

HOE 'BETROUWBAAR' IS

ARTIFICIËLE INTELLIGENTIE?

Wat betekent 'betrouwbare' AI? Wanneer kan je AI 'verantwoord' inzetten?

Om je met deze vragen te helpen, zijn er bepaalde richtlijnen voor de ontwikkeling en het gebruik van AI. Deze richtlijnen houden rekening met de ethische en maatschappelijke vraagstukken

en kwesties op het gebied van AI.

Het Kenniscentrum Data & Maatschappij vat [de 7 vereisten voor AI](#) samen, die werden opgesteld door de 'High-Level Expert Group on AI' (AI HLEG).

Per vereiste worden enkele vragen geformuleerd, die je helpen om te reflecteren over de betrouwbaarheid van jouw AI-systeem.

Als je behoefte hebt aan een meer gedetailleerde aanpak, ontwikkelde het Kenniscentrum Data & Maatschappij reeds [de AI Blindspots kaartenset](#). Met de

hulp van deze kaartenset denk je proactief na over (ethische en maatschappelijke) beslissingen en acties die je wil nemen aan het begin van je project of tijdens de ontwikkeling van je AI-systeem.

Disclaimer: de 7 vereisten overlappen elkaar, daarom lieten we overbodige vragen weg en en tonen we enkel de meeste dringende, dit om de complexiteit te verminderen.

Kenniscentrum Data & Maatschappij (2020). Hoe ethisch is AI? brAlInfood van het Kenniscentrum Data & Maatschappij, Brussel: Kenniscentrum Data & Maatschappij.

Dit document is beschikbaar onder een CC BY 4.0 licentie.

MENSELIJKE ACTIES & TOEZICHT

- Is de interactie tussen jouw AI-systeem en een mens betekenisvol en relevant?
- Wie neemt de finale beslissing, de machine of de mens? Als het de machine is, dan is er geen menselijk toezicht.
- Is jouw AI-systeem autonoom of zelflerend? Zo ja, zijn er toezichtsmechanismen?

TRANSPARANTIE

- Kan je aantonen hoe het algoritme is ontworpen en gebouwd, en hoe beslissingen worden genomen aan jouw team, maar ook aan iedereen die in contact komt met het systeem? Zorg dat iedereen begrijpt dat een AI-component deel uitmaakt van het systeem.

VERANTWOORDINGSPLICHT

- Heb je een overzicht van alle beslissingen en afwegingen die je hebt gemaakt om jouw systeem te creëren?
- Heb je de negatieve gevolgen voor alle betrokken partijen geïdentificeerd?
- Heb je de negatieve gevolgen kunnen beperken en zijn de genomen maatregelen gedocumenteerd?

TECHNISCHE STERKTE & VEILIGHEID

- Welke maatregelen zal je nemen als jouw AI-systeem wordt aangevallen of zich anders gedraagt dan verwacht, of als het wordt gebruikt voor een ander (ongewenst) doel?
- Is er een kans dat jouw AI-systeem onnauwkeurige voorspellingen geeft?
- Welke maatregelen neem je om onnauwkeurigheid aan te pakken?

DIVERSITEIT, NON-DISCRIMINATIE & EERLIJKHEID

- Hoe voorkom je oneerlijke partijdigheid in jouw AI-systeem? Kunnen anderen aangeven dat er sprake is van vooringenomenheid of discriminatie?
- Heb je nagedacht over hoe jouw AI-innovatie toegang tot je dienst voor gemarginaliseerde groepen verandert? Zo ja, biedt je een alternatieve dienst aan voor uitgesloten?
- Kunnen stakeholders deelnemen aan de ontwikkeling en het gebruik van jouw AI-systeem?

PRIVACY & DATABESTUUR

- Maakt jouw AI-systeem gebruik van persoonsgegevens? Zo ja, ben jij je bewust van de implicaties en vereisten bij het gebruik van persoonsgegevens? Kan je aantonen dat je voldoet aan de regelgeving inzake gegevensbescherming?
- Heb je toezichtsmechanismen voor het verzamelen, opslaan, verwerken en gebruiken van gegevens?
- Wie heeft toegang tot de gegevens van gebruikers? Heb je een protocol voor de toegang tot gegevens?

MAATSCHAPPELIJK & ECOLOGISCH WELZIJN

- Kan je de milieu-impact van de levenscyclus van jouw AI-systeem meten? Weet je hoe je deze kan verminderen?
- Zijn (eind)gebruikers zich bewust van de beperkte sociale interactie met jouw AI-systeem en de maatschappelijke impact van jouw AI-systeem?
- Kunnen andere groepen of individuen indirect worden beïnvloed door jouw AI-systeem, buiten jouw doelgroep om?

brAlInfood van het Kenniscentrum Data & Maatschappij



CLAIRE