

Het menselijke element in automatische systemen

Impressie eerste community meeting Ontwerpen voor publieke waarden in een digitale wereld op 16 december 2020

Ruim veertig onderzoekers, beleidsmakers en andere professionals uit verschillende maatschappelijke domeinen kwamen *online* samen rond het concept *Meaningful Human Control*. Toepassing hiervan kan helpen bij het borgen van belangrijke maatschappelijke waarden in automatische systemen in bijvoorbeeld auto's of wapens, maar ook bij allerlei vormen van (publieke) besluitvorming. Een korte impressie van de ochtend, waarin duidelijk werd dat het menselijke element in automatische systemen bewust vormgegeven moet en kan worden.

Ymkje de Boer (adviseur kennisdeling voor NWO) opende de bijeenkomst en vertelde dat de *community* is ontstaan binnen het NWO-MVI programma *Ontwerpen voor publieke waarden in een digitale wereld*. 'Vandaag is de eerste bijeenkomst. Hier zijn onderzoekers van de vijf onderzoeksprojecten binnen ons programma aanwezig, evenals beleidsmakers van de vijf overheden die het programma samen met NWO hebben opgericht. Er zijn vandaag ook mensen van buiten deze directe kring en dat is precies de bedoeling. We willen de *community* laten groeien en in de komende jaren een op kennis gebaseerde dialoog met elkaar voeren over urgente issues en nieuwe inzichten rond de borging van publieke waarden in automatische systemen. Ook is de meeting van vandaag bedoeld om nader kennis te maken met de vijf projecten.'

Waarden vanaf het begin meenemen in innovaties

Jeroen van den Hoven, hoogleraar *Ethics and Technology* aan de Technische Universiteit Delft en lid van de NWO-MVI-raad, trapte af. Hij lichtte de achtergronden van Maatschappelijk Verantwoord Innoveren (MVI) toe en gaf aan waarom het concept *Meaningful Human Control* hierin belangrijk kan zijn. 'Controle is een voorwaarde voor het kunnen nemen, toedelen en voelen van verantwoordelijkheid rond technologie. Nu we steeds meer te maken krijgen met geautomatiseerde systemen in allerlei domeinen, moeten we de vraag wie verantwoordelijk is voor verkeerde beslissingen of ongelukken wel kunnen blijven beantwoorden. Het ontwerpen voor verantwoordelijkheid past helemaal in de benadering van Maatschappelijk Verantwoordelijk Innoveren en *responsive design*: vanaf het begin neem je belangrijke waarden mee bij het ontwerpen van bijvoorbeeld soft- en hardware. Daarvoor is de inbreng van verschillende wetenschappelijke disciplines en van alle relevante partijen uit het beleids- en praktijkveld nodig.'

If you liked this target, you may also like this one!

Van den Hoven: 'AI kan onze kennis, controle, vrijheid en autonomie ondermijnen. Waar we naartoe moeten, is dat AI zo ontworpen wordt dat al deze waarden juist worden ondersteund. Wat we natuurlijk niet willen, is dat de algoritmen die bij Netflix, Spotify en Youtube leuke aanbevelingen doen op basis van je smaak straks in automatische wapensystemen ook zelfstandig doelen gaan opzoeken om te beschieten. Er staat dus nogal wat op het spel hier. En ons onderzoeksprogramma en deze *community* zijn ervoor bedoeld om daarover kennis en ervaringen uit te wisselen.'

Inald Lagendijk, hoogleraar *Computing-based Society* aan de Technische Universiteit Delft (bij [AItech](#)), gaf vervolgens als gastspreker meer inzicht in het concept *Meaningful Human Control*. 'Ik presenteer een denkrichting die we aan het uitproberen zijn, want AI wordt niet vanzelf 'mensgericht'. We willen geen 'ethical white washing' – een 'ethische sausje' dat achteraf over ontwerpen gegoten kan worden. Nee, het gaat om het inbouwen van de ethiek in het ontwerp van autonome systemen zelf.'

