Bundesnetzagentur

Trendthemen für DIN SPECs: Grundpfeiler der KI-Normung

Taras Holoyad

Workshop „Anforderungen an KI-Prüfungen: Standardisierungspotenziale identifizieren“

Berlin, 11.09.2023

DIN SPECs zu KI:

- Einführung
- Mehrwert der DIN SPEC
- Vorstellung der Normungsbedarfe
- Priorisierung

Normungsroadmap KI Version 2:

570 Fachleute

9 Arbeits-Gruppen/ Schwer-punkte

18 Mitglieder in der Koordinie-rungsgruppe

318 Autor*innen

116 Handlungs-bedarfe

6 Übergreifende Empfeh-lungen

448 Seiten

Normungsroadmap KI Version 2:

Umsetzung in Normung

Ausgewählte Bedarfe:

- Operationalisierung der „Erklärgüte“ von XAI-Methoden
- Entwicklung von Funktionalitätsklassen für KI-Technologien
- Entwicklung von Werkzeugkriterien für die Prüfung von KI-Systemen
- Standardisierung von Guidance-Dokumenten für die Abbildung von Risiken eines Systems, mit Fokus auf Funktionalität von KI-Komponenten
- Entwicklung von ineinandergreifenden Standards für KI-Systeme und notwendiger Konformitätsbewertungsverfahren
- Entwicklung von Qualifikationskriterien für Prüfer und Zertifizierter zu „Cybersecurity“ und „Privacy“ für KI
- Vernetzung aller Akteur*innen
- Definition von Kontrollpunkten

Ausgewählte Bedarfe:

- Operationalisierung der „Erklärgüte“ von XAI-Methoden
- Entwicklung von Funktionalitätsklassen für KI-Technologien
- Entwicklung von Werkzeugkriterien für die Prüfung von KI-Systemen
- Standardisierung von Guidance-Dokumenten für die Abbildung von Risiken eines Systems mit Fokus auf Funktionalität von KI-Komponenten
- Entwicklung von ineinandergreifenden Standards für KI-Systeme und notwendiger Konformitätsbewertungsverfahren
- Entwicklung von Qualifikationskriterien für Prüfer und Zertifizierter zu Cybersecurity und Privacy für KI
- Vernetzung aller Akteur*innen
- Definition von Kontrollpunkten

Mehrwert der DIN SPEC (1/2)

- **Deutscher Standard** zu **KI**
- **Nennung** der **Beteiligten** im Vorwort
- **Vollständige Finanzierung** durch Projekt „**Zertifizierte KI**“
- Erarbeitung: Ausschließlicher **Fokus** auf das Dokument
- Hohe **Geschwindigkeit**:
  - Keine öffentliche Kommentierung erforderlich
  - Mehrheitsentscheide & keine Konsenspflicht

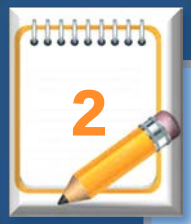
Mehrwert des Vorhabens (2/2)

Vorarbeit für weltweite Standardisierung!

- **Europa: CEN-CENELEC JTC 21 „Artificial Intelligence“**
- **International: ISO/IEC JTC 1/SC 42 „Artificial Intelligence“**

Themen für DIN SPECs

Operationalisierung der „Erklärgüte“ von XAI-Methoden



Entwicklung von Funktionalitätsklassen für KI-Technologien

Entwicklung von Werkzeugkriterien für die Prüfung von KI-Systemen

Vorstellung von Thema 1 (1/4)

Operationalisierung der „Erklärgüte“ von XAI-Methoden

Hintergund:

Explainable Artificial Intelligence (XAI) - Methoden:

- **Schaffung von Erklärbarkeit für Entscheidungsfindung & Wissensverarbeitung**
- **Unterstützung bei ethischer und rechtlicher Bewertung von KI-Systemen**

Operationalisierung der „Erklärgüte“ von XAI-Methoden

a) Anforderungen an XAI-Methoden

- Abhängig von
  - Anwendungsfällen & Interessierten
  - Was will man genau erklären?
  - Welche Entscheidung möchte man treffen?
  - Welche Handlungen kann man rechtfertigen?
  - Rolle einbezogener Daten

    z.B. Test, Trainingsdaten/Verteilung,

    Einfluss von Punkten auf ein Modell

Operationalisierung der „Erklärgüte“ von XAI-Methoden

b) Ground truth-Referenzdatensätze & Metriken

- Ground truth-Datenerzeugung:
  - mathematische Bildungsvorschriften,
    physikalische Simulation und Manipulation realer Daten
- Metriken für ground truth-Datensätze, z.B.
  - Precision/Recall
  - Metriken aus der Signaldetektionstheorie

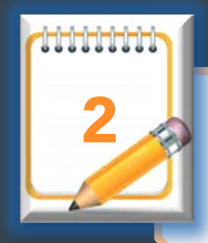
Vorstellung von Thema 1 (4/4)

