

rechtstreeks 2024 nr 2



Raad voor de
rechtspraak

Recht en tech

**Renée Kool en
Kim van der Kraats**

Interview met Christiaan
Baardman over wat rechters
moeten weten over cyberzaken

**Litska Strikwerda en
Anke van Gorp**

De inschatting van recidive met
behulp van algoritmen

Mireille Hildebrandt

AI en de toekomst van de
juridische methode

Tina van der Linden

Het gebruik van risicoprofielen
in de rechtspraak

Bart van der Sloot

De impact van AI op de
rechtspraak: is het bewijsrecht
klaar voor de 21^{ste} eeuw?

Recht en tech

Colofon

Rechtstreeks is een periodiek voor de Rechtspraak en richt zich op de praktijk en de ontwikkeling van de rechtspraak in Nederland. Het tijdschrift, verspreid door de Raad voor de rechtspraak, stelt zich ten doel wetenschappelijke inzichten en bijdragen aan het publieke debat over de rechtspraak ter kennis te brengen van allen die beroepshalve bij de rechtspraak betrokken zijn. Opname in *Rechtstreeks* betekent niet dat de inhoud het standpunt van de Raad voor de rechtspraak weergeeft.

Redactie

Dr. R.S.B. (Renée) Kool
Universitair hoofddocent bij het Willem Pompe Instituut voor strafrechtswetenschappen en het Utrecht Centre for Accountability and Liability Law (UCALL) aan de Universiteit Utrecht

Drs. A. (Harry) Koster
Kennisarchitect bij de Raad voor de rechtspraak

Mr. K.G.F. (Kim) van der Kraats
Rechter rechtbank Midden-Nederland en bijzonder hoogleraar rechtspraak aan de Universiteit Utrecht

Prof. mr. dr. A.T. (Bert) Marseille
Hoogleraar bestuurskunde, in het bijzonder de empirische bestudering van het bestuursrecht aan de Rijksuniversiteit Groningen

Dr. S.B.M. (Sebastiaan) Princen
Universitair hoofddocent bij Bestuurs- en Organisatiewetenschap (USBO) van de Universiteit Utrecht

Redactieraad

Prof. mr. T. Barkhuysen
Prof. mr. Y. Buruma
Prof. mr. dr. I. van Domselaar
Prof. mr. J.H. Gerards
Prof. mr. N.J.H. Huls

Redactieadres

Redactie *Rechtstreeks*
Raad voor de rechtspraak
Afd. Strategie
Postbus 90613
2509 LP Den Haag
E-mail: rechtstreeks@rechtspraak.nl

Oplage

3.770 exemplaren

ISSN 1573-5322

Uitgever

Xerox/OSAGE

Abonnementen

Rechtstreeks wordt gratis toegezonden aan hen die tot de doelgroep behoren. Wie meent voor toezending in aanmerking te komen wordt verzocht zijn naam, postadres en functie kenbaar te maken aan het secretariaat van *Rechtstreeks* (rechtstreeks@rechtspraak.nl).

Adresmutaties

Xerox Communicatie Service
t.a.v. het bedrijfsbureau
Kalvermarkt 53
2511 CB Den Haag
E-mail: info@xnloverheid.nl

Retouren

Bij onjuiste adressering verzoeken wij u gebruik te maken van de adresdrager en daarop de reden van retournering aan te geven.

© Staat der Nederlanden (Raad voor de rechtspraak)

Niets uit deze uitgave mag worden veelevoudigd, in een voor anderen toegankelijk gegevensbestand worden opgeslagen of worden openbaar gemaakt zonder voorafgaande schriftelijke toestemming van de Raad voor de rechtspraak. De toestemming wordt hierbij verleend voor het veelevoudigen, in een gegevensbestand toegankelijk maken of openbaar maken waarvoor geen geldelijke of andere tegenprestatie wordt gevraagd en ontvangen en waarbij deze uitgave als bron wordt vermeld.

Inhoud

Redactioneel	5
Over de auteurs	8
column Lost in translation <i>Joep Lindeman</i>	10
AI en de rechtspraak: de rechter als tienkamper? <i>Renée Kool en Kim van der Kraats</i>	13
AI en de toekomst van de juridische methode <i>Mireille Hildebrandt</i>	21
De impact van AI op de rechtspraak: is het bewijsrecht klaar voor de 21^{ste} eeuw? <i>Bart van der Sloot</i>	33
Drie aandachtspunten met betrekking tot de inschatting van recidive met behulp van algoritmen <i>Litska Strikwerda en Anke van Gorp</i>	41
Drie inzichten met betrekking tot gebruik van risicoprofielen <i>Tina van der Linden</i>	49

Redactioneel

‘Wij zijn als juristen ontzettend geletterd. Maar er zijn genoeg rechters die zonder enige gêne zeggen ‘met cijfers heb ik liever niet te maken, daar ben ik niet zo handig in’, aldus **Christiaan Baardman**, voorzitter van de cyberkamer van het gerechtshof Den Haag en coördinator van het Kenniscentrum Cybercrime, in het interview waarmee dit nummer van *Rechtstreeks* opent. Baardman vervolgt: ‘Als je op de stoel van de rechter zit, dan heb je gewoon te doen wat de zaak en jouw functie op dat moment van jou verlangen. Dan is er geen ruimte voor excuses.’

De beide citaten uit het interview vormen de aanzet voor een antwoord op twee samenhangende vragen die in dit nummer van *Rechtstreeks* centraal staan: wat heeft de rechter nodig om met kennis van zaken te kunnen reflecteren en oordelen over geschillen waarin informatietechnologie en AI-toepassingen een rol spelen en, voor zo ver dit niet het geval is: hoe zou dit probleem kunnen worden verholpen?

Over die eerste vraag zegt Baardman dat rechters altijd in staat moeten zijn de feiten waar de zaak om draait op waarde te schatten, ook als ze de hulp van een deskundige inroepen. Want ook dan geldt dat je als rechter de informatie van de deskundige juridisch op zijn essentie moet kunnen begrijpen. Als sprake is van lacunes, kan de rechtspraak de toerusting van rechters op verschillende manieren faciliteren. Niet iedereen kan echter alles. Het ‘probleem’ moet je daarom terugbrengen naar behapbare onderdelen: de samenstelling van zittingscombinaties, het kanaliseren van zaken, cursussen, werving en

selectie van rechters. De rechtspraak zou meer op zoek moeten naar ‘tienkampers’, die niet alleen goede juristen zijn maar ook ‘digitale empathie’ hebben. Daarnaast zou zij zich bewuster moeten zijn van de voordelen van de jeugd en van de kennis en vaardigheden van gerechtsjuristen. Het interview adresseert ook de vraag hoe de rechtspraak als organisatie kan profiteren van ICT- en AI-ontwikkelingen. Baardman is daar optimistisch over. Er zijn legio mogelijkheden om processen handiger in te richten, systemen te ontwikkelen om zittingen sneller voor te bereiden en om passende beslissingskaders aan te reiken voor het beoordelen van zaken.

De bijdrage van **Mireille Hildebrandt** (hoogleraar aan de Vrije Universiteit Brussel en de Radboud Universiteit) sluit prachtig aan bij de stelling van Baardman dat zodra rechters te maken krijgen met zaken waarin ICT- en AI-toepassingen een rol spelen, zij over de deskundigheid moeten beschikken om die op waarde te kunnen schatten. Hildebrandt zegt hetzelfde over de rol van juristen bij de inzet van AI als onderdeel van de juridische methode. Het gaat daarbij om zulke uiteenlopende zaken als rechtsvinding, het opstellen van contract of het totstandkoming van wet- en regelgeving. Juristen moeten zich bewust zijn van de normatieve implicaties van de ontwerpbeslissingen die daar aan de orde zijn. Ze hoeven weliswaar zelf geen softwareontwikkelaar te worden, maar moeten zich wel – of ze dat nu leuk vinden of niet – verdiepen in de techniek van AI, zodat ze de juiste vragen kunnen stellen bij de besluitvorming over de inzet daarvan.

Bart van der Sloot (associate professor bij Tilburg University) schrijft over de impact van AI op het bewijsrecht. De komende jaren zal er steeds meer AI-gegenereerde en -gemanipuleerde content verschijnen. Een aantal te verwachten gevolgen: partijen zullen vaker – met opzet of onbedoeld – met bewijsstukken komen die zijn gemanipuleerd door AI en zullen vaker het argument ‘ik dacht dat het waar was’ of juist ‘ik dacht dat het niet waar was’ aanvoeren en rechters kunnen niet afgaan op hun eigen waarneming bij het beoordelen van de authenticiteit van digitaal materiaal maar zullen ook niet zonder meer kunnen afgaan op AI-detectieprogramma’s, omdat die werken met betrouwbaarheidspercentages. Rechters zullen door dat alles vaker suboptimale beslissingen nemen. Wat te doen? Gelukkig laat Van der Sloot die vraag niet onbeantwoord, maar geeft hij een aantal suggesties voor reacties op een toekomst waarin bijna alle digitale content AI-gegenereerd is.

Litska Strikwerda en **Anke van Gorp** (beiden hoofddocent en onderzoeker aan de Hogeschool Utrecht) schrijven over de dilemma’s bij het gebruik van OxRec, een digitaal instrument dat een inschatting maakt van recidiverisico’s en dat wordt gebruikt door medewerkers van de reclassering. De ratio van risicobeoordelingen op basis van algoritmische big-data-analyse is dat die voor een nauwkeuriger beoordeling zorgen. Het verschilt per systeem hoeveel nauwkeuriger ze zijn dan risicobeoordelingen door mensen, en 100% nauwkeurigheid is per definitie niet haalbaar. Dat verklaart (tezamen met zorgen over privacy en etnisch profileren) de wens om de menselijke factor bij risicobeoordelingen niet uit te schakelen. Er zijn ten minste drie manieren waarop dat kan gebeuren: professionals kunnen volledige vrijheid krijgen

bij het wege van het risico, het voorspelde risico kan worden gezien als eerste stap in het beoordelingsproces, of juist als laatste, door de professional eerst zelf een oordeel te laten geven en dat vervolgens te laten vergelijken met de uitkomst van de algoritmische risicoanalyse. Strikwerda en Van Gorp spreken geen expliciete voorkeur uit, maar gezien hun opmerking dat risicobeoordelingen ondoorzichtig, ondoordringend en onbetwistbaar kunnen worden als te veel wordt overgelaten aan een algoritmisch hulpmiddel, lijkt hun voorkeur uit te gaan naar een procedure waarin de algoritmische risicoanalyse fungeert als check van de beoordeling van de professional.

Ook **Tina van der Linden** (universitair hoofddocent aan de Vrije Universiteit Amsterdam) schrijft over het gebruik van risicoprofielen, maar richt zich in het bijzonder op het risico van discriminatie en dan met name indirecte discriminatie als gevolg van het gebruik van proxy’s, afgeleide kenmerken (zoals deeltijdwerk (voor geslacht) en postcode (voor etniciteit)) die tot indirecte discriminatie kunnen leiden. Het gegeven dat bij het gebruik van risicoprofielen nooit van tevoren met zekerheid is te zeggen welke proxy als discriminatiegrond zal fungeren, pleit ervoor helemaal af te zien van het gebruik van risicoprofielen, aldus Van der Linden. Voor zover dat niet gebeurt, is de vraag of de Europese AI-verordening voldoende bescherming biedt tegen discriminatie door het gebruik van die profielen. Van der Linden is sceptisch, wat het des te urgenter maakt dat de rechter indringend toetst op discriminatoire effecten van het gebruik van risicoprofielen.

De titel van de column van **Joep Lindeman** (hoogleraar Universiteit Utrecht) ten slotte, 'Lost in translation', ziet níét op de verhouding tussen de wereld van de informatietechnologie en die van de juristen, maar op de verhouding tussen de rechter en diens 'publiek' en in het bijzonder op het gevaar dat de hoogste rechter, als de instantie die geroepen is het recht uit te leggen, zijn doel voorbijschiet door het toepassen

van oversubtiële onderscheidingen die worden verpakt in omstandige en onnavolgbare formuleringen. Die boodschap brengt hij, aan de hand van voorbeelden uit Nederland en Europa, glashelder voor het voetlicht.

Namens de redactie,
Bert Marseille

Over de auteurs



Mireille Hildebrandt

Prof. dr. mr. M. Hildebrandt is hoogleraar aan de faculteit Rechten en criminologie van de Vrije Universiteit Brussel en aan de faculteit Natuurwetenschappen, wiskunde en informatica van de Radboud Universiteit in Nijmegen. Haar onderzoek richt zich op de gevolgen van datagestuurde omgevingen voor recht en rechtsstaat. Recent rondde zij een vijfjarig onderzoeksproject af naar de invloed van zogenaamde ‘juridische technologieën’ op recht en rechtsstaat, waarbij juristen en informatici intensief samenwerkten (zie cohucicol.com).



Christiaan Baardman

Mr. Chr.A. Baardman is senior raadsheer bij het gerechtshof Den Haag en coördinator van het Kenniscentrum Cybercrime, dat hij in april 2009 mede oprichtte. Ook is hij voorzitter van de Cyberkamer van het hof Den Haag. Eerder werkte hij als advocaat bij (thans) Nauta Dutilh, en hij begon z'n carrière in de rechterlijke macht bij de rechtbank Haarlem.



Anke van Gorp

Dr. ir. A. van Gorp werkt als onderzoeker, hogeschoolhoofddocent en privacy officer op het gebied van privacy en informatieveiligheid bij de Hogeschool Utrecht. Ze doet onderzoek naar gebruik van persoonsgegevens en data in publieke veiligheid, met name als professionals uit zorg en veiligheid samenwerken. Zij promoveerde aan de TU Delft in ethiek en techniek.



Joep Lindeman

Prof. mr. J.M.W. Lindeman is als hoogleraar strafprocesrecht verbonden aan het Willem Pompe Instituut van de Universiteit Utrecht. Hij is een van de programmaleiders van het Montaigne Centrum voor Rechtsstaat en Rechtspleging. Zijn onderzoek richt zich vaak op de praktijk van de (buitengerechtelijke) strafrechtspleging in binnen- en buitenland en op de rechterlijke organisatie. Hij is een van de leden van het team van samenwerkende universiteiten dat de Tweede Kamer heeft geadviseerd over het wetsvoorstel Modernisering Strafvordering.



Tina van der Linden

Dr. mr. M. van der Linden is universitair hoofddocent law, ethics & technology aan het Amsterdam Law and Technology Institute (ALTI) van de Vrije Universiteit Amsterdam. Zij is al heel lang gefascineerd door de uitdagingen waar het gebruik van informatietechnologie het recht voor stelt. Haar missie is om studenten en juristen zich bewust te maken van het feit dat gebruik van techniek niet neutraal is, en dat juristen hierin een grote verantwoordelijkheid hebben.



Litska Strikwerda

Dr. L. Strikwerda LL.M is als hoofddocent en senior onderzoeker digitale veiligheid verbonden aan de Hogeschool Utrecht. In zowel haar onderwijs als haar onderzoek richt zij zich op het snijvlak tussen technologie, (straf-) recht en ethiek. Zij promoveerde aan de Universiteit Twente op een proefschrift *Virtual Acts, Real Crimes? A legal-philosophical analysis of virtual cybercrime*, waarvoor zij de Erasmus Praemium Dissertatieprijs ontving.



Bart van der Sloot

Dr. mr. B. van der Sloot werkt als associate professor aan het Tilburg Institute for Law, Technology, and Society, Tilburg University en specialiseert zich op het gebied van technologie en recht, artificiële intelligentie en mensenrechten. Hij is de hoofdredacteur van het European Data Protection Law Review. Hij is afgestudeerd in filosofie en recht. Hij heeft drie prestigieuze prijzen ontvangen: de NWO Top Talent beurs, de NWO Veni Grant en de KNAW Early Career Award.

Lost in translation

Joep Lindeman

‘Doe het met meer intensiteit.’ In de film *Lost in Translation* van regisseur Sofia Coppola speelt Bill Murray een Amerikaanse acteur (Bob Harris) die tijdens een werkbezoek aan Japan terecht komt in een maalstroom van surrealistische situaties, niet zelden veroorzaakt door allerlei mogelijke spraakverwarring. Zo ook in de scène waarin hij een reclamespotje moet opnemen voor een luxe whiskymerk (‘Suntory’). De Japanse regisseur geeft omstandige instructies die, tot Harris’ verbazing, door de tolk zeer bondig worden vertaald: *‘do it with more intensity’*. Harris: *‘Is that everything? It seemed like he said quite a bit more than that.’* Bij gebrek aan ondertiteling weet ook de kijker niet of de bondige vertaling recht doet aan de opdrachten van de regisseur, wat de wanhoop van Harris al te meer invoelbaar maakt.

Ik moet regelmatig aan die scène denken als ik arresten van het Hof van Justitie van de EU (HvJ EU) lees. Ook in het arrest 30 april 2024, ECLI:EU:C:2024:371, par. 63 was het weer raak. In een lange zin van ruim tweehonderd woorden beantwoordt het hof een door een Italiaanse rechter gestelde prejudiciële vraag. Die zeer omstandig is geformuleerd, zoals we dat van het HvJ EU gewend zijn. Onnavolgbaar ook, want het HvJ EU oordeelt dat het nationale recht een nationale rechter kan *verplichten* om in bepaalde strafzaken toegang te verlenen tot verkeers- of locatiegegevens, zolang die rechter de toegang maar kan *weigeren*. Ja, dat staat er echt. Mijn collega – een echte ‘Luxemburg-watcher’ – schudde meewarig het hoofd over zoveel wereldvreemdheid. Als een rechter verplicht is om iets te doen,

kan je toch niet tegelijk veronderstellen dat die rechter datgene toch kan weigeren?

Waar ging de zaak over? In de Italiaanse praktijk wordt de strafrechter geacht een vordering tot toegang tot verkeers- of locatiegegevens toe te wijzen als aan de formele wettelijke eisen is voldaan: er moet sprake zijn van verdenking van een strafbaar feit waarop drie jaar of meer gevangenisstraf staat en van een aantoonbaar opsporingsbelang. De Italiaanse wetgever had met die ‘drie jaar of meer’-eis bedoeld invulling te geven aan een zogenoemde evenredigheidseis: alleen bij verdenking van ernstiger misdrijven is het toewijzen van zo’n vordering toelaatbaar. Dat was dus buiten het HvJ EU gerekend. Dat verwijst naar het in artikel 52 lid 2 van het Handvest voor de Grondrechten aangehaalde evenredigheidsbeginsel en stelt dat de vordering toch geweigerd moet kunnen worden als er sprake is van een strafbaar feit dat ‘gelet op de maatschappelijke omstandigheden in de betrokken lidstaat, kennelijk niet ernstig is’.

Dat een rechter in dit soort gevallen het evenredigheidsbeginsel de doorslag moet kunnen laten geven lijkt logisch. Zowel het Handvest als het EVRM schrijven immers voor dat een in de wet geregelde inmenging in de persoonlijke levenssfeer alleen gerechtvaardigd is als deze (kort gezegd) noodzakelijk is en een legitiem doel dient. Deze column is natuurlijk niet de plek om dit arrest te analyseren. Ik laat daarom de vraag hoe het kan dat Italiaanse rechters hier kennelijk zo weinig discretionaire

ruimte hebben dan ook rusten. Ik houd het bij de constatering dat het HvJ EU verbazingwekkend veel woorden nodig heeft om dit duidelijk te maken: ‘Doe het met meer evenredigheid.’

