Een transparant debat over algoritmen

Gstrein, Oskar Josef; Zwitter, Andrej

Published in:
Bestuurskunde

DOI:
10.5553/Bk/092733872020029004004

Een transparant debat over algoritmen*

Oskar J. Gstrein & Andrej Zwitter

De politie gebruikt allerlei informatie om haar taken te vervullen. Waar het verzamelen en interpreteren van informatie traditioneel alleen door mensen kon worden gedaan, schept de opkomst van 'Big Data' nieuwe kansen en dilemma's. Enerzijds kunnen grote hoeveelheden gegevens worden gebruikt om algoritmen te trainen. Dit maakt het mogelijk om misdrijven als fietsendiefstal, inbraak of zelfs ernstige misdrijven als moord en terroristische aanslagen te 'voorspellen'. Aan de andere kant verdrinken zeer relevante vragen over doel, effectiviteit en legitimiteit van de toepassing van machinaal leren/'kunstmatige intelligentie' maar al te vaak in de oceaan van Big Data. Dit is met name problematisch als dergelijke systemen worden gebruikt in de publieke sector in democratieën waarin de rechtsstaat van toepassing is, en de verantwoordingsplicht en de mogelijkheid tot rechterlijke toetsing zijn gewaarborgd. In dit artikel onderzoeken we de rol die transparantie kan spelen om deze mogelijkheden en dilemma's met elkaar te verzoenen. Hoewel sommigen voorstellen om de systemen en de gegevens die zij zelf gebruiken transparant te maken, stellen wij voor om al tijdens het ontwerpproces een open en brede discussie over het doel en de doelstellingen te voeren. Dit zou effectiever kunnen zijn om ethische en juridische beginselen in de technologie te verankeren en de legitimiteit tijdens de toepassing te waarborgen.

Inleiding

De oorsprong van het woord 'algoritme' is terug te voeren op een islamitische wiskundige die in het begin van de negende eeuw in Bagdad werkte. Terwijl de naam van Mohammad ibn Musa al-Khwarizmi bijna vergeten was, waren zijn technieken om getallen te manipuleren zo indrukwekkend dat ze zich verspreidden naar de rest van de wereld en hedendaags nog steeds worden gebruikt. Zijn bijnaam al-Jabr werd 'algebra', en zijn oorspronkelijke naam al-Khwarizmi werd vertaald in het Latijn en andere talen, om uiteindelijk tot 'algoritme' te worden verbasterd (Durnová & Alberts, 2014, p. 101). Voortbouwend op deze wetenschappelijke erfenis kan een algoritme in grote lijnen worden omschreven als een wiskundige besluitvormingsprocedure die gebruikmaakt van variabelen om een numeriek resultaat of een uitkomst te produceren. Ondanks het feit dat algoritmen al meer dan duizend jaar oud zijn, heeft de opkomst van schijnbaar oneindige hoeveelheden data – ook wel omschreven als 'Big Data' (Zwitter, 2014, pp. 2-3; Furht & Villanustre, 2016) – er samen met aanzienlijk toegenomen

* Dr. O.J. Gstrein is universitair docent Governance & Innovation aan de Rijksuniversiteit Groningen, Campus Fryslân, Data Research Centre. Prof. dr. A.J. Zwitter is hoogleraar Governance & Innovation aan de Rijksuniversiteit Groningen, Campus Fryslân, Data Research Centre.

Bestuurskunde 2020 (29) 4
doi: 10.5553/Bk/092733872020029004004

mogelijkheden voor gegevensverwerking en geavanceerde data-analyse voor gezorgd dat complexe algoritmen de basis zijn geworden van geautomatiseerde besluitvormingssystemen. Deze geautomatiseerde systemen werken geheel of gedeeltelijk autonoom. Hoewel de particuliere sector vaak wordt beschouwd als de belangrijkste drijvende kracht achter de snelle grootschalige invoering van dergelijke geautomatiseerde besluitvormingssystemen, wordt in een rapport van de Duitse maatschappelijke organisatie Algorithmwatch uit 2019 benadrukt hoe wijdverbreid het gebruik van dergelijke systemen ook in de publieke sector in Europese landen is (Automating Society – Taking Stock of Automated Decision-Making in the EU, 2019), met name in tijden van COVID-19 (ADM Systems in the COVID-19 Pandemic, 2020; Zwitter & Gstrein, 2020). In toenemende mate analyseren academici en deskundigen deze ontwikkelingen en hun impact op het macro-, meso- en microniveau van de overheid (Veale & Brass, 2019, p. 142).

In dit artikel onderzoeken we het verband tussen deze toenemende wens om gebruik te maken van geautomatiseerde besluitvorming en de eis dat overheidsinstellingen transparant moeten handelen in een democratie. Daarom stellen we voor om transparantie niet als een doel op zich te beschouwen. Transparantie is veeleer een belangrijke waarborg voor de rechtsstaat, aangezien zij het scheppen van vertrouwen en legitimiteit bevordert. Daarom is onze discussie breder dan alleen het fenomeen van de zwarte doos van de algoritmische voorspelling, maar omvat deze ook kwesties van gegevensuitwisseling en procedurele transparantie. Transparantie dient de legitimiteit, aangezien de besluitvormingsprocessen in democratische instellingen die gebonden zijn aan de rechtsstaat, de geaccepteerde normen moeten volgen. Bovendien moeten hun beslissingen toegankelijk zijn voor 'checks and balances' (Zwitter & Hazenberg, 2020).

