



Europese Commissie

# DESKUNDIGENGROEP OP HOOG NIVEAU INZAKE KUNSTMATIGE INTELLIGENTIE



## ONTWERP ETHISCHE RICHTSNOEREN VOOR BETROUWBARE KI

SAMENVATTING

Werkdocument voor raadpleging van stakeholders

Brussel, 18 december 2018

# ONTWERP VAN ETHISCHE RICHTSNOEREN VOOR BETROUWBARE KI



Deskundigengroep op hoog niveau inzake kunstmatige intelligentie (AI HLEG)  
**Ontwerp van ethische richtsnoeren voor betrouwbare KI**

Europese Commissie  
Directoraat-generaal Communicatie

Contact Nathalie Smuha - Coördinator AI HLEG  
E-mail CNECT-HLG-AI@ec.europa.eu

Europese Commissie  
B-1049 Brussel

Document gepubliceerd op 18 december 2018, in het Engels.

**Dit werkdocument is opgesteld door de AI HLEG en laat onverlet dat zijn leden op specifieke punten een individueel standpunt kunnen hebben. Er wordt niet mee vooruitgelopen op de uiteindelijke tekst. Dit document is een werk in uitvoering en een definitieve versie ervan zal in maart 2019 worden gepresenteerd na raadpleging van de stakeholders via de Europese AI Alliance.**

De Europese Commissie of personen die namens de Commissie optreden, zijn niet aansprakelijk voor het gebruik dat eventueel van de volgende informatie wordt gemaakt. De inhoud van dit werkdocument valt uitsluitend onder de verantwoordelijkheid van de deskundigengroep op hoog niveau inzake kunstmatige intelligentie (AI HLEG). Hoewel personeel van de diensten van de Commissie heeft meegewerkt aan de totstandkoming van de richtsnoeren, geven de hierin geformuleerde meningen uitsluitend het standpunt van de AI HLEG weer en mogen zij in geen geval worden beschouwd als een officieel standpunt van de Europese Commissie. Dit is een ontwerp in het kader van de eerste opdracht van de AI HLEG. Een definitieve versie daarvan zal in maart 2019 aan de Commissie worden gepresenteerd. Een definitieve versie in het kader van de tweede opdracht – aanbevelingen formuleren voor een KI-beleid en -investeringen – zal medio 2019 worden gepresenteerd.

Meer informatie over de deskundigengroep op hoog niveau inzake kunstmatige intelligentie is online beschikbaar (<https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>).

Het beleid ten aanzien van hergebruik van documenten van de Europese Commissie is vastgelegd in Besluit 2011/833/EU (PB L 330 van 14.12.2011, blz. 39). Voor gebruik of overname van foto's of andere materialen die niet onder het auteursrecht van de EU vallen, moet u rechtstreeks toestemming vragen aan de houders van het desbetreffende auteursrecht.

## SAMENVATTING

Dit werkdocument bevat een ontwerp van de ethische richtsnoeren voor KI die zijn ontwikkeld door de deskundigengroep op hoog niveau inzake kunstmatige intelligentie van de Europese Commissie (AI HLEG). Een definitieve versie ervan wordt in maart 2019 verwacht.

Kunstmatige intelligentie (KI) is een van de grootste transformerende krachten van onze tijd en voorbestemd om onze samenleving ingrijpend te veranderen. Zij biedt enorme kansen om meer welvaart en groei te genereren die Europa moet zien te grijpen. De beschikbaarheid van grote hoeveelheden digitale gegevens, krachtige computersystemen en de ontwikkelingen op het gebied van KI-technologie, zoals machine-learning, hebben het afgelopen decennium grote vooruitgang mogelijk gemaakt. Belangrijke, op KI steunende ontwikkelingen op het vlak van zelfrijdende voertuigen, gezondheidszorg, robots in het huishouden en in de industrie, onderwijs of cyberbeveiliging zorgen voor een verbetering van de kwaliteit van ons dagelijks leven. KI is tevens cruciaal voor de aanpak van tal van grote uitdagingen waarmee de wereld wordt geconfronteerd, zoals gezondheid en welzijn, de klimaatverandering, betrouwbare rechtsstelsels en democratische systemen, en andere die in de doelstellingen voor duurzame ontwikkeling van de Verenigde Naties aan de orde zijn gesteld.

Maar naast het vermogen om individuen en de samenleving enorme voordelen bij te brengen, kleven aan kunstmatige intelligentie ook bepaalde risico's die goed moeten worden beheerd. Aangezien de voordelen van KI per saldo groter zijn dan de risico's, moeten we kiezen voor een parcours dat **de voordelen van KI maximaliseert en de risico's tot een minimum beperkt**. Met het oog daarop is **een benadering van KI nodig waarbij de mens centraal staat** en we voortdurend voor ogen moeten houden dat de ontwikkeling en het gebruik van KI geen doel op zichzelf is, maar een middel om het menselijke welzijn te vergroten. **Betrouwbare KI moet het leidende principe zijn**, aangezien mensen slechts zonder argwaan en ten volle de voordelen van KI zullen kunnen plukken als zij de technologie vertrouwen.

