

Informationsblatt für Anbieter:innen & Beiräte – Künstliche Intelligenz (KI)

Vorrang menschlichen Handelns:

- Menschliche Entscheidungsverantwortung
- Jederzeit möglicher Abbruch von Vorgängen
- Überprüfbarkeit der Interpretationen der KI
- Durch Menschen steuerbares Prüfsystem der verwendeten Daten
- Durch Menschen mögliche Ergänzung/Korrektur der verwendeten Daten

Ethik:

- Selbstbestimmte KI-Nutzung ohne Einfluss auf menschliche Entscheidungsfindung
- Wahrung gesellschaftlicher Werte, ethischer Grundsätze und Rechenschaftspflichten
- Faire und gleiche Behandlung aller Betroffenen
- Berücksichtigung der Interessen indirekt von KI betroffener Personen
- Berücksichtigung von Rückmeldungen der Nutzer:innen

Informationen und Transparenz gegenüber Nutzer:innen:

- Transparenz über Interagieren mit einer KI anstatt mit Menschen
- Zweck der eingesetzten KI
- Potenzielle und tatsächliche Risiken der KI
- Merkmale, Grenzen und potenzielle Mängel der KI
- Funktionsweise und Entscheidungen der KI
- Aktuell geltende rechtliche Rahmenbedingungen der KI

Zu beachtende rechtliche Vorgaben:

- KI-Verordnung: European Union Artificial Intelligence Act (EU AI Act)

Risikomanagement und Technische Aspekte:

- Absicherung gegen Angriffe, Unfälle, Fehler und Sicherheitsrisiken
- Kontroll- und Überwachungsprozesse der KI hinsichtlich Verzerrungen
- Vorkehrungen und Kontrollmechanismen der Nutzung und Löschung personenbezogener Daten
- Auffangplan für allgemeine Sicherheit: Berücksichtigung körperlicher Unversehrtheit von Menschen, möglichen Schäden, Sicherheitsrisiken für Umwelt und Tiere, Sicherheitsrisiken bei Ausfall oder Fehlern in der KI
- Basierend auf KI-Rahmenwerk einer vertrauenswürdigen Institution

Anwendbare Rahmenwerke und Institutionen:

- Deutsche „United Nations Educational Scientific and Cultural Organization“ (UNESCO) Kommission: UNESCO-Empfehlung zur Ethik Künstlicher Intelligenz. Bedingungen zur Implementierung in Deutschland
- Bundesministerium für Bildung und Forschung, Referat Künstliche Intelligenz: KI-Strategie Deutschland
- Bundesamt für Sicherheit in der Informationstechnik (BSI): Kriterienkatalog für KI-Clouddienste
- acatech - Deutsche Akademie der Technikwissenschaften e.V.: Plattform Lernende Systeme, Diverse Publikationen
- Deutsche Institut für Normung e. V. (DIN): KI-Normungsroadmap
- Fraunhofer Institut für Intelligente Analyse- und Informationssysteme (IAIS): KI-Prüfkatalog - Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz
- Institut der Wirtschaftsprüfer in Deutschland e. V. (IDW): Entwurf eines IDW Prüfungsstandards: Prüfung von KI-System
- Artificial Intelligence Ethics Impact (AIEI) Group: From Principles to Practice. An interdisciplinary framework to operationalise AI ethics
- Bayerisches Landesamt für Datenschutzaufsicht: Konferenz der unabhängigen Datenschutzbehörden des Bundes und der Länder