ZERTIFIZIERTE KI

Dr. Maximilian Poretschkin, Konsortialleitung

11. September 2023

Bildquelle: Gorodenkoff/stock.adobe.com

Ergebnisse der 1. Version Normungsroadmap KI auf Digital-Gipfel veröffentlicht

Zwei Deliverables zur Etablierung eines Prüfverfahrens:

- Prüfframework, das Vergleichbarkeit von Prüfungen garantiert (und kompatibel mit bestehenden IT-Prüfverfahren ist!)
  - Prozessprüfungen (Standards zur Entwicklung und Betrieb von KI-Systemen)
  - Produktprüfungen (Überprüfung von zugesicherten Eigenschaften)
  - Differenzierte Assurance Levels / Prüftiefen

- Kriterienwerke, welche Anforderungen an Vertrauenswürdigkeit operationalisieren und KI-spezifische Herausforderungen abbilden
  - Use Case Abhängigkeit bei der Formulierung ist Herausforderung (Metriken, Schwellwerte)
  - Völlig neue Prüfwerkzeuge und Prüfmethoden benötigt

Struktur Flagship-Projekt ZERTIFIZIERTE KI

ZERTIFIZIERTE KI
Qualität sichern. Fortschritt gestalten.
www.zertifizierte-ki.de

ZERTIFIZIERTE KI

Prüfgrundlagen	Bedarfsanalyse	Anwenderkreise	Prüfökosystem	Gesellschaftlicher Diskurs
▪ Prüfscope ▪ Kriterienwerke ▪ Prüftiefen ▪ Anforderungen Prüfwerkzeuge ▪ Konzept Prüfinfrastruktur	▪ Kundenanalyse ▪ Wirkungsanalyse ▪ Geschäftsmodell-entwicklung	▪ Initialisierung im ersten Jahr ▪ Spezifizierung entlang der Bedarfe	▪ Plattform für prototypische Prüfwerkzeuge ▪ Konzept Prüflabor ▪ Absicherungs-methoden für KI-Systeme	▪ Rechtliche, ethische und gesellschaftliche Fragestellungen ▪ Öffentliche Veranstaltungen

Breit angelegter Beteiligungsprozess

Partner:

Fraunhofer IAIS
Bundesamt für Sicherheit in der Informationstechnik
UNIVERSITÄT BONN
Universität zu Köln
RWTH AACHEN UNIVERSITY
DIN

KI-Prüfkatalog

- **Schritt 1: Risikoanalyse**
  Umfassende Risikoanalyse entlang der Dimensionen Fairness, Autonomie und Kontrolle, Transparenz, Verlässlichkeit, Sicherheit und Datenschutz
- **Schritt 2: Festlegung von Zielvorgaben**
  Festlegung objektiver, möglichst messbarer Zielkriterien, um die in Schritt 1 identifizierten Risiken abzuschwächen
- **Schritt 3: Auflistung von Maßnahmen**
  Systematische Auflistung von Maßnahmen entlang des Lebenszyklus einer KI-Anwendung, um die in Schritt 2 gesetzten Zielvorgaben zu erreichen
- **Schritt 4: Absicherungsargumentation**
  Erstellung einer stringenten Argumentation, dass die in Schritt 2 formulierten Ziele erreicht wurden

Prüfkatalog ist frei erhältlich unter:
https://www.iais.fraunhofer.de/de/forschung/kuenstliche-intelligenz/ki-pruefkatalog.html