In een arrest van 5 februari 2020 oordeelde de rechtbank Den Haag dat het gebruik van een van deze geautomatiseerde systemen op basis van algoritmen – 'SyRI' of Systeem Risico Indicatie – een inbreuk vormde op het recht op eerbiediging van het privé- en gezinsleven, zoals vastgelegd in artikel 8 van het Europees Verdrag voor de Rechten van de Mens (EVRM).[1] SyRI werd door de Nederlandse overheid gebruikt met het oog op het efficiënt opsporen van verschillende vormen van fraude met betrekking tot sociale uitkeringen en belastingen. Verschillende personen en maatschappelijke organisaties waren bezorgd over de mogelijkheid dat de gegevens die worden gebruikt om het algoritme en de daaruit voortvloeiende geautomatiseerde beslissingen te trainen, onevenredig gericht zijn op kansarme personen met een zwakke sociaaleconomische status of een bepaalde etnische achtergrond. De beslissingen van het systeem resulteerden dan ook in stigmatisering en discriminatie. De rechtbank besliste in haar vonnis dat de wetgeving die ten grondslag ligt aan SyRI, niet het door het EVRM vereiste eerlijke evenwicht heeft. Door gebruik te maken van SyRI heeft de overheid dan ook het recht op privéleven geschonden en de individuele en collectieve autonomie onevenredig beperkt. Belangrijk is dat de rechtbank constateert dat de toepassing

1 Rb. Den Haag 5 februari 2020, C-09-550982-HA ZA 18-388, ECLI:NL:RBDHA:2020:865.

van SyRI onvoldoende transparant en controleerbaar is. Als gevolg van de uitspraak moest de Nederlandse overheid stoppen met het gebruik van SyRI (Kager, 2020), wat ook kan worden gezien in het licht van de bredere bezorgdheid over discriminatie door publieke instellingen (Ministerie van Volksgezondheid, 2020).

Rekening houdend met deze en soortgelijke ontwikkelingen richt dit artikel zich voornamelijk op onze observaties over de implementatie van geautomatiseerde besluitvorming in het kader van de openbare veiligheid. Als partners in het EU-project Cutting Crime Impact (CCI)[2] hebben wij samengewerkt met *law enforcement agencies* (LEA's) uit Nederland, Nedersaksen in Duitsland, Greater Manchester in het Verenigd Koninkrijk, Estland, Catalonië en de politie van Lissabon in Portugal. Een van de vier aandachtsgebieden ('focus areas') van CCI is Predictive Policing (PP), dat kan worden gedefinieerd als het verzamelen en analyseren van gegevens over eerdere misdrijven voor de identificatie en statistische voorspelling van personen of gebieden met een verhoogde kans op criminele activiteiten om te helpen bij de ontwikkeling van politie-interventie- en preventiestrategieën en -tactieken (Meijer & Wessels, 2019, p. 1033). Hoewel deze toepassing van geautomatiseerde besluitvorming relatief nieuw is, heeft het veel discussie uitgelokt over de doelmatigheid, de eerlijkheid, de legitimiteit en de effectiviteit ervan (Richardson, Schultz & Crawford, 2019). Aangezien sommige deskundigen voorstellen om de systemen en de gegevens die ze gebruiken, transparant te maken (Hardyns & Rummens, 2018, p. 214), stellen we bovendien dat al tijdens het ontwerpproces een open en brede discussie over het doel en de doelstellingen van PP moet worden gevoerd. Dit is wellicht effectiever om ethische en juridische principes in de technologie te verankeren en de legitimiteit tijdens het gebruik te waarborgen.

Onderzoeksobservaties over algoritmisch beleid

Predictive Policing in het project Cutting Crime Impact

In het CCI-project nemen LEA's uit Nederland en Nedersaksen in Duitsland het voortouw. Het doel van het project is het ontwerpen en ontwikkelen van innovatieve 'toolkits' voor LEA's, die gebaseerd zijn op een 'state of the art review' en een onderzoek naar ethische, juridische en sociale uitdagingen van PP (Gstrein, Bunnik & Zwitter, 2019). Het ontwerpproces van het project is 'mensgericht', wat ook een van de belangrijkste eisen is van de 'High Level Expert Group on Artificial Intelligence' (AI-HLEG) van de Europese Unie, die in april 2019 ethische richtlijnen voor de ontwikkeling van AI-systemen heeft gepresenteerd (Weiser, 2019). Dit zet de CCI-aanpak in contrast met de praktijk van veel LEA's in de Verenigde Staten, het Verenigd Koninkrijk en Europa die 'ready-made solutions' kopen en exploiteren, ontworpen door niet-LEA-partijen zoals 'PredPol' of 'PreCobs' (Hardyns & Rummens, 2018, pp. 208-209). Het CCI-ontwerpproces bestaat uit drie uiteenlopende en convergerende fasen, waarin het landschap wordt

2 Zie www.cuttingcrimeimpact.eu, geraadpleegd op 17 september 2020.

Bestuurskunde 2020 (29) 4
doi: 10.5553/Bk/092733872020029004004

bestudeerd en herzien en waarin de ontwerpeisen worden vastgelegd door middel van interdisciplinair en transnationaal onderzoek.

