

AI in besluitvorming en toezicht bij de overheid

Deze zienswijze van Stichting Kafkabrigade beperkt zich tot kunstmatige intelligentie bij de overheid en daarbinnen tot gebruik bij individuele besluitvorming en toezicht.

Waar heeft de WRR het over ?

Hoewel het rapport van de WRR een definitie¹ van AI hanteert waarbinnen ook een thermostaat valt² – omdat ze zich niet wil beperken tot een specifiek onderdeel van de technologie en omdat ze zegt dat de overheid weinig gebruik maakt van nieuwe technieken als deep learning³ – richt het rapport zich toch vrijwel uitsluitend op “de dingen waarvan we nu onder de indruk zijn⁴” – te weten statistische methoden en met name zelflerende algoritmen – en niet op de toepassingen die nu massaal in gebruik zijn bij de overheid – rule-based algoritmen⁵. De definitie is functioneel in het licht van de ambities van het vakgebied.

Waar heeft de WRR het niet over?

Bij individuele beslissingen over burgers speelt een ander belangrijk onderscheid: wat is de aard van de gegevens die gebruikt worden? Feitelijke en individuele gegevens kunnen gebruikt worden voor individuele beslissingen. Zachte of geaggreerde gegevens kunnen slechts een signaalfunctie hebben voor onderzoek waarvan de burger geen last heeft. De reden daarvoor is gelegen in de breed gedeelde ideeën in de Westerse cultuur over wat rechtvaardig en onrechtvaardig, terecht en onterecht, redelijk en onredelijk⁶. Al rijden veel mensen van mijn leeftijd en geslacht te hard, dat rechtvaardigt geen verkeersboete. Daarvoor moet een individueel feit zijn vastgesteld. Ook de eis tot herhaalde verantwoording, omdat ik afwijk van een statistiek of daaraan juist voldoe vinden we niet

1 “Systemen die intelligent gedrag vertonen door hun omgeving te analyseren en - met enige graad van autonomie - actie te ondernemen om specifieke doelen te bereiken” van de High-Level Expert Group on Artificial Intelligence. Zie p. 44 en 48 van het rapport.

2 Zoals ze zelf schrijft op pagina 44.

3 p. 47

4 Bostrom, 2016 in het rapport op pagina 46.

5 De term wordt exact één keer genoemd, p 63.

6 Widlak, A.C., Van Eck, B.M.A. & Peeters, R. (2021). Towards Principles of Good Digital Administration. In: R. Peeters & M. Schuilenburg (red.). The Algorithmic Society: Technology, Power and Knowledge. (pp. 67-83). Routledge, London.

acceptabel, of dit nu racistisch is of een ander willekeurig groepskenmerk blijkt. In veel gevallen is minder belangrijk of we van een techniek onder de indruk zijn, zoals van *machine learning*, of niet, zoals bij bestandsvergelijkingen. Een belangrijker onderscheid – binnen de context van besluitvorming – is meestal de aard van de gebruikte data. En daarnaast hoe die techniek gebruikt wordt, wat de plaats van die techniek is in het uitvoeringsproces en met welk gevolg die data gebruikt wordt.

Een nadere demistificatie

De WRR duidt als aanjagers van de huidige doorbraken van AI de wetenschappelijke doorbraken (1), de grotere rekenkracht (2) en de beschikbare data(3)⁷. Wat niet benadrukt wordt in het rapport is de specifieke onderzoeksopzet van de bekende doorbraken waarvan we onder de indruk zijn⁸. Daarin kan een computer zichzelf trainen. Een computer kan reeds bepalen wanneer een zet of strategie in een spel – zoals Go of een Atari computerspel – uiteindelijk leidt tot winst of verlies. Daarom kan een zelflerend algoritme worden voorzien van glasheldere feedback door de computer zelf, snel en zonder dataverzameling. Niet voor elk probleem gelden deze uitzonderlijke omstandigheden. Vergelijkbaar succes is daarom niet verzekerd.

