E DIN EN 18229-2:2027-01 (E)

Erscheinungsdatum: 2026-11-27

AI trustworthiness framework - Part 2: Accuracy and robustness; English version prEN 18229-2:2026

Inhalt

Seite

European foreword 4

1 Scope 6

2 Normative references 6

3 Terms and definitions 6
3.1 General terms 7
3.2 Terms related to AI 7
3.3 Terms related to risk management and quality management 10
3.4 Terms related to usability 10

4 Abbreviations 11

5 Transparency in design and development 11
5.1 Transparency objectives 11
5.2 Transparency mechanism selection 12
5.2.1 General 12
5.2.2 Step 1 — Identification of information necessary for the deployer to interpret AI system's output 12
5.2.3 Step 2 — Identification of transparency gaps 14
5.2.4 Step 3 — Selection of transparency mechanisms 14
5.2.5 Step 4 — Closure of the transparency assessment 15
5.3 Output interpretability and human-machine interface design 15
5.3.1 Output design 15
5.3.2 Human-machine interface design 16
5.3.3 Interpretation tools 16
5.4 Verification and validation of transparency mechanisms 17
5.4.1 General 17
5.4.2 Transparency evaluation criteria 17
5.4.2.1 General 17
5.4.2.2 Form of the transparency evaluation criteria 18
5.4.3 Verification 18
5.4.4 Validation 18
5.4.5 Cross-method determinations 18

6 Documentation 19
6.1 Instructions for use 19
6.1.1 Purpose of the instructions for use 19
6.1.2 Format and quality requirements 19
6.1.2.1 Format 19
6.1.2.2 Content quality and communication 19
6.1.2.3 Version control and updates 20
6.1.3 Content of the instructions for use 20
6.1.3.1 General information on the AI system provider identification 20
6.1.3.2 AI System characteristics, capabilities and limitations 20
6.1.3.3 Pre-determined changes 23
6.1.3.4 Human oversight measures 23
6.1.3.5 Computational and hardware resources, expected lifetime and maintenance 24
6.1.3.6 Logging mechanisms 25
6.2 Technical documentation 25

Annex A (informative) Validation methods for transparency mechanisms **27**
A.1 General **27**
A.2 Usability-testing method **27**
A.3 Sample-size method **28**
Annex B (informative) Content of the instructions for use for hypothetical AI systems **29**
B.1 General **29**
B.2 Intended purpose **29**
B.3 Accuracy, robustness and cybersecurity performance **29**
B.4 Performance regarding specific persons or groups **29**
B.5 Input and training data specifications **29**
B.6 Output interpretation guidance **30**
B.7 Built-in technical human oversight measures **30**
B.8 Procedural human oversight measures recommended for the deployer **30**
B.9 Competencies required of persons assigned to human oversight **30**
B.10 Resource requirements **30**
B.11 Expected lifetime **31**
B.12 Maintenance and care measures **31**
Annex ZA (informative) Relationship between this European Standard and the requirements of Regulation (EU) 2024/1689 aimed to be covered **32**
Bibliography **33**

Tables

Table 1 — Example of clinical testing **13**
Table ZA.1 — Correspondence between this European Standard and Regulation 2024/1689 ..... **32**