In de Europese Unie loopt de Nederlandse politie voorop in het gebruik van PP, te beginnen met de ontwikkeling van een eigen 'Crime Anticipation System' (CAS) in 2013 (Hardyns & Rummens, 2018, p. 207; Gstrein et al., 2019, p. 85). Wereldwijd is de geschiedenis van PP terug te voeren op Los Angeles in Californië, waar in 2008 de eerste PP-systemen werden ontwikkeld en getest. Deze systemen waren gebaseerd op de veronderstelling dat statistische methoden voor het voorspellen van aardbevingen ook nuttig zouden kunnen zijn voor het voorspellen van criminaliteit. PP is met gretigheid geïmplementeerd vanwege de beloften van verminderde criminaliteit door het preëmptief inzetten van agenten. Het zou op zijn beurt de LEA's efficiënter maken en helpen om de kosten te drukken. Men kan PP als volgt categoriseren (Perry, McInnis, Price, Smith & Hollywood, 2013):

- Methoden voor het voorspellen van misdaden: dit zijn benaderingen die gebruikt worden om plaatsen en tijden te voorspellen met een verhoogd risico op criminaliteit.
- Methoden voor het voorspellen van daders: deze benaderingen identificeren individuen die het risico lopen om in de toekomst delinquent te worden.
- Methoden voor het voorspellen van de identiteit van daders: deze technieken worden gebruikt om profielen te creëren die mogelijke daders nauwkeurig vergelijken met specifieke misdrijven uit het verleden.
- Methoden voor het voorspellen van slachtoffers van misdrijven: vergelijkbaar met de methoden die zich richten op daders, misdaadlocaties en tijden van verhoogd risico, worden deze benaderingen gebruikt om groepen of, in sommige gevallen, individuen te identificeren die waarschijnlijk het slachtoffer van een misdrijf zullen worden.

De meeste PP-systemen die momenteel in gebruik zijn, zijn locatiegericht en passen de zogenaamde 'near-repeat-approach' toe. Deze theorie is gebaseerd op criminologische inzichten die suggereren dat het misdrijf opnieuw in dezelfde buurt gaat plaatsvinden (Gstrein et al., 2019, p. 86). Door de combinatie van historische criminaliteitsdata met aanvullende bronnen – zoals de weersomstandigheden, straatrasterpatronen of het gemiddelde inkomen in een buurt – wordt een kaart met 'hot spots' berekend. Deze 'hot spots' geven locaties aan waar misdrijven, zoals inbraak of autodiefstal, kunnen plaatsvinden. Zoals geschetst kunnen PP-systemen zich echter ook richten op individuen (Sommerer, 2020, pp. 29-115). Aan het ontwerp en het gebruik van conventionele PP-systemen zijn verschillende aan elkaar gerelateerde ethische, juridische en sociale uitdagingen verbonden (Gstrein et al., 2019, pp. 86-93).

Van voorspellen naar preventie, naar het delen van informatie

Met het oog op praktische ethische bezwaren zijn er vier punten die aandacht behoeven:

- 'selffulfilling prophecies';
- discriminatie door algoritmen;

Tabel 1 Samenvatting gebaseerd op Gstrein et al., 2019, pp. 86-93

Ethisch domein	Transparantie en verantwoording *Kunnen LEA's nog steeds uitleggen waarom ze agenten sturen om een specifieke locatie te monitoren, of vertrouwen ze op een 'black box'? Wie is verantwoordelijk voor de beslissingen?* Gegevensselectie en machinebevooroordeeldheid *Hoe ziet het proces eruit om de juiste/legitieme gegevens te selecteren? Zullen de gegevens volledig en accuraat zijn? Welk soort gegevens is relevant? Meer data hoeft niet per se meer inzicht te geven.* Visualisatie en interpretatie van prognoses *Het is belangrijk dat LEA's gegevens kunnen interpreteren en begrijpen wat het systeem daadwerkelijk als een prognose oplevert.* Tijd en effectiviteit *Onduidelijk blijft hoe de situatie zich zou hebben ontwikkeld als PP niet zou zijn gebruikt; moeilijk empirisch te meten effecten.* Stigmatisering van individuen, omgevingen en gemeenschapsgebieden *Het gebruik van gegevens zou de reeds bestaande vooroordelen ten aanzien van bepaalde gebieden of groepen kunnen versterken.*
Juridisch domein	Focus op de bescherming van het individu *Voorspellingen zijn gebaseerd op geaggregeerde (Big) Data, die niet aan een specifiek individu kunnen worden gekoppeld. Daarom is het mogelijk dat het kader voor gegevensbescherming niet van toepassing is.* Gebrek aan effectieve wettelijke waarborgen en rechtsmiddelen *Zelfs als een individu kan worden getroffen, is het onduidelijk wat een recht op herziening van geautomatiseerde individuele besluitvorming in detail inhoudt.*
Sociaal domein	Sociaal contract *Toenemende discriminatie door 'feedbackloop'; meer gegevens over bepaalde groepen resulteert in meer LEA-aanwezigheid.* Hoelang is historische criminaliteit relevant? *Het is moeilijk om perspectieven voor sociale ontwikkeling in te bedden in algoritmen. Hoelang is historische criminaliteit relevant en onder welke omstandigheden kan deze worden weggelaten voor toekomstige voorspellingen?* Vertrouwen winnen door transparantie? *Hoe om te gaan met burgers en 'moeilijk te bereiken' groepen? Vertrouwen winnen door de essentiële kenmerken van en keuzes voor een algoritme open te stellen?*

– ongepaste inmenging in de privésfeer (*nulla poena sine crimine*);
– zwarte doos; gebrek aan transparantie leidt tot problemen bij de beoordeling van bewijsmateriaal.