Voor betrouwbare KI moeten **twee voorwaarden** vervuld zijn: (1) de grondrechten, de toepasselijke regelgeving en fundamentele beginselen en waarden die een **“ethisch doel”** waarborgen, moeten worden gerespecteerd, en (2) de **technologie moet robuust** en betrouwbaar zijn, want een gebrekkige beheersing van de technologie – zelfs al zijn de bedoelingen goed – kan ongewild schade aanrichten.

Deze richtsnoeren schetsen daarom **een kader voor betrouwbare KI**:

- **Hoofdstuk I** gaat over het **waarborgen van het “ethische doel” van KI**, en meer bepaald de grondrechten, beginselen en waarden waaraan zij moet voldoen.
- Vertrekkende van die beginselen, worden in **hoofdstuk II richtsnoeren voor de totstandbrenging** van betrouwbare KI vastgesteld die zowel het ethische doel als de technische robuustheid behandelen. Daartoe worden de vereisten voor betrouwbare KI opgesomd en wordt een overzicht gegeven van de technische en niet-technische methoden om deze tot stand te brengen.
- In **hoofdstuk III** tot slot, worden die vereisten **geoperationaliseerd** door middel van een concrete, maar niet-uitputtende controlelijst voor betrouwbare KI. Deze lijst wordt vervolgens aangepast voor gebruik in specifieke gevallen.

In tegenstelling tot andere documenten over ethische KI, zijn de richtsnoeren er met andere woorden niet op gericht de fundamentele onderliggende waarden en beginselen voor KI nogmaals op te sommen, maar wel een leidraad te verschaffen voor de concrete implementatie en operationalisering ervan in KI-systemen. Die leidraad wordt gegeven voor drie abstractieniveaus, van het meeste abstracte in hoofdstuk I (grondrechten, beginselen en waarden) tot het meest concrete in hoofdstuk III (controlelijst).

De richtsnoeren zijn gericht tot **alle stakeholders die KI ontwikkelen, installeren of gebruiken** en die

bedrijven, organisaties, onderzoekers, overheidsdiensten, instellingen, individuele personen of andere entiteiten omvatten. In de definitieve versie van deze richtsnoeren zal een mechanisme worden voorgesteld om de belanghebbenden in staat te stellen ze op vrijwillige basis te onderschrijven.

Er zij op gewezen dat het niet de bedoeling is dat deze richtsnoeren in de plaats komen van beleid of regulering in welke vorm dan ook (komt aan bod in het kader van de tweede opdracht van de AI HLEG – aanbevelingen formuleren voor beleid en investeringen, verwacht in mei 2019), noch dat zij de invoering ervan ontmoedigen. De richtsnoeren moeten bovendien worden gezien als een dynamisch document, dat mettertijd periodiek moet worden bijgewerkt om relevant te blijven naarmate de technologie en onze kennis ervan zich ontwikkelen. Dit document moet derhalve dienen als uitgangspunt voor het debat over “**betrouwbare KI, made in Europe**”.

Terwijl Europa zijn boodschap van ethische benadering van KI slechts kan uitdragen op voorwaarde dat het op wereldvlak meespeelt, is de **ethische benadering een eerste vereiste voor een verantwoordelijke concurrentie**, om vertrouwen te wekken bij de gebruiker en KI op grotere schaal ingang te doen vinden. Deze richtsnoeren zijn niet bedoeld om KI-innovatie in Europa aan banden te leggen, maar op grond van ethische motieven een uniek soort KI tot stand te brengen, dat de belangen van het individu en de gemeenschap beschermt en voorop stelt. Hierdoor kan Europa zichzelf profileren als voortrekker van geavanceerde, veilige en ethische KI. Alleen als betrouwbaarheid gegarandeerd is, zullen de Europese burgers ten volle kunnen profiteren van de voordelen van KI.

Tot slot wordt met deze richtsnoeren beoogd om buiten Europa **de bezinning en het debat** over een ethisch kader voor KI **op mondiaal niveau** aan te zwengelen.

