

BIG DATA BESLUITVORMINGSPROCESSEN EN SLUIPWEGEN VAN DISCRIMINATIE

Hans Lammerant, Peter Blok & Paul de Hert *

Samenvatting | Het gebruik van big data heeft een grote vlucht genomen in zowel de private als de publieke sector. In dit artikel wordt de verhouding tussen big-datatoepassingen en discriminatie onderzocht. Het inzetten van big-datatoepassingen kan discriminatie tegengaan, maar kan ook via verschillende sluipwegen discriminatie veroorzaken. Het artikel laat zien hoe de gelijke-behandelingswetgeving en het fundamentele non-discriminatierecht de problematische kanten van op big data gebaseerde besluitvorming kan opvangen.

Trefwoorden | Big data, geautomatiseerde besluitvorming, discriminatie, non-discriminatie, gelijke behandeling, gelijkheid.

Navigator van Wolters Kluwer | NTM-NJCMBull. 2018/01

1 Inleiding

Big data wordt gezien als een *disruptive technology*, dat wil zeggen een techniek die in vele maatschappelijke sectoren enorm waardevol kan zijn, maar ook dingen op zijn kop zet.¹ Big-datatoepassingen leiden er onder meer toe dat steeds vaker en fijnmaziger onderscheid wordt gemaakt tussen personen en dat bij besluitvorming over personen in toenemende mate wordt vertrouwd op de uitkomsten van big-data-analyses. In *NRC-Handelsblad* werd onlangs bijvoorbeeld een recruiter geciteerd die liet weten dat hij 'bijna blind' vertrouwt op het oordeel van zijn big-datasysteem over de geschiktheid van een kandidaat voor een bepaalde functie.²

Gegeven die impact van big data is het van belang na te gaan of de toepassing van die technologie de bescherming van fundamentele rechten aantast. Daarbij gaat de aandacht vaak uit naar het privacyrecht. Die focus op het privacyrecht is in zoverre terecht dat big data vaak *big personal data* zijn, dat wil zeggen gegevens over personen. Voor de bescherming van personen

■ Mr. H. Lammerant is als onderzoeker verbonden aan de Law, Science, Technology & Society onderzoeksgroep van de Vrije Universiteit Brussel; Prof. mr. P.H. Blok is hoogleraar octrooirecht en privacy aan de Universiteit Utrecht en raadsheer in het Gerechtshof Den Haag; Prof. dr. P. de Hert is als hoogleraar verbonden aan de vakgroep Interdisciplinaire Studies van het Recht en Criminologie van de Vrije Universiteit Brussel en universitair hoofddocent aan de Universiteit Tilburg.

1 Zie onder meer McKinsey Global Institute, *Big Data: The next frontier for innovation, competition and productivity*, 2011 (online publiek); Mayer-Schönberger, V. & K. Cukier, *Big Data. A revolution that will transform how we live, work and think*, Londen: John Murray Publishers 2013; OESO, *Data-driven innovation: Big Data for growth and well-being*, Parijs: OECD Publishing 2014 (online publiek); White House, *Big Data: Seizing opportunities, preserving values*, Washington, DC: Executive Office of the President 2014 (online publiek).

2 C. Don, 'Op sollicitatie bij een computer', *NRC Handelsblad* 27 juli 2017, p. 9.

op wie de gegevens betrekking hebben, is echter niet alleen het privacyrecht van belang.³ De verschillen in behandeling die het resultaat zijn van big-datatoepassingen kunnen ook botsen met de regels met betrekking tot discriminatie en gelijke behandeling. Wat bijvoorbeeld als blijkt dat het systeem waarop de recruiter uit het hiervoor genoemde voorbeeld bijna blind vertrouwt, personen van een bepaald ras aanzienlijk minder vaak als kandidaat voor een baan aanbeveelt dan anderen? Biedt het feit dat dit onderscheid wordt gemaakt op basis van een analyse van gegevens waarin ras geen rol speelt een voldoende rechtvaardiging voor dit onderscheid?

In dit artikel zullen dergelijke vragen worden beantwoord aan de hand van een onderzoek naar de verhouding tussen big data en het non-discriminatierecht. Daarbij zal eerst uiteen worden gezet wat big-datatechnieken zijn en hoe ze worden toegepast bij de besluitvorming (sectie 2) en hoe de toepassing daarvan bedoeld of onbedoeld kan leiden tot discriminatie (sectie 3). Vervolgens zal de toepassing van big-datatechnieken worden getoetst aan het discriminatieverbod. Daarbij zal eerst worden gekeken naar de specifieke regels van de gelijke-behandelingswetgeving (sectie 4) en vervolgens naar de meer algemene normen van het in de Grondwet en mensenrechtenverdragen verankerde non-discriminatierecht (sectie 5). Geëindigd wordt met een reflectie over de noodzaak van een plicht tot het tot stand brengen van *non-discrimination by design* (sectie 6) en conclusies (sectie 7).

