E DIN EN 18282:2026-11 (D/E)

Erscheinungsdatum: 2026-10-09

Künstliche Intelligenz - Cybersicherheitsspezifikationen für KI-Systeme; Deutsche und Englische Fassung prEN 18282:2026

Artificial intelligence - Cybersecurity specifications for AI Systems; German and English version prEN 18282:2026

Inhalt

Seite

Europäisches Vorwort 7

Einleitung 8

1 Anwendungsbereich 10

2 Normative Verweisungen 10

3 Begriffe 10
3.1 Begriffe aus der EU-KI-Verordnung und anderen EU-Verordnungen 10
3.2 Andere zugehörige Begriffe 13

4 Abkürzungen 20

5 Rahmenwerk für die Cybersicherheit des KI-Systems 20
5.1 Allgemeine Anforderungen 20
5.2 Zusammenhang zwischen Bedrohungen, Schwachstellen und Maßnahmen 21
5.3 Geforderte Ergebnisse 21
5.4 Dokumentation des Rahmenwerks für Cybersicherheit 22

6 Bestimmung relevanter Umstände für KI-Cybersicherheit 22
6.1 Allgemeine Anforderungen 22
6.2 Anwendungsbereich und operativer Kontext 23
6.3 Für die Cybersicherheit relevante Werte 23
6.4 Betriebs- und Bereitstellungsumstände 24

7 Identifizierung von KI-spezifischen Schwachstellen in der Cybersicherheit 24
7.1 Allgemeine Anforderungen 24
7.2 Prozess der Identifizierung von Schwachstellen 24
7.3 Anwendungsbereich KI-spezifischer Schwachstellen 24
7.4 Verknüpfung zu Bedrohungen und Maßnahmen 25

8 Identifizierung von relevanten Cybersicherheitsbedrohungen 25
8.1 Allgemeine Anforderungen 25
8.2 Prozess zur Identifizierung von Bedrohungen 25
8.3 Anwendungsbereich KI-spezifischer Bedrohungen 26
8.4 Verknüpfung zu Schwachstellen und Maßnahmen 27

9 Bestimmung relevanter Risiken für KI-Cybersicherheit 28
9.1 Allgemeine Anforderungen 28
9.2 Bestimmung der Relevanz von Cybersicherheitsrisiken 28
9.3 Risikoakzeptanz und Verknüpfung mit Maßnahmen zur Risikokontrolle bei KI-Systemen 29

10 Maßnahmen auswählen und umsetzen 29
10.1 Data Poisoning 29
10.1.1 Verhüten 29
10.1.2 Erkennen 30
10.1.3 Reagieren 31
10.1.4 Beseitigen 31
10.1.5 Kontrollieren 31

10.2 Model Poisoning ... 32
10.2.1 Verhüten ... 32
10.2.2 Erkennen ... 33
10.2.3 Reagieren ... 33
10.2.4 Beseitigen ... 34
10.2.5 Kontrollieren – Eingrenzung von Auswirkungen ... 34
10.3 Feindliche Angriffe oder Modellumgehung ... 35
10.3.1 Allgemeines ... 35
10.3.2 Verhüten ... 35
10.3.3 Erkennen ... 36
10.3.4 Reagieren ... 37
10.3.5 Beseitigen ... 37
10.3.6 Kontrollieren ... 37
10.4 Angriffe auf die Vertraulichkeit ... 38
10.4.1 Verhüten ... 38
10.4.2 Erkennen ... 39
10.4.3 Reagieren ... 40
10.4.4 Beseitigen ... 40
10.4.5 Kontrollieren ... 40
10.5 Modellmängel ... 41
10.5.1 Verhüten ... 41
10.5.2 Erkennen ... 41
10.5.3 Reagieren ... 42
10.5.4 Beseitigen ... 42
10.5.5 Kontrollieren ... 43
10.6 Bedrohungen im Zusammenhang mit generativen KI-Modellen (einschließlich LLMs, en: large language models) ... 43

11 Verifizierungs- und Testanforderungen ... 44
11.1 Allgemeine Anforderungen ... 44
11.2 Auslöser im Lebenszyklus ... 44
11.3 Arten von Cybersicherheitstests ... 45
11.4 Tests der Widerstandsfähigkeit gegenüber Poisoning ... 45
11.5 Adversarial Testing ... 45
11.5.1 Allgemeine Anforderungen ... 45
11.5.2 Identifizierung von relevanten feindlichen Angriffen ... 46
11.5.3 Entwicklung und Durchführung von Tests ... 46
11.5.4 Annahmen zum Angreiferwissen ... 46
11.6 Vertraulichkeitstests ... 47
11.7 Tests zur Modell- oder Algorithmusausnutzung ... 47
11.8 Kombinations- und Red-Team-Tests ... 47
11.9 Abnahmekriterien ... 48

12 Dokumentationsanforderungen für KI-Cybersicherheit ... 48
12.1 Allgemeine Anforderungen ... 48
12.2 Dokumentation relevanter Umstände ... 48
12.3 Dokumentation von Schwachstellen ... 48
12.4 Dokumentation von Bedrohungen ... 49
12.5 Dokumentation von Cybersicherheitsrisiken ... 49
12.6 Dokumentation von technischen Maßnahmen und Tests ... 49
12.7 Dokumentation der Betriebsanleitungen ... 50

Anhang A (informativ) Erläuternde Leitlinien zu Abschnitt 5 bis Abschnitt 12 ... 51
A.1 Zweck und Nutzung ... 51
A.2 Überblick über die KI-Cybersicherheit während des Lebenszyklus ... 51
A.3 Zusammenhang zwischen Abschnitten dieses Dokuments ... 51
A.4 Darstellung von KI-Cybersicherheitsaktivitäten ... 52
A.5 Anwendung von KI-Cybersicherheitsmaßnahmen ... 53
A.6 Interaktion mit anderen Normen ... 53

A.7 Kontinuierliche Verbesserung und Weiterentwicklung von Bedrohungen 54
A.8 Beispiele für Umstände, Kontext und Umgebung 54
A.9 Beispiele für KI-spezifische Bedrohungen 54
A.10 Beispiele für KI-spezifische Schwachstellen 55
A.11 Beispiele für Maßnahmen 56
Anhang B (informativ) Empfohlene organisatorische Kontrollen 57
B.1 KI-Managementsystem 57
B.2 Anwendung bewährter Prozesse aus dem Software-Engineering bei der KI-Entwicklung 57
B.3 Einbeziehung von KI-Überlegungen in das Informationssicherheitsmanagementsystem 57
Anhang C (informativ) Andere einschlägige Normen 58
Anhang D (normativ) KI-spezifische Werte 59
Anhang ZA (informativ) Zusammenhang zwischen dieser Europäischen Norm und den grundlegenden Anforderungen der abzudeckenden Verordnung 2024/1689 61
Literaturhinweise 64

Bilder

Bild A.1 — KI-Cybersicherheitsaktivitäten über den gesamten Systemlebenszyklus 52
Bild A.2 — Beispiele für Bedrohungen 55

Tabellen

Tabelle D.1 — Liste von KI-spezifischen Werten 59
Tabelle ZA.1 — Zusammenhang zwischen dieser Europäischen Norm und Verordnung (EU) 2024/1689 Artikel 15 Absatz 5 61