In Nederland staat sinds ‘Harderwijk’ het toetsen aan het evenredigheidsbeginsel door de bestuursrechter nadrukkelijk op de kaart. De Afdeling bestuursrechtspraak van de Raad van State zocht in de Harderwijk-uitspraak, na conclusie van AG’s Wattel en Widdershoven, aansluiting bij de Unierechtelijke evenredigheids-toets.¹ Sindsdien hebben Nederlandse rechters echter gedemonstreerd het soms moeilijk te vinden om die toetsingsruimte van de rechter op inzichtelijke manier onder woorden te brengen. In hun annotatie spreken Schlössels & Albers van een heus ‘evenredigheidsfeuilleton’.² De rechtspraak omtrent de evenredigheidstoetsing heeft zich volgens hen in rap tempo ontwikkeld tot ‘mandarijnenwetenschap’: er zal geen gewone burger zijn die deze rechtspraak kan volgen en zelfs voor juristen is het flink werken om er chocola van te maken (punt 3 van de annotatie). Ook Meijerman bezigt in een ongeveer gelijk-tijdig verschenen artikel toevallig de term mandarijnenwetenschap in verband met de bestuursrechtelijke evenredigheidsrechtspraak. Hij waarschuwt voor ‘dogmatische kathedralen’.³

Een waar gevaar voor ‘lost in translation’ dus. Hoewel ik het wat ironisch vind dat termen als ‘mandarijnenwetenschap’ en ‘dogmatische kathedralen’ nu ook weer niet uitblinken in

duidelijkheid. Zelfs het Genootschap Onze Taal blijkt de herkomst en betekenis van het begrip ‘mandarijnenwetenschap’ niet specifiek te kunnen duiden.⁴

Wat nu te denken van die jurisprudentiële woordenbrij? Dat het HvJ EU soms veel woorden nodig heeft om zijn oordeel over het voetlicht te brengen is wellicht nog enigszins te begrijpen: het hof moet een antwoord geven op een vraag van een nationale rechter en moet dat antwoord zó formuleren dat niet alleen die rechter, maar alle EU-lidstaten er in beginsel mee uit de voeten kunnen. De hiervoor aangehaalde uitspraak is gedaan door de Grote Kamer van het hof, waarin vijftien rechters zitting hebben. Zie dan nog maar tot een bondige uitspraak te komen... (misschien illustratief: Coppola won voor het door haar alleen geschreven script voor *Lost in Translation* een Oscar. Bij het script voor de film *The Flintstones* waren tien jaar eerder meer dan dertig screenwriters betrokken; zij wonnen de ‘Golden Raspberry’ voor het slechtste filmscript van dat jaar.)

Ik heb er minder begrip voor dat juridische fijnslijperij en omstandig taalgebruik zo hardnekkig aanwezig lijken te blijven in onze nationale rechtspraak. In 2017 is begonnen met het jaarlijks uitreiken van de ‘Klare Taalbokaal’, bij de meeste medewerkers van de Rechtspraak hopelijk een bekend fenomeen. Ook in *Rechtstreeks* is er al vaker aandacht geweest voor begrijpelijke rechtspraak. Zie bijvoorbeeld de column van

¹ RvS 2 februari 2022, ECLI:NL:RVS:2022:285.

² CBB 26 maart 2024, ECLI:NL:CBB:2024:190, *Jurisprudentie Bestuursrecht 2024/73*.

³ V.L. Meijerman, ‘Monumentaal menselijk bestuursrecht’, *Weekblad fiscaal recht 2024/131*.

⁴ Zo leid ik af uit de beantwoording van een vraag op het (toen nog) Twitter-account van Onze Taal: ‘Het lijkt vooral een juridisch jargonwoord te zijn. We kunnen niet direct iets over de herkomst vinden. We vermoeden verwantschap met de verstarde ambtenaar of formalist.’

Van der Bruggen in een eerdere uitgave van deze publicatie.⁵ Van der Bruggen schreef daarnaast met Pander Maat en Van Lent het boek *Spreekt het recht tot iedereen?*, waarna zij met Verburg in 2023 een van de preadviezen voor de Nederlandse Juristenvereniging schreef.⁶ *Klare Taal* was dat jaar het onderwerp van de NJV-jaarvergadering. Van der Bruggen en Verburg laten er geen misverstand over bestaan: het is de verantwoordelijkheid van de Rechtspraak als organisatie (en dus niet alleen van ‘individuele enthousiastelingen’) om te bevorderen dat er in begrijpelijke taal wordt rechtgesproken.

U begrijpt: ik was benieuwd wat daarover in het jaarverslag van de Rechtspraak over 2023 te lezen zou zijn. Dat werd nogal een deceptie. Ten eerste is er in het hele jaarverslag geen woord te vinden over het preadvies! Op de website van de Rechtspraak trouwens ook niet. Dat is nog eens stank voor dank. En na een obligaat stukje over het vergroten van de begrijpelijkheid van de rechtspraak neemt het jaarverslag een verrassende

afslag: rechters kunnen er eigenlijk lang niet altijd iets aan doen dat hun uitspraken niet begrijpelijk zijn. Het ligt aan de *wetgeving!* Die heeft vaak een ‘hoog technisch/specialistisch karakter’. Invoeringswetgeving van EU-wetgeving en jurisprudentie worden in dit verband onder andere als boosdoener genoemd (p. 11). Dan wordt het dus nooit wat, met die begrijpelijke rechtspraak, lijkt de Raad voor de rechtspraak te willen zeggen. Het is, gezien de door mij hiervoor aangehaalde voorbeelden, wel te begrijpen dat de Nederlandse rechter soms de wanhoop nabij is. Maar daarmee is niet gerechtvaardigd dat de Raad voor de rechtspraak aan zo’n mooi preadvies voorbijgaat om vervolgens de hete aardappel door te spelen aan de wetgever.

Hoe dan ook: het helpt niet, die ondoorgrondelijke wetgeving en omstandig gemotiveerde jurisprudentie van de EU. Hoe zou je ‘Klare Taal’ vertalen in de 23 andere officiële talen van de EU? Sofia Coppola (ze heeft Italiaanse voorouders) zou misschien zeggen: *‘Parlare chiaro!’*

⁵ G. van der Bruggen, ‘Het strafvonnis als onderonsje’, *Rechtstreeks* 2021/1.

⁶ Handelingen Nederlandse-Vereniging 2023, Deventer: Kluwer 2023.

AI en de rechtspraak: de rechter als tienkamer?

Interview met Christiaan Baardman

Renée Kool en Kim van der Kraats

Tekst: Sandra van Noord

Terwijl het voor veel rechters nog redelijk ver van hun bed is, is de cyberwereld voor raadsheer Christiaan Baardman geen grote onbekende. Naast senior raadsheer bij het gerechtshof in Den Haag is hij voorzitter van de cyberkamer bij het hof, die strafzaken over computercriminaliteit in hoger beroep behandelt. Ook is hij coördinator van het Kenniscentrum Cybercrime, waar rechters terecht kunnen voor informatie over dit onderwerp. Renée Kool en Kim van der Kraats van de redactie van *Rechtstreeks* spraken met hem, over wat rechters moeten weten over cyberzaken, cyberkamers en de onontkoombaarheid van digitalisering. Baardman sprak daarover vanuit een strafrechtelijk perspectief.

Het gesprek vindt plaats via Teams en de interviewers zijn bij aanvang een beetje zenuwachtig of het allemaal gaat lukken. Christiaan Baardman springt daar meteen op in: ‘Laat ik beginnen met een observatie. Iets wat heel typisch voor de Rechtspraak is en raakt aan het onderwerp van ons gesprek: zoals jij net begon, dat het al heel wat is dat de Teamsconnectie lukt. Dat past bij de Rechtspraak: wij beginnen vaak, aan een buitenstaander die ons niet kent, een signaal af te geven, ‘wij zitten eigenlijk op onbekend terrein. Ik voel me niet helemaal in mijn element. Ik ben al bezig om mij in een voor mij vreemde wereld te bewegen.’ Dat is precies waarover dit onderwerp gaat.

Mooi, dus we zijn al goed begonnen. Een eerste eenvoudige vraag: bent u wel eens geconfronteerd met onbekende ICT-toepassingen?

Dat is een vraag als: ben je vandaag uit bed gestapt? Want dat is zo vaak het geval, zeker in deze digitale wereld, waar alles toch vrij stormachtig verandert. Alleen je realiseert je het niet zo, je staat er niet zo bij stil. Voor iedere rechter geldt: ik heb nu met iets te maken wat ik niet eerder heb gehad, want het bestond eerder niet.

Het hangt er wel van af of we het over cyber of over digitaal bewijs hebben. Een probleem voor de rechtspraak, als het puur om cyber gaat, is dat niet zo veel rechters hiermee in aanraking komen, zeker niet op dagelijkse basis. Maar als je daar als rechter mee te maken krijgt, en je weet er weinig van af en je hebt er geen affiniteit mee, dan zit je er snel erg mee in je maag. Digitaal bewijs is een andere kwestie, dat komt in ontzettend veel zaken wel voor, dat is gewoon dagelijks leven geworden. Je ziet het in allerlei ICT-toepassingen, algoritmes en AI. Zoals verkeersdelicten, dat je nu via systemen kunt bepalen of de bestuurder wel of niet in een app heeft gekeken. Elke dag lees je weer in de krant

over toepassingen waarvan je denkt ‘heb ik dat nou eerder gelezen?’ Soms wel, maar heel vaak zijn het dingen waarvan je eerder niet wist dat dat ook al kan. Dat heeft ook consequenties voor de kwalificaties die aan rechters worden gesteld. Het is lastig te eisen dat ze hun werk goed blijven doen, hun vakbekwaamheid blijven etaleren in die zaken die ze voorgelegd krijgen, als er steeds andere toepassingen zijn.

Rechters moeten er de tijd voor hebben om hierover te leren. Zo konden we vroeger een deskundige van het NFI uitnodigen op zitting, die een ‘college’ gaf over DNA om met die materie om te kunnen gaan.

DNA is mooie vergelijking. Om het nog iets breder te maken: waar we het over hebben, is een ontmoeting van twee werelden, de wereld van de geletterden en de cijferaars met de wereld van de bits en bytes.

Wij zijn als jurist ontzettend geletterd. Maar er zijn genoeg rechters die zonder enige gêne zeggen ‘met cijfers heb ik liever niet te maken, daar ben ik niet zo handig in’. Dat kan natuurlijk niet! Het is tot daar aan toe er zo over te denken, maar je kunt het niet uitspreken. Want als we dat etaleren, dan weet ook de verdachte of de procespartij dat wat er verder gaat gebeuren eigenlijk hoogst onvoorspelbaar is. Want die rechter heeft daarmee duidelijk gemaakt dat hij hier moeite mee heeft, het misschien niet snapt of ook niet leuk vindt. En ondertussen is die rechter wel degene die gaat beslissen op het probleem van de verdachte.

Dat is dus een heel fundamenteel iets en het gekke is dat wij daar vaak helemaal niet zo sensitief op zijn, want we durven het gewoon te zeggen, zonder ons bewust te zijn dat dat totaal fout is. Als je op de stoel van de rechter zit, dan heb je gewoon te doen wat de zaak en jouw functie op dat moment van jou verlangen. Dan is er geen ruimte voor excuses. Ik vind dus dat rechters zich daar bewuster van moeten zijn, dat ze ook in die digitale zaak gewoon hun gezag tonen.

Dus die rechter neemt zijn verantwoordelijkheid en verdiept zich als hij zo’n cyber-crimezaak krijgt extra bij de voorbereiding. Wat mag hij daarbij vanuit de organisatie verwachten?

Dat heeft twee kanten. De rechter moet zijn rug niet alleen rechten in de voorliggende zaak naar de procespartijen toe, maar moet ook zijn rug rechten naar de organisatie. En dat wil zeggen dat hij moet zeggen: als ik mij niet geëquipeerd voel door tijdgebrek omdat ik een te scherp rooster heb, ofwel omdat ik eigenlijk meer opleiding nodig heb om zo’n type zaak tot een goed einde te brengen en echt recht te spreken, dan moet de organisatie zorgen dat ik daartoe in staat gesteld word. Omgekeerd, de organisatie, die meer tijd heeft om in algemene zin na te denken over de uitdagingen waar individuele rechters voor staan, moet nadenken hoe zij die rechters daarin kan faciliteren. Want het is ook een ontmoeting tussen een behoefte en een vraag.

Ik ken de Rechtspraak als een organisatie die wat opleidingen betreft toch betrekkelijk vrijblijvend is, er komt veel rechterlijke autonomie bij kijken. Het is een beetje een piepsysteem (behalve als rechters van afdeling wisselen, want dan is dat echt anders): rechters mogen, zeker als ze al langere tijd in een afdeling werkzaam zijn, zélf uitmaken welke cursussen zij leuk vinden om te volgen. En rechters

hebben de neiging veel dienstjaren te maken. Ik denk dat bij dit onderwerp een andere wind past: die vrijblijvendheid moet plaatsmaken voor veel meer regie. Gerechten moeten zeggen: ‘Met de uitdagingen waar de mensen uit onze afdeling voor staan, kan het niet zo zijn dat er niemand naar een cursus cybercrime gaat. Dat kán niet!’ Het maakt niet uit wie er dan gaat, maar er zal een aantal mensen een cursus moeten volgen. Er zullen veel rechters zijn die dan gaan sputeren: ‘Moet ik zo’n cursus doen? Ik kom die zaken nauwelijks tegen. Veel te moeilijk voor mij.’ Als je het over cybercrime hebt, dan begrijp ik dat wel, want dat gaat ook om geringe aantallen. Tegelijk, als je die zaak tegenkomt en je gaat toch op die stoel zitten – want je bent ingedeeld en het komt op jouw zitting te staan – dan heb je te doen wat die zaak verlangt, of die nou één keer in het jaar voorkomt, één keer in de vijf jaar of één keer in de week. Je moet het dan gewoon laten zien. En dat is een verantwoordelijkheid die best lastig is.

Een deel zal basale kennis zijn die iedereen die een beetje meedoet aan het maatschappelijk leven wel eigen is. Misschien moet de Rechtspraak daarnaast een soort basisscholing aanbieden. Maar wanneer wordt iets een deskundigheid waarover een rechter niet hoeft te beslissen? Hij kan bijvoorbeeld een IT-expert inschakelen, om te kijken of de e-mails echt zijn. Waar ligt die grens dan precies?

Die ligt volgens mij dieper dan wat jij zegt. Je moet als rechter namelijk wel snappen wat die deskundige zegt. En je moet beseffen dat je dan dus aan jezelf en aan procespartijen duidelijk maakt dat je extra kennis nodig hebt.

Als jij niet in staat bent om die informatie van de deskundige in ieder geval juridisch op zijn essentie te snappen en in jouw juridische systeem te plaatsen, dan heb je wel formeel het vinkje mogen zetten bij 1: merken waar jouw grenzen liggen, en 2: deskundigheid inroepen omdat het buiten die grenzen ligt. Maar dan weet ik niet of je het vinkje mag zetten bij iets waar het uiteindelijk om gaat en dat is dat je er iets goeds mee gedaan hebt. En ik zou wel eens willen weten – ik weet niet of daar veel onderzoek naar is gedaan – of deskundigen die wij inschakelen in zaken, die als ze de uitspraak zien, het vinkje zetten: ‘Die rechter snapte mij’.

Als we het hebben over de mogelijkheden van AI: dat is eigenlijk net het heelal, zo uitgebreid als de AI-omgeving is. Wat zijn de grenzen van waar een rechter verantwoordelijk voor wil en kan zijn?

Het is immens groot en wat wij dus ook moeten bedenken: we moeten ook goed omgaan met dat in- en uitzoomen. We kunnen ons heel erg moedeloos laten maken, want ‘het is allemaal zo ingewikkeld, we gaan het nooit snappen en vandaag weet ik het en morgen komt er misschien weer iets nieuws bij, ik word ouder en ik beweeg me niet op social media’. Maar je moet het ook weer kleiner maken en niet groter dan wat de zaak vereist. Het is een strategische uitdaging en die pak je niet op in het detail, maar op het heelalniveau. En dat moet je vervolgens terugbrengen naar datgene wat de hapklare onderdelen zijn: de samenstelling van zittingscombinaties, het centraliseren of kanaliseren van zaken, cursussen, alsmede de werving en selectie van rechters. Het is het hele palet waar we het naar vernauwen en verkleinen vanuit het perspectief dat het heel, heel groot is.

Zou zo iets als de Ondernemingskamer bij het Hof Amsterdam ook iets zijn voor digitale kennis? Een speciale kamer daarvoor, met rechters met meer kennis en ervaring over dit onderwerp?

Ik heb daar wel eens langer over nagedacht en dan stuit ik toch wel op een aantal problemen. Wat die Ondernemingskamer als voordeel heeft, is dat heel concreet is geregeld voor welke geschillen de Ondernemingskamer bevoegd is, dat is keurig vastgelegd in wetsartikelen. Het is niet zo makkelijk om voor 'digitale zaken', als ik dat even algemeen houd, te definiëren: wat maakt een zaak nou zo specifiek dat die het beste kan worden voorgelegd aan een cyberachtige kamer? Het vergt heel veel om precies te verwoorden om wat voor type zaak het moet gaan. Toch verdient die optie nadere overdenking, al was het omdat de Ondernemingskamer met de inbedding van leken zich heeft ontwikkeld tot een zeer gezaghebbend college.

Dat is een praktische vraag bij de uitvoering. Maar als het gaat om de benodigde kennis?

Ik zit in zo'n kamer (de cyberkamer van het Haagse hof, *red.*) en zie dus de voordelen ervan, om allerlei redenen. We hebben een betrekkelijk constante instroom van cyberzaken, waarmee we onze kennis kunnen blijven aanscherpen. De kamer bestaat uit mensen die het rechtsgebied leuk vinden, niet bang zijn voor een beetje pionieren en een nieuwsgierigheid hebben naar en affiniteit met digitale technieken. Bovendien maken mijn vaste collega Jan-Willem van den Hurk en ik deel uit van het Kenniscentrum Cybercrime en houden wij langs die weg de ontwikkelingen, juridisch maar ook technisch, nauwlettend bij. En dat geeft overwicht op de zaak omdat je niet buitenspel komt te staan als het echt inhoudelijk gaat worden op het technisch vlak. Een ander voordeel is dat de advocaten-generaal die hier ook in gespecialiseerd zijn, cyber-AG's zoals ze heten, graag aan onze kamer vragen voorleggen vanwege onze ervaring en kennis. Ons antwoord wordt sneller serieus genomen, als ik het zo mag zeggen.

Zou je dat dan structureler moeten maken? Ik denk dat je daar wel aan moet denken, maar dat vergt ook nogal wat uitwerking. Het verdient serieuze overdenking, omdat het een kolossale uitdaging is waar we als rechtspraak voor staan en daar bestaat geen pasklare oplossing voor. Ik kom met regelmaat fantastische rechterlijke uitspraken tegen van, als ik dat zo mag zeggen, 'gewone' zittingscombinaties. Of dat het gevolg is van toevallige factoren als complementaire kennis en ervaring van rechters plus de gerechtsjurist en goede input van de advocaat en/of officier van justitie, kan ik moeilijk inschatten. In die gevallen gaat het dus goed zonder dat er sprake is van allerlei vangnetten. Maar met grotere regelmaat komen wij als kenniscentrum rechterlijke uitspraken tegen die in de motivering en/of de bewezenverklaring ervan blijik geven dat het niet helemaal goed is gegaan of in elk geval niet goed genoeg. De vraag is: zijn dat nou de uitzonderingen of zijn de onvolkomenheden te talrijk en moet ons dat dus reden tot zorg geven?

