

Bouwen aan betrouwbare AI

Nathalie Smuha

KU Leuven, Faculteit Rechten, onderzoekster en assistent-docent (FWO)

De Europese benadering: context

EUROPESE STRATEGIE ROND ARTIFICIËLE INTELLIGENTIE

Gepubliceerd in april 2018

De ontwikkeling en
opname van AI
'boosten'

De socio-economische
veranderingen
aanpakken

Een adequaat ethisch
en juridisch kader
verzekeren



In die context: oprichting van een expertengroep op hoog niveau inzake
Artificiële Intelligentie (*AI HLEG*)

Expertengroep – mandaat

Voorzitter:
Pekka Ala-Pietilä

52 leden:



Industrie



Onderzoek



Burger-
maatschappij

Twee opdrachten:

1. Ethische richtlijnen voor AI
2. Beleidsaanbevelingen

+ Interactie met de ‘European AI Alliance’

- Platform met >3500 leden (*open voor iedereen*) om het Europese beleid inzake AI te bespreken

Ethische richtlijnen – voor betrouwbare AI



- De mens staat centraal: AI is slechts een middel, geen doel op zich
- Gebaseerd op fundamentele rechten
 - “Wij geloven in een benadering van AI-ethiek op basis van de grondrechten die zijn vastgelegd in de EU verdragen, het Handvest van de grondrechten van de EU en de internationale mensenrechtenwetgeving. Eerbiediging van grondrechten, binnen een kader van democratie en de rechtsstaat, biedt de meest veelbelovende grondslagen voor het vaststellen van abstracte ethische beginselen en waarden die in het kader van KI kunnen worden geoperationaliseerd.”
- ‘Betrouwbare AI’ als fundamentele ambitie, bestaande uit drie elementen:

Wettige AI

Ethische AI

Robuuste AI

Ethische richtlijnen – fundamentele rechten

5 'families' van fundamentele rechten

- 1) Respect voor menselijke waardigheid
- 2) Vrijheid van het individu
- 3) Respect voor democratie, justitie en de rechtsstaat
- 4) Gelijkheid, non-discriminatie en solidariteit
- 5) Burgerrechten

-> 4 ethische principes



Respect voor
menselijke autonomie



Preventie van
schade



Rechtvaardigheid



Verklaarbaarheid

Ethische richtlijnen – 7 essentiële vereisten



Menselijke controle en toezicht



Diversiteit, non-discriminatie en rechtvaardigheid



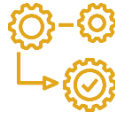
Technische robuustheid en veiligheid



Maatschappelijk en ecologisch welzijn



Privacy en data-governance



Verantwoording



Transparantie

Te evalueren en te verzekeren gedurende de hele levenscyclus van een AI-systeem