Tracking & tracing gezamenlijk vormgeven

Lagendijk illustreerde zijn betoog aan de hand van een voorbeeld over de geautomatiseerde selectie van sollicitanten voor een baan: zijn ze geschikt genoeg om voor een gesprek te worden uitgenodigd op basis van hun CV? 'Een AI-systeem zal dit nooit 100% goed doen, maar bijvoorbeeld wel in 80 of 90% van de gevallen. De zogenoemde 'reward-function' is dan bijvoorbeeld 0.87. Daar kun je tevreden mee zijn, maar als nu blijkt dat één specifieke groep sollicitanten stelselmatig en onterecht minder goed door de selectie komt, dan is er iets mis. De procedure moet immers eerlijk gaan en daarvoor is een menselijke afweging nodig: wat vinden we eigenlijk precies eerlijk? En wie maakt dat uit? Twee begrippen spelen nu een belangrijke rol: *tracking* en *tracing*. *Tracking* betekent dat herleidbaar is hoe het systeem zich verhoudt tot belangrijke waarden, zoals hier eerlijkheid; *tracing* betekent dat het gedrag, de capaciteiten en de uitkomsten van het systeem op z'n minst tot één menselijke actor moeten worden kunnen teruggeleid. Om *tracking* en *tracing* goed te kunnen vormgeven voor daadwerkelijke *Meaningful Human Control* moeten relevante stakeholders met elkaar aan tafel.'

Moral Operational Design Domain

Lagendijk: 'Het is dus niet genoeg om alleen een *Operational Design Domain* te hebben – de technische specificaties van waar een systeem aan moet voldoen. Nee, er is ook een *Moral Operational Design Domain* bij nodig. Daarin leg je de waarden, normen en grenzen vast waarbinnen de AI moet functioneren. Dat is iets dat je in dialoog met elkaar doet. Je kunt dan ook aangeven dat sommige beslissingen nu eenmaal niet door de machine kunnen worden genomen, maar alleen door de betrokken mensen.

We willen deze denkrichting verder ontwikkelen zodat we tot een soort checklist kunnen komen voor het ontwerp van met name de systemen die een grote maatschappelijke impact kunnen hebben. Denk hierbij zeker ook aan de technologieën waaraan in de vijf onderzoeksprojecten van dit NWO-programma wordt gewerkt.' Een deelnemer vroeg of we op deze manier aan AI geen hogere eisen stellen dan aan gewoon menselijk gedrag, dat immers ook feilbaar is – bijvoorbeeld in sollicitatieprocedures. Lagendijk: 'Menselijke feilbaarheid ontslaat ons niet van de plicht om de verantwoordelijkheid voor geautomatiseerde beslissingen goed in te bouwen.' En het articuleren van wat nu echt eerlijk is – is dat wel te doen? Lagendijk: 'Het is een proces waarbij iedereen aan tafel moet zitten. In andere domeinen zijn we dat allang gewend. Bij het vormgeven van automatische systemen moet dat net zo goed.'

Recht op zinvolle menselijke tussenkomst

Jurgen Goossens, universitair hoofddocent staats- en bestuursrecht aan Tilburg University en projectleider van een van de vijf projecten, was de laatste spreker. In het project *CHAIN* (*Connecting Humankind through Algorithms and Information Networks*) over de inzet van de gedistribueerde technologie blockchain en het gebruik van algoritmen (smart contracts) door de overheid, nemen de onderzoekers een constitutionele perspectief in rond geautomatiseerde overheidsbesluitvorming. Goossens heeft daar samen met collega's recent enkele belangrijke publicaties over uitgebracht. 'Stel dat de overheid in hoge mate geautomatiseerde besluiten gaat nemen, dan zou er toch sprake moeten zijn van een recht op zinvolle menselijke tussenkomst, zoals dit nu bijvoorbeeld ook al bestaat in de wetgeving rond verwerking van persoonsgegevens. We hebben zowel te maken met beslissingsondersteuning door AI – in dat geval vooral bij de zogenoemde discretionaire bevoegdheden van de overheid –, als met *volledig geautomatiseerde* beslissingen, bijvoorbeeld over belastingtoeslagen of studiefinanciering via DUO. Dat laatste zijn 'gebonden bevoegdheden' van de overheid. Wat met name in dat laatste geval kan gebeuren, is dat beslissingen 'lost in automation' raken of dat er heel rigide en gede-individualiseerde beslissingen ontstaan. Dan willen we toch wel dat er nog een mens is die kan ingrijpen.'