Een andere dimensie betreft de werving en selectie. Als ik even een vergelijking naar sport mag maken en ik neem atletiek, dan heb je daar bijvoorbeeld mensen die heel hard kunnen lopen. Maar niet elke hardloper is een goeie discuswerper of een goede hoogspringer. Ondertussen maakt de tienkamp duidelijk dat er ook mensen zijn die én kunnen hardlopen én hoogspringen én

verspringen én kogelstoten én horden lopen. Ik denk dus dat het zinvol is dat wij dit voor wat betreft onze werving en selectie ook eens overdenken, en meer op zoek gaan naar ‘tienkampers’ in de rechtspraak. Dus niet alleen naar iemand die een proefschrift heeft geschreven over ouderschap en over weet ik veel wat voor knappe onderwerpen – wat ik ze echt niet zal nadoen! – maar ook naar mensen die nadrukkelijker ervan blijf geven die, wat ik maar noem, ‘digitale empathie’ hebben. Want we hebben ze hard nodig.

Het zit ‘m ook in cursussen. En misschien moeten we óók c.q. méér eisen gaan stellen aan procespartijen die hun standpunten baseren op de uitkomsten van uiteenlopende toepassingen van AI en algoritmes, in die zin dat zij rechterlijke toetsing van die uitkomsten mogelijk maken of op z’n minst vergemakkelijken, bijvoorbeeld op straffe van het passeren van die standpunten. Dat zit ‘m dus in regelgeving. Al met al gaat het om een veelheid van denkrichtingen en aanpassingen. We kunnen niet creatief genoeg zijn om echt iets te doen, want de ingezette digitalisering gaat nooit meer veranderen. Het is een onomkeerbare ontwikkeling en het wordt vast nog veel ingewikkelder dan het nu al is. En dat is niet het probleem van morgen, dat is het probleem van gisteren.

Ik vind trouwens dat we ons als Rechtspraak sowieso ook scherper bewust moeten zijn van de voordelen van de jeugd van de gerechtsjuristen. Feit is immers dat rechters betrekkelijk oud zijn, al was het maar in de zin dat zij op de universiteit hoegenaamd niets hebben geleerd over de digitale revolutie en de betekenis daarvan voor de juridische wereld. Rechterlijke oordelen in cyberzaken en zaken waarin het aankomt op de waardering van digitaal bewijs berusten dan ook niet op universitaire kennis, maar op bijscholing en door het werk verkregen inzichten. Dat is best een beangstigend besef. Dat is voor de gerechtsjuristen misschien al een beetje anders, omdat zij wel iets geleerd hebben tijdens de bachelor- of masterfase. Belangrijker is dat die gerechtsjuristen veel vertrouwder zijn met het gebruik van digitale mogelijkheden. Zij weten niet beter. Met die kennis moeten we ons voordeel doen. We moeten ons dus meer bewust zijn dat we nuttige kennis wel al in huis hebben, in de juridische toga, en dat we die beter moeten gebruiken dan dat we dat nu doen.

Stel dat we zo’n speciale kamer hebben en de rechter zit daar op zijn plek en hij weet wat hij zou willen/moeten weten of wat een partij als verweer zou kunnen voeren. Maar dat wordt niet gevoerd omdat de advocaat geen kennis van zaken heeft. Moet de rechter zo’n advocaat dan helpen?

Uiteindelijk wil geen enkele rechter een verdachte ten onrechte veroordelen. Dus als het zo is dat die verdachte een niet goed ingevoerde advocaat heeft, zal de rechter zich niet gemakkelijk voelen als hij toch tot een veroordeling komt, omdat ergens een verweer niet gevoerd is. De balie heeft zich keurig georganiseerd in allerlei specialistenverenigingen, maar cybercrime ontbreekt voorsnog. En dat vind ik heel opmerkelijk. Ik zou het toejuichen als de balie zich ook op dit terrein zou specialiseren. Maar dat raakt jouw probleem niet, want het gaat erom dat wat ook die advocaat zegt, gespecialiseerd of niet, het niet is gezegd. Moet ik als rechter dan helpen?

Dat raakt mij in ieder geval professioneel altijd wel. De verdachte kán het gedaan hebben maar als het niet uit het dossier blijkt, dan kan en mag je hem niet veroordelen.

Er wordt altijd gezegd: elke verkeersdode is er een teveel. Ik zou zeggen: in de rechtspraak is elke ten onrechte veroordeelde of ten onrechte in het ongelijk gestelde procespartij maar ook elke ten onrechte vrijgesproken verdachte er een teveel. Dat is de lat extreem hoog leggen. Maar als wij ons werk gewetensvol en goed willen doen, dan moet die lat heel hoog liggen en dan moeten we op alle vlakken alles uit de kast halen om dat te doen. In feite verschilt het antwoord op deze vraag niet van het klassieke dilemma hoe de rechter moet omgaan met de kennislacune van de verdediging in onverschillig welke andere strafzaak.

Maakt de achterstand van de Rechtspraak haar kwetsbaar ten opzichte van ‘de andere kant’, in ieder geval bij de criminele organisaties? Zij lopen ver voor, de rechtspraak kan hen nooit inhalen, bij wijze van spreken.

Wij zullen hen nooit inhalen, al was het maar omdat wij nu eenmaal altijd geschillen voorgelegd krijgen die in het verleden hebben plaatsgevonden. Neem bijvoorbeeld de PGP-communicatie (waarbij tekst wordt versleuteld en onleesbaar is voor anderen, *red.*) en de onwaarschijnlijk grote hoeveelheid informatie die dankzij ontsluiting van de vrijelijk uitgewisselde communicatie is blootgelegd. Dat onderwerp bestond tot een aantal jaar geleden niet, het heeft geleid tot een ontzaglijk aantal grote zaken (ondermijning, liquidaties) die hun bewijs vinden in die cryptocommunicatie en is daarmee uitgegroeid tot een betrekkelijk regulier onderdeel van strafzaken waarmee elke rechter te maken kan krijgen. Dat is een voorbeeld van een megaontwikkeling, voor de rechtspraak én in het criminele milieu. Misschien met deze kanttekening dat het criminele milieu, dat zich veilig waande achter cryptotelefoons en applicaties met versleutelde communicatie, met een kwetsbaarheid werd geconfronteerd door een onvermoed opsporingssucces dat die communicatie tóch leesbaar wist te maken en ineens achterliep door het bewijs dat dat tegen hen opleverde. Dat laat onverlet dat die criminelen natuurlijk allang op zoek zijn naar andere communicatietechnieken die nog beter versleuteld zijn én blijven. En het is weer een uitdaging voor politie en Openbaar Ministerie om ook die nieuwe vormen van versleutelde communicatie te ontrafelen. Dus zullen we weer een poos achter de feiten aanlopen. We lopen per definitie achter zaken aan, dat is nou eenmaal inherent aan het recht.

Mireille Hildebrandt noemt in haar bijdrage [verderop in dit nummer] ook door ChatGPT gedicteerde vonnissen.

Dan kom ik weer terug bij wat ik eerder zei. Elke toepassing, of het nu ChatGPT is of wat dan ook, ontslaat ons niet van onze primaire taak, en die is te bedenken of de voorliggende uitkomst er een is waar ik mijn handtekening onder wil zetten.

Je krijgt dan vragen als ‘hoe is dat product van ChatGPT tot stand gekomen?’, ‘Staat er precies in wat in mijn ogen relevant is, heb ik de goede zoektermen ingegeven?’ Het is misschien soms een beetje naïef van ons rechters dat we ons heel modern en hip voelen als we iets als ChatGPT inzetten.

Maar waar het eigenlijk op aankomt is dat je als rechter vervolgens zegt: dan ben ik niet klaar. Het introduceert namelijk weer een nieuwe vraag: ben ik toegerust om de toepassing waar ik nu gebruik van maak ook op zijn waarde te schatten? Je verlegt het probleem dus naar iets wat nog een grotere uitdaging voor jezelf is.

En dan kom ik bij iets uit wat ik wel treffend vind. Jongeren gebruiken ChatGPT inmiddels voor het maken van collegepapers, proefwerken, verslagen, enzovoorts. Maar uit een recent onderzoek blijkt dat ze het níét gebruiken voor sinterklaasgedichten, want ze kwamen er allemaal achter dat met de toepassing van die chatbot het gedicht voor mama of broer de essentie miste en daarmee waardeloos was. Dat de *ziel* mist waar het uiteindelijk met Sinterklaas om te doen is. En die ziel is in ons werk: recht doen. Als we dat dus niet doen, dan zijn die toepassingen hartstikke leuk en aardig en ze zullen natuurlijk veel beter worden dan ze nu al zijn. Maar wij moeten waarmaken dat wij, bij welke toepassing dan ook, uiteindelijk de hoeders van de ziel van ons werk zijn. En dat zal nooit een robot zijn en dat zal nooit het systeem zijn. Dat is onze al dan niet natte handtekening.

Heeft de digitalisering ook voordelen?

De Rechtspraak houdt zich met digitalisering in twee domeinen bezig: datgene wat wij in zaken tegenkomen en aan de andere kant digitalisering in onze systemen om ons werk efficiënter te doen. Zodat we sneller de zaken eruit vissen die efficiënter en sneller kunnen worden afgedaan. Onze processen handiger inrichten, systemen ontwikkelen die ons helpen om de zittingen sneller voor te bereiden, passende beslissingskaders aanreiken. Met andere formats werken waardoor we makkelijker verbanden kunnen leggen bovenop alle bestaande mogelijkheden om in hele dikke dossiers met al die ordners sneller de informatie te vinden waarvoor we vroeger het hele dossier door moesten worstelen. Het heeft dus ook evidente voordelen, laten we daar ook eerlijk over zijn en daar vol op inzetten.

Met behulp van AI wordt er ook gekeken naar uitspraken van rechters, hoe vaak ze de werkgever of de werknemer in het gelijk stellen. Je zou kunnen laten voorspellen wat een bepaalde rechter zeer waarschijnlijk gaat beslissen. Rechters kunnen er ook hun eigen beslissingen mee analyseren. De Raad voor de rechtspraak zegt hierover dat het niet de bedoeling is dat de buitenwereld op rechtersniveau gaat voorspellen welke kant het in een bepaalde zaak waarschijnlijk op zal gaan. Hoe kijkt u daar tegenaan?

Wij kunnen als rechtspraak wel zeggen: dat willen we niet en dat gaan we niet doen. Maar als een groot advocatenkantoor of binnenkort misschien dankzij betaalbare beschikbaarheid een individuele advocaat dat nou gewoon doet? We moeten niet de illusie hebben dat wij het allemaal zelf kunnen uitmaken. We kunnen niet altijd de wereld vormen zoals wij die het liefst zien. Het is een kwestie van tijd, en dan gaat dat veel goedkoper worden, minder ingewikkeld en makkelijker toegankelijk – of wij dat nu leuk vinden of niet. Dus laten we ook niet naïef zijn. Dit is juist weer die immensheid van dit onderwerp.

Wat misschien interessanter is, is of je een systeem kunt maken waar technische misslagen van rechters uitkomen. Ook dat is een uitdaging. Je kunt zo méér te weten komen dan alleen of rechters sterkere voorkeuren voor de ene of andere kant hebben.

Ik ben maar een operationele man die zittingen doet en weliswaar op kennisniveau ook meer in een soort ondersteunende rol iets doet, maar de organisatiekennis in dat opzicht ontbeer ik. Ik vind het nu heel normaal dat ik echt papierloos mijn werk kan doen, in kleinere maar ook grote zaken, dat had ik een paar jaar geleden niet voor mogelijk gehouden. Ik behoor tot de generatie die dit allemaal nooit op school of de universiteit heeft geleerd. Ook al kom ik misschien minder ver dan iemand die tien jaar of twintig jaar jonger is, dat weerhoudt mij er niet van om zover als mijn polsstok reikt mij efficiëntere werkwijzen eigen te maken. Als we dat allemaal met onze talenten – want die hebben we! – meenemen, dan ontwikkelen we ons allemaal individueel, en daarmee de Rechtspraak als geheel. Je krijgt dan de som van de enkele delen en dat is best een mooi iets. Zo'n suggestie van jou zou daar ook een voorbeeld van kunnen zijn, als de organisatie zou gaan kijken naar de voordelen.

AI en de toekomst van de juridische methode

Mireille Hildebrandt

1 Inleiding

Vier mannen die ervaring opdeden bij de datingapp Tinder (als ‘tech veteranen’) haalden vorig jaar 2 miljoen euro op voor hun start-up LegalFly, die onder meer ook de Nederlandse markt bedient.¹ Deze bijdrage richt zich op de inzet van AI als deel van de juridische methode, of het nu gaat om de rechtsvinding, het opstellen van contracten of het voorbereiden van wet- en regelgeving. Die inzet wordt mogelijk gemaakt door allerhande juridische technologieën, grotendeels aangeboden door de traditionele juridische uitgevers, maar inmiddels proberen dus ook tech start-ups de markt te betreden.² In paragraaf 2 bespreek ik kort de juridische methode zoals we die sinds jaar en dag kennen en beoefenen, gevolgd door een overzicht van verschillende typen van juridische technologie in paragraaf 3, waarna paragraaf 4 een methode beschrijft om de inzet van AI-technologie in het recht in kaart te brengen, de onderliggende methodieken te analyseren en vergelijken en ten slotte de impact op het recht te evalueren. De strekking van deze bijdrage is dat juristen zich zullen moeten verdiepen in de normatieve implicaties van ontwerpbeslissingen, niet om informaticus of software-ontwikkelaar te worden, maar om bij aanschaf en inzet van dit type technologieën de juiste vragen te kunnen stellen. Paragraaf 5 sluit af met concluderende opmerkingen.

2 De hermeneutiek van de rechtspraktijk

Er is een tijd geweest dat juridische wetenschappers zich hardop afvroegen of de rechtswetenschap wel een ‘wetenschappelijke’ methode heeft en in het verlengde daarvan hardop twijfelden of het recht wel een wetenschap is.³ Het recht zelf is uiteraard geen wetenschap maar een specifieke professionele en maatschappelijke praktijk die het object vormt van de rechtswetenschappen. Voor zover het gaat om de ‘traditionele’ rechtswetenschap zijn recht en rechtswetenschap voortdurend met elkaar in gesprek. Dat gesprek verloopt bijvoorbeeld via de juridische doctrine, die bijdraagt aan het systematiseren en begripen van zowel de wet als de jurisprudentie, die op hun beurt rekening

¹ “Onze product-first-mindset, verrijnd door onze ervaringen bij Tinder en Match Group, drijft ons om oplossingen te creëren die precies voldoen aan de behoeften van de gebruiker”, aldus Miessen, die rotsvast overtuigd is van zijn startup”, zie Visterin 2023 en verder legalfly.ai/nl/company.

² Zie par. 3.2.1 en voetnoot 10. De term AI verwijst hier naar een breed palet van computationele technieken, die verder gaan dan ‘digitalisering’. Over de term AI als PR-stunt uit 1956, zie par. 3.1.

³ Van Delft 2003. Zie voor verder onderzoek naar de juridische methode bijvoorbeeld het tijdschrift *Recht en Methode*, zie het redactioneel van Taekema & Klink 2011.

houden met ontwikkelingen in de doctrine. De juridische methode is dan ook 'gedeeplinkt' met de hermeneutiek, want het gaat steeds om het uitleggen van bindende juridische teksten in het licht van de feiten – en van de feiten in het licht van bindende juridische teksten. Die teksten leunen op het ongeschreven recht ofwel het geheel van ongeschreven vooronderstellingen die betekenis geven aan en tegelijk tot uitdrukking komen in het geschreven recht. 'Tekst' verwijst naar geschreven taal-handelingen die uit de aard der zaak multi-interpretabel zijn – ook al betekent dat niet dat teksten om het even hoe kunnen worden uitgelegd. Die multi-interpretabiliteit is ook eigen aan de gesproken taal, maar wordt versterkt wanneer de *spreker* een *auteur* wordt, die niet noodzakelijkerwijs aanwezig is wanneer de aangesproken persoon of groep de boodschap ontvangt. De auteur, in het recht bijvoorbeeld de wetgever, kan de lezers, in het recht de aangesproken rechtssubjecten, niet meer corrigeren als ze de tekst verkeerd verstaan. Daarmee ontwikkelt het recht zich tot een relatief autonoom fenomeen, waar uiteindelijk de rechter – en niet de wetgever – het laatste woord heeft. Ook die rechter kan het recht niet om het even hoe uitleggen; de rechter is gebonden aan het recht en gehouden het op integere wijze uit te leggen. Daarbij valt integriteit niet samen met ethiek of logische consistentie; de rechter mag de eigen morele voorkeuren niet zomaar opleggen aan anderen en logische consistentie is afhankelijk van de uitleg (die er tegelijk door wordt beperkt, dat wel). Belangrijk is dat de rechter institutioneel onafhankelijk kan spreken en door een hecht verankerd systeem van checks-and-balances wordt weerhouden van willekeurige uitspraken (denk enerzijds aan hoger beroep en cassatie en anderzijds aan de selectie en opleiding van rechters, waarin een cruciale rol is weggelegd voor de professionele ethiek).⁴

Multi-interpretabiliteit maakt contestatie mogelijk, roept die zelfs op. Daarmee is het recht een argumentatieve praktijk, waar niet alleen hoor en wederhoor spelen maar ook de vraag welke argumenten, door wie, wanneer en onder welke voorwaarden door de rechter weerlegd of gehonoreerd moeten worden. Juridisch redeneren is geen kwestie van het ontwerpen en toepassen van argumentatieschema's, maar een permanente anticipatie van tegenspraak. Recht is geen vooraf vast te leggen monoloog maar een complexe dialoog waar een veelheid van normen en feiten gerespecteerd moeten worden. Het is geen gesloten systeem van logische afleidingen, maar een genormeerde *praktijk* waarin ongelijkheidscompensatie voor de zwakkere partij, redelijke voorzienbaarheid, respect voor de democratische wetgever, en praktische en effectieve rechtsbescherming leidend zijn. Dit alles betekent dat de kloof tussen juridische normen en hun toepassing geen *bug* is maar een *feature*;⁵ het is die kloof die contestatie en daarmee de rechtsstaat mogelijk maakt. Waarbij de taak van de rechter uiteindelijk is om die kloof met gezag te sluiten, op een manier die – in het verlengde van voorgaande uitspraken – recht doet aan het individuele geval, en tegelijk richting kan geven aan volgende zaken. Voorwaar geen sinecure en zeker niet algoritmisch te voorzien. Dat wil echter niet zeggen dat AI geen toegevoegde waarde kan hebben. Dat zal afhangen van zowel het ontwerp als het gebruik van de betreffende AI-systemen.

⁴ Over de verhouding tussen recht, code en ethiek: Hildebrandt 2020a.

⁵ Dat wil zeggen 'geen nadeel maar een voordeel'.

3 De inzet van AI in de rechtsvinding en bij de wetgeving

3.1 Overzicht

De term AI of kunstmatige intelligentie is als marketingconcept geboren tijdens de Dartmouth Conferentie van 1956.⁶ Tijdens die conferentie werd de term voorgesteld als middel om financiering te vinden voor wat Herbert Simon 'complexe informatieverwerking' noemde. Zijn collega's lachten hem uit, omdat er voor zoiets saais nooit geld zou worden vrijgemaakt. En zo geschiedde, voortaan heette 'complexe informatieverwerking' kunstmatige intelligentie. Lange tijd verwees de term naar regelgestuurde computersystemen die op basis van vooraf ingegeven kennis (de input) conclusies en beslissingen (de output) produceerden, door die kennis op een welbepaalde manier vatbaar te maken voor automatisering. Die 'welbepaalde manier' verwijst naar formalisering, die vereist dat iedere vorm van ambiguïteit wordt uitgesloten, omdat computersystemen daar geen weg mee weten. Uiteindelijk zijn alle huidige computersystemen gegrond in de zogenaamde *logic gates*,⁷ die bepalen hoe kennis en informatie moet worden omgezet om verwerking mogelijk te maken. Kennis en inzichten die zich niet lenen voor zo'n omzetting passeren de poort niet en doen dus niet mee.