2 Big data: gehanteerde technologie en toepassing bij besluitvorming

2.1 Drie ontwikkelingen die geleid hebben tot de huidige big-datatechnologie

Het verwerken van gegevens in grote hoeveelheden is niet nieuw.⁴ Reeds onder de Romeinse keizers werden geregeld volkstellingen georganiseerd. Napoleon creëerde rond 1800 een bevolkingsregister waarin onder meer gegevens over doop, huwelijk, belastingheffing, dienstplicht en overlijden werden vastgelegd en met de opkomst van de computer in de jaren zestig van de vorige eeuw heeft het verwerken van gegevens een grote vlucht genomen in vrijwel elke organisatie. In de loop der tijd zijn die gegevensverwerkingen steeds omvangrijker en complexer geworden. Die ontwikkeling is het afgelopen decennium in een stroomversnelling geraakt. Dat is vooral het gevolg van een samenloop van de volgende drie technologische ontwikkelingen, die elkaar onderling versterken.

Ten eerste, zijn de mogelijkheden om gegevens te verzamelen de afgelopen jaren exponentieel gegroeid.⁵ Zo is het door een verschuiving van veel handelingen en contacten van de offline-

3 Zie voor een overzicht van de diverse relevante rechtsgebieden, zoals het gegevensbeschermingsrecht, het mededingingsrecht, het telecomrecht en het intellectueel eigendomsrecht, onder meer P.H. Blok (red.), *Big data en het recht*, Den Haag: Sdu 2017.

4 Zie voor een overzicht de historische ontwikkeling van gegevensverzameling onder meer: T.J. Barnes, 'Big data, little history', *Dialogues in Human Geography* 2013, afl. 3, p. 297-302; F. Bongers, C.-J. Jager en R. te Velde, *Big Data in onderwijs en wetenschap: Inventarisatie en essays*, Utrecht: Dialogic 2015, p. 11-13 en Wetenschappelijke Raad voor het Regeringsbeleid, *Big Data in een vrije en veilige samenleving*, Den Haag: WRR 2016, p. 37-39.

5 Zie het overzicht in onder meer OECD 2014 (*supra* noot 1), p. 8-9; WRR 2016 (*supra* noot 4), p. 40-41; Executive Office of the President, President's Council of Advisors on Science and Technology, *Report to the President: Big*

wereld naar het internet mogelijk geworden al die handelingen en contacten eenvoudig vast te leggen. Waar het voorheen bijvoorbeeld een enorme opgave was om de vrienden- en kennissenkring van iemand in kaart te brengen en de wijzigingen daarin bij te houden, zijn de onderlinge verbanden tussen miljarden personen tegenwoordig relatief eenvoudig te vinden in de gegevensbestanden van sociale media. Daarnaast worden steeds meer producten en apparaten uitgerust met sensoren die hun omgeving en hun eigen functioneren vastleggen. Op die manier kunnen ook over de offline-wereld steeds gemakkelijker gegevens worden verzameld. Bovendien worden die producten en apparaten, zoals zelfrijdende auto's, horloges en medische apparaten, vaak aangesloten op het internet (*Internet of Things*). Daarmee komen de verzamelde gegevens beschikbaar voor allerlei toepassingen. Zo is er al een tandenborstel te koop die bijhoudt hoe de gebruiker poetst. Via een internetverbinding kan het apparaat worden gekoppeld aan een app op een telefoon, die de gebruiker feedback geeft over zijn poetsgedrag.⁶

Ten tweede, is de capaciteit om gegevens op te slaan en te verwerken gigantisch toegenomen.⁷ De zogeheten Wet van Moore, die in 1965 voorspelde dat het aantal componenten per chip, en daarmee het vermogen om gegevens te verwerken, elke anderhalf jaar zou verdubbelen is decennialang uitgekomen en blijkt nog altijd een redelijk goede indicatie van de groei van de rekenkracht van computers.⁸ Ook de mogelijkheden om gegevens op te slaan zijn exponentieel toegenomen. Naar verluidt verdubbelt de opslagcapaciteit elke drie jaar, de laatste jaren onder meer door de komst van datacenters waarin tienduizenden servers en harddisk-drivers staan.⁹ McKinsey heeft uitgerekend dat in 2010 al meer dan 1000 exabytes aan nieuwe gegevens werden opgeslagen, dat wil zeggen 1000×10^{18} bytes.¹⁰ Deskundigen gaan ervan uit dat de opslag aan het einde van dit decennium nog een factor 40 hoger zal zijn en dat de totale hoeveelheid opgeslagen data dan dus zal zijn gegroeid tot minstens 40 zettabytes, dat wil zeggen 40×10^{21} bytes.¹¹ Bovendien wordt die enorme opslagcapaciteit vaak via clouddiensten aangeboden. Daardoor staat de infrastructuur om gegevens op te slaan en te verwerken niet alleen ter beschikking aan grote organisaties en ondernemingen, maar ook aan mkb-bedrijven en particulieren. De kosten van gegevensopslag vormen ook steeds minder een belemmering. In 1998 kostte een disk om een gigabyte gegevens op te slaan nog circa 50 EUR. Sindsdien is de prijs jaarlijks verminderd met gemiddeld veertig procent en is inmiddels dus nagenoeg nihil.¹² Tegenwoordig bieden bedrijven vele gigabytes opslag dan ook gratis aan.

