Tussen utopie en dystopie: hoe professionals artificiële intelligentie actief (kunnen) vormgeven

Jan-Luuk Hoff en Stephan Grimmelikhuijsen

Inleiding

'Wat betekent de opkomst van algoritmen voor de rechtspraak?' Deze vraag stond centraal op de bijeenkomst die het hof Amsterdam organiseerde over rechtspraak en artificiële intelligentie (AI). Als bestuurskundige onderzoekers en relatieve 'buitenstaanders' bij dit debat willen wij de lezingen in perspectief plaatsen en ons begrip van AI in de rechtspraak vergroten, vooral door het 'begrip' AI te problematiseren. Dat doen we aan de hand van sociologische en bestuurskundige inzichten. We zetten de twee uiterste interpretaties van AI – 'utopisch' versus 'dystopisch' – tegenover elkaar, betrekken die interpretaties meer verfijnd op elkaar, gelinkt aan context en handelende actoren. Want zonder context en acties geen AI.

Uitersten

In het publieke debat worden vaak twee uiterste visies op AI tegenover elkaar gesteld:

- de *utopische* visie op AI, die de belofte van een waardenvrij systeem vooropstelt, met eerlijkere en efficiëntere beslissingen vergeleken met menselijke beslissingen;
- de *dystopische* visie, die het discriminatoire en oncontroleerbare karakter van AI benadrukt: als er eenmaal iets fout zit in het systeem is het moeilijk dit te achterhalen en corrigeren.[1]

Wat beide visies gemeen hebben is dat AI een monolithisch karakter wordt toegekend. Met andere woorden, AI als een coherente, eenduidige en vastomlijnde ontwikkeling die onze

1 Ziewitz, M. (2016). Special Issue Introduction Governing Algorithms: Myth, Mess, and Methods. *Science, Technology, & Human Values*, 1, 3-16.

maatschappij – en daarbinnen de rechtspraak – zal beïnvloeden. Zo wordt er in de rechtspraak als snel gesproken over de komst van 'de robotrechter' die rechterlijke besluitvorming zal gaan verbeteren, efficiënter maken en – ooit – rechters overbodig zal maken. AI is echter veel meer en tegelijkertijd minder dan een robotrechter. We moeten uitkijken voor het toekennen van eenduidige eigenschappen en invloeden aan AI in de rechtspraak. AI kent namelijk veel verschillende gedaanten, krijgt pas betekenis in de rechtspraktijk, en kan deels actief door de professie worden vormgegeven.

Menselijke en professionele praktijken

Floris Bex, hoogleraar *Data science* aan de Tilburg University, liet aan de hand van de 'slimme' *chatbot* die hij ontwikkelt voor het doen van online aangifte bij handelsfraude, tijdens zijn lezing zien dat de AI-technologie die schuilgaat achter deze chatbot geen coherent geheel is, maar een assemblage van codes, datastromen, modellen, processen en systemen. Volgens ons laat de presentatie van Bex ook zien dat momenteel zelfs het bouwen van een relatief simpele toepassing, zoals het beoordelen van potentiële online aangifte van handelsfraude, mensenwerk is. Zo vraagt het ontwikkelen van de juiste algoritmen en beslisbomen veel menselijke en waarde-geladen beslissingen.
Dit impliceert dat AI-technologie in het ene domein niet zomaar kan worden overgedragen naar een ander domein. Met andere woorden, het feit dat AI zorgt dat 'de computer' nu beter *Go* kan spelen dan de mens, betekent nog niet dat AI ook in andere domeinen – laat staan de rechtspraak – slimmere beslissingen kan nemen. Iedere toepassing staat op zich en het is goed mogelijk dat er voorlopig geen slimme robotrechter zal komen, omdat de data of de technologie dit nog niet mogelijk maakt.