Een van de grootste nadelen van PP is dat er geen eenduidig empirisch bewijs is dat voorspellingen leiden tot lagere criminaliteitscijfers (Gstrein et al., 2019, pp. 89-90). Bovendien zijn voorspellingen alleen mogelijk als er voldoende gegevens beschikbaar zijn voor een bepaald gebied of een bepaalde misdaad, wat belangrijke redenen zijn om te begrijpen waarom PP onlangs in Zwitserland zeer kritisch is beoordeeld (Kayser-Bril, 2020). Deze aspecten, samen met de bezorgdheid over de versterking van de bestaande vooringenomenheid in historische misdaadgegevens, en de angst voor 'selffulfilling prophecies' als gevolg van meer agenten die permanent op zoek zijn naar misdaden in dezelfde buurten, hebben

Bestuurskunde 2020 (29) 4
doi: 10.5553/Bk/092733872020029004004

de aantrekkelijkheid van PP ernstig aangetast. Sommige deskundigen concluderen dat de belofte van PP de complexiteit van de criminaliteit en de daarmee samenhangende maatschappelijke aspecten niet erkent (Richardson et al., 2019, pp. 225-227).

In een onderzoeksrapport van Amnesty International van september 2020 wordt het 'Sensing-project' van de Nederlandse politie in de stad Roermond kritisch geëvalueerd op het punt van discriminatie. Het rapport beschrijft hoe de ontwikkelde en geteste instrumenten ter bestrijding van kleine criminaliteit (winkeldiefstal) rond een winkelcentrum gebruikmaken van data en algoritmische modellen om het risico te beoordelen dat een misdrijf door een bepaalde persoon of op een bepaalde locatie wordt gepleegd. De auteurs suggereren dat er ernstige tekortkomingen zijn in het ontwerp van het systeem, de data-administratie, de evaluatie en de databanken. Het PP-systeem werkt met dergelijke generieke profielen (een Duitse auto met meerdere passagiers op weg naar het winkelcentrum) dat het systeem veel 'valse positieven' creëert. Bovendien komen op deze manier grote aantallen mensen in extra politiedatabases terecht en worden hun gegevens verwerkt en opgeslagen (Amnesty International, 2020).

Hoewel dergelijke overwegingen typerend zijn voor debatten onder leiding van het maatschappelijk middenveld en academici, vond ons onderzoek in het CCI-project bovendien uitdagingen met betrekking tot de effectieve implementatie van gegevensgerichte systemen in gevestigde praktijken en procedures. Simpel gezegd willen veel actoren niet zozeer vertrouwen op een op gegevens gebaseerde voorspelling als wel op hun instinct en ervaring. Bovendien kunnen schijnbaar goedaardige aspecten, zoals beschikbaarheid en gemakkelijke toegankelijkheid van een tablet of een smartphone om data in de juiste context te plaatsen, cruciaal blijken om PP operationeel te maken. Dergelijke problemen zijn vaak moeilijk aan te pakken via algemene ethische richtsnoeren, zoals de ethische richtsnoeren voor betrouwbare AI die zijn opgesteld door de EU-deskundigengroep (AI-HLEG) op hoog niveau voor kunstmatige intelligentie (Weiser, 2019). Het is moeilijk te begrijpen wat de zeven beginselen, zoals 'Human Agency and Oversight' of 'Privacy and Data Governance', in specifieke contexten in de praktijk moeten betekenen. Dit is ook een van de belangrijkste redenen waarom het kader voor ethische richtsnoeren vergezeld moet gaan van specifieke regelgeving die bevoegdheden, individuele rechten en plichten omvat.

Tot slot dient zich de vraag aan of de focus van *data-driven policing* eigenlijk wel moet liggen op de waarde en nauwkeurigheid van voorspellingen. Deze voorspellingen lossen op zichzelf geen taak van LEA's op. Zij zouden veeleer kunnen helpen om criminaliteit te voorkomen, wat ook met andere middelen zou kunnen worden bereikt, zoals een belangrijkere rol voor wijkagenten. Terwijl dreigingsmodellen kunnen worden aangepakt met preventief gebruik van politiediensten om de kans op criminaliteit te verkleinen, biedt dit weinig soelaas voor de onderliggende oorzaken van criminaliteit, waarmee rekening moet worden gehouden bij het ontwerpen van interventies. Idealiter zou een model voor causale criminaliteit het ook mogelijk maken om meer preventieve middelen voor criminaliteits-

bestrijding op lange termijn in te zetten, waarbij niet alleen de proximale oorzaken worden aangepakt, maar ook de intermediaire en diepere oorzaken.

