

Moeten we computers laten rechtspreken?

Jaap Hage

Herneming van een oud debat

De rechterlijke macht en haar organisatie liggen Ton Hol na aan het hart. Maar voor zover mij bekend heeft Ton zich nog niet in een publicatie uitgelaten over de rol van computers of computerprogramma's als mogelijke rechters. Misschien kan het onderstaande hem uitdagen om zich in een debat over computerrechtspraak te mengen.

In de jaren 90 is er in Nederland een kortstondig debat geweest over de mogelijkheid van computerrechtspraak en in de geest van die tijd ging dat vooral over wat computers al dan niet konden, of in de toekomst zouden kunnen. In deze bijdrage zal de nadruk uiteindelijk meer liggen op wat mensen kunnen en of mensen bijzondere capaciteiten hebben die hun rechtspraak voor altijd buiten het bereik van computers zal houden. De bijdrage is bedoeld om te prikkelen en om het debat over computerrechtspraak leven in te blazen. Want dat dit debat (opnieuw) moet worden gevoerd lijkt me duidelijk, nu we steeds meer kennis krijgen over kunstmatige intelligentie en over de beperkte rationaliteit van menselijke besluitvorming.

Wat geschiedenis

Aan het eind van de jaren 80 en in het begin van de jaren 90 waren zogenaamde 'expertsystemen' een belangrijk onderwerp in de ontwikkeling van de kunstmatige intelligentie en ook in het toen opkomende onderzoeksgebied van recht en kunstmatige intelligentie. De gedachte achter een expertsysteem is dat een computerprogramma kennis zou kunnen bevatten, bijvoorbeeld in de vorm van een aantal regels, en dat er daarnaast een redeneermechanisme in zou zitten dat deze kennis zou kunnen toepassen op feiten die door de gebruiker van het systeem zouden worden ingevoerd. Op veel terreinen zou de benodigde kennis afkomstig moeten zijn van experts of het betreffende vakgebied, maar in het recht zouden de benodigde regels voor een groot deel ook ontleend kunnen worden aan de gangbare rechtsbronnen zoals regelgeving en rechtspraak. Juist omdat de benodigde regels in het recht als het ware voor het grijpen zouden liggen, leek het erop dat het recht een geschikt domein zou zijn voor de ontwikkelingen van expertsystemen. En als zulke systemen eenmaal ontwikkeld zouden zijn, zouden ze geschikte hulpmiddelen zijn om goedkope en onpartijdige rechtsbeslissingen te genereren. Met andere woorden, expertsystemen zouden rechters kunnen gaan vervangen in alle behalve de moeilijkste gevallen.

Zo is het dus niet gegaan. Waarom niet? Eén antwoord is dat juridisch redeneren anders werkt dan 'gewoon' redeneren, bijvoorbeeld omdat regels met elkaar in conflict kunnen zijn, of omdat er uitzonderingen op regels mogelijk zijn. Als reactie op dit mogelijke antwoord is er veel onderzoek gedaan naar de logica van juridisch redeneren. Dit onderzoek heeft geleid tot belangrijke doorbraken op het gebied van de logica van het recht, maar deze doorbraken hebben de rechtspraktijk nauwelijks bereikt en bovendien hebben ze er niet toe geleid dat computers op grote schaal zijn ingezet om in concrete gevallen recht te doen.

Een ander antwoord is dat de toepassing van regels op concrete gevallen interpretatie van de regels en kwalificatie van die concrete gevallen veronderstelt. Kun je bijvoorbeeld een regel over dierenwelzijn toepassen op kippen in kippen'fabriek'? Dat lijkt duidelijk het geval, maar als een computer tot die conclusie moet komen, dan moet hij wel 'weten' dat kippen dieren zijn en dat het verblijf in zo'n 'fabriek' implicaties heeft voor het welzijn van deze dieren. Deze kennis wordt wel

'wereldkennis' genoemd, in onderscheid van specialistische 'domeinkennis', en om goede beslissingen te nemen is er veel van dergelijke wereldkennis nodig. In de vorige eeuw konden computers nog niet beschikken over grote hoeveelheden wereldkennis, maar tegenwoordig hebben wij, maar ook computers, het internet en kunnen computers heel aardig natuurlijke taal 'begrijpen'. Het probleem van de wereldkennis lijkt daarmee grotendeels opgelost. Toch is er nog nauwelijks computerrechtspraak.