Corien Prins, voorzitter van de Wetenschappelijke Raad voor Regeringsbeleid (WRR), liet zien dat de opkomst van AI-technologie belangrijke vragen opwerpt voor de Rechtspraak. De wereld digitaliseert in snel tempo en de mogelijkheden van AI worden al gebruikt in de commerciële advocatuur. Het is daarom volgens haar een zaak van professionele verantwoordelijkheid dat de Rechtspraak verkent of en hoe AI toegepast kan worden in de werkpraktijk. Bij die verkenning kan het door de WRR ontworpen kompas behulpzaam zijn, dat focust op de betrouwbaarheid en transparantie van AI, het beschermen van fundamentele rechten van cliënten en de onafhankelijkheid en behoorlijkheid van rechtspleging.
Opvallend in de lezing van Corien Prins was dat AI niet zozeer als technologisch fenomeen werd geconceptualiseerd, maar veel meer als een bredere technologische beweging van buitenaf, sterk verbonden met digitalisering en de toenemende machtspositie van grote commerciële partijen.

Impliciet laat deze lezing, net als die van Bex, zien dat we moeten oppassen AI te zien als een gesloten, afgebakend systeem dat wordt geïmplementeerd in een voorspelbare omge-

ving die losstaat van deze verstoring.[2] Zoals Prins terecht aanstipte, kan AI de rechtspraak bedreigen wanneer deze de toegankelijkheid, transparantie en rechtvaardigheid van de rechtspraak aantast. Echter, het grootste gevaar van het algoritme zit daarbij niet *in de technologie zelf*, maar in de *wijzen waarop technologie betekenis krijgt* in bestaande praktijken, in dit geval professionele praktijken, zoals de rechtspraak.
In die bestaande professionele praktijken lijkt de aandacht vooral uit te gaan naar het technische ontwerp van AI en de kwaliteit van data, met in het achterhoofd het geloof dat de werking van AI-technologie controleerbaar is, wanneer dit maar goed gereguleerd is. Door deze nadruk op de technische aspecten van AI-systemen gaat het besef verloren dat technologie niet aan de *tekentafel* betekenis krijgt, maar in de *praktijken* waarin ze interacteert met de professionals waar ze voor ontworpen is.

Interacties in context

In dit licht geeft het werk van Wanda Orlikowski over de dualiteit van technologie en organisaties interessante inzichten.[3] Volgens Orlikowski moeten we technologie als zodanig niet zien als alleen 'hardware' (data, systemen, algoritmen), maar ook niet puur als uitkomst van menselijk handelen. Technologie kent immers zekere eigenschappen die het menselijk handelen beïnvloeden en begrenzen. Tegelijkertijd stuurt menselijk handelen de werking van technologie. Daarom moeten we de werking van technologie zien als product van de interactie van mens en technologie, welke weer wordt beïnvloed door de institutionele context. Kortom, AI krijgt vorm en betekenis in de institutionele, rechtstatelijke context waarin deze interacteert met rechters.

Een gevolg van deze redenering is wel dat we bij de ontwikkeling van AI moeten uitgaan van de onvoorspelbaarheid van deze technologie.[4] We weten immers niet hoe rechters en technologie zich gaan gedragen wanneer zij interacteren in de praktijk. Een goed voorbeeld hiervan is te vinden in de medische wereld. Daar werd een algoritme ontworpen dat op basis van genetisch materiaal, klinische signalen en de medische geschiedenis van patiënten een voorspelling gaf van de verwachte effectiviteit van medicijnen tegen borstkanker. Artsen zagen op hun scherm een lijst met medicijnen en daarachter de verwachte effectiviteit van het medicijn,

2 Gillespie, T. (2016). #trendingistrending: when algorithms become culture – Culture Digitally. In R. Seyfert & J. Roberge (Eds.), *Algorithmic Cultures: Essays on Meaning, Performance and New Technologies* (pp. 53-75), Routledge.

3 Orlikowski, W. J. (1992). The duality of technology: Rethinking the concept of technology in organizations. *Organization science*, *3*(3), 398-427.