Zeker, het aanvullen van bestaande menselijke veronderstellingen en instincten met nauwkeurige gegevens die passen bij de context van een situatie, zou kunnen helpen bij het verminderen van potentiële vooroordelen. Dit kan echter alleen worden bereikt als de gegevens nauwkeurig worden gecreëerd, geanalyseerd, gedeeld en geïnterpreteerd, met een duidelijke focus en doelgerichtheid. In plaats van de nadruk te leggen op de omvang van de gegevens die kunnen worden gebruikt om voorspellingen te doen met behulp van algoritmen, is het daarom wellicht beter om na te gaan hoe de juiste informatie op het juiste moment kan worden gedeeld tussen de relevante eenheden van een LEA om het politiewerk effectief en legitiem te maken.

Herdefiniëring van de eisen

Het geval van PP toont aan dat de invoering van geautomatiseerde besluitvorming in het openbaar bestuur zou kunnen leiden tot de ontwikkeling van technische systemen die gebaseerd zijn op veronderstellingen die in werkelijkheid niet kloppen. Dit blijkt enerzijds uit de waarde van de voorspellingen zelf. Het is praktisch onmogelijk om te begrijpen hoe nauwkeurig ze zijn, en de complexe creatie door middel van algoritmen kan alleen door deskundigen worden begrepen. Deze combinatie zou kunnen leiden tot het ontstaan van een 'black box society' (Pasquale, 2015). Aan de andere kant is de inbedding van technologie in de gevestigde praktijken van LEA's niet eenvoudig. Veel praktische elementen moeten worden overwogen en aangepakt. Agenten zullen alleen op technologie vertrouwen als ze dat echt kunnen, wat in de eerste plaats vragen oproept over de investering in dergelijke systemen.

Een antwoord op deze dilemma's kan liggen in de zorgvuldige beoordeling van de ontwerpeisen. Als het gaat om het gebruik van gegevens en de totstandbrenging van een gezonde gegevenscultuur, kunnen de lidstaten van de Europese Unie zelfs een voordeel hebben ten opzichte van andere regio's in de wereld, aangezien de kaders voor gegevensbescherming en gegevensbeheer goed ontwikkeld zijn (Gstrein et al., 2019, p. 87). Daarnaast zijn er steeds meer richtlijnen beschikbaar gekomen over de ethische uitdagingen die gepaard gaan met het gebruik van Big Data (Zwitter, 2014; Weiser, 2019; Završnik, 2019). Hoewel de belofte van Big Data ligt in de verborgen patronen van data die vanwege de hoeveelheid informatie misschien alleen herkenbaar zijn voor algoritmen, is het toch raadzaam om heel goed te bedenken in welke richting de antwoorden gezocht moeten worden. Het stellen van goede en duidelijke onderzoeksvragen kan immers leiden tot nuttiger antwoorden dan vage aannames en instructies. In die zin blijft het principe van doelmatigheid, dat een van de hoekstenen is van de traditionele gegevensbeschermingswetgeving (Ukrow, 2018, pp. 242-243), cruciaal voor systeemgeavanceerde algoritmen.

Bestuurskunde 2020 (29) 4
doi: 10.5553/Bk/092733872020029004004

De rol van transparantie

Gebrek aan juridische waarborgen & belang van de ethiek
Hoewel sommige aspecten van PP de mogelijkheden van LEA kunnen vergroten, leidt het gebruik van algoritmen en Big Data ook tot bezorgdheid over het principe van *nulla poena sine crimine* (geen straf zonder misdaad), de rechtsstaat en de vrije en onafhankelijke evaluatie van bewijsmateriaal in een rechtbank die haar beslissingen zou baseren op nauwelijks te evalueren resultaten door middel van algoritmen. In dit verband heeft de Wetenschappelijke Raad voor het Regeringsbeleid in 2016 een rapport gepubliceerd over 'Big Data in een vrije en veilige samenleving', waarin de nadruk wordt gelegd op mogelijke vooroordelen in datasets en waarin wordt gepleit voor voorzichtigheid bij de inzet van autonome en semi-autonome besluitvorming in het veiligheidsdomein (Ministerie van Algemene Zaken, 2016).

Daarnaast kan het gebruik van Big Data, verzameld door het monitoren van sociale media en open bronnen ('OSINT'), enorme inzichten geven in zowel maatschappelijke processen als individuele voorkeuren (Gstrein & Ritsema van Eck, 2018, pp. 70-74). Tegelijkertijd kan het leiden tot massale privacyschendingen en manipulatie voor politieke doeleinden, zoals het geval van Cambridge Analytica heeft aangetoond (Hu, 2020). Veel gebruikers van digitale diensten zijn zich totaal niet bewust van dergelijke mogelijkheden, wat ook de vraag oproept hoe transparant de overheid en LEA's moeten zijn bij het gebruik van deze methoden. Voor een effectief gebruik van Big Data voor het voorspellen van criminaliteit op individueel niveau is het in eerste instantie nodig om gegevens te verzamelen voordat er enige verdenking over het individu wordt geuit. Daarom zijn er algemene procedurele waarborgen die invallen alleen rechtvaardigen als er een redelijke grond voor is. Aangezien PP gewoonlijk gebaseerd is op geaggregeerde gegevens en patroonanalyses, zijn dergelijke zorgen met betrekking tot individuele rechten echter niet voldoende om de 'beoogde' subjecten te beschermen. Zo kan het gebruik van gegevens over misdrijven en slachtoffers in het verleden in samenhang met de levenscyclus van slachtoffers, het weer, het tijdstip van de dag, het verkeer en demografische gegevens, leiden tot de identificatie van hotspots voor misdrijven. Het academische debat van de afgelopen jaren heeft ook vragen verkend over de privacy van groepen en de bedreiging van de collectieve autonomie die dergelijke praktijken met zich mee kunnen brengen (Taylor, Van der Sloot & Floridi, 2017). Dit vereist zeker een bredere maatschappelijke discussie over hoe de potentiële schade kan worden beperkt.