Juridische informatica levert sinds de jaren zeventig van de vorige eeuw een vloed aan wetenschappelijke papers en *executable code* waarmee *automated decision making* (ADM) mogelijk wordt gemaakt, via eenvoudige of meer complexe IFTTT (if this then that)-beslisbomen.⁸ De geclaimde voordelen van IFTTT-systemen zijn schaalvergroting, snelheid, voorspelbaarheid en narekenbaarheid. Dit type software is inmiddels wat uit beeld geraakt als AI, wat overigens niet betekent dat het geen rol meer speelt in actuele beslisprocessen. Alle aandacht gaat nu naar datagestuurde systemen, die beweerdelijk nieuwe kennis afleiden uit grote databestanden, onder de noemer van 'machinaal leren' (ML). Voor veel mensen staat AI nu gelijk aan ML-systemen, waaronder zoekmachines, recommender-systemen, voorspelsystemen en generatieve AI-systemen die vragen beantwoorden in vloeiend Engels, samenvattingen maken en allerhande schriftelijke stukken opstellen. Hoewel ML-systemen werken met waarschijnlijkheden in plaats van logica, blijft de onderliggende logica die van de *logic gates*. De waarschijnlijkheid wordt afgeleid uit de distributie van de inputdata en die data moet altijd eerst worden bewerkt tot wat zich leent voor bewerking in een *logic gate*. Daarin ligt dan ook de beperking van wat een computersysteem wel en niet kan. Hetzelfde geldt voor de wiskundige

⁶ Hildebrandt 2020b.

⁷ *Logic gates* verwijzen naar de fysieke grondslag van de huidige computersystemen, die materieel gebaseerd zijn op elektrische knoppen (bijv. transistors) die aan dan wel uit staan en aldus logische operaties mogelijk maken in overeenstemming met Booleaanse algebra (die getallen weergeeft door combinaties van twee cijfers: 0 en 1). Een 'operatie' is een logische bewerking van 'input', resulterend in 'output'. De beschikbare logische operaties zijn AND, OR, NOT, NAND, NOR, XOR en XNOR, die in waarheidstafels kunnen worden weergegeven. Meer of andere logische operaties zijn niet mogelijk, en iedere input moet zijn omgezet in wat zich leent voor dit type operaties (en dus zelf ook weer bestaan uit een combinatie van de cijfers 0 en 1).

⁸ Hildebrandt 2012.

modellen die de output van ML-systemen bepalen.⁹ Veel geavanceerde computersystemen zijn overigens een combinatie van regel- en datagestuurde systemen, waarbij de mate waarin en de wijze waarop beide type systemen worden geïntegreerd doorslaggevend zijn voor de relevantie van de output.

In de rechtsvinding kan AI verschillende rollen spelen: als zoekmachine, als voorspelmachine en/of als tekstontwerpmachine. Daarnaast kan AI een rol spelen bij het ontwerpen van wet- en regelgeving, door wetgeving die zich daarvoor leent niet alleen in natuurlijke taal maar tegelijk in een domeinspecifieke programmeertaal te schrijven.¹⁰ Dat is relevant voor de rechtsvinding voor zover het de interpretatieruimte van de rechter inperkt. Die inperking is onvermijdelijk vanwege de vertaalslag van een in natuurlijke taal geformuleerde juridische norm naar een regel die door een computersysteem kan worden uitgevoerd. Daarbij worden termen of regels die voor meerdere uitleg vatbaar zijn omgezet naar eenduidige termen of regels. Die omzetting impliceert een keuze, hetgeen de speelruimte inperkt die eigen is aan het recht. Het is immers die speelruimte die de uitleg en de toepassing van het recht aanvechtbaar houdt en rechtsontwikkeling mogelijk maakt.

3.2 De bijdrage van AI-systemen aan het ‘maken’ van recht

3.2.1 AI als juridische zoekmachine

De inhoud van mogelijk relevante juridische teksten is op geen enkele manier door een persoon te bevatten; er is geen tekort aan juridische informatie maar een teveel. Dat maakt de voorvraag welke informatie relevant is tot een steeds lastiger te beantwoorden vraag. Welke rechterlijke uitspraken, welke wet- en regelgeving, verdragen en rechtswetenschappelijke artikelen moeten worden geselecteerd en bestudeerd voordat een oordeel kan worden geveld inzake een voorliggende zaak? Welke documenten moeten worden onderzocht om een strategie te ontwerpen ten aanzien van een bepaald type zaken? Allerhande software staat al sinds jaar en dag ter beschikking om wegwijs te worden in het woud van mogelijk relevante teksten, op basis van verschillende technieken. Het aanbod wordt in toenemende mate bepaald door het verdienmodel van de aanbieders, vaak grote juridische uitgeverij,¹¹ maar kan ook door de rechterlijke macht zelf worden bepaald (zoals in geval

⁹ Voor een kritische bespreking van de implicaties voor regelgestuurde systemen zie Dreyfus 1992. voor datagestuurde systemen Smith 2019.

¹⁰ Merigoux, Chataing & Protzenko 2021. Waddington 2020. In Nederland wordt bij de Belastingdienst gewerkt met Regelspraak, zie Wolferink e.a. 2020.

¹¹ Interessante uitzondering in Nederland is bijvoorbeeld recht.nl, die onder meer een Juridische Tracker heeft ontwikkeld voor rechterlijke uitspraken, zie ‘Jurisprudentie Tracker nu beschikbaar voor alle abonnees Recht.nl’, recht.nl, 20 november 2022. Verder gaat het om *Legal Intelligence* van Kluwer, zie legalintelligence.com/nl en om *Rechtsorde* van Sdu Juridisch, zie rechtsorde.nl/wat-is-rechtsorde, hoewel Sdu zeer recent ook GenAI-L in de markt heeft gezet, zie sdu.nl/genial, in samenwerking met de Europese juridische uitgever Lefebvre Sarrut. Zie ook Visterin 2023.

van de zoekmachines van de beide Europese hoven) of door de wetgever (zie bijvoorbeeld het EU LEOS-project om Europese wetgeving ‘interoperabel’ te maken).¹²

3.2.2 AI als juridische voorspelmachine

De idee dat de toekomst zich laat voorspellen op basis van statistische patronen in historische dataverzamelingen voedt al jarenlang het ontwikkelen van *predictive AI* als voorspeller van de uitkomst van rechtszaken.¹³ De gedachte is dat hiermee een eerste indicatie kan worden gegeven van de meest waarschijnlijke uitkomst van een casus, maar uit nader onderzoek blijkt dat wat op technische gronden wordt aangeboden als een voorspelling vaak niet meer dan een identificatie van de uitkomst is of van indicatoren die correleren met de uitkomst.¹⁴ De inzet van dit soort voorspellingen bij het terugdringen van de achterstand van nog te behandelen zaken bij rechtbanken of hoven wordt daarmee een hachelijke zaak, omdat zelfs vanuit het perspectief van de informatica dit soort applicaties niet betrouwbaar zijn. Los daarvan is het natuurlijk zo dat het recht geen statisch gegeven is en juist beoogt in te spelen op zowel feitelijke veranderingen als voortschrijdend inzicht. Computationale voorspellingen zijn uiteindelijk altijd gebaseerd op de vooronderstelling dat de distributie van toekomstige data gelijk is aan die van de historische data waarop het systeem is getraind. Die vooronderstelling hoort in het recht niet op te gaan. Daarmee is zo’n voorspelling niet waardeloos. Het gaat er meer om die op waarde te schatten. Voor een methode om dit type AI te analyseren en evalueren verwijs ik naar paragraaf 4. Die analyse en evaluatie kunnen er natuurlijk wel toe leiden dat blijkt dat een concreet systeem, dan wel een bepaald type systeem, geen toegevoegde waarde heeft, dan wel een negatieve impact heeft op recht en rechtsstaat.

3.2.3 AI als uitspraakontwerpmachine

De doorbraak van generatieve AI gaat niet onopgemerkt voorbij aan het recht.¹⁵ De verleiding lijkt groot om tijd te besparen, competitief voordeel te behalen of nieuwe inzichten op te doen door een variant van ChatGPT in te zetten bij het voorbereiden van rechterlijke uitspraken. Hoewel de output gecontroleerd kan worden om ‘hallucinaties’ te verwijderen, eventueel correcte bronvermelding toe te voegen en de voorgestelde motivering in detail te onderzoeken, zou een dergelijke inzet een doorslaggevende impact hebben op de rechtsvinding. Daarover moet worden nagedacht op basis van een nuchtere analyse van hoe dit type systemen werken, op basis van welke vooronderstellingen en wat de inzet betekent voor de expertise van juristen (zie paragraaf 4). Het is vooral ook zaak om aandacht te besteden aan de omgang met generatieve AI binnen de juridische opleidingen, bijvoor-

¹² LEOS staat voor Legislation Editing Open Software, zie LEOS - Open Source software for editing legislation op joinup.ec.europa.eu. Het maakt gebruik van *markup language* om wetgeving beter doorzoekbaar en *linkable* te maken. Zie ook de mapping, analyse en evaluatie in de *Typology for Legal Technologies*: publications.cohubicol.com/typology/leos.

¹³ Bijvoorbeeld in de Verenigde Staten Katz e.a. 2017; in Europa Chalkidis e.a. 2019 en Chalkidis e.a. 2021.

¹⁴ Medvedeva, Wieling & Vols 2022; Medvedeva & McBride 2023.

¹⁵ Zie wat Thomson Reuters, de uitgever van een van de grootste juridische platforms in de Verenigde Staten Westlaw, daarover zegt, Shepherd en Lomax 2024.

beeld door studenten, advocaten en rechters in opleiding te leren werken met dit type instrumenten en daarbij steeds alle output te laten onderzoeken op basis van betrouwbare bronnen.¹⁶

3.2.4 AI als wet- en regelgevingsmachine

Ten slotte besteed ik hier aandacht aan de inzet van domeinspecifieke programmeertalen die dienst kunnen doen bij het omzetten van wet- en regelgeving in executable code, waarmee wet- en regelgeving die zich daarvoor leent op een consistente wijze zou kunnen worden uitgevoerd.¹⁷ Denk aan subdomeinen van de belastingwetgeving en de sociale zekerheid waar het gaat om berekeningen op basis van het al dan niet vervuld zijn van een aantal specifieke voorwaarden. Dit type technologie wordt gepropageerd onder de noemer van Rules as Code (RaC),¹⁸ een wereldwijde beweging om met name administratieve wet- en regelgeving meer voorspelbaar en hanteerbaar te maken, zowel voor de handhaving ervan (aan de zijde van de overheid) als voor geautomatiseerde compliance (aan de zijde van bedrijven). In Nederland gaat het bijvoorbeeld om Regelspraak, een softwaresysteem ontwikkeld binnen de Belastingdienst om de consistentie bij de afhandeling van aangiften te automatiseren en aldus schaalvergroting bij de afhandeling op een verantwoorde wijze vorm te geven.¹⁹ Ook dit type technologie roept vele vragen op, ook voor de rechtspraak. Een belangrijke vraag daarbij is of, en zo ja in hoeverre, rechtsgevolg kan worden toegekend aan de output van zo'n systeem en wat dat zou betekenen voor de rechterlijke toetsing (denk aan het vertrouwensbeginsel en het bestuursrechtelijke beginsel van de formele rechtskracht). Maar zelfs als er geen sprake is van het toekennen van rechtsgevolg aan de output van zo'n systeem, kan het voor de rechter lastig worden om te achterhalen of de software de bedoeling van de wetgever juist heeft vertaald naar een individueel geval en hoe de uitleg die met de output van het systeem is gegeven, zich verhoudt tot de juridische interpretatiemethodiek. De combinatie van grammaticale, wet-systematische, historische en teleologische interpretaties valt immers niet noodzakelijk samen met de historische bedoeling van de wetgever. Daar komt uiteraard nog bij dat de rechter wellicht zal moeten toetsen aan mensenrechtelijke verdragen en aan de invulling van het legaliteitsbeginsel.

4 Juridische methodologie in het tijdperk van AI

4.1 Een online webtool als geïntegreerde methode

Het is niet ongebruikelijk om de inzet van juridische technologie hetzij te omarmen, hetzij af te wijzen. In beide gevallen lijkt het niet nodig zich in technische details te verdiepen. Dat leidt ertoe dat anderen beslissen over het ontwerp van dergelijke technologie en in dat kader allerlei ontwerp-beslissingen nemen die consequenties kunnen hebben voor de aard van het recht. Omdat die beslissingen juristen aangaan, heeft een team van juristen en informatici een methodologie

¹⁶ Hildebrandt 2023.

¹⁷ Merigoux, Chataing & Protzenko 2021.

¹⁸ Waddington 2020.

¹⁹ Corsius e.a. 2021.

ontworpen om de betrouwbaarheid en de juridische houdbaarheid te evalueren van de output van relevante juridische technologieën. Die methodologie is verankerd in een online webtool: de *Typologie van juridische technologieën* (hierna: de *Typologie*), die praktische handvatten biedt aan zowel de gebruikers (praktijkjuristen) als de ontwikkelaars van dit type technologieën en daarnaast de broodnodige verdieping biedt door gelinkte studies en relevante achtergrond informatie.²⁰ Doel is om 1) het labyrint van juridische technologieën in kaart te brengen, 2) onderlinge vergelijking mogelijk te maken en 3) zowel het nut voor als de impact op het recht te evalueren. De meest relevante type systemen zijn in kaart gebracht (zie par. 4.2), zodat de gehanteerde methoden om deze systemen te analyseren en te evalueren (zie par. 4.3) ook relevant zijn voor de vele systemen die geen deel uitmaken van de *Typologie*. Sinds de publicatie van de *Typologie* zijn door de ontwerpers van de *Typologie* twee nieuwe systemen geanalyseerd en geëvalueerd, aan de hand van de voorgestane methode (de Juridische Tracker van Recht.nl en CoCounsel van Casetext).²¹

4.2 Mapping

Na gedetailleerd onderzoek naar een groot aantal juridische technologieën zijn dertig daarvan geselecteerd voor een diepgaand onderzoek. Het gaat om een aantal toegepaste systemen, enkele cruciale wetenschappelijke papers en enkele relevante datasets. Bij het invoegen in de typologie gaat het niet om het ontwerpen van een taxonomie, waarbij elkaar uitsluitende concepten worden gebruikt om een domein omvattend te beschrijven (alsof het om statische informatie zou gaan). Het gaat juist om overlap zichtbaar te maken en dus om een aantal analytische onderscheidingen te maken die bij kunnen dragen aan een beter begrip, bijvoorbeeld: code- en/of datagestuurde systemen, open source en/of door IP-rechten beschermde systemen, ontworpen door technische en/of juridische experts, gericht op advocaten, rechters en/of wetgevingsjuristen, nuttig voor strategie-bepaling en/of voor het afhandelen van concrete zaken. Daarbij wordt ook gekeken naar de herkomst-jurisdictie van de ontwerpers en de geclaimde relevantie voor andere jurisdicties en naar de mate waarin de beoogde gebruiker van het systeem de betrouwbaarheid kan achterhalen en/of onderzoeken.

4.3 Analyse en evaluatie

De technische en juridische analyse en de evaluatie van de geselecteerde systemen vormen een belangrijke aanvulling op de traditionele juridische methode. Daarbij staan in eerste instantie twee hoofdvragen voorop: 1) wat is de geclaimde functionaliteit? en 2) hoe wordt die claim gesubstantieerd?

De geclaimde functionaliteit wordt in de *Typologie* eerst kort samengevat, zodat de gebruiker van het webtool snel weet waar het om gaat. Vervolgens wordt die functionaliteit onderzocht aan de hand van drie subvragen: 1) wat zijn volgens de aanbieders de belangrijkste functionele kenmerken? 2) wat zijn volgens de aanbieders de onderliggende rationale en te behalen voordelen? en 3) welke

²⁰ Diver e.a. 2022..

²¹ Zie beide besprekingen: McBride & Medvedeva 2023; Van den Hoven, Meessen & Medvedeva 2022. Saillant detail is dat Casetext inmiddels is overgenomen door Thompson Reuters, de eigenaar van Westlaw.

ontwerpkeuzes hebben de aanbieders volgens eigen zeggen gemaakt? De aldus beschreven *geclaimde* functionaliteit is volledig gebaseerd op publiekelijk toegankelijke informatie, verstrekt door de aanbieder van een applicatie of dataset, dan wel door de auteur van de betreffende wetenschappelijke paper.²²

Voorbeelden van geclaimde functionaliteiten zijn: het sneller vinden van relevante jurisprudentie, het voorspellen van de uitkomst van een geding, het rangschikken van zaken in geval van achterstand bij rechterlijke instanties, het omzetten van regelgeving in geautomatiseerde besluitvorming, het in kaart brengen van toepasselijke regelgeving, het nakijken van complexe contracten, het analyseren van argumentatiestructuren, het geautomatiseerd adviseren inzake de aanpak van een specifieke casus, het uitlichten van foutief gebruik van rechtelijke uitspraken (bijvoorbeeld omdat er nieuwe ontwikkelingen in de rechtspraak zijn die niet zijn meegenomen), het vertalen van regelgeving in executable code die zonder menselijke tussenkomst beslissingen kan genereren en ten slotte het schrijven van conceptdocumenten, van pleidooi tot rechterlijke uitspraak met gebruik van zogenaamde generatieve AI. De *Typologie* biedt verschillende handvatten om die functionaliteiten snel in kaart te brengen en systemen met elkaar te vergelijken.

Vervolgens is, in nauwe samenwerking met de informatici, onderzocht of de aanbieder dan wel auteur de claims heeft onderbouwd op een manier die onderzoek naar de betrouwbaarheid mogelijk maakt. Bij commerciële applicaties is dat meestal niet het geval, omdat het verdienmodel rust op geheimhouding. In dat geval is onderzoek gedaan naar het waarschijnlijke ontwerp van het betreffende systeem, op grond waarvan de informatici uitleggen in hoeverre zo'n type ontwerp de geclaimde functionaliteit wel of niet kan waarmaken en waar eventuele technische kwesties kunnen spelen. Die kwesties kunnen bijvoorbeeld zien op de bewering dat de uitkomst van rechtszaken wordt voorspeld, terwijl dat op basis van de input van het systeem niet mogelijk is.²³ Het kan ook gaan om het feit dat systemen die zijn gebouwd met het oog op een specifieke jurisdictie of rechtsdomein, worden aangeboden voor andere jurisdicties of rechtsdomeinen.

Ten slotte is gekeken naar de impact van eventuele toepassing op het recht zelf. Daarbij is onderzocht in hoeverre rechtsgevolg wordt toegekend aan de output van een systeem, dan wel invloed op de toekenning van rechtsgevolg verwacht moet worden. Voor zover systemen niet betrouwbaar zijn is dat evident problematisch, maar ook wanneer zij in technische zin wel betrouwbaar zijn roept mogelijke impact op de totstandkoming van het recht velerlei vragen op. Die vragen hebben te maken met de beperkingen die inherent zijn aan computersystemen en met name aan de inperking van de interpretatieruimte van de rechter, de wetgever en het bestuur – alsmede het bemoeilijken of onmogelijk maken van contestatie door de justitiabele. Het belang van de *Typologie* schuilt in het feit dat het hier niet gaat om het debiteren van algemeenheden, zoals 'computers zijn beperkt in hun

²² De claims zijn gedocumenteerd met citaten en hyperlinks naar de relevante pagina's in het Internet Archive: [web.archive.org](https://archive.org).