4 Barley, S. R. (1986). Technology as an occasion for structuring: Evidence from observations of CT scanners and the social order of radiology departments. *Administrative science quarterly*, 78-108.

uitgedrukt in een numerieke score. Uiteindelijk was het wel de arts die een beslissing moest nemen over de behandeling. De ontwerpers van het algoritme waren ervan overtuigd dat het gebruik van dit algoritme de behandelkeuzes van professionals effectiever en rationeler zou maken, omdat deze gebaseerd zouden worden op precieze en persoonlijke data van patiënten. Echter, de artsen die in de praktijk met het algoritme gingen werken worstelden met de uitkomsten van het algoritmen. De uitkomsten van het algoritme weken regelmatig af van de voor hen bekende professionele normen en routines, wat voor verwarring zorgde bij artsen over de manier waarop zij de score van het algoritme moesten interpreteren. Zij hadden onvoldoende kennis van de werking van het algoritme, welke data er werd gebruikt en hoe de output van het algoritme geïnterpreteerd kon worden. Daardoor verwerd de output van het algoritme tot een onzekere en ambigue bron van informatie, waardoor het algoritme niet het 'gewenste' effect had op de besluitvorming.[5]

Relaties, in en rond de professie

Dit voorbeeld laat zien dat het belangrijk is om meer aandacht te hebben voor de manier waarop AI-technologie zijn betekenis krijgt in de praktijk en intervenieert in *relaties* tussen rechtszoekenden, rechters, politiek en samenleving. Gezien de druk vanuit politiek en samenleving op de toegang van de rechtspraak en de efficiëntie van rechtsgangen, is het van belang dat de Rechtspraak zich actief inzet om AI-technologie onderdeel te maken van hun praktijken. Juist wanneer het initiatief komt vanuit de professie zelf, kunnen AI-toepassingen ontworpen worden die niet alleen op de tekentafel veelbelovend zijn, maar ook in de praktijk van waarde zijn.

Het is bijvoorbeeld niet alleen belangrijk om uit te kunnen leggen hoe een algoritme tot een advies voor een rechtelijke uitspraak komt, maar nog veel belangrijker om te verklaren hoe een rechter vervolgens met dit advies omgaat en het gebruikt. Alleen dan kunnen rechterlijke uitspraken efficiënter of transparanter worden. Het is daarom noodzakelijk dat rechters begrijpen hoe algoritmen redeneren en dat ontwerpers van algoritmen snappen hoe rechters redeneren. Dat vereist een open en verbindende houding van de rechters, zodat AI-toepassingen in samenspraak met professionals ontwikkeld en geïmplementeerd kunnen worden. Dit moet tevens goed ingebed worden op institutioneel niveau. Opleidingen moeten meer aandacht hebben voor de toepassing en het gebruik van AI-technologie in de rechtspraak en moeten studenten de vaardigheden meegeven om hun kennis en expertise te verbinden aan nieuwe vormen van expertise zoals AI.

5 Chorev, N. E. (2019). Data ambiguity and clinical decision making: A qualitative case study of the use of predictive information technologies in a personalized cancer clinical trial. *Health Informatics Journal*, 1-11.

Tot slot

Het onderzoek dat promovenda en senior rechter Manuella van der Put bij het hof Amsterdam presenteerde is een duidelijk voorbeeld van de verbindende houding die wij bepleiten. In haar onderzoek kijkt zij hoe AI-technologieën kunnen bijdragen aan een efficiëntere rechtspraak. Dit doet ze niet alleen theoretisch, door te kijken op welke gebieden AI en rechters elkaar aanvullen, maar ook in de praktijk, door in samenspraak met de professionele praktijk te werken aan een AI-kennissysteem dat rechters moet gaan ondersteunen in hun werk.

Het zijn dit soort initiatieven die cruciaal zijn in het maken van een constructieve verbinding tussen de expertise van artificiële intelligentie en de expertise van rechters. Dit is noodzakelijk want AI is geen eenduidige, onveranderbare ontwikkeling die de rechtspraak eenzijdig gaat veranderen, maar een meerkoppig 'monster' waaraan rechters en de rechtspraak als instituut *actief* vorm kunnen en moeten geven.