Een ander probleem met betrekking tot de verzamelde gegevens betreft de maatschappelijke vooroordelen. Het COMPAS-algoritme (Correctional Offender Management Profiling for Alternative Sanctions) bijvoorbeeld, software die is ontwikkeld als intellectueel eigendom om het recidivecijfer van vrijgelaten gevangenen te voorspellen, vertoonde duidelijke tekenen van vooringenomenheid ten opzichte van Afro-Amerikaanse gevangenen. De resultaten van deze software werden door rechtbanken in Florida en andere staten gebruikt als bewijs om de

waarschijnlijkheid te voorspellen of een gevangene uit de gevangenis zou worden vrijgelaten. Niet alleen discrimineerde het algoritme zwarte gevangenen, rechters hadden ook geen mogelijkheden om de resultaten te verifiëren door het algoritme en de gebruikte gegevens te beoordelen, omdat de software intellectueel eigendom was van Northpointe, Constellation Software (Angwin, Larson, Mattu & Kirchner, 2016; Gstrein, 2016). Deze zaak illustreert dat private software en intellectuele eigendomsrechten op gespannen voet staan met de transparantie-eisen van de rechtsstaat. Algoritmische transparantie is dus een belangrijk element van algoritmische rechtvaardigheid. Uitlegbare kunstmatige intelligentie is echter nog ver verwijderd van elk betrouwbaar gebruik in het domein van justitie en veiligheid.

De onmogelijkheid om maatschappelijke aspiraties te verankeren

Een brede en transparante discussie over de aard van een systeem dat gebruik maakt van algoritmen is niet alleen noodzakelijk om schade te voorkomen, maar ook om duidelijke doelen en doelstellingen voor het gebruik van het systeem te definiëren. Wat betekent het om te bereiken? Wanneer is het gebruik ervan succesvol? Hoe kan een dergelijk succes worden geëvalueerd en opnieuw worden gedefinieerd?

Deze discussie kan in verband worden gebracht met de filosofische strijd tussen legaliteit en rechtvaardigheid. Zoals rechtsfilosoof Hans Kelsen in de 'Reine Rechtslehre' van 1934 benadrukte, is het mogelijk om een rechtssysteem als puur positivistisch voor te stellen, gebaseerd op een strikte hiërarchie van toepasselijke wetten (Kelsen, 2008). Dit maakt schijnbaar heldere en transparante beslissingen mogelijk, waardoor lange beraadslagingen over ethiek, moraal, rationaliteit, bewijsvoering en rechtvaardigheid achterhaald zijn. Jean-Jacques Rousseau daarentegen eist dat het uiteindelijke doel van een rechtssysteem niet de 'correcte' toepassing van normen is. In zijn visie is het doel van het 'droit naturel' het bereiken van rechtvaardigheid. Op deze manier voldoen individuen en civiele autoriteiten aan hun morele verplichting ten opzichte van elkaar (Wokler, 2012, pp. 89-97).

Legaliteit is slechts een middel om gerechtigheid te bereiken, het is niet het einddoel van het rechtssysteem als zodanig. Met andere woorden, het zoeken naar gerechtigheid vereist ook een normatieve afweging van hoe de wereld eruit zou moeten zien. Een gezonde en rechtvaardige democratie hangt af van deze moeilijke, subjectieve en niet-kwantificeerbare overwegingen. Zelfs als we complexe algoritmen begrijpen, en nog meer als we dat niet doen, moeten we belangrijke beslissingen in het domein van veiligheid en inlichtingen niet alleen aan algoritmen overlaten. De verklaringen die algoritmes produceren, zijn uitsluitend gebaseerd op hoe de wereld (waarschijnlijk) is, gebaseerd op wat is geëxtrapoleerd uit historische gegevens. Dit zegt niets over wat een samenleving nastreeft. Algoritmen zijn van nature blind voor het stellen van zulke ambitieuze doelen. Als we eenmaal erkennen dat er een kloof bestaat tussen empirische analyse en normatieve rechtvaardigheid, is de essentiële vraag: hoe kan een systeem dat gebaseerd

Bestuurskunde 2020 (29) 4
doi: 10.5553/Bk/092733872020029004004

is op algoritmische beslissingen 'standaard legitimiteit' opleveren? Het is de taak van wetgevers en beleidsmakers om deze vraag te beantwoorden door middel van discours en op maat gemaakte effectbeoordelingen, niet van machines of degenen die deze voornamelijk ontwikkelen om producten en diensten te kunnen verkopen (Nemitz, 2018; Gstrein, 2019).