Het laatste antwoord waar we op in zullen gaan is dat computers, anders dan mensen, geen rechtvaardigheids- of rechtsgevoel hebben en dat dit gevoel wezenlijk is voor goede rechtspraak. Dit laatste lijkt misschien onwaar, omdat rechtspraak vaak bestaat uit het rechttoe rechtaan toepassen van wettelijke regels. Maar ook als het daar uiteindelijk op neer komt, zal een menselijke rechter, al dan niet impliciet, nagaan of wetstoepassing tot onwenselijke resultaten zou leiden. Het oordeel over wat wenselijk is, is - zo zou men kunnen betogen - aan mensen voorbehouden. Computers ontberen het daarvoor noodzakelijke rechtsgevoel en kunnen daarom niet goed rechtsbeslissingen nemen, zelfs niet in eenvoudige gevallen. Immers, het oordeel of een geval eenvoudig is, hangt af van de rechtvaardigheidstoets, waartoe computers niet in staat zijn.

Machine learning

Tot zover de stand van de discussie zoals die in de jaren 80 en 90 van de vorige eeuw was. Maar in diezelfde tijd begon een ontwikkeling in de kunstmatige intelligentie die het laatste decennium tot volle bloei is gekomen. De kennis van een expertsysteem werd nog door mensen in het systeem gestopt. Tegenwoordig hebben we computerprogramma's die zelfstandig leren (*machine learning*) en de resultaten van dit leerproces zijn soms spectaculair en ze zijn niet voor mensen te voorspellen. Zou het niet mogelijk zijn om computers zelf te laten leren hoe ze moeten rechtspreken, net zoals ze binnen een dag hebben geleerd hoe ze moeten schaken of go spelen, of hoe ze de afbeeldingen van foto's - en ook menselijke gezichten - kunnen herkennen?

Misschien is dat mogelijk, maar dan is het wel nodig dat we eerst weten wat een goede rechtsbeslissing is. Als computers zelf iets leren, zijn er twee hoofdvarianten van hoe dat kan: zonder of met toezicht. Bij leren zonder toezicht krijgt de computer gegevens en probeert daarin zelf samenhangen te ontdekken. Het leerproces bestaat dan uit het aanleren van welke samenhangen er bestaan. Zo kan een computer bijvoorbeeld leren dat donkere wolken samen gaan met regen. Bij leren met toezicht krijgt de computer feedback als hij een conclusie trekt uit bepaalde informatie. Dus als hij een foto van Ton krijgt en concludeert dat het een foto van Tineke is, dan krijgt hij negatieve feedback en als hij concludeert dat het een foto van een mens, van een man, of zelfs van Ton is, krijgt hij positieve feedback. Bij positieve feedback wordt de kennis op basis waarvan de conclusie werd getrokken sterker gemaakt, eventueel tot er uiteindelijk een harde regel is gevormd; bij negatieve feedback wordt die 'kennis' zwakker gemaakt, eventueel tot die 'kennis' geheel uit het systeem is verwijderd.

Voor het leren van recht zonder toezicht zou de computer moeten leren om samenhang te zien tussen constellaties van feiten en rechtsgevolgen. Bij het leren met toezicht zou de computer correcte feedback moeten krijgen als hij rechtsgevolgen verbindt aan bepaalde feiten. In beide gevallen is het nodig dat de computer geconfronteerd wordt met de correcte rechtsgevolgen bij een casus. Een complicatie in dit verband is dat het in het recht regelmatig mogelijk is om serieus van mening te verschillen over wat het juiste rechtsgevolg bij een casus is. Voor machinaal leren van recht is het nodig is dat een leercasus een correct antwoord heeft en het lastige bij recht is dat er

regelmatig niet één correct antwoord is. Vanuit rechtstheoretisch perspectief is hierover veel te zeggen, maar hoe boeiend dat ook moge zijn, voor wie een computerprogramma het recht wil laten leren is dit een lastige complicatie.

Menselijke rechters

Als het zo lastig is om recht te leren, hoe kunnen mensen dat dan leren? Wie eerlijk is, zal moeten toegeven dat we dat niet weten. We hebben aan universiteiten onderwijsprogramma's waarin studenten leren wat de regels van het geldende recht zijn, wat goede (en slechte) juridische redeneringen zijn, aan welke 'bronnen' argumenten ontleend kunnen worden, en nog veel meer. Het bestaan van deze programma's draagt eraan bij dat er een sociale groep van juristen ontstaat (of voort bestaat) waarvan de leden het grotendeels eens zijn over wat correcte rechtsbeslissingen zijn. Maar het lijkt erop dat de rechtsbeslissingen correct zijn omdat de juristen het er over eens zijn en dat het niet zo is dat ze het eens zijn omdat de beslissingen (onafhankelijk) correct zijn. Zo lijkt het in het algemeen te zijn, maar soms wordt een afwijkende opvatting zo goed beargumenteerd, dat relevante personen - we zullen ze 'rechters' noemen - overtuigd raken van iets dat bij een minder effectieve argumentatie niet aanvaard zou worden.