## **RICHTSNOEREN**

De richtsnoeren willen een leidraad bieden aan alle stakeholders die KI ontwikkelen, installeren of gebruiken. Hierna wordt een samenvatting gegeven van de verschillende hoofdstukken:

### **Hoofdstuk I: Essentiële richtsnoeren om het ethische doel te waarborgen:**

- Zorg ervoor dat **de mens centraal staat**: KI moet worden ontwikkeld, geïnstalleerd en gebruikt met een “ethisch doel”, steunend op de grondrechten, maatschappelijke normen en waarden en de ethische beginselen van *Beneficence* (goed doen), *Non Maleficence* (niet schaden), *de autonome wil van de mens, rechtvaardigheid en verantwoording*. Dit is essentieel om tot **betrouwbare KI** te komen.
- Toets de mogelijke effecten van KI op het individu en de gemeenschap anticiperend in het licht van de grondrechten, ethische beginselen en waarden. Schenk **bijzondere aandacht** aan situaties waarin **kwetsbaardere groepen**, zoals kinderen, personen met een beperking of minderheden, betrokken zijn of aan situaties waarin er **ongelijkheid is wat betreft macht of beschikking over informatie**, bijvoorbeeld tussen werkgevers en werknemers of tussen bedrijven en consumenten.
- Erken en wees u ervan bewust dat KI de mens en de samenleving grote voordelen kan opleveren, maar ook negatieve gevolgen kan hebben. Verlies gebieden met een kritisch belang niet uit het oog.

### **Hoofdstuk II: Essentiële richtsnoeren om betrouwbare KI tot stand te brengen:**

- Houd **vanaf de vroegste ontwerpfase** rekening met de **vereisten voor betrouwbare KI**: verantwoordingsplicht, data governance, design for all, beheer van KI-autonomie (menselijk toezicht), non-discriminatie, eerbiediging van de autonome wil van de mens, respect voor de privacy, robuustheid, veiligheid en transparantie.
- Bestudeer technische en niet-technische methoden om te waarborgen dat die vereisten in het KI-

systeem worden geïmplementeerd. Houd daarenboven met die vereisten rekening bij de samenstelling van het team dat aan het systeem werkt en bij de bouw van het systeem zelf, de testomgeving en de potentiële toepassingen van het systeem.

- Verstrek op een proactieve manier duidelijke **informatie aan stakeholders** (klanten, werknemers enz.) over de mogelijkheden en beperkingen van een KI-systeem, om onrealistische verwachtingen te voorkomen. Het garanderen van de **traceerbaarheid** van het KI-systeem is in dit verband essentieel.
- Incorporeer betrouwbare KI in **de cultuur van de organisatie** en informeer stakeholders over de manier waarop betrouwbare KI wordt geïmplementeerd in het ontwerp en gebruik van KI-systemen. Betrouwbare KI kan tevens aan bod komen in deontologische of gedragscodes van organisaties.
- Zorg voor deelname aan en het **betrekken van stakeholders** bij het ontwerp en de ontwikkeling van een KI-systeem. Zorg daarnaast voor **diversiteit** bij het samenstellen van de teams die het product ontwikkelen, installeren en testen.
- Tracht de **controleerbaarheid van KI-systemen te vergemakkelijken**, met name in kritieke omgevingen of situaties. Zet het systeem zo op dat individuele besluiten kunnen worden teruggetraceerd naar uw verschillende inputs, data, aangeleerde modellen enz. Definieer tevens **uitleggingsmodellen** voor het KI-systeem.
- Stel een specifieke procedure van **accountability governance** in.
- Voorzie in **opleidingen en scholing** en zorg ervoor dat managers, ontwikkelaars, gebruikers en werkgevers weten wat betrouwbare KI inhoudt en hoe zij ermee moeten omgaan.
- Wees bedacht op conflicterende doelstellingen (transparantie kan de deur openzetten voor misbruik; identificeren en corrigeren van discriminatie kan botsen met de privacybescherming). Communiceer over deze afwegingen en documenteer ze.
- Stimuleer onderzoek en innovatie om het voldoen aan de vereisten voor betrouwbare KI te bevorderen.

### **Hoofdstuk III: Essentiële richtsnoeren om betrouwbare KI te evalueren:**

- Stel een **controlelijst** voor betrouwbare KI vast bij de ontwikkeling, de installatie of het gebruik van KI en pas deze aan aan de specifieke situatie waarin het systeem wordt gebruikt.
- Bedenk dat een controlelijst **nooit uitputtend** zal zijn en dat het voor betrouwbare KI niet volstaat vakjes aan te kruisen, maar een continu proces nodig is van vaststellen van behoeften, afwegen van oplossingen en waarborgen van betere resultaten gedurende de hele levenscyclus van het KI-systeem.

Deze richtsnoeren maken deel uit van een visie op en een benadering van kunstmatige intelligentie waarin de mens centraal staat, die Europa in staat moet stellen op wereldvlak een voortrekkersrol te spelen inzake ethische, veilige en geavanceerde KI. Zij zijn erop gericht **“Betrouwbare KI, made in Europe”** te bevorderen, teneinde het welzijn van de Europese burgers te vergroten.