Onder druk

In alle gevallen van overheidsbeslissingen moet er een verbinding gelegd worden tussen algemene regels en individuele beslissingen via een proces van toepassing. Goossens: 'In principe is het toepassen van regels op individuele gevallen mensenwerk. Daarbij horen onder meer de algemene beginselen van behoorlijk bestuur, en de democratische en rechtelijke controle. Maar hoe meer onderdelen in het besluitvormingsproces worden vervangen door AI, des te meer de beginselen van de democratische rechtsstaat en beginselen van behoorlijk bestuur onder druk komen te staan. Hoe transparant is geautomatiseerde overheidsbesluitvorming bijvoorbeeld nog voor het parlement of de rechter als zij controle uitoefenen? En worden de algemene beginselen misschien zelfs niet door AI getransformeerd?'

Juridisch en institutioneel werk aan de winkel

Goossens noemde verschillende gevaren van deze ontwikkeling, waaronder beslissingen die worden gebaseerd op data met een '*bias*' of op ongecorrigeerde en verouderde data. 'Zowel vanuit constitutioneel als bestuursrechtelijk oogpunt moeten we opnieuw kijken deze situatie en het belang van een '*human-centred AI*' borgen. We concluderen dan dat de mens altijd '*in the loop*' moet blijven. Die mens moet een rol spelen als verantwoordelijke als beslisser of op zijn minst als – zinvolle – controleur. De mens moet beslissen hoe en wanneer algoritmen worden ingezet en gecontroleerd in geautomatiseerde besluitvormingsprocessen en deze goed ontwerpen en 'trainen'. Mijn stelling voor het debat zou zijn dat de wetgever in het geval van algoritmische besluitvorming een recht van zinvolle menselijke tussenkomst en verplichting tot zinvolle menselijke controle door het bestuur uitdrukkelijk verplicht zou moeten stellen.' Is de huidige wetgeving hiervoor afdoende, vroeg een deelnemer. 'De algemene beginselen zijn wel robuust en kunnen mee-ontwikkelen; het gaat dan om een herijking en herinterpretatie ervan. In de bestuurlijke praktijk kunnen er dingen misgaan met algoritmische besluitvorming en dan moet de burger wel voldoende rechten hebben.' Is het dan niet vooral een institutioneel probleem in plaats van een juridisch probleem? 'Beide,' stelde Goossens. 'We kijken ook in ons onderzoek niet alleen naar de juridische kant, maar ook naar de betrokken actoren of instituties, de operationalisering van wet- en regelgeving in de uitvoeringspraktijk en hoe het daadwerkelijk wordt ervaren door burgers.'

Na dit plenaire deel van de meeting ging een deel van het gezelschap uiteen in groepen om door te praten over de thematiek en nader met elkaar kennis te maken.

De volgende meeting is in de eerste helft van 2021, waarschijnlijk ook nog online. Wilt u op onze verzendlijst? Mail mvi@mba.nl.

Gehouden presentaties:

- Jeroen van den Hoven ([PDF](#))
- Inald Lagendijk ([PDF](#))
- Jurgen Goossens ([PDF](#))
- Algoritmische beslisregels en geautomatiseerde feitenvaststelling: gevolgen voor de fundamentele uitgangspunten van het Nederlandse constitutionele recht, Preadvies Staatsrechtconferentie 2020, Jurgen Goossens, Ernst Hirsch Ballin, Eva van Vugt ([PDF](#))
- Effectieve rechtsbescherming bij algoritmische besluitvorming in het bestuursrecht, Jurgen de Poorter en Jurgen Goossens ([PDF](#))