Wat maakt nu dat bepaalde argumentatie zo effectief is? Eén antwoord is dat de effectieve argumentatie rechters de zaak vanuit een nieuw perspectief laat zien en dat het onverwachte antwoord vanuit dit nieuwe perspectief het correcte lijkt te zijn. Een ander antwoord is dat sommige pleiters retorisch zo begaafd zijn dat ze rechters iets kunnen laten denken dat ze niet gedacht zouden hebben als ze wat rationeler zouden zijn geweest.

Gelukkig worden rechters in Nederland goed opgeleid en we mogen daarom toch verwachten dat ze niet eenvoudig misleid zullen worden? Toch niet, als Nederlandse rechters mensen zijn. De laatste decennia komen er steeds meer onderzoeksresultaten beschikbaar die laten zien dat mensen in hun denken bijzonder vatbaar zijn voor denkfouten (*biases*). Dit soort onderzoek is ook gedaan voor rechters en die blijken dezelfde fouten te maken als 'gewone' mensen. Dat geldt nog steeds als ze weten welke denkfouten vaak gemaakt worden. Een cursus logica of drogredenen op de universiteit of in de opleiding voor de rechterlijke macht zal het probleem daarom niet oplossen. Mensen zijn, anders dan ze mogelijk denken, niet de baas over welke conclusies ze trekken uit informatie die ze krijgen.

Het trekken van conclusies is in grote mate een onbewust proces en de factoren die daarin een rol spelen onttrekken zich in belangrijke mate aan de controle van degene die de conclusie trekt. Het is namelijk niet alleen zo dat mensen structureel geneigd zijn tot het maken van denkfouten; ze weten vaak ook niet wat hun beweegt. Als iemand probeert zijn eigen gedrag te verklaren, is de kans groot dat een foutieve verklaring wordt gegeven. De 'introspectie' waarmee mensen hun eigen beweegredenen menen te kunnen achterhalen is onbetrouwbaar. Dit betekent ook dat als rechters hun beslissingen motiveren, ze daarmee misschien aangeven waarom hun beslissing (in hun ogen) de juiste is, maar het betekent niet dat ze daarmee aangeven waarom ze hun beslissing hebben genomen. Dat zou niet erg zijn als de beslissing maar correct is onafhankelijk van hoe hij tot stand kwam. Maar als de correctheid van de beslissing afhangt van wat rechters plegen te beslissen en niet andersom, is het proces dat tot de beslissing leidt juist heel belangrijk.

Terug naar computerrechtspraak

Waarom is de al dan niet gebrekkige rationaliteit van rechters van belang voor computerrechtspraak? Omdat deze gebrekkige rationaliteit doorwerkt in de inhoud van het recht, aangezien die inhoud wordt bepaald door wat rechters plegen te beslissen. Men zou hier kunnen tegenwerpen dat de wetgever de inhoud van het recht in belangrijke mate bepaalt, niet de rechter. Maar deze tegenwerping ziet over het hoofd dat het rechters, of - meer in het algemeen - de sociale praktijk, zijn die de producten van de wetgever de status van recht toekennen. De sociale praktijk bepaalt wat recht is, en die praktijk wordt bewoond door mensen met beperkte rationaliteit, die bovendien maar beperkt door hebben wat hen drijft.

Voor de wenselijkheid van computerrechtspraak lijkt het van belang dat computers (bijna) hetzelfde kunnen als wat mensen kunnen. De mens, in deze context vaak in de persoon van de rechter, is de maat van goede rechtspraak. Computers mogen enkel ook meedoen als ze bijna net zo goed zijn als menselijke rechters. (En mogelijk zelfs dan nog niet, maar dat komt later.) Maar menselijke beslissers zijn minder rationeel dan ze zelf vaak denken. In sommige opzichten - wereldkennis, bijvoorbeeld - kunnen computers misschien nog niet wat mensen wel kunnen, maar in andere opzichten - vatbaarheid voor denkfouten - zijn ze waarschijnlijk nu al beter. Het bezwaar tegen computerrechtspraak is niet per se dat computers het slechter doen dan menselijke beslissers, maar dat ze het mogelijk anders doen.