²³ Zie noot 14.

vermogen de toekomst te voorzien', maar om het zorgvuldig in kaart brengen van de details die ertoe doen bij het ontwerp en gebruik van specifieke systemen.

4.4 Verdiepende methodologie

Zoals hiervoor aangegeven zijn de geselecteerde systemen ingebracht in een online webtool, waarmee juristen en ontwikkelaars snel en gemakkelijk een eerste indruk kunnen krijgen van een systeem, vergelijkbare systemen kunnen vinden, en vervolgens door kunnen klikken naar de technische en juridische analyse en evaluatie. De idee is dat de juridische methode zich bij onderzoek naar dit soort technologieën niet kan beperken tot traditionele tekstgestuurde publicatie van de opgedane inzichten, die immers niet alleen tekst betreffen maar ook programmeertalen, datasets, algoritmes en modellen. Die inzichten vragen van juristen de bereidheid om zich te verdiepen in de methode en conceptualisering van een andere discipline. Om overzicht te behouden is een geïntegreerde presentatie ontworpen waarin op verschillende manieren en niveaus kan worden gezocht, zodat de complexiteit van de materie zichtbaar wordt, zonder dat de 'gebruiker' de draad kwijtraakt.

Als methode is de *Typologie* is gelinkt aan twee vocabulaires: 1) een set van juridische grondbegrippen zonder welke het niet mogelijk is om het recht te verstaan, denk bijvoorbeeld aan het begrip 'rechtsgevolg' maar ook aan 'rechtssubject', 'positief recht' of 'jurisdictie'²⁴ en 2) twee sets van relevante termen uit de computerwetenschappen,²⁵ namelijk een set voor codegestuurde en een set voor datagestuurde termen. Deze vocabulaires bieden, samen met twee *Research Studies*,²⁶ de mogelijkheid om zich verder te verdiepen in de achtergronden en context van de *Typologie*, en nodigen vooral ook uit om de methodologie van beide disciplines te onderzoeken. De onderliggende methode is gericht op wederzijds begrip en beoogt expliciet een interdisciplinaire soep te vermijden.

De methode is inmiddels ook ingezet bij juridisch onderwijs en leent zich goed voor professionele training van rechters en advocaten.²⁷

5 Concluderende opmerkingen

De juridische methode is constitutief voor het recht zelf, want die methode bepaalt hoe rechterlijke uitspraken en wet- en regelgeving wel en niet geïnterpreteerd kunnen en/of moeten worden. Het gaat daarbij niet om logica of statistiek maar om taalhandelingen en respect voor de fundamentele multi-interpretabiliteit van het recht als menselijk artefact. Dat linkt de juridische methode aan de contestabiliteit die de kern vormt van de rechtsstaat. De integratie van verschillende type AI-systemen

²⁴ Zie publications.cohubicol.com/vocabularies/law.

²⁵ Zie publications.cohubicol.com/vocabularies/cs.

²⁶ Diver e.a. 2023 en McBride & Diver 2024.

²⁷ Hildebrandt 2023. Zie ook syllabus *Master course: Critical Reflections on Computational Law* op publications.cohubicol.com, tutorial *Method for Evaluating Legal Technologies (MELT)*, op publications.cohubicol.com en 'Hildebrandt presents at the CJEU "Innovation Days" Conference (23 May 2023)', cohubicol.com, 22 mei 2023.

in het recht zal zonder meer impact hebben op de juridische methode. Zowel het ontkennen van die impact als het categorisch omarmen of afwijzen daarvan leidt gemakkelijk tot onzichtbare integratie en daarmee tot ondoordachte reconfiguratie van de juridische methode. In deze bijdrage wordt een lans gebroken voor een nuchtere analyse en evaluatie van verschillende typen juridische AI-systemen, gegrond in respectvolle samenwerking tussen juristen en computerwetenschappers. Voor de inhoud van die analyse en evaluatie verwijs ik niet alleen naar paragraaf 4 hiervoor, maar vooral naar de webtool zelf. Toegegeven, dat betreft een type publicatie waar juristen niet aan gewend zijn, maar dat is dan ook het punt. Wie zich serieus wil bezinnen op de juridische implicaties van de inzet van AI in het recht zelf, zal buiten de eigen comfortzone moeten treden, zonder daarbij overigens de eigen waarde van recht en rechtsbescherming te verkwanselen.

Literatuur

Chalkidis e.a. 2021

I. Chalkidis e.a., *LexGLUE: A Benchmark Dataset for Legal Language Understanding in English* (SSRN Scholarly Paper), Rochester: Social Science Research Network 2021, papers.ssrn.com.

Chalkidis e.a. 2019

I. Chalkidis, I. Androutsopoulos & N. Aletras, 'Neural Legal Judgment Prediction in English', *arXiv:1906.02059 [cs]* 2019, arxiv.org.

Corsius e.a. 2021

M. Corsius e.a., 'RegelSprak: a CNL for Executable Tax Rules Specification', in: T. Kuhn e.a. (red.), *Proceedings of the Seventh International Workshop on Controlled Natural Language (CNL 2020/21)*, Amsterdam 2021.

Van Delft 2003

D. van Delft, 'Geléerd zijn jullie wel', *NRC Handelsblad* 12 april 2003.

Diver e.a. 2022

L. Diver e.a., *Typology of Legal Technologies*, COHUBICOL 2022, publications.cohubicol.com/typology.

Diver e.a. 2023

L. Diver e.a., *Research Study on Text Driven Law*, Brussels 2023, met dank aan de ERC Advanced Grant 'Counting as a Human Being in the Era of Computational Law' (COHUBICOL) van Hildebrandt, vanwege de European Research Council (ERC) onder het HORIZON2020 Excellence of Science program ERC-2017-ADG No 788734 (2019-2024), publications.cohubicol.com/research-studies/text-driven-law.

Dreyfus 1992

H.L. Dreyfus, *What computers still can't do: a critique of artificial reason*, Cambridge/London: MIT Press 1992.

Hildebrandt 2012

M. Hildebrandt, 'Oordeelsvorming door mens en machine: heuristieken, algoritmes en legitimatie', in: E.T. Feteris e.a. (red.), *Gewogen oordelen. Essays over argumentatie en recht*, Den Haag: Boom Juridische uitgevers 2012.

Hildebrandt 2020a

M. Hildebrandt, 'Chapter 11. Closure: on ethics, code and law', in: *Law for Computer Scientists*, Oxford University Press 2020.

Hildebrandt 2020b

M. Hildebrandt, 'Smart technologies', *Internet Policy Review* 2020/4.

Hildebrandt 2023

M. Hildebrandt, 'Grounding computational 'law' in legal education and professional legal training', in: B. Brozek, O. Kanevskaia & P. Palka (red.), *Research Handbook on Law and Technology*, Edward Elgar Publishing 2023, p. 99-127.

Van den Hoven, Meessen & Medvedeva 2022

E. Van den Hoven, P. Meessen & M. Medvedeva, 'Caselaw revisited: Recht.nl's case law tracker assessed with the Typology of Legal Technologies', onderzoeksblog cohubicol.com, 23 december 2022.

Katz e.a. 2017

D.M. Katz, M.J.B. Li & J. Blackman, 'A general approach for predicting the behavior of the Supreme Court of the United States', *PLOS ONE* 2017/4.

McBride & Diver 2024

Pauline McBride & Laurence Diver, *Research Study on Computational Law* (Brussels 2024), met dank aan de ERC Advanced Grant 'Counting as a Human Being in the Era of Computational Law' (COHUBICOL) van M. Hildebrandt, vanwege de European Research Council (ERC) onder het HORIZON2020 Excellence of Science program ERC-2017-ADG No 788734 (2019-2024), publications.cohubicol.com/research-studies/computational-law.

McBride & Medvedeva 2023

P. McBride & M. Medvedeva, 'Casetext's CoCounsel through the lens of the Typology', onderzoeksblog cohubicol.com, 4 juli 2023.

Medvedeva & McBride 2023

M. Medvedeva & P. McBride, 'Legal Judgment Prediction: If You Are Going to Do It, Do It Right', in: D. Preoțiu-Pietro e.a. (red.), *Proceedings of the Natural Legal Language Processing Workshop 2023*, Singapore: Association for Computational Linguistics 2023.

Medvedeva, Wieling & Vols 2022

M. Medvedeva, M. Wieling & M. Vols, 'Rethinking the field of automatic prediction of court decisions', *Artificial Intelligence and Law* 2022.

Merigoux, Chataing & Protzenko 2021

D. Merigoux, N. Chataing & J. Protzenko, 'Catala: a programming language for the law', in *Proceedings of the ACM on Programming Languages* 2021/5, afl. ICFP, nr. 77.

Shepherd & Lomax 2024

T. Shepherd & S. Lomax, 'Generative AI in the legal industry: The 3 waves set to change how the business works', blog posting thomsonreuters.com, 27 februari 2024.

Smith 2019

B.C. Smith, *The promise of artificial intelligence: reckoning and judgment*, Cambridge: The MIT Press 2019.

Taekema & Klink 2011

S. Taekema & B. van Klink, 'Een nieuwe impuls aan het methodendebat', *Law and Method* 2011/1.

Visterin 2023

W. Visterin, 'LegalFly haalt 2 miljoen binnen', blog posting Computable.be, 29 november 2023.

Waddington 2020

M. Waddington, 'Rules as Code', *Law in Context. A Socio-legal Journal* 2020/1.

Wolferink e.a. 2020

A.J.C. Wolferink e.a., 'RegelSpraak: een brug tussen wetgeving en ICT', *RegelMaat* 2020/1.

De impact van AI op de rechtspraak: is het bewijsrecht klaar voor de 21^{ste} eeuw?

Bart van der Sloot

Introductie

Dit artikel zoomt in op artificiële intelligentie (AI), en met name een variant daarvan, namelijk generatieve artificiële intelligentie. Kort gezegd gaat het dan om computerprogramma's die zeer realistische, maar door de computer gegenereerde, content produceren. Denk aan *large language models* als ChatGPT die een gedicht over Donald Trump kunnen schrijven in de stijl van William Shakespeare, Patricia Lockwood of je buurvrouw; denk aan *deepfake*-technologie waarmee je een video van Mark Rutte kan maken en hem iets kan laten doen of zeggen wat hij nooit heeft gedaan of gezegd; denk aan virtuele avatars die zo realistisch zijn dat ze fans en sponsorcontracten hebben.¹ Deze synthetische tekst-, video- en audioproducties zijn niet meer van echt te onderscheiden en omdat burgers zelf toegang hebben tot synthetische technologieën is de verwachting dat er de komende jaren steeds meer AI-gegenereerde en -gemanipuleerde content zal verschijnen. Dat kan de nodige implicaties hebben voor het bewijsrecht en de bewijspraktijk in rechtszaken. Daarom bespreekt dit artikel kort de uitgangspunten van het bewijsrecht in het civiel en strafrecht, geeft het daarna een schets van generatieve artificiële intelligentie en de synthetische content die daarmee kan worden gegenereerd, analyseert het de wijzen waarop deze technologische ontwikkelingen een impact zullen hebben op de rechtspraak en geeft het tot slot een aantal punten waarop het bewijsrecht aangepast zou kunnen worden.

Bewijsrecht

Binnen het civiel recht is bewijs in feite alle informatie die relevant is voor de rechterlijke procedure.² De rechter beoordeelt het geleverde bewijs op relevantie en authenticiteit. De hoofdregel is 'wie eist, bewijst' (art. 150 Wetboek van Burgerlijke Rechtsvordering (Rv)). De Hoge Raad heeft in 2020 nogmaals op een rijtje gezet welke mogelijkheden de rechter heeft in het algemeen ten aanzien van de stelplicht en de bewijslast tussen partijen:

¹ Van der Sloot 2024.

² Deze paragraaf is gebaseerd op Van der Sloot, Wagenveld & Koops 2021. Yvette Wagenveld heeft bijgedragen aan de tekst over het bewijsrecht.

1. Hoofregel is dat de stelplicht en de bewijslast ter zake van het *condicio sine qua non*-verband op de benadeelde rusten. Dit geldt ook bij besluitaanprakelijkheid. 2. De rechter kan in de bijzondere omstandigheden van het geval aanleiding vinden om te oordelen dat op de wederpartij van de benadeelde een verzwaarde motiveringsplicht rust (dat wil zeggen een verplichting om voldoende feitelijke gegevens te verstrekken ter motivering van haar betwisting, teneinde de benadeelde voldoende aanknopingspunten te bieden voor het nader onderbouwen en zo nodig bewijzen van de door hem gestelde feiten). 3. Als een partij haar betwisting van de stellingen van de andere partij onvoldoende motiveert, kan de rechter aan die betwisting voorbijgaan, zodat de gestelde feiten vaststaan. 4. Daarnaast kan de rechter, indien de wederpartij de stellingen van de benadeelde ter zake van het *condicio sine qua non*-verband voldoende heeft betwist, op grond van zijn waardering van de wederzijdse stellingen en het voorhanden bewijsmateriaal de betwiste stelling voorshands bewezen achten, behoudens tegenbewijs. 5. Ten slotte kan de rechter oordelen dat in de bijzondere omstandigheden van het geval uit de eisen van redelijkheid en billijkheid een omkering van de bewijslast voortvloeit.³

Het bewijsrecht is onlosmakelijk verbonden met de waarheids- en volledigheidsplicht (art. 21 Rv). Gevolg hiervan kan zijn dat partijen feiten moeten aandragen die niet in hun eigen belang zijn. De rechter heeft de mogelijkheid om partijen te bevelen een bepaalde stelling toe te lichten of bescheiden over te leggen (art. 22 Rv). Verder onderzoekt de rechter de zaak en beslist op basis van hetgeen partijen aan hun vordering, verzoek of verweer ten grondslag leggen (art. 24 Rv). De negatieve zijde daarvan houdt in dat de rechter feitelijke gronden van de vordering, het verzoek of het verweer niet mag aanvullen. De positieve zijde dat de rechter moet onderzoeken of de feiten die in de procedure zijn komen vast te staan het gevorderde ook kunnen dragen. Het uitgangspunt bij bewijs is dat het kan worden geleverd door alle middelen (art. 152 Rv). Partijen zullen actief de authenticiteit of integriteit ter discussie moeten stellen, anders zal de rechter doorgaans geen aanleiding zien om het bewijsmiddel ter onderzoek te stellen. Het bewijsrecht richt zich met name op de klassieke informatiebronnen, zoals het geschrift, en de verklaringen van getuigen en de deskundige, al zijn daar later de elektronische gegevens (digitale handtekeningen) aan toegevoegd.

Binnen het strafrecht geldt dat nadat de rechter de formele vragen van artikel 348 Wetboek van Strafvordering (Sv) langs is gelopen de eerste materiële hoofdvraag is: is bewezen dat het tenlastegelegde feit door de verdachte is begaan?⁴ Bewijzen kan worden gezien als het aantonen dat in redelijkheid niet kan worden getwijfeld aan de juistheid van het verwijt dat aan de verdachte wordt gemaakt. Het negatief-wettelijk bewijsstelsel houdt onder meer in dat er sprake is van bewijsminima, waarbij de rechterlijke overtuiging een belangrijke rol speelt. Op de verdachte rust geen bewijslast en zodoende ook geen bewijsrisico, deze rust op het Openbaar Ministerie. Het bewijsrisico houdt in dat

³ HR 25 september 2020, ECLI:NL:HR:2020:1510.

⁴ Corstens 2020.

een veroordeling uitblijft als de rechter van oordeel is dat wettig en overtuigend bewijs ontbreekt. De bewezenverklaring dient te zijn gefundeerd op basis van wettige bewijsmiddelen (zie art. 339 lid 1 en art. 340 tot en met 344a Sv).

Onder wettige bewijsmiddelen vallen de eigen waarneming van de rechter, verklaringen van de verdachte, een getuige of een deskundige en schriftelijke bescheiden. Via de eigen waarneming van de rechter kunnen 'nieuwere' bewijsbronnen worden geïntroduceerd, zoals bewijs in de vorm van video- en geluidsopnamen. De eigen waarneming van de rechter dient alleen ter sprake te worden gebracht 'indien de procespartijen door het gebruik van die waarneming voor het bewijs zouden worden verrast omdat zij daarmee geen rekening behoeften te houden. Of daarvan sprake is, is afhankelijk van de omstandigheden van het geval, zoals het procesverloop, de aard van de waarneming en het verband van die waarneming met het voorhanden bewijsmateriaal'.⁵ Uitzondering op het vereiste dat de waarneming ter terechtzitting moet zijn gedaan, zijn de foto's en videobanden die bij het dossier zijn gevoegd.

Artificiële intelligentie

Kunstmaticige intelligentie en dan met name generatieve artificiële intelligentie kan worden gebruikt om materiaal te vervaardigen. Je kunt ChatGPT bijvoorbeeld de e-mails van je ex laten lezen en vervolgens vragen om een e-mail te schrijven in de stijl van je ex, waarin belastende verklaringen worden gedaan. Door middel van *voice cloning* kun je op basis van een aantal audiofragmenten van een persoon hem vervolgens iedere zin laten zeggen die je wil; zo leek president Biden in de Amerikaanse voorverkiezingen kiezers op te roepen thuis te blijven.⁶ Dat geldt ook voor video's, waarvan vooralsnog de belangrijkste toepassing is het genereren van *deep porn*. Daarin lijkt het alsof bekende vrouwen, zoals Welmoed,⁷ maar ook in toenemende mate schoolgaande kinderen,⁸ naakt poseren of in een pornofilm spelen, wat grote impact kan hebben op hun sociale status en zelfbeeld. Ook kan AI nieuwe personen genereren. De foto's, video's en geluidsfragmenten van niet-bestaande personen zijn zo realistisch dat je zou zweren dat ze echt bestaan.⁹ Sterker nog, uit onderzoek blijkt dat mensen synthetische personen doorgaans betrouwbaarder vinden dan mensen van vlees en bloed.¹⁰

⁵ HR 15 december 2009, ECLI:NL:HR:2009:BJ2831, r.o. 3.5.3 en HR 25 september 2012, ECLI:NL:HR:2012:BX4990, r.o. 3.6.

⁶ Alex Seitz-Wald & Mike Memoli, 'Fake Joe Biden robocall tells New Hampshire Democrats not to vote Tuesday', nbcnews.com, 22 januari 2024.

⁷ Welmoed en de sexfakes, WNL, terug te zien op NPO Start.

⁸ 'Spaans dorp in shock door AI-gegenereerde naaktfoto's van meisjes', rtlnieuws.nl.

⁹ Zie bijvoorbeeld thispersondoesnotexist.com.

¹⁰ Nightingale & Farid 2022.

Deze deepfakes zijn voor het menselijk oog dan ook niet meer van echt te onderscheiden. Twee of drie jaar geleden waren vaak nog sporen van manipulatie te zien, zoals dat de ogen van een gemanipuleerd persoon niet knipperden, maar die beperkingen zijn inmiddels overwonnen. Experts die beelden controleren op authenticiteit zijn het lang niet altijd met elkaar eens, wat soms tot lange en complexe discussies leidt. Computerprogramma's die beelden en audio op authenticiteit controleren halen nu nog zo'n 60% van de *fakes* eruit, maar experts verwachten dat dat percentage de komende jaren zal dalen. Bovendien geven veel van dat soort detectieprogramma's slechts een betrouwbaarheidspercentage: de kans dat dit beeld is gemanipuleerd is 32% of de kans dat dit audiofragment niet is gemanipuleerd is 81%.

