DIN CEN/CLC ISO/IEC/TS 12791:2025-03 (D)

Informationstechnik - Künstliche Intelligenz - Behandlung von unerwünschtem Bias bei Klassifizierungs- und Regressionsaufgaben des maschinellen Lernens (ISO/IEC TS 12791:2024); Deutsche Fassung CEN/CLC ISO/IEC/TS 12791:2024

Inhalt

Seite

Europäisches Vorwort 7

Vorwort 8

Einleitung 9

1 Anwendungsbereich 10

2 Normative Verweisungen 10

3 Begriffe 10
3.1 Allgemeines 10
3.2 Künstliche Intelligenz 12
3.3 Bias 13
3.4 Prüfung 15

4 Abkürzungen 16

5 Behandlung von unerwünschtem Bias im Lebenszyklus von KI-Systemen 17
5.1 Anfangsphase 17
5.1.1 Identifizierung der interessierten Parteien 17
5.1.2 Erfordernisse und Anforderungsdefinition der interessierten Parteien 18
5.1.3 Beschaffung 19
5.1.4 Datenquellen 21
5.1.5 Integration mit Risikomanagement 22
5.1.6 Abnahmekriterien 23
5.2 Gestaltung und Entwicklung 23
5.2.1 Merkmalsdarstellung 23
5.2.2 Angemessenheit der Metadaten 24
5.2.3 Datenannotationen 24
5.2.4 Anpassung von Daten 25
5.2.5 Methoden zur Handhabung identifizierter Risiken 25
5.3 Verifizierung und Validierung 25
5.3.1 Allgemeines 25
5.3.2 Statische Prüfung von in der Entwicklung verwendeten Daten 26
5.3.3 Dynamische Prüfung 27
5.4 Reevaluierung, kontinuierliche Validierung, Betrieb und Überwachung 28
5.4.1 Allgemeines 28
5.4.2 Externe Veränderung 29
5.5 Entsorgung 30

6 Techniken zum Umgang mit unerwünschtem Bias 30
6.1 Allgemeines 30
6.2 Algorithmische Techniken und Trainingstechniken 31
6.2.1 Allgemeines 31
6.2.2 Vortrainierte Modelle 32
6.3 Datentechniken 32

7 Umgang mit unerwünschtem Bias im Lebenszyklus eines verteilten KI-Systems 33

Anhang A (informativ) Darstellung der Lebenszyklusprozesse 35

Anhang B (informativ) Mögliche Auswirkungen von unerwünschtem Bias auf verschiedene spezifische Arten von Benutzern **37**
B.1 Erfordernisse von Beschäftigten **37**
B.2 Vulnerable Verbraucher **37**
Literaturhinweise **39**

Tabellen

Tabelle A.1 — Darstellung des Lebenszyklus von KI-Systemen **35**