ISO/PAS 8800:2024-12 (E)

Road vehicles - Safety and artificial intelligence

Contents

Page

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
9.5.2 Derivation of refined AI safety requirements to manage uncertainty 47
9.5.3 Refinement of the input space definition for AI safety lifecycle 50
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
14.4.1 Safety risk of the AI system during operation phase ... 99
14.4.2 Safety activities during the operation phase ... 99
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**