Juist gebrek aan data maakt dat statistische technieken – zoals AI – niet per se gaan leiden tot de verwachtingen die we daarvan nu lijken te hebben. Een voorbeeld buiten de overheid is de verwachting van minder dodelijke ongelukken door een zelfrijdende auto. Er zijn simpelweg te weinig dodelijke ongelukken per miljard kilometers autoverkeer om dat überhaupt vast te kunnen stellen op een afzienbare termijn⁹. En binnen de overheid had de IND ook al voordat zij ORKA¹⁰ ontwikkelde kunnen berekenen dat er te weinig voorbeelden van fraude zijn om tot een betrouwbaar resultaat te komen.

Het is belangrijk te beseffen dat AI, in de betekenis van machine learning, uiteindelijk patroonherkenning is. Het is het vinden – met trail-and-error – van een verzameling gewichtjes waarmee het ene patroon – de input – zo goed mogelijk vertaald kan worden

7 P. 79

8 Waarbij overigens een bricolage van technologieën ingezet wordt gericht op een specifiek en smal doel.

9 In 2017 reden we in Nederland 119,1 miljard kilometer en waren er 201 verkeersdoden door een personen- of bestelauto.

10 <https://www.nrc.nl/nieuws/2022/05/06/ind-verzweeg-een-dikke-error-met-kennismigranten-a4123661>

in het andere patroon – de output. Het is daarom dat AI in deze betekenis niet uitlegbaar is. Voor ons mensen heeft die verzameling getalletjes geen betekenis. Het is toevallig de verzameling getalletjes die het ene patroon het beste vertaalt in het andere. Het is geen reden, geen argument en geen regel die in woorden is uit te drukken.

Bovenstaande maakt inzichtelijk dat het idee om AI “uitlegbaar” te maken door hier argumenten aan te verbinden misleidend is. Er is geen “regel” of “argument” gevonden. Het blijft statistiek. Ook een chat is juist niet de plek om deze technologie in te zetten. Als burgers menselijke intelligentie verwachten en geconfronteerd worden met een geheel ander soort intelligentie, waarmee niet op een menselijke manier te praten valt, schep je vervreemding.

Dit zijn geen argumenten om de betekenis van AI te trivialisieren, zeker niet. Het is niet moeilijk om allerlei toepassingen te bedenken, ook binnen de overheid. De ANPR camera, die nu alleen een kenteken herkent, is in allerlei varianten denkbaar, zoals het vasthouden van een telefoon bijvoorbeeld. Een zeer hoge accuratesse is dan wellicht voldoende argument om controle op fouten te verplaatsen naar de burger ten behoeve van efficiëntievoordelen voor de overheid¹¹. Maar het zijn wel argumenten om de trendgevoeligheid van de overheid voor nieuwe middelen tegenwicht te geven in het licht van hun opgave.

Positioneren van AI en statistische toepassingen in overheidsbesluitvorming

In de afbeelding hieronder is onder de kop “Dienstverlening” het primaire proces te zien van elke digitale uitvoeringsorganisatie of -afdeling. Dit proces is vaak volledig geautomatiseerd volgens wat het WRR-rapport duidt als de “rule-based” benadering van AI, maar verder buiten beschouwing laat. Het gaat hier om logische regels van het type “als X, dan Y”, waarbij feitelijke en individuele gegevens gebruikt worden. Van deze techniek zijn we niet meer onder de indruk, is *schijnbaar* eenvoudig, in gebruik voor de meerderheid van de individuele besluiten en heeft grote impact op de rechten van burgers, door omvang, financiële gevolgen en problemen bij het herstel van fouten¹².