Conclusie

Hoe veelbelovend de geautomatiseerde besluitvorming op basis van geavanceerde algoritmen ook is, dergelijke systemen zullen niet onvermijdelijk alomtegenwoordig worden. Ze zullen hun toegevoegde waarde op lange termijn moeten bewijzen voor elk doel en elke toepassing. In de context van PP lijkt het erop dat conventionele benaderingen van het ontwerp meer dan vereenvoudigd zijn en gebaseerd op verkeerde en niet-aspiratorische aannames, wat ertoe leidt dat gemeenschappen in steden als Los Angeles eisen om ze op te geven (Ryan-Mosley & Strong, 2020). Dit geval is bijzonder opmerkelijk, aangezien PP voor het eerst werd ontwikkeld in Los Angeles en daar een relatief lange gebruiksgeschiedenis heeft. Daarnaast zijn soortgelijke op algoritmen gebaseerde technologieën, zoals gezichtsherkenning, recentelijk verboden in verschillende Amerikaanse steden (Lecher, 2019), en een nieuw rapport van de Canadese maatschappelijke organisatie Citizen Lab roept op tot het instellen van een moratorium op het gebruik van technologie die gebaseerd is op algoritmische verwerking van historische massapolitiegegevens door LEA's (Robertson, Khoo & Song, 2020).

In dit artikel hebben we de ethische, juridische en sociale uitdagingen gedeeld die verbonden zijn aan het gebruik van conventionele PP-systemen. We zien de noodzaak om de eisen aan dergelijke systemen te herdefiniëren, wat kan worden bereikt door interdisciplinaire en mensgerichte ontwerpprocessen toe te passen. Deze zijn in staat om rekening te houden met uiteenlopende standpunten en aspiraties van een breed scala aan belanghebbenden, historische feiten en maatschappelijke doelstellingen. Naar onze mening moet er een meer holistische discussie komen over algoritmische transparantie die verder gaat dan het fenomeen van de zwarte doos en die procedurele, praktische en organisatorische elementen omvat. Het is duidelijk dat de vraag naar transparantie wijst op een gebrek aan vertrouwen. Dit vertrouwen zou echter kunnen worden bereikt met een duidelijke focus op een op de mens gericht ontwerp. Bovendien vereist vertrouwen dat bedrijven die systemen ontwikkelen met behulp van algoritmen, volledig toetsbaar moeten zijn. Specifiek, alle software en AI in dienst van democratische staten moet volledig controleerbaar zijn in termen van hun functie en hun manieren om gegevens te beoordelen en er conclusies uit te trekken. In de loop van deze exercitie zou het duidelijk kunnen worden dat de focus van dergelijke systemen moet verschuiven van voorspelling naar preventie en het delen van informatie. Uiteindelijk, en mits zorgvuldig ontworpen, zouden algoritmen een betere besluitvorming kunnen ondersteunen op basis van een granulaire verzameling van feiten en een uitgebreide evaluatie en interpretatie, die in staat is

rekening te houden met de context. Transparante kunstmatige intelligentie die uit te leggen is en die rekening kan houden met hoe de wereld eruit zou moeten zien, is echter nog ver weg. De momenteel beschikbare systemen hebben tot nu toe hun betrouwbaarheid niet bewezen. Daarom zijn ze nog niet klaar om belangrijke actoren op het gebied van justitie en veiligheid te worden.

Literatuur

ADM Systems in the COVID-19 Pandemic: A European Perspective. (2020). AlgorithmWatch. https://algorithmwatch.org/en/project/automating-society-2020-covid19/

Amnesty International. (2020, September 29). *We sense trouble: Automated discrimination and mass surveillance in the Netherlands*. https://www.amnesty.org/en/documents/document/?indexNumber=eur35%2f2971%2f2020&language=en

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). *Machine bias*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=MJx_BFsEFMNT2bSeAG2YZISppjWRS64u

Automating Society – Taking Stock of Automated Decision-Making in the EU. (2019, January 29). AlgorithmWatch. https://algorithmwatch.org/en/automating-society/

Durnová, H., & Alberts, G. (2014). Was Algol 60 the first algorithmic language? *IEEE Annals of the History of Computing, 36*(4), 104-104. https://doi.org/10.1109/MAHC.2014.63

Furht, B., & Villanustre, F. (2016). Introduction to Big Data. In B. Furht & F. Villanustre (Eds.), *Big Data Technologies and Applications* (pp. 3-11). Springer International Publishing. https://doi.org/10.1007/978-3-319-44550-2_1

Gstrein, O.J. (2016, October 24). How to approach technology, human rights and personality in the digital age? – A few thoughts. *Privacy and Personality*. https://www.privacyandpersonality.org/2016/10/how-to-approach-technology-human-rights-and-personality-in-the-digital-age-a-few-thoughts/

Gstrein, O.J. (2019, November 13). Predictive Policing: Positivism is not enough to rule our complex world. *About:Intel*. https://aboutintel.eu/predictive-policing-complex-world/

Gstrein, O.J., Bunnik, A., & Zwitter, A. (2019). Ethical, legal and social challenges of Predictive Policing. *Católica Law Review, 3*(3), 77-98.