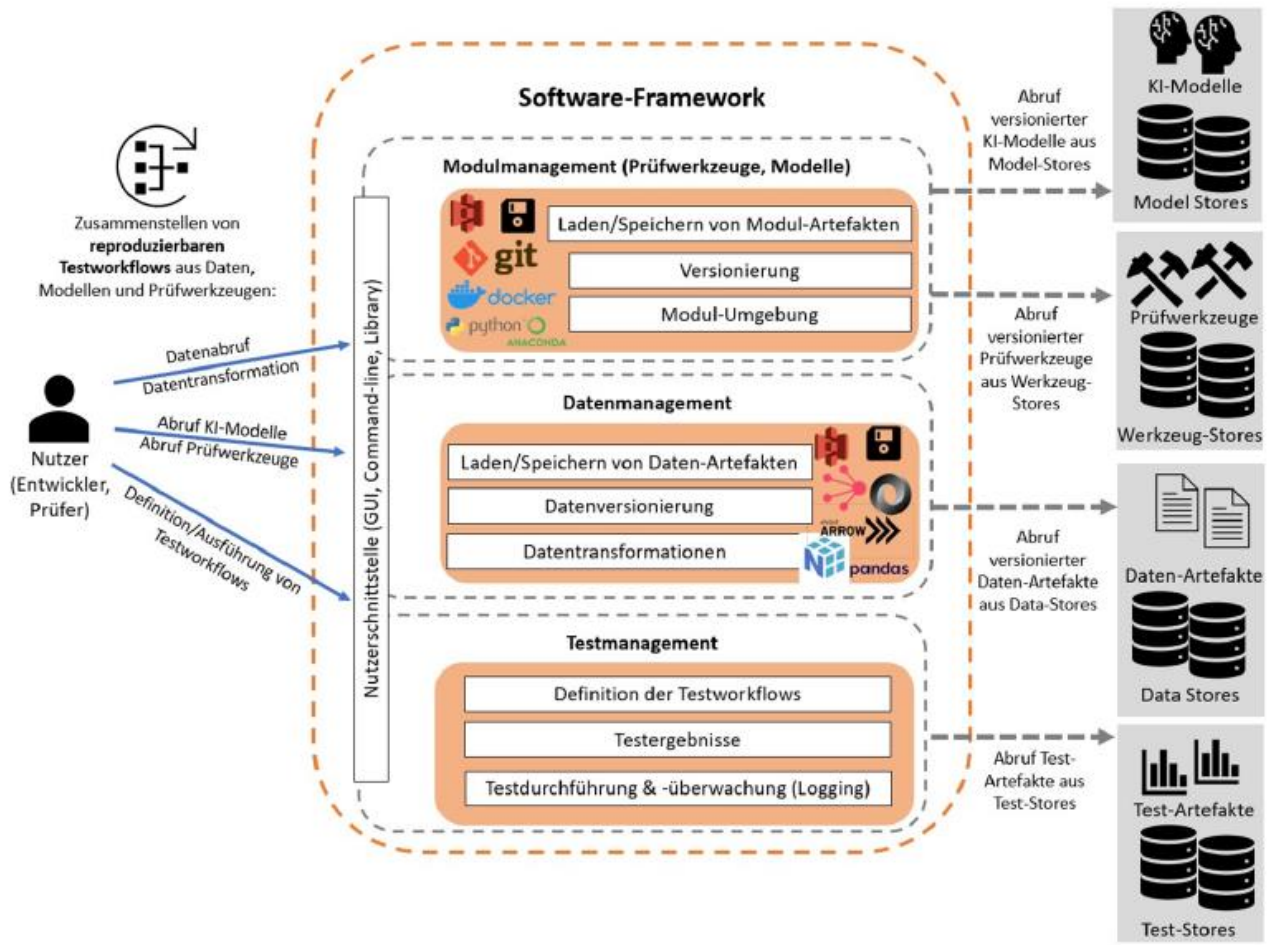
Einsatzbereiche

Unser KI-Prüfkatalog unterstützt

- Entwickler*innen bei der Gestaltung und
- KI-Prüfer*innen bei Evaluation und Qualitätssicherung

von KI-Anwendungen.

Automatisierung von KI-Prüfungen



Software-Framework für reproduzierbare und vergleichbare Tests

- Versionierte Ablage von
  - Daten
  - Testwerkzeugen (Toolsuite)
  - Modellen
  - Pipelines
  - Tests
- Interoperable, containerisierte Module für Modelle und Prüfwerkzeuge
- Cloud-kompatibler Software-Stack

Anwenderkreis „Vertrauenswürdige KI-Cloud-Services“

20. Juni 2022: Kick-off

6. September 2022: KI-Qualität und Prüfverfahren im Finanzsektor

8. Mai 2023: AI Risk Management Frameworks

Oktober 2023: Implementierung der KI-Verordnung mit horizontalen Standards

Anwenderkreis „Fundamentalmodelle“

20. Juni 2023: Foundation Models – Chancen und Herausforderungen

27. September 2023: Multimodale Foundation Models – Funktionsweise und Einsatz als Werkzeug zum semantischen Testen

2023: Datenqualität – Grundlage für Vertrauenswürdigkeit von verschiedenster KI-Systemen

2023: Prüfframeworks und Infrastruktur als allgemeine Basis zur Gestaltung von vertrauenswürdiger KI

Standardisierungsaktivitäten

- **DIN/TS 92004 Artificial intelligence — Quality requirements and processes — Risk scheme for AI systems along the entire life cycle**
  Die DIN/TS 92004 enthält Anforderungen an die Risikoanalyse und -behandlung für die Entwicklung und den Betrieb von Systemen der künstlichen Intelligenz (KI), die Komponenten des maschinellen Lernens (ML) enthalten.
- **DIN SPEC 92001-3 Artificial intelligence — Life cycle processes and quality requirements — Part 3: Explainability**
  This is the third document in a series, and it aims to ensure that AI systems are developed, deployed, and used efficiently, responsibly, and in a trustworthy way. It focuses on "Explainability" – the ability to understand how AI makes decisions. This DIN SPEC 92001-3 provides a domain-independent guide on promoting explainability throughout the AI system's life cycle.
- **DIN SPEC 92005 Uncertainty Quantification in ML**
  Die DIN SPEC 92005 legt allgemeine Leitfäden und Anforderungen für die Entwicklung und Nutzung von Methoden zur Quantifizierung von Unsicherheiten im Maschinellen Lernen (ML) fest. definiert Dieser Standard definiert grundlegende Begriffe für die Quantifizierung der Unsicherheit für ML und spezifiziert den Zweck, die Verwendung und die Notwendigkeit dieser Analysen.

Fraunhofer IAIS

DIN