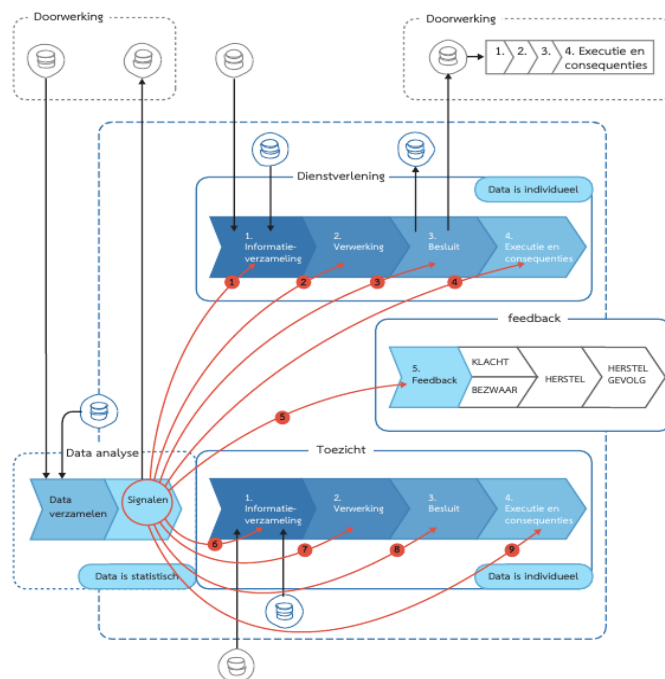
11 Hoe dit ook fout kan gaan – en niet hersteld bleek te kunnen worden – bleek in 2013 toen campers en fietsenrekken werden aangezien voor aanhangwagens. Zie: Raad van State, 2018, Ongevraagd advies over de effecten van de digitalisering voor de rechtsstatelijke verhoudingen, p. 6.

12 Hierop wijst ook het Rathenau Instituut in het slotwoord van “Algoritmes afwegen”. (17 mei 2022)

Statistische methoden, zoals AI in de betekenis van machine-learning, zijn te vinden onder de kop “Data-analyse”. In de praktijk worden die nu (vrijwel) alleen¹³ gebruikt om het toezicht te richten¹⁴. Daarom staat dit voor de toezichtskolom. En deze kop staat op de grens van de organisatie en de buitenwereld, omdat uitvoerders dit vaak in samenwerking doen. Hier zijn de gegevens zacht en/of statistisch van aard.

De rode lijnen in de afbeelding zijn de nieuwste en potentiële toepassingen van risicoselectie in de nabije toekomst. Elk van deze lijnen vraagt om explicitering van de normstelling in termen van behoorlijk bestuur. Mag een signaal, gegeneerd door een AI of een andere statistische methode, leiden tot automatische “uitworp”, waardoor een burger niet meegaat in het automatische proces van besluitvorming, maar een andere behandeling krijgt? Is dit geen probleem, omdat dit leidt tot een behandeling door mensen? Of is dit juist wel een probleem, omdat de uitworp leidt tot een verwachtingspatroon bij de behandelaren dat er wel iets mis moet zijn? Mag het leiden tot uitsluiting van een bepaald kanaal, zoals de website? Als een AI of een andere statistische methode, leidt tot een selectie voor het toezicht, mag dat leiden tot opschorting van dienstverlening? Mag het leiden tot onderzoek waar de burger last van heeft? Mag het leiden tot (herhaalde) verantwoording? Mag dat leiden tot een gegeven in een registratie die gedeeld wordt met andere organisaties? Of is dit dwaallichtinformatie, die leidt tot onterechte verdachtmakingen?

Het zijn vragen die gelden voor AI, niet alleen omdat de selectie een black box is, maar omdat de basis geen feitelijk of individueel gegeven is of beide. AI heeft dit overeen met allerlei andere statistische technieken. Daarmee is niet gezegd dat deze techniek niet bruikbaar is, maar wel dat een normatief vacuüm niet bijdraagt aan behoorlijk bestuur.



13 Zie het WRR working paper “Big Data voor fraudebestrijding”, WRR, 2016.

14 In het debat van alledag praten we over fraude, waarbij het in de praktijk doorgaans gaat om het overtreden van de inlichtingenplicht van burgers.