Deepfakes worden al ingebracht in de rechtszaal, soms met het expliciete doel de rechter te misleiden,¹¹ soms ook niet: denk aan een drillrapper die om stoer te doen een deepfake van zichzelf maakt waarin hij iemand lijkt te vermoorden. Hoe groot het deel van alle digitale content op dit moment al is dat is gegenereerd of gemanipuleerd door AI is moeilijk precies vast te stellen. Het gaat vermoedelijk reeds om een zeer groot percentage, zo groot dat veel AI-bedrijven hun modellen niet meer trainen op post-2022 datasets, omdat er dan een feedbackloop zou ontstaan.¹² Daarbij komt dat steeds meer materiaal per definitie is gemanipuleerd door AI. Bij digitaal (video)bellen worden bijvoorbeeld automatisch de hoge tonen eruit gefilterd en de huid geëgaliseerd. Fotocamera's werken tegenwoordig met AI, die 'weet' hoe objecten eruitzien. Als bossen in brand staan en je daarvan een foto maakt kan het bijvoorbeeld zijn dat de bossen er op de foto toch groen uitzien; als je op een mistige avond een foto van de maan maakt kan het zo zijn dat je de maan glashelder ziet.¹³ Al met al schatten experts in dat over vier jaar meer dan 90% van alle online content AI-gegenereerd of -gemanipuleerd is.¹⁴

Gevolgen voor de rechtspraak

Het is duidelijk dat deze ontwikkeling grote gevolgen zal hebben voor de rechtspraak en de aannames die aan het huidige bewijsrecht ten grondslag liggen onder druk zetten. Een aantal niet-limitatieve voorbeelden:

1. Procespartijen zullen in toenemende mate onbedoeld bewijsstukken aandragen die gemanipuleerd zijn door AI. Dat zal vaak gaan om kleine manipulaties, die evenwel relevant kunnen zijn. Een verdachte kan bijvoorbeeld een duidelijk herkenbaar litteken hebben, dat niet op de video-beelden is te zien: is dat omdat de verdachte niet op de beelden is te zien, maar een ander,

¹¹ Rachel Curry, 'AI deepfakes are poised to enter court proceedings at time of low trust in legal system', cnbc.com, 6 maart 2024.

¹² Een feedbackloop, in dit verband, betekent dat AI veel content genereert die op het internet is te vinden. Omdat AI-systemen met name leren van data die van het internet worden 'geschraapt', leren ze in toenemende mate van zichzelf in plaats van mensen, hun gedragingen en de feitelijke wereld. Deze feedbackloop zou tot een zogenoemde *system collapse* kunnen leiden: het systeem ontwikkelt zich niet meer maar bijt zich in zijn staart (Shumailov e.a. 2023).

¹³ James Vincent & Jon Porter, 'Samsung caught faking zoom photos of the Moon', theverge.com, 13 maart 2023.

¹⁴ Schick 2020.

of omdat de software de huid egaler heeft gemaakt? In civiele zaken kunnen beide procespartijen bijvoorbeeld fotobewijs aanleveren, waarbij AI beide foto's op een andere wijze heeft gemanipuleerd, omdat de twee camera's van een ander merk zijn en op andere software draaien.

2. Procespartijen zullen in toenemende mate bewijsstukken aandragen die met opzet zijn gemanipuleerd met als doel de rechter te misleiden. Denk daarbij ook aan criminele organisaties die niet zozeer bezig zijn met het manipuleren van bewijs met het oog op een specifieke rechtszaak, maar al jaren bezig zijn met het leggen van digitale broodkruimels naar dode sporen of naar rivaliserende bendes. Denk in civiele disputen aan beide ex-partners in voogdijzaken die een audiofragment als bewijs aandragen waarop de ander tegen de kinderen lijkt te schreeuwen en waarvan beide de authenticiteit van elkaars materiaal betwisten. De vraag voor de rechter is dan: zijn beide fragmenten waar, zijn beiden onwaar of is slechts een van de fragmenten waar, en zo ja, welke dan?
3. Rechters kunnen niet op hun eigen waarneming afgaan bij het beoordelen van de authenticiteit van digitaal materiaal. Dit kan leiden tot een nog grotere rol van deskundigengetuigen, waarbij elke partij weer zijn eigen deskundige zal invoeren en waarbij de rechter lang niet altijd voldoende kennis van de materie heeft om de getuigenverklaringen op betrouwbaarheid en relevantie te kunnen beoordelen.
4. AI-detectieprogramma's geven vaak betrouwbaarheidspercentages. Dat roept de vraag op wat rechters moeten met een video die 67% scoort op betrouwbaarheid. Is dat voldoende of niet, dienen er andere betrouwbaarheidspercentages te gelden binnen het bestuursrecht, het civiel recht en het strafrecht en zo ja, welke en noopt dit tot een nog grotere nadruk op ondersteunend bewijs, wat rechtszaken vaak nog complexer maakt?
5. Het bewijsrecht en de standaarden die daarin gelden zijn tamelijk algemeen. Onduidelijk is wat partijen die de authenticiteit van elkaars bewijsmiddelen betwisten precies moeten bewijzen en welke motiveringsdrempel zij moeten halen en hoe zich dat bijvoorbeeld vertaalt naar een AI-detectie-betrouwbaarheidspercentage. Nog lastiger wordt het als het gaat om door AI-gegenereerde personen. Stel, een man staat voor de rechter vanwege beelden waarop te zien is hoe hij op zijn jaarlijkse vakantie naar de Filipijnen met een jong meisje naar zijn slaapkamer gaat. Normaal gesproken zou de man nog zijn onschuld kunnen aantonen door het meisje te horen voor de rechter, om te verklaren dat er niets onoorbaars is gebeurd; als het gaat om een levensecht virtueel meisje is die mogelijkheid er echter niet.
6. Bovendien zal niet iedereen de authenticiteit van synthetisch materiaal betwisten. We kennen al het probleem van de valse bekentenis; denk ook de eerdergenoemde drillrapper die wellicht de schande van een ontkenning groter vindt dan de consequenties van een bekentenis. Ook kan de reputatie of nalatenschap van personen na hun dood worden bezoedeld door fakemateriaal te verspreiden; de persoon zelf zal zwijgen als het graf. Denk ook aan veroordelingen *in absentia*, waarbij de verdachte niet aanwezig is, en aan mensen die lijden aan een mentale stoornis. Een voogd zal niet altijd in staat zijn om namens een persoon te ontkennen, zeker niet als het tentoongespreide gedrag op een fakevideo niet *prima facie* onwaarschijnlijk is, wat ook geldt voor iemand die stomdronken was en daar geen herinnering meer aan heeft en die door een medeburger wordt beschuldigd van het vernietigen of beschadigen van eigendom.

7. Het 'ik dacht dat het waar was'-argument zal in toenemende mate worden gevoerd. Ik dacht dat mijn baas mij in een videocall toestemming gaf om het koffiezetapparaat mee naar huis te nemen; ik dacht dat de premier op het NOS-journaal, waarvan een clipje op Facebook circuleerde, echt zei dat de lockdown over was; ik dacht dat de video van mijn ex-partner waarin die vreemd ging waar was en daarom heb ik ...
8. Het 'ik dacht dat het niet waar was'-verweer zal in toenemende mate worden gevoerd. Denk aan illegale porno waarvan de verdachte beweert niet te hebben geweten dat het echt was; denk aan nieuws met een oproep van de burgermeester of politie om niet naar een bepaald gebied te gaan; denk aan beelden waarvan ouders op de hoogte zijn waarop het lijkt alsof hun puberzoon een klasgenoot molesteert.
9. Synthetische realiteit beïnvloedt daarbij vaak de echte realiteit. Binnen het strafrecht, bijvoorbeeld, is de verdachtmaking soms al voldoende voor een stigma. Denk aan beelden die online verschijnen, waarop het OM een onderzoek instelt, van een politicus die met een Russische spion lijkt te overleggen. Denk binnen het civiel recht aan een video van ouders die gewelddadig lijken te zijn naar hun kinderen, waarop ze uit huis worden geplaatst. Als het onderzoek naar de authenticiteit van het beeldmateriaal na twee jaar uitsluitel geeft dat het materiaal niet echt was is het niet altijd in het belang van de kinderen om weer terug bij hun ouders te worden geplaatst.
10. Tot slot zullen rechters in toenemende mate suboptimale beslissingen nemen. Omdat er geen goed systeem is om bewijs op authenticiteit te controleren geldt dat hoe meer synthetisch materiaal in omloop is hoe waarschijnlijker het wordt dat rechters iemand veroordelen of een partij in het gelijk stellen op basis van synthetisch bewijs. Daarnaast zal ook het omgekeerde gelden: dat rechters aarzelen over de authenticiteit van bewijs of een feit niet onomstotelijk vast kunnen stellen en daarom een verdachte vrijspreken of een claim ten onrechte niet toewijzen. De toename in synthetisch materiaal maakt ook dat mensen die terecht strafrechtelijk zijn veroordeeld of een civielrechtelijk dispuut hebben verloren het dwaalspoorargument zullen gebruiken: de rechter is in fakebewijs getrapt. Een argument dat hoe langer hoe meer niet zomaar terzijde kan worden geschoven.

Conclusie

Gegeven de ontwikkelingen als geschetst in deze bijdrage is het wellicht raadzaam om het bewijsrecht en de rechtspraak te overdenken. Het belangrijkste is daarbij de basisgedachte. Die is nu dat de rechter in principe aanneemt dat bewijs authentiek is, tenzij er contra-indicaties zijn. Zo'n contra-indicatie kan eruit bestaan dat er bij de rechter zelf twijfel is over de echtheid van een bewijsstuk of dat de wederpartij twijfels plaatst. Rechters kunnen echt niet meer van nep onderscheiden en procespartijen zullen niet altijd hun twijfels te berde brengen. Omdat de voorspelling is dat over een aantal jaren meer dan 90% van alle digitale content AI-gegenereerd of-gemanipuleerd is lijkt een fundamentele omkering van die aanname op zijn plaats. Bewijs is vermoedelijk synthetisch en de relevante vraag is: door wie is het gemanipuleerd, waarom, hoe en is de manipulatie relevant voor de zaak. Dit roept de vervolgvraag op wie deze vraag moet beantwoorden en hoe. Moet de rechterlijke macht zelf de capaciteit opbouwen om dit te doen? Moet er een onafhankelijk onderzoekscentrum

voor komen? Of moeten partijen zelf aantonen dat hun bewijs niet gemanipuleerd is? Het antwoord is: waarschijnlijk allemaal.

Om fouten te voorkomen kan het raadzaam zijn als de rechterlijke macht al het ingediende bewijs standaard te laten controleren op authenticiteit door een AI-detectieprogramma. Die haalt er met een stofkam een deel van het gemanipuleerde materiaal uit en plaatst bij andere rode vlaggen. Het zou goed zijn als domeinexperts die rode vlaggen kunnen opvolgen. Ten tweede zou er meer aandacht moeten zijn in de opleidingen van rechters, officieren en advocaten om zo meer bewustwording aangaande de synthetische realiteit te genereren. Daarnaast is de vraag of het NFI niet een grotere rol moet spelen bij het beoordelen van digitaal materiaal; het is zich nu al aan het oriënteren op deze positie.¹⁵ Een vierde optie zou zijn om in het bewijsrecht op te nemen dat slechts digitaal materiaal mag worden ingediend als dat een erkend watermark heeft.¹⁶ Een watermark is in feite een logboek: persoon A heeft foto X gemaakt en vervolgens de kleuren helderder gemaakt; persoon B heeft vervolgens de bladeren aan de boom groener gemaakt; persoon C heeft de personen in de foto ogenschijnlijk een aantal jaar jonger gemaakt; etc.¹⁷ Ten vijfde zou kunnen worden overwogen om de sancties bij (malicious) ingediend fakebewijs te verhogen; nu zijn die sancties doorgaans beperkt tot het terzijde schuiven van bewijs of het seponeren van een zaak. Zeker bij professionele partijen als advocaten en het Openbaar Ministerie zou een vergrote zorgplicht kunnen worden opgelegd om bewijsmateriaal dat zij indienen zelf op authenticiteit te testen.

¹⁵ 'Nieuwe methodes NFI en UvA voor herkennen deepfakes', nfi.nl, 16 november 2022.

¹⁶ Santiago Lyon, 'Leica Launches World's First Camera with Content Credentials Built-in – Delivering Authenticity at the Point of Capture', blog.adobe.com, 26 oktober 2023.

¹⁷ Zie verder: Jacobs 2024.

Literatuur

Corstens 2020

G.J.M. Corstens, *Het Nederlands strafprocesrecht*, Deventer: Kluwer 2020.

Jacobs 2024

B. Jacobs, 'The authenticity crisis', *Computer Law & Security Review* (53) 2024, 105962.

Nightingale & Farid 2022

S. J. Nightingale & H. Farid, 'AI-synthesized faces are indistinguishable from real faces and more trustworthy', *Proceedings of the National Academy of Sciences* 2022/8.

Schick 2020

N. Schick, *Deepfakes: The coming infocalypse*, Twelve 2020.

Shumailov e.a. 2023

I. Shumailov e.a., *The curse of recursion: training on generated data makes models forget*, arXiv preprint arxiv:2305.17493, 2023.

Van der Sloot 2024

B. van der Sloot, *Regulating the Synthetic Society: Generative AI, Legal Questions, and Societal Challenges*, Oxford: Hart 2024.

Van der Sloot, Wagenveld & Koops 2021

B. van der Sloot, Y. Wagenveld & B.J. Koops, *Deepfakes: De juridische uitdagingen van een synthetische samenleving*, Den Haag: WODC 2021.

Drie aandachtspunten met betrekking tot de inschatting van recidive met behulp van algoritmen

Litska Strikwerda en Anke van Gorp

Inleiding

Onze maatschappij ‘verdigitaliseert’. Er wordt steeds meer gebruikgemaakt van digitale middelen en digitale data om besluiten te nemen of te ondersteunen.¹ Binnen de rechtspraak is dat niet anders. Bijvoorbeeld: als de rechtbank of officier van justitie de reclassering opdracht geeft om een inschatting te maken van het recidiverisico van een verdachte wordt deze inschatting opgesteld met behulp van OxRec (voluit: *Oxford Risk of Recidivism Tool*)², een digitaal recidive-inschattinginstrument op basis van een algoritme.³ Hoe OxRec precies werkt, wordt in de volgende paragraaf uiteengezet.

De introductie van OxRec moet worden begrepen tegen de achtergrond van bredere maatschappelijke ontwikkelingen. Onze samenleving beweegt zich steeds meer in de richting van een ‘*pre-crime society*’, waarin criminaliteit wordt gezien als een risico waartegen preventief moet worden opgetreden.⁴ Tegelijkertijd is er sprake van een ‘actuariële wending’⁵, die omschreven kan worden als een tendens om risicobeoordeling steeds minder over te laten aan de menselijke beoordelingsvrijheid en steeds meer te laten berusten op algoritmische big-data-analyse. Het idee hierachter is dat risicobeoordelingen op basis van algoritmische big-data-analyse nauwkeuriger zijn dan risicobeoordeling door mensen. Er is empirisch bewijs dat suggereert dat dit inderdaad zo is, maar het is onduidelijk hoe significant het verschil is.⁶ En er is ook een keerzijde. In algemene zin zijn er drie aandachtspunten indien gebruikgemaakt wordt van algoritmen om risico’s in te schatten. Allereerst kunnen groepen burgers via algoritmen op een negatieve manier worden geprofileerd. Daarnaast zijn er zorgen over de privacy van de burgers op wie de gebruikte data betrekking hebben en de beoordelingsvrijheid van de

¹ Versendaal e.a. 2023.

² www.reclassering.nl > over-de-reclassering > wat-wij-doen > risc.

³ Oxrisk.com.

⁴ Zedner 2007, p. 262.

⁵ McKay 2019.

⁶ Van Dijk 2022.

professionals die deze instrumenten gebruiken (hebben zij nog wel de regie?).⁷ In dit artikel bespreken we deze aandachtspunten specifiek met betrekking tot OxRec.

Uiteindelijk komen we tot de conclusie dat de genoemde aandachtspunten samenhangen met belangrijke juridische en ethische waarden, namelijk gelijkheid, privacy en autonomie. Er moet een evenwicht gevonden worden tussen de bescherming van deze waarden en het belang dat wordt gehecht aan de nauwkeurigheid die algoritmische recidive-inschattingsinstrumenten bieden. De recent door het Europese parlement aangenomen Wet op de artificiële intelligentie zal in de toekomst richting kunnen geven aan deze belangenafweging.⁸

Hoe werkt OxRec?

OxRec werd ontwikkeld door de Forensic Psychiatry and Psychology Group van de Universiteit van Oxford. Dit algoritmische recidive-inschattingsinstrument is vrij beschikbaar op het internet⁹ en wordt sinds 2019 door Reclassering Nederland gebruikt.¹⁰ OxRec is gebaseerd op wetenschappelijk onderzoek dat in Zweden¹¹, Nederland¹² en Engeland¹³ is gevalideerd. In Nederland werd OxRec geïntroduceerd als onderdeel van RISC (Recidive Inschattingsschalen). Dit algoritmische recidive-inschattingsinstrument is bedoeld om reclasseringswerkers te ondersteunen bij het geven van een gestructureerd professioneel oordeel.¹⁴ Zij moeten in OxRec gegevens invoeren over: het geslacht, de leeftijd, de burgerlijke staat, de hoogst genoten opleiding, het werk, het besteedbaar inkomen en de migratieachtergrond¹⁵ van de delinquent, de duur van de gevangenisstraf voor het meest recent begane delict, of er sprake is van een gewelddadig indexdelict (van toepassing als het meest recente delict moord, aanranding, beroving, brandstichting, een seksueel delict, bedreiging of intimidatie was), of er sprake is van eerdere gewelddadige delicten (vóór het indexdelict), of de delinquent in een achterstandswijk woont (er moet een buurtscore worden ingevuld gebaseerd op het aantal bijstandsgerechtigden, de werkloosheid, het aantal migranten, het percentage echtscheidingen, het opleidingsniveau, de woonmobiliteit, de misdaadcijfers en het besteedbaar inkomen in de wijk)

⁷ Versendaal e.a. 2023. Over deze onderwerpen schreef de eerste auteur van dit artikel ook in een hoofdstuk van het nog te verschijnen *Research Handbook in Data Science and Law*.

⁸ Verordening van het Europees Parlement en de Raad tot vaststelling van geharmoniseerde regels betreffende artificiële intelligentie (Wet op de artificiële intelligentie) en tot wijziging van bepaalde wetgevingshandelingen van de Unie, COM/2021/206 final.

⁹ Oxrisk.com.

¹⁰ Van Dijk 2020a.

¹¹ Fazel e.a. 2016.

¹² Fazel e.a. 2019.

¹³ Beaudry e.a. 2023.

¹⁴ www.reclassering.nl > over-de-reclassering > wat-wij-doen > risc.

