E DIN EN 18283:2026-11 (E)

Erscheinungsdatum: 2026-10-23

Artificial Intelligence - Concepts, measures and requirements for managing bias in AI systems; English version prEN 18283:2026

Contents

Page

European foreword 5

Introduction 6

1 Scope 7

2 Normative references 7

3 Terms and definitions 7
3.1 General AI and bias related terms 7
3.2 Terms related to the EU AI Act 8
3.3 General terms related to risk 9
3.4 Groups and their identification 11
3.5 Bias management processes 13
3.6 Terms related to data 14
3.7 Terms related to testing 15

4 Abbreviated terms 16

5 Overview 17

6 Bias profile 20
6.1 General 20
6.2 Inputs to the bias profile 21
6.3 External inputs to the bias profile 22
6.4 Bias scenario construction 23
6.4.1 General 23
6.4.2 Bias scenario contents 23
6.4.3 Bias scenario example 24
6.5 Bias profile output to external processes 24

7 Bias management 25
7.1 The overall process 25
7.1.1 Bias and risk management 25
7.1.2 Bias management activities 25
7.1.3 Bias management in the context of the AI Act 27
7.2 Identification of relevant groups 28
7.2.1 General 28
7.2.2 Identification of relevant categories – requirements 29
7.2.3 Derivation of relevant groups - requirements 31
7.3 Identification of bias-related hazards 31
7.4 Determination of metrics 32
7.4.1 General 32
7.4.2 Quality assessment methods 32
7.4.3 Bias metrics 33
7.4.4 Confidence levels and the family-wise error rate 34
7.5 Reviewing and updating the bias profile 35

8 Bias estimation 36
8.1 General 36
8.2 Availability of testing data 36
8.3 Data set analysis 37

8.3.1 General ... 37
8.3.2 Information for data set bias testing ... 37
8.3.3 Data set bias testing activities ... 38
8.4 AI component output analysis ... 38
8.4.1 General ... 38
8.4.2 Information for AI component bias testing ... 38
8.4.3 AI component bias testing activities ... 39
8.5 AI system output analysis ... 40
8.5.1 General ... 40
8.5.2 Information for AI system output bias testing ... 40
8.5.3 AI system output bias testing activities ... 40
8.6 AI system outcome analysis ... 41
8.6.1 General ... 41
8.6.2 Information for AI system outcome bias testing ... 41
8.6.3 AI system outcome bias testing activities ... 41
9 Bias evaluation ... 42
10 Metrics ... 43
10.1 General ... 43
10.2 Metric notation ... 44
10.2.1 Notation for Metric calculations ... 44
10.2.2 Probabilistic notation ... 44
10.3 Classification metrics ... 45
10.3.1 Demographic Disparity Difference (DDD) ... 45
10.3.2 Demographic Disparity Ratio ... 46
10.3.3 Conditional Demographic Disparity Difference (CDDD) ... 48
10.3.4 Predictive parity ... 50
10.3.5 Equal opportunity ... 51
10.3.6 Predictive equality ... 52
10.4 Regression metrics ... 54
10.4.1 Mean output distance (MOD) ... 54
10.4.2 Loss difference (LD) ... 55
10.5 Ranking metrics ... 57
10.5.1 Averaged representation difference (ARD) ... 57
10.6 Outcome metrics ... 59
10.6.1 General ... 59
10.6.2 Outcome demographic disparity difference (ODDD) ... 59
10.6.3 Outcome demographic disparity ratio (ODDR) ... 60
10.7 Data metrics ... 62
10.7.1 General ... 62
10.7.2 Relevant group information ... 62
10.7.3 Label imbalance in data sets ... 63
10.8 Meta-Metrics ... 65
10.8.1 General ... 65
10.8.2 Requirements for selecting the conditioning features ... 65
10.8.3 Stratified conditional measurements ... 65
10.8.4 Distribution-conditioned measurements ... 66
11 Metric template ... 69
12 Bias mitigation ... 70
12.1 General ... 70
12.2 Categorizing bias mitigation methods ... 70
12.3 Template for bias mitigation methods ... 71
12.4 Bias mitigation methods ... 71
12.4.1 Personnel training ... 71
12.4.2 Personnel diversity ... 72
12.4.3 Adhering to interaction design and user experience requirements ... 72
12.4.4 Adhering to user interface design requirements ... 73

12.4.5 Re-sampling ... 73
12.4.6 Re-weighting ... 74
12.4.7 Distribution transformations ... 74
12.4.8 Synthetic data generation ... 75
12.4.9 Adversarial debiasing ... 75
12.4.10 Threshold adjustment ... 75
12.4.11 Calibration ... 76
12.4.12 Re-ranking or output modification ... 76
12.5 Documentation and justification ... 76
12.6 Bias mitigation and risk management ... 77

13 Information gathering and consultation ... 77
13.1 Collection of information and evidence ... 77
13.2 Consultation of interested parties ... 78

Annex A (informative) Identification of the relevant groups for bias testing Use case: AI-supported consumer-loan eligibility assessment system ... 80
A.1 General ... 80
A.2 Intended purpose of the AI system and potentially affected persons ... 80
A.3 Data description ... 81
A.3.1 General ... 81
A.3.2 Loan-application data ... 81
A.3.3 Credit bureau records ... 81
A.3.4 Open-banking data ... 81
A.3.5 Utility and telecommunications payment data ... 81
A.3.6 Identity and address-verification data ... 81
A.3.7 Application-channel data ... 81
A.3.8 Geographic data ... 82
A.3.9 Historical lender records ... 82
A.4 Potentially negatively affected persons (including vulnerable groups) ... 82
A.5 Risk scenarios involving bias-related hazards ... 82
A.6 Post-market monitoring information ... 83
A.7 Real-world testing results ... 83
A.8 Results from previous bias assessments ... 83
A.9 Categories protected under the EU Charter ... 83
A.10 Derivation of relevant groups ... 84
A.11 Identification of intersectionality ... 86
A.12 State of the art review ... 87
A.13 Conclusion ... 87

Annex B (informative) Possible causes of bias ... 88

Annex ZA (informative) Relationship between this European Standard and the requirements of Regulation (EU) 2024/1689 aimed to be covered ... 90

Bibliography ... 91

Tables

Table ZA.1 — Correspondence between this European Standard and Regulation (EU) 2024/1689 ... 90