to achieve an acceptably low level of residual risk associated with the use of AI systems. Therefore, in addition to providing guidance for tailoring or extending the ISO 26262 series and ISO 21448, this document focuses on the principles that support the creation of a project-specific assurance argument for the safety of the AI elements within on-board vehicle systems. This includes proposing risk reduction measures during the design and operation phases using an iterative approach to reducing risk as outlined in ISO/IEC Guide 51.

Hazard analysis and risk analysis are beyond the scope of this document. These are considered a part of the vehicle level systems safety engineering activities described in the ISO 26262 series and ISO 21448, or in application of specific standards such as ISO TS 5083.

ISO/IEC TR 5469 provides generic guidance for the application of AI technologies as part of safety functions, independent of specific industry sectors. Many of the concepts outlined in ISO/IEC TR 5469 can be applied in the context of road vehicles. There is therefore a close relationship to concepts described within this document and ISO/IEC TR 5469.

ISO/IEC TR 5469 provides classification schemes to determine the safety requirements on the AI/ML function. These include the usage level and AI technology class.

The usage level is related to the nature of the task being performed by the engineered AI system.

NOTE 2 The usage levels are described in ISO/IEC TR 5469:2024, 6.2.

The technology class is related to the problem complexity and the transferability of existing standards to demonstrating an adequate level of safety based on properties of the target function and the AI technology used.

NOTE 3 For the technology classes, see ISO/IEC TR 5469:2024, 6.2.

This document does not explicitly call out the classes and usage levels of ISO/IEC TR 5469.

EXAMPLE 2 For some AI technology, the application of ISO 26262 is deemed to be sufficient. This corresponds to Class I of ISO/IEC TR 5469.

The guidance outlined within this document is relevant for all usage of AI for which safety requirements can foreseeably be allocated either through:

- a) the use of AI for the functionality itself;

This is a preview of ISO/PAS 8800:2024. [Click here to purchase the full version from the ANSI store.](#)

NOTE 4 These usages correspond to the usage levels A1, A2, C of ISO/IEC TR 5469. In all cases, the applicability of the guidance provided within this document can be determined by the allocation of safety requirements to the AI technology, whereas the usage levels of ISO/IEC TR 5469 can be used to support the requirements elicitation process.

This document is aligned with standards and documents developed by ISO/IEC JTC1/SC42. AI-specific definitions are used from ISO/IEC 22989, unless in conflict with safety-specific definitions.

Other documents developed within ISO/IEC JTC1/SC42 can be used to provide additional guidance on specific aspects of AI that are relevant to safety-related properties. Examples of such documents include ISO/IEC TR 24027 and ISO/IEC TR 24029-1.

This document harmonizes the concepts already described in ISO 21448:2022, Annex D.2 and ISO/TS 5083:20—¹⁾, Annex B whilst extending these with specific guidance regarding the definition of safety requirements of machine learning (ML), ML safety analyses and the creation of associated safety evidence during the development and deployment lifecycle.

ISO/TS 5083:20—, Annex B is an application of this document to automated driving systems (ADS).

The relationship with the above-mentioned documents is summarized in [Table 1-1](#).

Table 1-1 — How this document relates to other publications on AI safety

Publication	Relationship with this document
ISO/IEC 22989	AI-specific definitions are used from ISO/IEC 22989, unless in conflict with safety-specific definitions. Safety-related properties are a subset of generic AI properties described in ISO/IEC 22989.
ISO/IEC TR 5469	This document does not explicitly call out the classes and usage levels of ISO/IEC TR 5469. This document considers and adapts to road vehicles the general framework described in ISO/IEC TR 5469 on safety properties, virtual testing and physical testing, confidence in use of AI development frameworks and architectural redundancy patterns.
ISO 26262	This document is a tailoring or extension of ISO 26262 for AI elements of the system. See Clause 5 for details.
ISO 21448	This document is a tailoring or extension of ISO 21448 for AI elements of the system. See Clause 5 for details.
ISO TS 5083:20—	ISO TS 5083:20—, Annex B is an application of this document to automated driving systems (ADS).

This document adds the following contents with respect to the documents listed in [Table 1-1](#):

- tailoring or extensions of ISO 26262 and ISO 21448 required specifically for AI elements of the system (referred to as AI systems);
- a conceptual model for reasoning about errors and their causes specific to AI systems;
- a reference AI safety lifecycle;
- the safety assurance argument for AI systems;
- a method for deriving AI safety requirements for AI systems;
- considerations for the design of safe AI systems;
- considerations on data management for the AI systems;

1) Under preparation. Stage at the time of publication: ISO/DTS 5083.

This is a preview of ISO/PAS 8800:2024. [Click here to purchase the full version from the ANSI store.](#)

- a safety analysis approach for AI systems (focused on insufficiencies);
- activities during operation required to ensure the continuous AI safety.