¹⁵ De Nederlandse versie van OxRec maakt geen gebruik van gegevens over migratieachtergrond. Zie: Brief van de minister van Justitie van 18 augustus 2020 (*Kamerstukken II 2019/20, Aanhangsel Handelingen*, nr. 3805).

en of er sprake is van alcoholmisbruik, drugsmisbruik of psychische stoornissen.¹⁶ OxRec analyseert deze gegevens door middel van regressie.¹⁷ Regressie is de meest voorkomende vorm van voorspellende analyse. Regressiealgoritmen schatten of voorspellen de waarde van een variabele voor de entiteit waarover de gegevens worden verzameld.¹⁸ In het geval van OxRec gaat het om een voorspelling van de kans op recidive binnen een of twee jaar. Deze kans wordt weergegeven aan de hand van risiconiveaus (laag, gemiddeld of hoog).¹⁹ Of OxRec een vorm van artificiële intelligentie (AI) is, hangt af van de definitie die men daarvan hanteert. De term AI wordt in de praktijk vaak gebruikt voor zelflerende computerprogramma's.²⁰ Maar de recent aangenomen Wet op de artificiële intelligentie verstaat onder AI ook statistische benaderingen,²¹ zoals de door OxRec gebruikte regressie.

Het eerste aandachtspunt: profileren en discriminatie

In de inleiding kwam aan de orde dat groepen burgers via algoritmen op een negatieve manier kunnen worden geprofileerd. Een specifieke vorm van negatief profileren via algoritmen is etnisch profileren. Hierover gaat de bijdrage van Tina van der Linden elders in deze uitgave van *Rechtstreeks*. Specifiek met betrekking tot algoritmische recidive-inschattingsinstrumenten bestaan grote zorgen over etnisch profileren en daarover is dan ook veel literatuur verschenen.²² Daarom is dit het eerste aandachtspunt dat we met betrekking tot OxRec bespreken. Etnisch profileren kan worden omschreven als het gebruik van criteria die duiden op een bepaalde afkomst of herkomst, zoals huidskleur of ras, om mensen te selecteren, bijvoorbeeld voor een controle door de politie of de Koninklijke Marechaussee. Etniciteit mag alleen gebruikt worden als selectiecriterium indien daar bijzonder zwaarwegende redenen voor zijn. Zijn die er niet, dan is er sprake van discriminatie.²³ Algoritmische recidive-inschattingsinstrumenten kunnen op twee manieren etnisch profileren.²⁴ Direct, indien er afkomst- of herkomst-risico-indicatoren zijn of criteria die daarop duiden, en indirect, indien zij gebruikmaken van risico-indicatoren waaruit de afkomst of herkomst kan worden afgeleid.²⁵ Hoe dat werkt, leggen we hier uit.

¹⁶ Oxrisk.com.

¹⁷ Fazel e.a. 2016.

¹⁸ Provost & Fawcett 2013.

¹⁹ Fazel e.a. 2016.

²⁰ [inzicht-in-ai.nl > wat-zijn-ai-en-algoritmes](https://inzicht-in-ai.nl/wat-zijn-ai-en-algoritmes).

²¹ Artikel 3 (1) en Annex I.

²² Zie bijvoorbeeld Sachoulidou 2023; Braverman e.a. 2016; Van Dijk 2020a; Van Dijk 2022; Kulk & Van Deursen 2020.

²³ Gerechtshof Den Haag, 14 februari 2023, ECLI:NL:GHDHA:2023:173.

²⁴ Professionals kunnen natuurlijk ook last hebben van biases en discrimineren en dan kan het gebruik van een systematische risico-analysetool wellicht biases of discriminatie helpen voorkomen.

²⁵ Van Dijk 2020b.

In het geval van OxRec is geen sprake van direct etnisch profileren. In OxRec hoeven geen gegevens te worden ingevoerd over afkomst, herkomst of daarop duidende criteria zoals etniciteit, huidskleur of ras.²⁶ In andere landen moeten wel gegevens worden ingevoerd over de migratieachtergrond van de delinquent. Voor de Nederlandse versie van OxRec mogen, zoals hiervoor vermeld, geen gegevens over migratieachtergrond worden gebruikt.²⁷ Maar er moet wel worden ingevoerd of de delinquent in een achterstandswijk woont (de zogenaamde buurtscore). Wat een achterstandswijk is, wordt onder andere bepaald aan de hand van het aantal migranten dat er woont, zo werd in de vorige paragraaf uitgelegd. Indirect kan deze risico-indicator daarom wel leiden tot etnisch profileren, zo stelt Gijs van Dijk in een inmiddels veelbesproken artikel in het *Nederlands Juristenblad*.²⁸ In antwoord op Kamervragen over dit artikel, gaf de minister van Justitie en Veiligheid aan dat de buurtscore in OxRec 'niet op basis van ras, etniciteit, religie of afkomst' selecteert, 'maar op basis van de sociaaleconomische status' van de delinquent en dat dit legitiem is, omdat uit wetenschappelijk onderzoek blijkt dat deze van invloed is op het recidiverisico.²⁹ De impact die de buurtscore heeft op de uiteindelijke risicoscore is echter minimaal: gemiddeld 0 tot 2%.³⁰ Naast de buurtscore kunnen ook andere risico-indicatoren in OxRec indirect leiden tot etnisch profileren. Inkomen, werk en relatiestatus kunnen bijvoorbeeld onbedoeld iets anders meten dan alleen recidivegevaar. 'Indien blijkt dat personen met een migratieachtergrond slechtere kansen hebben op de arbeidsmarkt, vaker relaties verbreken en/of vaker worden opgepakt dan personen zonder migratieachtergrond', is een algoritmisch recidive-inschattinginstrument dat die gegevens gebruikt als risico-indicatoren 'niet veel meer dan een afspiegeling en reproductie van in de maatschappij ingebakken ongelijkheden'.³¹ Het gevaar hiervan is dat het risico op recidive dan ten aanzien van bepaalde etnische groepen hoger wordt ingeschat. Ook dat is discriminatie.³²

Het tweede aandachtspunt: privacy

Naast de hiervoor besproken zorgen over etnisch profileren en discriminatie zijn er zorgen over de privacy van de burgers waarop de data betrekking hebben die algoritmische recidive-inschatting-instrumenten gebruiken. Zo worden er bijzondere persoonsgegevens verwerkt, bij OxRec zijn dat in ieder geval data over alcoholmisbruik, drugsmisbruik en eventuele psychische stoornissen.³³ Deze bijzondere persoonsgegevens worden gebruikt om een voorspelling te doen over het toekomstig recidiverisico. Hierin schuilt een gevaar. Middelengebruik en psychische stoornissen zijn dynamische factoren, ze veranderen in tijd. Als je kijkt naar een individu dan kan een beter afgesteld medicijngebruik

²⁶ Oxrisk.com; Maas, Legters & Fazel 2020, p. 2058.

²⁷ Oxrisk.com; Maas, Legters & Fazel 2020, p. 2058.

²⁸ Van Dijk 2020a.

²⁹ Brief van de minister van Justitie van 18 augustus 2020 (*Kamerstukken II 2019/20, Aanhangsel Handelingen, nr. 3805*).

³⁰ Maas, Legters & Fazel 2020, p. 2058.

³¹ Van Dijk 2020b.

³² Van Dijk 2020b.

³³ Oxrisk.com.

of het beter volgen van medicijnvoorschriften een grote invloed hebben op de psychische stoornis. Bij het invullen van OxRec worden deze gegevens echter al snel statische persoonsgegevens: iemand heeft een gediagnosticeerde psychische stoornis of niet. Daarnaast kan een gebrek aan transparantie ontstaan. Bij een regressieanalyse zoals nu gehanteerd wordt in OxRec, kan nog uitgelegd worden welke factoren een rol spelen en wellicht ook de weging van die factoren. Als reclasseringsmedewerkers OxRec invoeren dan zien ze deze weging overigens niet.³⁴ Als OxRec en andere recidive-voorspellende algoritmes in de toekomst zelflerend zouden worden, dan wordt het steeds lastiger voor reclasseringswerkers en zelfs voor de ontwikkelaars van een systeem om uitleg te geven waarom iemand ingedeeld wordt in een bepaalde risicocategorie. Dit kan op gespannen voet komen te staan met het recht op informatie dat degene op wie de persoonsgegevens betrekking heeft op grond van artikel 18 van de Wet justitiële en strafvorderlijke gegevens heeft. Transparantie en uitlegbaarheid zijn ook vereisten in de recent aangenomen Wet op artificiële intelligentie.³⁵

Het derde aandachtspunt: professionele beoordelingsvrijheid

Tot slot kunnen algoritmische recidive-inschattingsinstrumenten de autonomie van de professionals die ermee moeten werken in het gedrang brengen, omdat ze hun discretionaire bevoegdheid kunnen inperken. In het geval van OxRec gaat het dan om de discretionaire bevoegdheid van de reclasseringswerkers die door dit instrument worden ondersteund bij het geven van een professioneel oordeel, maar ook om de discretionaire bevoegdheid van de rechters die dit oordeel vervolgens meenemen in de beoordeling van de zaak die hen voorligt. Hun beoordelingsvrijheid is niet onbeperkt, maar wordt gekaderd door wetgeving, jurisprudentie en richtlijnen. Het gevaar van algoritmische recidive-inschattingsinstrumenten schuilt erin dat ze het oordeel van professionals binnen de vrije ruimte die ze wel hebben, en die ze ook nodig hebben om algemene regels goed te kunnen afstemmen op het individuele geval dat hen voorligt, steeds meer gaan sturen.³⁶ Indien algoritmische recidive-inschattingsinstrumenten het recidiverisico met 100% zekerheid zouden kunnen inschatten, zou dat geen probleem zijn, maar zoals eerder aan de orde is gekomen, is dat niet het geval. Algemeen uitgangspunt, dat ook ten grondslag ligt aan de recent aangenomen Wet op de artificiële intelligentie, is dan ook dat de mens in het spel (*human-in-the-loop*) gehouden moet worden. Met betrekking tot algoritmische recidive-inschattingsinstrumenten zijn er grofweg drie manieren waarop dit bewerkstelligd kan worden. Ten eerste kan volledige beoordelingsvrijheid en flexibiliteit worden gegeven aan professionals door hen te laten beslissen of en hoe het voorspelde recidiverisico wordt beoordeeld en gewogen. Ten tweede kan het voorspelde recidiverisico gezien worden als een eerste stap die dient als een basislijn die de professional kan volgen of waarvan deze kan afwijken. Tot slot kan recidiverisico-beoordeling worden gezien als een iteratief proces waarin een professional eerst het recidiverisico beoordeelt en dit vervolgens vergelijkt met de voorspelling van een recidivebeoordelingsinstrument. De professional kan dan de voorspelling van het beoordelingsinstrument gebruiken om na te denken

³⁴ Oxrisk.com.

³⁵ Art. 13 Wet op artificiële intelligentie.

³⁶ McKay 2019.

over de oorspronkelijke beoordeling en eventueel de beoordeling aan te passen.³⁷ Het nadeel van elk van deze manieren om de beoordelingsvrijheid van professionals die met algoritmische recidive-inschattingsinstrumenten werken te waarborgen, is dat ze een zekere mate van willekeur in de hand werken.³⁸ Echter, indien risicobeoordelingen te veel worden overgelaten aan een algoritmisch hulpmiddel, kunnen deze ondoorzichtig, ondoorgroendelijk en onbetwistbaar worden.³⁹ In het bijzonder is het dan moeilijk om potentiële vertekeningen te identificeren en aan te pakken⁴⁰, bijvoorbeeld omdat de gebruikte data niet beschikbaar zijn of onduidelijk is hoe het beoordelingsinstrument precies werkt (het blackboxprobleem). Dit is onwenselijk.

Tot besluit

OxRec is een algoritmisch recidive-inschattingsinstrument dat bedoeld is om reclasseringswerkers te ondersteunen bij het geven van een professioneel oordeel over het recidiverisico van delinquenten. In onze huidige samenleving heerst het door empirisch bewijs ondersteunde idee dat risicobeoordelingen op basis van algoritmische big-data-analyse nauwkeuriger zijn dan risicobeoordeling door mensen. Hoe significant dit verschil is, is echter onduidelijk. Vaststaat dat ook risicobeoordelingen op basis van algoritmische big-data-analyse het recidiverisico niet met 100% zekerheid kunnen inschatten. In dit artikel hebben we aandacht gevraagd voor drie risico's met betrekking tot de inschatting van recidive met behulp van algoritmen. Allereerst kan dit leiden tot (indirect) etnisch profileren en discriminatie. Daarnaast zijn er zorgen over de privacy van de burgers op wie de gebruikte data betrekking hebben en de beoordelingsruimte van de professionals die deze instrumenten gebruiken. Deze aandachtspunten hangen alle drie samen met belangrijke juridische en ethische waarden: gelijkheid, privacy en autonomie. Er moet een evenwicht gevonden worden tussen de bescherming van deze waarden en het belang dat wordt gehecht aan de nauwkeurigheid die algoritmische recidive-inschattingsinstrumenten bieden. In de nabije toekomst zal de Europese Wet op de artificiële intelligentie hier richting aan kunnen geven. Algoritmische recidive-inschattingsinstrumenten worden door deze wet gecategoriseerd als AI-toepassingen met een hoog risico.⁴¹ Dit betekent dat ze zullen moeten voldoen aan strenge regels. In het licht van de waarden gelijkheid en privacy is van belang dat de Wet op de artificiële intelligentie vereist dat AI-systemen met een hoog risico voldoen aan kwaliteitscriteria met betrekking tot non-discriminatie en de omgang met persoonsgegevens.⁴² Ook de waarde autonomie wordt door de wet gewaarborgd omdat deze vereist dat er altijd sprake is van menselijk toezicht op AI-systemen met een hoog risico.⁴³

³⁷ Van Dijk 2022.

³⁸ Van Dijk 2022.

³⁹ Hildebrandt 2018.

⁴⁰ Van Dijk 2022.

⁴¹ Art. 6 (2) en Annex III Wet op de artificiële intelligentie.

⁴² Art. 10 Wet op de artificiële intelligentie.

⁴³ Art. 14 Wet op de artificiële intelligentie.

Literatuur

Beadry e.a. 2023

G. Beaudry e.a., 'Prediction of violent reoffending in people released from prison in England: External validation study of a risk assessment tool (OxRec)', *Journal of Criminal Justice* 2023, 102061.

Braverman e.a. 2016

D.W. Braverman e.a., 'OxRec model for assessing risk of recidivism: ethics', *The Lancet Psychiatry* 2016/9, p. 808.

Van Dijck 2020a

G. van Dijck, 'Algoritmische risicotaxatie van recidive. Over de Oxford Risk of Recidivism tool (OXREC), ongelijke behandeling en discriminatie in strafzaken', *Nederlands Juristenblad* 2020/25, p. 1784-1790.

Van Dijck 2020b

G. van Dijck, 'Naschrift', *Nederlands Juristenblad* 2020/28, p. 2059-2060.

Van Dijck 2022

G. van Dijck, 'Predicting Recidivism Risk Meets AI Act', *European Journal on Criminal Policy and Research* 2022/3, p. 407.

Fazel e.a. 2016

S. Fazel e.a., 'Prediction of violent reoffending on release from prison: derivation and external validation of a scalable tool', *Lancet Psychiatry* 2016/6, p. 535.

Fazel e.a. 2019

S. Fazel e.a., 'Prediction of violent reoffending in prisoners and individuals on probation: a Dutch validation study (OxRec)', *Scientific Reports* 2019/1.

Hildebrandt 2018

M. Hildebrandt, 'Law as computation in the era of artificial legal intelligence: Speaking law to the power of statistics', *University of Toronto Law Journal* 2018/supplement 1, p. 12-35, 34.

Kulk & Van Deursen 2020

S. Kulk & S. van Deursen, *Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek*, Utrecht: Mouton Centrum voor Rechtsstaat en Rechtspleging/Universiteit Utrecht 2020.

McKay 2019

C. McKay, 'Predicting risk in criminal procedure: actuarial tools, algorithms, AI and judicial decision-making', *Current Issues in Criminal Justice* 2019/1.

Maas, Legters & Fazel 2020

M. Maas, E. Legters & S. Fazel, 'Professional en risicotaxatie-instrument hand in hand. Hoe de reclassering risico's inschat', *Nederlands Juristenblad* 2020/28.

Provost & Fawcett 2013

F. Provost & T. Fawcett, *Data Science for Business: What you need to know about data mining and data-analytic thinking*, O'Reilly Media 2013.

Sachoulidou 2023

A. Sachoulidou, 'Going beyond the 'common suspects': to be presumed innocent in the era of algorithms, big data and artificial intelligence', *Artificial Intelligence and Law* 2023

Versendaal e.a. 2023

J. Versendaal e.a., *Structuurnota voor continuering van het Lectoraat Betekenisvol Digitaal Innoveren*, Hogeschool Utrecht, 2023.

Zedner 2007

L. Zedner, 'Pre-crime and post-criminology', *Theoretical Criminology*, 2007/2, p. 261-281.

Drie inzichten met betrekking tot gebruik van risicoprofielen

Tina van der Linden

Inleiding

De stelling die centraal staat in deze bijdrage is dat mensen in het algemeen, juristen in het bijzonder en rechters bij uitstek, zeer kritisch moeten zijn met betrekking tot de uitkomsten van artificiële intelligentie (AI). Dat geldt voor generatieve AI (zoals *large language models* waarvan ChatGPT de bekendste is), en dat geldt ook voor wat minder spectaculaire 'AI' waar we al meer ervaring mee hebben: het gebruik van risicoprofielen. Om zo kritisch te kunnen zijn moet je snappen hoe het werkt én hoe gemakkelijk het mis kan gaan. Je ziet het pas als je het door hebt. Deze bijdrage gaat alleen over risicoprofielen. De bijdrage van Litska Strikwerda en Anke van Gorp elders in deze uitgave gaat uitgebreider in op een specifiek systeem voor het inschatten van het recidiverisico van een bepaalde verdachte, en bespreekt in detail de inherente problemen met een dergelijk systeem. Gelukkig zijn we het eens en sluiten onze bijdragen ook inhoudelijk goed op elkaar aan – en niet alleen de titels.

Gebruik van risicoprofielen

Laat ik eerst duidelijk maken wat ik met de verschillende begrippen bedoel. Een risicoprofiel is een 'verzameling van een of meer selectiecriteria op basis waarvan een bepaald risico op normovertreding wordt ingeschat en een selectiebeslissing wordt gemaakt'.¹ Voorbeelden zijn de risicoprofielen die gebruikt zijn in de kinderopvangtoeslagaffaire,² in het SyRI-systeem,³ en door DUO om 'verdachte' uitwonende studenten te controleren.⁴ Ras is een juridische verzamelterm voor huidskleur, afkomst en nationale of etnische afstamming.⁵ Directe discriminatie is direct onderscheid maken op een van de bekende discriminatiegronden, zoals verboden door onder meer artikel 1 van onze Grondwet, artikel 14 EVRM en artikel 1 van Protocol 12 bij het EVRM. Van indirecte discriminatie is sprake als er niet direct op basis van een van deze gronden onderscheid gemaakt wordt, maar het gemaakte onderscheid wél een door zo'n grond beschermde groep in disproportionele mate treft.⁶ Proxy's zijn

¹ College voor de Rechten van de Mens 2021, p. 8.

² CBS 2024.

³ Rechtbank Den Haag 5 februari 2020, ECLI:NL:RBDHA:2020:865.

⁴ DUO 2024.

⁵ College voor de Rechten van de Mens 2021, p. 7.

