WIE „VERTRAUENSWÜRDIG" IST

KÜNSTLICHE INTELLIGENZ?

Was bedeutet „vertrauenswürdige" KI? Wann kann man KI „verantwortungsvoll" einsetzen?

Um dich bei diesen Fragen zu unterstützen , gibt es bestimmte Leitlinien für die Entwicklung und den Einsatz von KI. Diese Richtlinien konzentrieren sich auf die ethischen und sozialen Themen und Fragen im Zusammenhang mit KI.

Das Knowledge Center Data & Society hat die 7 Anforderungen an KI zusammengefasst, die von der High-Level Expert Group on AI (AI HLEG) entwickelt wurden.

Zu jeder Anforderung wurden eine Reihe von Fragen formuliert, die es dir ermöglichen, über die Vertrauenswürdigkeit deines KI-Systems nachzudenken.

Für einen detaillierteren Ansatz hat das Knowledge Center Data & Society kürzlich die KI Blindspots-Karten entwickelt. Mit diesem Tool reflektierst du proaktiv die (ethischen und sozialen) Entscheidungen und Handlungen, die du zu Beginn deines Projekts oder während der Entwicklung deines KI-Systems treffen möchtest.

Disclaimer: die 7 Anforderungen überschneiden sich, daher haben wir redundante Fragen eliminiert und die dringendsten Fragen ausgewählt, um die Komplexität zu reduzieren.

Knowledge Center Data & Society (2020). Wie ethisch ist KI? brAInfood vom Knowledge Centre Data & Society. Brüssel: Knowledge Centre Data & Society

BrAInfood ist ein gemeinsames Projekt von D&M, CLAIRE, DFKI, und ZHAW, das die gemeinsame EU-Vision von #AI4Good und #AI4All vorantreibt

MENSCHLICHES HANDELN & AUFSICHT

- Ist die Interaktion zwischen deinem KI-System und einem Menschen sinnvoll und relevant?
- Wer trifft die letzte Entscheidung, die Maschine oder der Mensch? Wenn es die Maschine ist, dann gibt es keine menschliche Aufsicht.
- Ist das KI-System autonom oder selbstlernend? Wenn ja, gibt es Kontrollmechanismen?

TRANSPARENZ

- Kannst du deinem Team, aber auch jedem, der mit dem System in Kontakt kommt, zeigen, wie der Algorithmus entworfen und aufgebaut ist und wie Entscheidungen getroffen werden? Stelle sicher, dass jede/r versteht, dass eine KI-Komponente Teil des Systems ist.

RECHENSCHAFTSPFLICHT*

- Hast du einen Überblick über alle Entscheidungen und Abwägungen, die du getroffen hast, um dein System zu entwickeln?
- Hast du negative Konsequenzen für alle Beteiligten identifiziert?
- Konntest du negative Konsequenzen reduzieren und sind diese Maßnahmen dokumentiert?

*accountability

TECHNISCHE ROBUSTHEIT & SICHERHEIT

- Welche Maßnahmen ergreifst du, wenn dein KI-System angegriffen wird oder sich anders verhält als erwartet oder wenn es für einen anderen (ungewollten) Zweck eingesetzt wird?
- Besteht die Möglichkeit, dass dein KI-System ungenaue Vorhersagen macht?
- Welche Maßnahmen gibt es, um Ungenauigkeiten zu vermeiden?

VIELFALT, NICHT-DISKRIMINIERUNG & FAIRNESS

- Wie wirst du unfaire Voreingenommenheit in deinem KI-System vermeiden? Ist es für andere möglich, auf Voreingenommenheit oder Diskriminierung zu identifizieren?
- Hast du berücksichtigt, wie deine KI-Innovation den Zugang zu deinem Service für marginalisierte Gruppen verändert? Wenn ja, bietest du einen alternativen Service für die Benachteiligten an?
- Können sich deine Stakeholder an der Entwicklung und Nutzung deines KI-Systems beteiligen?

brAInfood of the Knowledge Centre Data & Society

DATENSCHUTZ & DATA GOVERNANCE

- Verwendet dein KI-System personenbezogene Daten? Wenn ja, bist du dir der Auswirkungen und Anforderungen bei der Verwendung von personenbezogenen Daten bewusst? Kannst Du nachweisen, dass du die datenschutzrechtlichen Bestimmungen einhältst?
- Verfügst du über Kontrollmechanismen für die Erhebung, Speicherung, Verarbeitung und Nutzung von Daten?
- Wer kann auf die Daten der Nutzer zugreifen? Verfügst du über ein Datenzugriffsprotokoll?

GESELLSCHAFTLICHES & ÖKOLOGISCHES WOHLBEFINDEN

- Kannst du die Umweltauswirkungen des Lebenszyklus deines KI-Systems messen? Weißt du, wie du sie reduzieren kannst?
- Sind sich die (End-)Nutzer über die Grenzen der sozialen Interaktion mit deinem KI-System und die gesellschaftlichen Auswirkungen deines KI-Systems bewusst?
- Könnten über deine Zielgruppe hinaus auch andere Gruppen oder Einzelpersonen indirekt von deinem KI-System betroffen sein?