Gstrein, O.J., & Ritsema van Eck, G.J. (2018). Mobile devices as stigmatizing security sensors: The GDPR and a future of crowdsourced 'broken windows'. *International Data Privacy Law, 8*(1), 69-85. https://doi.org/10.1093/idpl/ipx024

Hardyns, W., & Rummens, A. (2018). Predictive Policing as a new tool for law enforcement? Recent developments and challenges. *European Journal on Criminal Policy and Research, 24*(3), 201-218. https://doi.org/10.1007/s10610-017-9361-2

Hu, M. (2020). Cambridge Analytica's black box. *Big Data & Society, 7*(2), 2053951720938091. https://doi.org/10.1177/2053951720938091

Kager, J. (2020, February 5). *Algoritmesysteem SyRI in de ban na uitspraak rechtbank*. https://www.fnv.nl/nieuwsbericht/algemeen-nieuws/2020/02/algoritmesysteem-syri-in-de-ban-na-uitspraak-recht

Kayser-Bril, N. (2020, July 22). *Swiss police automated crime predictions but has little to show for it*. AlgorithmWatch. https://algorithmwatch.org/en/story/swiss-predictive-policing/

Kelsen, H. (2008). *Reine Rechtslehre* (M. Jestaedt, Ed.). Mohr Siebeck. https://doi.org/10.1628/978-3-16-156465-9

Lecher, C. (2019, July 17). *Oakland city council votes to ban government use of facial recognition*. The Verge. https://www.theverge.com/2019/7/17/20697821/oakland-facial-recogntiion-ban-vote-governement-california

Meijer, A., & Wessels, M. (2019). Predictive Policing: Review of benefits and drawbacks. *International Journal of Public Administration, 42*(12), 1031-1039. https://doi.org/10.1080/01900692.2019.1575664

Ministerie van Algemene Zaken. (2016, 28 april). *Big Data in een vrije en veilige samenleving – Rapport – WRR*. Ministerie van Algemene Zaken. https://www.wrr.nl/publicaties/rapporten/2016/04/28/big-data-in-een-vrije-en-veilige-samenleving

Ministerie van Volksgezondheid, Welzijn en Sport. (2020, 2 april). *Ruim een kwart van de Nederlanders ervaart nog steeds discriminatie – Nieuwsbericht – Sociaal en Cultureel Planbureau*. Ministerie van Volksgezondheid, Welzijn en Sport. https://www.scp.nl/actueel/nieuws/2020/03/26/ruim-een-kwart-van-de-inwoners-van-nederland-ervaart-nog-altijd-discriminatie

Nemitz, P. (2018). Constitutional democracy and technology in the age of artificial intelligence. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376*(2133), 20180089. https://doi.org/10.1098/rsta.2018.0089

Pasquale, F. (2015). *The Black Box Society* (pp. 1-18). Harvard University Press; JSTOR. https://www.jstor.org/stable/j.ctt13x0hch.3

Perry, W.L., McInnis, B., Price, C.C., Smith, S., & Hollywood, J.S. (2013). *Predictive Policing: Forecasting crime for law enforcement*. https://www.rand.org/pubs/research_briefs/RB9735.html

Richardson, R., Schultz, J., & Crawford, K. (2019). *Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice* (SSRN Scholarly Paper ID 3333423). Social Science Research Network. https://papers.ssrn.com/abstract=3333423

Robertson, K., Khoo, C., & Song, Y. (2020, September 1). *To surveil and predict: A human rights analysis of algorithmic policing in Canada*. The Citizen Lab. https://citizenlab.ca/2020/09/to-surveil-and-predict-a-human-rights-analysis-of-algorithmic-policing-in-canada/

Ryan-Mosley, T., & Strong, J. (2020, June 5). The activist dismantling racist police algorithms. *MIT Technology Review*. https://www.technologyreview.com/2020/06/05/1002709/the-activist-dismantling-racist-police-algorithms/

Sommerer, L. (2020). *Personenbezogenes Predictive Policing*. Nomos.

Taylor, L., Sloot, B. van der, & Floridi, L. (2017). Conclusion: What do we know about group privacy? In L. Taylor, L. Floridi, & B. van der Sloot (Eds.), *Group privacy: New challenges of data technologies* (pp. 225-237). Springer International Publishing. https://doi.org/10.1007/978-3-319-46608-8_12

Ukrow, J. (2018). Data protection without frontiers? On the relationship between EU GDPR and Amended CoE Convention 108. *European Data Protection Law Review, 4*(2), 239-247. https://doi.org/10.21552/edpl/2018/2/14

Veale, M., & Brass, I. (2019). Administration by algorithm? Public management meets public sector machine learning. In *Algorithmic Regulation*, Yeung, K. & Lodge, M. (Eds.), Oxford University Press. http://oxford.universitypressscholarship.com/view/10.1093/oso/9780198838494.001.0001/oso-9780198838494-chapter-6

Weiser, S. (2019, April 3). *Building trust in human-centric AI*. FUTURIUM – European Commission. https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines

Wokler, R. (2012). Rousseau, the age of enlightenment, and their legacies. In *Rousseau, the Age of Enlightenment, and Their Legacies*, Garsten, B. (Ed.), Princeton University Press. http://www.degruyter.com/princetonup/view/title/507406

Završnik, A. (2019). Algorithmic justice: Algorithms and big data in criminal justice settings. *European Journal of Criminology*, 1477370819876762. https://doi.org/10.1177/1477370819876762

Zwitter, A. (2014). Big Data ethics. *Big Data & Society, 1*(2), 2053951714559253. https://doi.org/10.1177/2053951714559253

Zwitter, A., & Gstrein, O.J. (2020). Big data, privacy and COVID-19 – learning from humanitarian expertise in data protection. *Journal of International Humanitarian Action, 5*(1), 4, s41018-020-00072-00076. https://doi.org/10.1186/s41018-020-00072-6

Zwitter, A., & Hazenberg, J. (2020). Governance, blockchain, cyberspace: How technology implies normative power and regulation. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3660795