⁶ HvJ 13 mei 1985, Bilka Kaufhaus GmbH t. Karin Weber von Hartz, C170/84.

afgeleide kenmerken, waarmee indirect gediscrimineerd kan worden; bekende voorbeelden zijn deeltijdwerk (voor geslacht) en postcode (voor ras, in de hiervoor genoemde juridische zin). Om indirect te kunnen discrimineren hoeven geslacht en ras dus niet expliciet bekend te zijn, proxy's volstaan. De zojuist door het Europees Parlement goedgekeurde AI Verordening⁷ definieert een AI-systeem als: 'een machine-gebaseerd systeem dat is ontworpen om met verschillende niveaus van autonomie te werken en dat na implementatie aanpassingsvermogen kan vertonen, en dat, voor expliciete of impliciete doelstellingen, uit de input die het ontvangt outputs afleidt zoals voorspellingen, inhoud, aanbevelingen of beslissingen die fysieke of virtuele omgevingen kunnen beïnvloeden.'⁸ Ik ga ervan uit dat alle op risicoprofielen gebaseerde systemen waar het hier over gaat, kwalificeren als AI-systeem in de zin van de AI Verordening.

In de rechtstoepassing worstelen we al langer met het gebruik van risicoprofielen; wat explicieter sinds de kinderopvangtoeslagaffaire, maar in de rechtspraak daarvoor ook al.⁹ In november 2021 publiceerde het College voor de Rechten van de Mens een toetsingskader met het oog op discriminatie door (het gebruik van) risicoprofielen, gevolgd door een kamerbrief van de minister van Binnenlandse Zaken¹⁰ en een Handreiking voor non-discriminatie by design.¹¹ De heersende leer was altijd: we onderkennen dat gebruik van risicoprofielen tot indirecte discriminatie kan leiden, dus is het riskant om ze te gebruiken, maar dat gebruik is wel nodig en als we maar héél voorzichtig en alert zijn dan kan het wel. En ook (proxy's voor) ras mogen in risicoprofielen gebruikt worden, als ze maar niet het enige criterium zijn.

Met de baanbrekende uitspraak van het hof Den Haag van 14 februari 2023 veranderde dat: 'De KMar onderwerpt in voorkomend geval personen aan een controle vanwege, bijvoorbeeld, hun huidskleur, waar andere personen, die niet dezelfde huidskleur hebben maar wel aan de overige indicatoren voldoen, niet gecontroleerd worden. Dit is discriminatie op grond van ras.'¹² De noot bij dit arrest verduidelijkt: 'Als de combinatie van indicatoren is 'man + goed gekleed + donkere huidskleur', dan zullen goed geklede witte mannen niet worden gecontroleerd en goed geklede mannen van kleur wel. Dat de andere indicatoren óók beslissend zijn (en goede geklede vrouwen van kleur dus ook niet

⁷ Het oorspronkelijke voorstel: Verordening van het Europees Parlement en de Raad tot vaststelling van geharmoniseerde regels betreffende artificiële intelligentie (Wet op de artificiële intelligentie) en tot wijziging van bepaalde wetgevingshandelingen van de Unie, COM/2021/206 final. De meest recente versie is die van het Europees Parlement van 19 april 2024. Ik verwijs naar de artikelen uit de laatste (Engelstalige) versie, maar ik gebruik wel de Nederlandse terminologie.

⁸ Art. 3 sub 1 AI Verordening.

⁹ Een simpele zoekopdracht op rechtspraak.nl op 'risicoprofiel' levert ruim 1450 resultaten op, maar daarbij zitten ook veel zaken over het risicoprofiel van investeerders in financiële zaken – voor hier niet relevant.

¹⁰ Kamerstukken II 2021/22, 30950, nr. 281.

¹¹ Van der Sloot e.a. 2021.

¹² Hof Den Haag 14 februari 2023, ECLI:NL:GHDHA:2023:173, r.o. 8.24

worden gecontroleerd) maakt dat niet anders.¹³ Het eruit gepikt worden op basis van huidskleur is discriminatie – en dat wordt volgens mij niet op een of andere manier ‘goedgemaakt’ door het feit dat vervolgens misschien wel bij die persoon gevonden wordt waar op werd gecontroleerd.¹⁴

In de zaak van het hof ging het over directe discriminatie, omdat huidskleur als criterium werd gebruikt. Ik zou nog verder willen gaan: ook proxy’s voor ras (of een andere discriminatiegrond) zouden niet in risicoprofielen gebruikt moeten worden, omdat het risico op ongerechtvaardigde indirecte discriminatie te groot is. De uiterste consequentie zou zijn dat risicoprofielen eigenlijk helemaal niet meer gebruikt zouden moeten worden.¹⁵ Het is vooraf onmogelijk om te zien welk gegeven (wellicht in combinatie met andere gegevens) als proxy voor een discriminatiegrond fungeert. Zo zou het feit dat een ‘uitwonende’ student bij een familielid woont als proxy voor ras kunnen fungeren.¹⁶ Een bekend voorbeeld is ook dat het ineens overgaan op ongeparfumeerde verzorgingsproducten kan wijzen op zwangerschap.¹⁷ Wij kunnen zoïets vooraf niet verzinnen – maar dit soort verbanden blijken wel uit *data analytics*, en kunnen leiden tot onbedoelde ‘proxy discriminatie’.¹⁸

Ik heb drie inzichten om de stelling dat we ook geen proxy’s voor ras in risicoprofielen moeten gebruiken te onderbouwen.

Drie inzichten

Ten eerste. Het feit dat een risicoprofiel tot empirisch juiste voorspellingen leidt is juridisch niet relevant. Vrouwen hebben nog steeds minder kans om door te stromen naar een topfunctie.¹⁹ Dat wil niet zeggen dat het eerlijk (niet discriminerend) is om daarom vrouwen niet uit te nodigen voor een sollicitatiegesprek.²⁰ Met betrekking tot ras: het is ongetwijfeld empirisch juist dat jongeren met een migratieachtergrond uit een bepaalde achterstandswijk een grotere kans hebben om in de criminaliteit terecht te komen.²¹ Maar het is apert onrechtvaardig om hen daarom als potentiële criminelen te behandelen. We horen mensen te behandelen en te beoordelen op basis van hun eigen gedrag, waar ze zelf verantwoordelijk voor zijn, en niet op basis van groepsgebonden kenmerken waar ze zelf niet voor gekozen hebben en dus niet verantwoordelijk voor zijn (artikel 1 Grondwet).

¹³ Djwalapersad & De Vries 2023, p. 615.

¹⁴ Hetgeen in de zaak van het hof Den Haag overigens niet aan de orde was.

¹⁵ Van der Linden 2024.

¹⁶ Zoals het geval was bij de indirecte discriminatie door DUO, Van der Linden 2024.

¹⁷ Hill 2012.

¹⁸ Prince & Schwarcz 2020.

¹⁹ Merens 2022.

²⁰ Goodman 2018.

²¹ Bezemer & Leerkens 2021.

Ten tweede. Als mensen op een bepaalde manier behandeld worden gaan ze zich daar ook naar gedragen. De medische wetenschap heeft het over een ‘nocebo-effect’ (als tegenhanger van placebo): lage verwachtingen hebben een negatief effect op de werking van een behandeling. Het plaatsen van kinderen (van een gelijk beginniveau) in een ‘goede’ en een ‘slechte’ klas leidt tot goede respectievelijk slechte leerresultaten.²² Als jongeren met een migratieachtergrond uit een achterstandswijk als (potentiële) criminelen behandeld worden dan wordt daarmee de kans vergroot dat ze zich ook werkelijk crimineel gaan gedragen.²³

Ten derde. Als door datagedreven opsporing een bepaalde bevolkingsgroep meer gecontroleerd wordt (bijvoorbeeld op fraude), dan wordt bij die bevolkingsgroep natuurlijk ook meer fraude gevonden. Als vervolgens de statistieken aangepast worden en op basis daarvan die bevolkingsgroep nog meer gecontroleerd wordt, loopt het al heel gauw uit de hand.²⁴ Dat mechanisme leidt tot oververtegenwoordiging van bepaalde groepen in de fraudestatistieken: het gaat om zelfversterkende feedbackloops oftewel selffulfilling prophecy's. En zelfs als iedereen die betrapt is wel daadwerkelijk fraude heeft gepleegd, is er toch sprake van discriminatie als mensen uit die bevolkingsgroep een veel grotere kans hebben om gecontroleerd te worden.²⁵

Natuurlijk is het contra-intuïtief om, gegeven de altijd schaarse opsporingscapaciteit, die niet in te zetten daar waar je, niet alleen op basis van een onderbuikgevoel maar zelfs op basis van ‘objectieve’ data, verwacht iets te vinden. Toch zal het moeten, als we het institutionele racisme waar al onze instituties en ook wijzelf van doordrenkt zijn,²⁶ écht willen aanpakken.²⁷

Het gaat hier over iets anders dan de invulling van ‘een redelijk vermoeden van schuld’ uit artikel 27 Wetboek van Strafvordering (Sv). Hier gaat het over de invulling van de discretionaire ruimte die aan ambtenaren gegeven wordt om te bepalen wie gecontroleerd wordt, waarbij in theorie iedereen gecontroleerd zou mogen worden. Artikel 27 Sv bepaalt tegen wie en onder welke voorwaarden (grondrechtenbeperkende) dwangmiddelen ingezet mogen worden.

Hoog risico AI, bias en de (super)human-in-the-loop

Maar de AI Verordening heeft toch allerlei bepalingen die ons tegen schadelijke effecten van gebruik van AI-systemen moeten beschermen? Zeker. Maar tegen discriminatie door het gebruik van risicoprofielen biedt de AI Verordening maar weinig soelaas.

²² Merton 1948; Bregman 2019, p. 29-32.

²³ Wiley & Esbensen 2016.

²⁴ Smeets 2020.

²⁵ Hof Den Haag 14 februari 2023, ECLI:NL:GHDHA:2023:173, r.o. 8.24

²⁶ Nieuwsbericht Rijksoverheid: 'Aandacht voor institutioneel racisme binnen toezicht Belastingdienst en Toeslagen', rijksoverheid.nl 30 mei 2022, College voor de Rechten van de Mens, zonder datum.

²⁷ Kamerbrief over aanpak (institutioneel) racisme, 25 januari 2023, 2023-000048297.

In de AI Verordening worden AI-systemen ingedeeld in risico-categorieën: AI-systemen die een onaanvaardbaar risico opleveren worden verboden (art. 5), voor een hoog risico gelden uitgebreide vereisten, en voor systemen die weinig of geen risico opleveren gelden lagere of geen eisen. De AI die gebruikt wordt in het soort zaken waar we het hier over hebben, zal veelal kwalificeren als een AI-systeem met een hoog risico. Artikel 6 verwijst naar Annex 3 waar een opsomming staat. Voor ons zijn van belang sub 4 (werkgelegenheid, personeelsbeheer en toegang tot zelfstandige arbeid), sub 5 (toegang tot en gebruik van essentiële particuliere diensten en openbare diensten en uitkeringen), sub 6 (rechtshandhaving), sub 7 (migratie, asiel en beheer van grenscontroles) en sub 8 (rechtsbedeling en democratische processen). Dat betekent dat die AI-systemen aan de gestelde eisen moeten voldoen. Die eisen hebben onder andere betrekking op risicobeheer (art. 9), data en databeheer (art. 10), transparantie en informatieverstrekking aan gebruikers (art. 13), menselijk toezicht (human-in-the-loop, art. 14) en nauwkeurigheid, robuustheid en cyberbeveiliging (art. 15). Als vervolgens bij zelf uitgevoerde conformiteitsbeoordelingsprocedures blijkt dat een AI-systeem inderdaad aan die eisen voldoet, moet er een CE-markering opgeplakt worden (art. 48). Ik ga kort in op de eisen van artikelen 10 en 14.

Artikel 10 bepaalt dat de datasets die gebruikt worden voor het trainen, valideren en testen van AI-systemen onderworpen moeten worden ‘aan passende praktijken op het gebied van databeheer’. In het bijzonder moet er gelet worden op ‘het opstellen van relevante aannames, met name met betrekking tot de informatie die de data moeten meten en vertegenwoordigen’ en op ‘een beoordeling met het oog op mogelijke vertekeningen die (...) een negatieve impact kunnen hebben op fundamentele rechten of tot verboden discriminatie kunnen leiden.’ Er moeten passende maatregelen worden genomen om dergelijke vertekeningen (biases) op te sporen, te voorkomen en te verminderen (art. 10 lid 2 sub d, f en g). Het probleem is alleen dat de data de bestaande maatschappelijke ongelijkheden correct weergeven, zie het eerste inzicht dat ik hiervoor noemde (inclusief eventuele vertekeningen door de zelfversterkende feedbackloop van het derde inzicht). Kunnen we dan wel zeggen dat er sprake van een vertekening, een bias? Zo ja, dan is een eerste uitdaging om die als zodanig waar te nemen. En als die dan al gecorrigeerd zou kunnen worden (bijvoorbeeld door het toevoegen van synthetische data om ongelijkheden te compenseren),²⁸ dan zou dat betekenen dat de dataset onjuist wordt. En dat is ook een probleem. Kortom: AI-systemen zullen bestaande diepgewortelde discriminatoire patronen (zoals institutioneel racisme en de nog steeds bestaande ongelijkheid tussen mannen en vrouwen) voortzetten en legitimeren.

Artikel 14 bepaalt dat er menselijk toezicht moet zijn. Mensen zijn heel vatbaar voor ‘automation bias’,²⁹ het verschijnsel dat output van een computersysteem meer vertrouwd wordt dan diezelfde uitkomst als die door een mens gesuggereerd wordt. Ik ben bang dat het er met een goedgekeurde AI niet beter op gaat worden: het welbekende ‘computer says no’ verwordt misschien wel tot ‘certified AI says no’.

²⁸ Ferrara 2024.

²⁹ Ruschemeier & Hondrich 2024.

Ga daar maar eens tegenin! Ja, er is menselijk toezicht – maar hoe reëel is dat bij een hoge werkdruk en bij gebrek aan organisatorische prikkels om van het ‘advies’ van het systeem af te wijken?³⁰

Conclusie

De rechter is de laatste hoop voor een gedupeerde. Dus de rechter moet heel goed opletten om niet zelf ook aan die automation bias ten prooi te vallen, heel goed kijken wat er nou precies in een bepaalde zaak gebeurd is, welke rol een risicoprofiel gespeeld heeft, en of de human-in-the-loop zijn of haar werk wel goed gedaan heeft (en heeft kunnen doen).³¹ De rechter moet niet alleen marginaal toetsen, niet volstaan met de vaststelling dat de juiste procedure gevolgd is. Al helemaal in gevallen waarin de persoon om wie het gaat ‘toevallig’ tot een groep behoort die door een discriminatiegrond beschermd wordt. Is er ergens in de beslissing om deze persoon aan controle te onderwerpen, een risicoprofiel gebruikt? Kan iemand uitleggen op basis waarvan het risicoprofiel juist deze persoon eruit gepikt heeft? Is dat op basis van iets dat als (proxy voor) een discriminatiegrond gezien zou kunnen worden? Dan is de kans groot dat die selectiebeslissing al neerkomt op discriminatie.³² Heel kritisch zijn dus, superhuman-in-the-loop!

Literatuur

Afdeling bestuursrechtspraak van de Raad van State 2021

Afdeling bestuursrechtspraak van de Raad van State, *Lessen uit de kinderopvangtoeslagzaken, Reflectierapport van de Afdeling bestuursrechtspraak van de Raad van State*, Den Haag: Raad van State 2021.

Bezemer & Leerkes 2021

W. Bezemer & A. Leerkes, *Oververtegenwoordiging verder ontcijferd. Een kwantitatief onderzoek naar sociale verschillen in verdenkingskans en zelfgerapporteerd crimineel gedrag onder jongeren in Nederland*, Den Haag: Politie en Wetenschap/Sdu 2021.

Bregman 2019

R. Bregman, *De meeste mensen deugen - Een nieuwe geschiedenis van de mens*, De Correspondent 2019.

CBS 2024

CBS, *Haalbaarheidsstudie toeslagenaffaire-onderzoek*, 15 maart 2024, cbs.nl.

College voor de Rechten van de Mens zonder datum

College voor de Rechten van de Mens, *Institutioneel racisme*, mensenrechten.nl.

³⁰ Constantino 2022.

³¹ Rapport Werkgroep Reflectie Toeslagenaffaire Rechtbanken 2021, Afdeling bestuursrechtspraak van de Raad van State 2021.

³² Hof Den Haag 14 februari 2023, ECLI:NL:GHDHA:2023:173, r.o. 8.24.

College voor de Rechten van de Mens 2021

College voor de Rechten van de Mens, *Discriminatie door risicoprofielen, Een mensenrechtelijk Toetsingskader*, 30 november 2021, mensenrechten.nl.

Constantino 2022

J. Constantino, 'Exploring Article 14 of the EU AI Proposal: Human in the Loop Challenges When Overseeing High-Risk AI Systems in Public Service Organisations', *Amsterdam Law Forum* 2022/3.

Djwalapersad & De Vries 2023

V. Djwalapersad & K.M. de Vries, 'Noot bij Hof Den Haag 14 februari 2023, ECLI:NL:GHDHA:2023:173', *Jurisprudentie Vreemdelingenrecht* 2023/6.

DUO 2024

DUO, *Intern onderzoek controle uitwonendenbeurs*, 1 maart 2024, rijksoverheid.nl.

Ferrara 2024

E. Ferrara, 'Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies', *Sci* 2024/3.

Goodman 2018

R. Goodman, 'Why Amazon's Automated Hiring Tool Discriminated Against Women', aclu.org, 12 oktober 2018.

Hill 2012

K. Hill, 'How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did, forbes.com, 16 februari 2012.

Van der Linden 2024

T. van der Linden, 'En nou is het afgelopen!' *Tijdschrift voor Internetrecht* 2024/2.

Merens 2022

A. Merens, *Een lange weg, De ondervertegenwoordiging van vrouwen in management en top nader verklaard*, Den Haag: Sociaal en Cultureel Planbureau 2022.

Merton 1948

R. Merton, 'The Self-Fulfilling Prophecy', *The Antioch Review* 1948/2.

Prince & Schwarcz 2020

A.E.R Prince & D. Schwarcz, 'Proxy Discrimination in the Age of Artificial Intelligence and Big Data', *Iowa Law Review* 2020/3.

Rapport Werkgroep Reflectie Toeslagenaffaire Rechtbanken 2021

Rapport Werkgroep Reflectie Toeslagenaffaire Rechtbanken, *Recht vinden bij de rechtbank, Lessen uit kinderopvangtoeslagzaken*, rechtspraak.nl 2021

Ruscheimer & Hondrich 2024

H. Ruschemeier & L. Hondrich, 'Automation Bias in Public Administration - an Interdisciplinary Perspective From Law and Psychology', papers.ssrn.com, 23 februari 2024.

Van der Sloot e.a. 2021

B. van der Sloot e.a., Handreiking non-discriminatie by design, Rijksoverheid.nl, 10 juni 2021.

Smeets 2020

I. Smeets, '510 staandehoudingen en de zelfversterkende feedback-loop', *de Volkskrant* 12 mei 2020.

Wiley & Esbensen 2016

S.A. Wiley & F.A. Esbensen, 'The effect of police contact: Does official intervention result in deviance amplification?' *Crime & Delinquency* 2016/3.

Eerder verschenen

- | | |
|-------------|---|
| 2024, nr. 1 | Reflecties op het Visitatierapport van de Rechtspraak 2022-2023 |
| 2023, nr. 3 | (On)bewuste vooroordelen in de rechtspraak |
| 2023, nr. 2 | Juridische ondersteuning van de rechter |
| 2023, nr. 1 | Vertrouwen in de rechtspraak |