Operationalisierung der „Erklärgüte“ von XAI-Methoden

c) **Verifizierung:** XAI bei Qualitätssicherung

- Nutzen von XAI für Qualitätssicherung bei ML-basierten Systemen
- Praktische Relevanz und Auswirkungen
- Benchmarking-Ansätze zur Beurteilung der Qualitätssicherung
- Generierung von Antworten mittels XAI-Methoden
  - ➢ Breite Bewertbarkeit von XAI-Methoden

Vorstellung von Thema 2 (1/4)



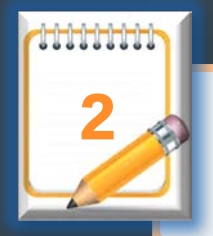
Entwicklung von Funktionalitätsklassen für KI-Technologien

Hintergund:

KI-Produkte: Spezifische Anforderungen an Vertrauen

- **Erfüllung von Anforderungen an KI-Systeme mittels technischer Funktionen, z.B.**
  - **Fehlererkennung**
  - **Abwehr von Angriffen**

Vorstellung von Thema 2 (3/4)



Entwicklung von Funktionalitätsklassen für KI-Technologien

a) Definition und Klassifizierung von Funktionalitäten:

- Beschreibung von Funktionalitäten im Zusammenhang mit dem Kontext eines KI-Systems

- Funktionalitätsspezifische Formulierung von Anforderungen, z.B. an Fehlererkennung

Entwicklung von Funktionalitätsklassen für KI-Technologien

b) Charakteristiken für Vertrauen in Funktionalitäten:

- Bewertung von Korrektheit und Wirksamkeit von Funktionalitäten

- Überprüfbarkeit der Funktionalitäten im Entwicklungs- und Betriebsprozess

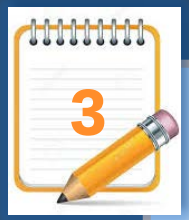
Entwicklung von Funktionalitätsklassen für KI-Technologien

c) Baukasten für Funktionalitätsklassen

- Überblick über Funktionalitäten,
  die KI-Systeme bei Entwicklung & Betrieb begleiten können

- In Abhängigkeit von KI-System und Kontext:
  - ➢ Überblick über umsetzbare Funktionalitäten
  - ➢ Einheitlich strukturierte, umsetzbare Anforderungen

Vorstellung von Thema 3 (1/3)

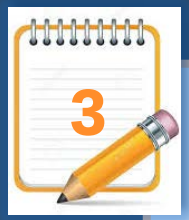


Entwicklung von Werkzeugkriterien für die Prüfung von KI-Systemen

Hintergund:

Charakterisierung von Werkzeugen für KI-Prüfungen:

- **Erfordernis:**
  - Entwicklung spezifischer Kriterien für Prüfung und Zertifizierung der Werkzeuge
- **Teil eines KI-Zertifizierungsprogramms**



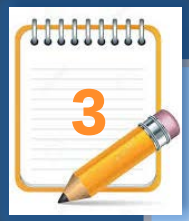
Entwicklung von Werkzeugkriterien für die Prüfung von KI-Systemen

a) Beschreibung von Werkzeugen für KI-Prüfungen:

- Charakterisierung messbarer Gütekriterien bei KI-Systemen

- Festlegung von Gütecharakteristiken für Werkzeuge, um hohe Aussagekraft und Zuverlässigkeit sicherzustellen

Vorstellung von Thema 3 (3/3)



Entwicklung von Werkzeugkriterien für die Prüfung von KI-Systemen

b) Prüfverfahren für Werkzeuge:

- Festlegung grundlegender Anforderungen an Prüfung und Zertifizierung der Werkzeuge
- Beschreibung von Prüfkriterien und Prüfmethoden
- Konzipierung von Testabläufen

Vorstellung von Thema 3 (3/3)

I **Operationalisierung der „Erklärgüte“ von XAI-Methoden**

a) Anforderungen an XAI-Methoden
b) Ground truth-Referenzdatensätze & Metriken
c) Verifizierung: XAI bei Qualitätssicherung

II **Entwicklung von Funktionalitätsklassen für KI-Technologien**

a) Definition und Klassifizierung von Funktionalitäten
b) Charakteristiken für Vertrauen in Funktionalitäten
c) Baukasten für Funktionalitätsklassen

III **Entwicklung von Werkzeugkriterien für die Prüfung von KI-Systemen**

a) Beschreibung von Werkzeugen für KI-Prüfungen
b) Prüfverfahren für Werkzeuge

Bundesnetzagentur

Trendthemen für DIN SPECs:
Grundpfeiler der KI-Normung

Taras Holoyad
Standardisierung künstlicher Intelligenz

06131/18 1459
Taras.Holoyad@bnetza.de