DIN/TS 92004:2024-07 (E)

Artificial intelligence - Quality requirements and processes - Risk identification and analysis for AI systems along the entire life cycle; Text in English

Inhalt

Seite

Foreword 4

Introduction 5

1 Scope 6

2 Normative references 6

3 Terms and definitions 6

4 Introduction to AI risk assessment 7
4.1 Notion of risk 7
4.2 AI risk assessment along the lifecycle 8

5 Reliability specifications for AI systems 9
5.1 General 9
5.2 Risk identification 9
5.3 Risk analysis 11
5.3.1 General 11
5.3.2 Reliability in standard cases 11
5.3.3 Robustness 14
5.3.4 Intercepting errors at model level 17
5.3.5 Uncertainty estimation 20

6 Unwanted bias 21
6.1 General 21
6.2 Risk identification 21
6.3 Risk analysis 22
6.3.1 General 22
6.3.2 Identifying sensitive characteristics of (potentially disadvantaged) groups/individuals 23
6.3.3 Analysing and quantifying data bias 23
6.3.4 Checking for legal or regulatory relevance of the AI system's potential biases 23
6.4 Best practices 24

7 Autonomy and control 24
7.1 General 24
7.2 Risk identification 24
7.3 Risk analysis 26
7.3.1 General 26
7.3.2 Appropriate and responsible task distribution between humans and AI application 26
7.3.3 Information and empowerment of users and affected persons 31

Annex A (informative) Performance metrics for Machine Learning tasks 37
A.1 Reliability metrics 37
A.2 Unwanted bias metrics 38

Annex B (informative) Begriffe 39

Bibliography 40

Figures

Figure 1 — Example AI system life cycle model with AI system-specific processes [1] 8

Figure 2 — Risk management process [2] 9

Tables

Table 1 — Reliability risk categorization 10

Table 2 — Unwanted bias risk categorization 22

Table 3 — Autonomy and control risk categorization 25