

## Bijdrage Rathenau Instituut aan het rondetafelgesprek Artificiële intelligentie in het recht

Aan: Vaste Commissie voor Veiligheid en Justitie

Van: Rinie van Est, Jurrien Hamer, Linda Kool | 26 maart 2018

Deze bijdrage van het Rathenau Instituut gaat niet specifiek in op artificiële intelligentie (AI) in het recht, maar gaat over de inzet van AI door de overheid. De analyse en kwesties die we aanstippen, gelden echter ook voor de toepassing van AI in het recht. Hieronder behandelen we de relatie tussen AI en publieke waarden en fundamentele mensenrechten. We laten zien dat het belangrijk is om naar de gehele datawaardenketen te kijken. Tenslotte geven we aan hoe er gewerkt kan worden aan een betekenisvolle publieke en menselijke controle over AI.

### 1. AI en publieke waarden

Algoritmen zijn mensenwerk. Bij AI gaat het om (zelf)lerende algoritmen. Bij het trainen van AI spelen data een belangrijke rol. Met name op het terrein van *deep learning* algoritmen is de afgelopen jaren veel vooruitgang geboekt. De beschikbaarheid van *big data* en snelle computers jaagt die trend aan. Deze ontwikkeling is veelbelovend aangezien algoritmen kunnen helpen bij het efficiënt en doordacht nemen van besluiten.

AI raakt aan tal van publieke waarden en grondrechten. De laatste twee jaar zijn er wereldwijd vele initiatieven geweest om die in kaart te brengen. Die initiatieven laten een brede consensus zien met betrekking tot de waarden en fundamentele mensenrechten die bij AI op het spel kunnen staan. De kwesties sluiten sterk aan bij het geheel aan ethische en maatschappelijke kwesties die het Rathenau Instituut in het rapport [Opwaarderen: Borgen van publieke waarden in de digitale samenleving](#) (Kool et al. 2017) benoemt. De kwesties die spelen zijn: Privacy, veiligheid, rechtvaardigheid, menselijke waardigheid, autonomie en, niet te vergeten, controle en macht over technologie.

Het is daarom van belang om het gebruik en verdere ontwikkeling van AI op een maatschappelijk verantwoorde manier vorm te geven. Betekenisvolle menselijke en publieke controle over AI is daarbij cruciaal. Met name op het gebied van de rechtspraak is publieke controle over de data en slimme systemen belangrijk. Met andere woorden, publieke technologische soevereiniteit is bij AI in het recht gewenst.

### 2. Betekenisvolle controle over de gehele datawaardenketen

Deze controle moet gerealiseerd worden over de gehele datawaardenketen: Bij het verzamelen van de data die geanalyseerd worden, bij de manier waarop we die data analyseren en bij het toepassen daarvan.

# Rathenau Instituut

## Dataverzameling

AI heeft grote datasets nodig om te kunnen leren. Deze datasets zijn vaak incompleet, gekleurd of bevatten andere fouten. De inzet van AI – in welk domein dan ook – vraagt daarom om hoogkwalitatieve data, anders werken de fouten in de dataset door in de beslissingen ('*garbage in, garbage out*'). In het Rathenau rapport '[Datagedreven samenleving](#)' (Kool et al. 2016) wijzen we daarom op het belang van 'data stewardship': afspraken en instrumenten om de kwaliteit van data vast te leggen, zodat partijen die data delen of inkopen, weten wat ze in handen hebben.

## Data-analyse

AI is mensenwerk: het zijn mensen die algoritmes ontwerpen en interpreteren. Ontwerpers, technici, en organisaties die AI inzetten, maar ook beleidsmakers en toezichhouders dragen dus verantwoordelijkheid voor hoe AI wordt gebruikt. Dit betekent ten eerste dat bedrijven en organisaties die AI inzetten zelf betekenisvolle controle moeten hebben over de algoritmen die zij ontwikkelen, en uitleg kunnen geven over de beslissingen die algoritmen nemen. Zij dienen daarom richtlijnen op te stellen voor de verantwoorde ontwikkeling en inzet van AI. Transparantie en accountability van algoritmen zijn belangrijke uitgangspunten. Ten tweede betekent deze verantwoordelijkheid dat de maatschappij – burgers, beleidsmakers, toezichhouders – betekenisvolle controle heeft over de inzet van AI. Burgers of consumenten moeten weten welke beslissingen het systeem over hen maakt. Zij moeten ook bezwaar kunnen aantekenen tegen deze beslissingen. Ook een toezichhouder moet in staat zijn (of worden gesteld) om betekenisvolle controle op AI uit te oefenen. En dit geldt ook voor rechters.