Enkele relevante aandachtspunten voor de commissie Digitale Zaken

1. Bij AI als artificiële neurale netwerken (en andere statistische methoden)

Het gebruik van statistische methoden kan een plaats hebben om het toezicht te richten, mits er voldoende bewustzijn is dat het niet gaat om een verdenking, maar om signalen die tientallen verklaringen kunnen hebben¹⁵, waarvan een fout van of fraude door de burger er slechts één is. De omkering van de bewijslast, waarbij een burger zich moet verantwoorden op basis van een signaal, omdat de overheid haar gegevenshuishouding onvoldoende op orde heeft, is hier het aandachtspunt.

Signalen worden ook wel gebruikt om te zorgen voor zogenaamde “gewenste uitval”. Dit is rode lijn “2” in de afbeelding. Een burger gaat dan niet mee in het geautomatiseerde proces, maar krijgt een andere behandeling. Voor burgers is dit onzichtbaar. Dat geldt niet voor rode lijn “1”, waar een burger wordt uitgesloten van een bepaald kanaal, zoals een aanvraag via de website. Of we het gebruik van AI voor deze zaken acceptabel vinden of niet, is een politieke vraag.

Een deel van het antwoord op deze vraag ligt natuurlijk in de rol van de mens in het proces. Juist bij uitworp is er een mens die uiteindelijk de beslissing neemt. Echter de rol die deze ambtenaar heeft gekregen is totaal van karakter veranderd. In welke mate wordt hij beïnvloed door de selectie die met AI gemaakt wordt? In welke mate kan hij die selectie beoordelen of corrigeren? In hoeverre maakt hij gebruik van signalen (rode lijn “7”) bij zijn beoordeling en is er dan sprake van informatie of eerder van een dwaallicht?

2. Bij AI als rule-based systemen

Terecht stelt de WRR dat een “black box”-karakter ook kan optreden in andere domeinen,¹⁶ zoals bij rule-based algoritmen. Dit is echter een ander soort “black box”, omdat de oorzaak anders is. Zo is de omvang van deze “rules” dusdanig dat zij niet meer toetsbaar¹⁷ is, voor een rechter bijvoorbeeld. Transparantie en toetsbaarheid zou bereikt kunnen worden de beslisregels van dergelijke algoritmen toegankelijk en doorzoekbaar te maken. Dan zou – bijvoorbeeld – een ondersteunende afdeling van de rechtspraak de

15 Algemene Rekenkamer, 2014, Basisregistraties, p. 20 e.v.

16 p. 47

17 Van Eck, 2018, Geautomatiseerde Ketenbesluiten & Rechtsbescherming. p. 199.

relevante beslisregels bij een geautomatiseerd besluit naar voren kunnen halen voor de rechter, net zoals ook relevante wetgeving en jurisprudentie wordt aangereikt via ondersteuning.

3. De rol van data bij zowel rule-based AI als machine learning AI

De rol van data voor de kwaliteit van de uitkomsten is bij beide vormen van AI cruciaal. De mogelijkheid om in de Nederlandse gegevensinfrastructuur correcties te verwerken of überhaupt waar te nemen dat er sprake is van een correctie is (zeer) beperkt. Ook de meta-datering en logging laat te wensen over. Wanneer onduidelijk is welke feiten actueel waren, gebruikt zijn en waarvan ze afkomstig zijn, is onduidelijk wat de conclusies van een AI systeem waard zijn.

4. De rol van overzicht van gegevensstromen

AI – in beide vormen – zien we ook wel als socio-technisch systeem, waarin niet alleen het algoritme een rol speelt, maar ook de mens en ook de instituties. Eén onderontwikkelde institutie is de administratie van gegevensstromen. Door het gebrek aan overzicht van welke organisaties gegevens afkomstig zijn – de input van een AI systeem – en waar ze heen gaan – de output van een AI systeem – kunnen allerlei zogenaamde “ripple effects” optreden. Dat wil zeggen dat er gevolgen optreden buiten de organisatie of set van organisaties die we in ogenschouw nemen. Met name gebruik van de output van AI systemen door (vele) organisaties (ver) buiten de organisatie waar het systeem tot een bepaalde conclusie kwam en zonder institutionele actor die deze ketens overziet, leidt makkelijk tot het verlies van context. Wat informatie lijkt, blijkt dan een dwaallicht.