## Datatoepassingen

Vervolgens is het belangrijk om te kijken welke effecten de inzet van AI in de praktijk heeft. AI is gebaseerd op statistiek, op kansberekening. Daarom is er altijd sprake van een foutmarge. Er zal altijd een (kleine) groep mensen getroffen zijn door het systeem omdat beslissingen in hun geval niet juist blijken. De ervaring met 'traditionele' ICT-systemen leert dat ICT-systemen in de praktijk nooit foutloos zijn en dat eenmaal gemaakte fouten zeer hardnekkig-, en moeilijk te corrigeren zijn. De gevolgen voor burgers of consumenten kunnen zeer schrijnend zijn, zie bijvoorbeeld het Rathenau rapport '[Databases: Over ICT-beloftes, informatiehonger en digitale autonomie](#)' (Munnichs et al, 2009). Het is daarom noodzakelijk om werkbare correctieprocedures te ontwikkelen, en tijdig het systeem te evalueren. Tenslotte kan veel menselijk werk, bijvoorbeeld in de zorg, het onderwijs, de militaire sector alsook in de rechtspraak, niet zomaar door een computer worden overgenomen.

### 3. Naar betekenisvolle publieke en menselijke controle over AI

Bij de inzet van AI voor publieke taken, zoals de rechtspraak, is het van belang publieke waarden goed te borgen. Met betrekking tot de inzet van algoritmen is er behoefte aan transparantie van algoritmen en verheldering van verantwoordelijkheden.

# Rathenau Instituut

Het Rathenau Instituut beveelt daarom de volgende acties aan:

## **Bestendigen van waarden**

Verhelder via onderzoek en dialoog welke publieke waarden en fundamentele mensenrechten van belang zijn bij de inzet van AI (o.a. in het recht) en zodoende gewaarborgd dienen te worden. In de studie [Human rights in the robot age](#) (Van Est & Gerritsen 2017) geeft het Rathenau Instituut aan dat met betrekking tot de rechtspraak het recht op een eerlijk proces, uitlegbaarheid van het vonnis en de verantwoordelijkheid van de menselijke rechter belangrijke kwesties zijn. Gelet op het feit dat vele AI systemen uit andere landen komen, is het voor het vertrouwen in de Nederlandse rechtsspraak van belang dat de Staat op het gebied van AI in het recht technologisch soeverein is.

## **Ethische richtlijnen voor specifieke toepassingen**

Vanuit het oogpunt van *value sensitive design* kan het voor bepaalde toepassingen van algoritmen van belang zijn om de praktische stap te maken van publieke waarden naar ethische richtlijnen, die programmeurs concrete handvatten bieden bij het maken van de algoritmen. Een goed voorbeeld is het werk van de [Duitse ethische commissie](#) op het gebied van geautomatiseerd en verbonden rijden. Het lijkt een goed idee om op het terrein van AI in het recht na te gaan tot welke ethische richtlijnen de hierboven genoemde relevante ethische kwesties dienen te leiden.

## **Maak algoritmes transparant, verklaarbaar en toerekenbaar**

Overheden moeten aan burgers kunnen uitleggen hoe AI die wordt ingezet, werkt en het leven van burgers beïnvloedt. De overheid dient dus zorg te dragen voor *algorithmic accountability* (zie hierboven). Dit kan via het verplicht stellen van een Algoritmische Impact Assessment (AIA). Maak ook helder wie waarvoor verantwoordelijk is.

## **Zorg voor effectief toezicht**

Hoezeer burgers zich ook informeren, ze kunnen maar beperkt controleren of algoritmes maatschappelijk verantwoord zijn. Hier ligt dus een rol voor expertcommissies en toezichthouders.

Hierboven lieten we de algemene principes zien voor borging van publieke belangen in alle stappen van de datawaardeketen. Op dezelfde wijze kunnen we ook ons denken over AI in het recht op gang brengen.

## **Literatuur**

Kool, L., J. Timmer, L. Royakkers & R. van Est (2017) 'Opwaarderen: het borgen van publieke waarden in de digitale samenleving.' Den Haag: Rathenau Instituut.

Kool, L., J. Timmer & R. van Est (2016) 'Datagedreven samenleving: Achtergrondstudie.' Den Haag: Rathenau Instituut.

Munnichs, G., M. Besters & M. Schuijff (2009). 'Databases: Over ICT-beloftes, informatiehonger en digitale autonomie.' Den Haag: Rathenau Instituut.

Van Est, R. & J.B.A. Gerritsen, m.m.v. L. Kool (2017) 'Human rights in the robot age.' The Hague: Rathenau Instituut.