Ten derde, zijn de mogelijkheden om verschillende gegevens met elkaar in verband te brengen (*data fusion*) en daaruit waardevolle informatie te destilleren (*datamining*) de afgelopen jaren

Data and Privacy: A Technological Perspective, mei 2014, p. 30 (online publiek); B. Custers, 'Big data en big data-technologie', in Blok 2017 (*supra* noot 3), p. 24-26.

6 A. Potenza, 'This new Bluetooth-connected toothbrush brings a dentist into your bathroom', *The Verge* 9 juni 2016 (online publiek).

7 Zie het overzicht in onder meer OECD 2014 (*supra* noot 1), p. 9-11; Whitehouse 2014A (*supra* noot 1), p. 30-32; Custers 2017 (*supra* noot 5), p. 26-28.

8 Zie onder meer McKinsey 2011 (*supra* noot 1), p. 2; OECD 2014 (*supra* noot 1), p. 10.

9 White House 2014A (*supra* noot 1), p. 30.

10 McKinsey 2011 (*supra* noot 1), p. 19.

11 IDC, *The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East*, December 2012 (online publiek).

12 OECD 2014 (*supra* noot 1), p. 9.

sterk toegenomen.¹³ *Data fusion* ervoor zorgt dat ongestructureerde of heterogene data uit verschillende gegevensbronnen (bijvoorbeeld GPS-data, videobeelden, klantenbestanden en gegevens over gebruik van sociale media) worden omgezet in een gestructureerde homogene vorm, zodat de gegevens kunnen worden vergeleken en geanalyseerd. Op dat vlak zijn de afgelopen jaren vele technologische hobbels uit de weg geruimd.¹⁴ Ook op het gebied van *datamining* zijn grote sprongen gemaakt, vooral door de ontwikkeling van zelflerende systemen. Waar de techniek voorheen bijvoorbeeld werd gebruikt om een door de mens verondersteld verband te verifiëren aan de hand van gegevens, kunnen de systemen nu zelf patronen en verbanden ontdekken.¹⁵ Deze zogeheten datagedreven analyses kunnen onverwachte, waardevolle nieuwe informatie opleveren. Zo resulteerde de analyse van een enorme hoeveelheid biologische en medische gegevens in een veelbelovend medicijn tegen een bepaalde vorm van kanker.¹⁶

De combinatie van deze drie ontwikkelingen maakt dat vandaag gegevensverwerking in ten minste drie opzichten *big is*, te weten in omvang, variëteit en snelheid. Daarom wordt big data wel gekarakteriseerd aan de hand van drie V's: *volume*, *variety* en *velocity*.¹⁷ Steeds grotere hoeveelheden gegevens komen beschikbaar (*volume*). Vanwege de uiteenlopende herkomst van de gegevens kunnen de aard en het format van de gegevens sterk variëren (*variety*) en vanwege de alomtegenwoordigheid van gegevensverzamelingen worden data voortdurend geactualiseerd en toegevoegd (*velocity*). Het bijzondere, en in zekere zin revolutionaire aan big data is dat die drie V's in technologisch opzicht nauwelijks meer een belemmering vormen.

2.2 De toepassing van big data in de besluitvormingsfase én de daaraan voorafgaande trainingsfase

De hiervoor beschreven big-datatechnieken kunnen onder meer worden toegepast om een onderscheid te maken in de beoordeling, behandeling of benadering van verschillende individuen of groepen. Bekend zijn de toepassingen waarmee reclameboodschappen, zoals banners op een website, worden afgestemd op persoonlijke kenmerken van de bezoeker.¹⁸ Financiële instellingen benutten big data om bijvoorbeeld fraude op te sporen en kredietwaardigheid te beoordelen.¹⁹ Verzekeraars bieden autoverzekeringen aan waarbij de hoogte van de premie afhankelijk is van het rijgedrag van de verzekeringnemer²⁰ en, zoals in de inleiding al werd opgemerkt, hebben ook recruiters de waarde van big-datatechnieken ontdekt. Ze gebruiken de techniek om te bepalen welke kandidaat het meest geschikt is voor een vacature.

13 Zie het overzicht in onder meer McKinsey 2011 (*supra* noot 2), p. 27-36; White House 2014B (*supra* noot 6), p. 24-29; Custers 2017 (*supra* noot 5), p. 28-31.

14 Zie White House 2014A (*supra* noot 1), p. 25.

15 *Ibid.*

16 Zie J. Medeiros, 'The startup fighting cancer with AI', *Wired*, 22 maart 2016 (online publiek).

17 Zie onder meer White House 2014A (*supra* noot 2); OECD 2014 (*supra* noot 1), p. 11; WRR 2016 (*supra* noot 4), p. 126.

18 Zie voor voorbeelden onder meer McKinsey 2011 (*supra* noot 1), p. 67-68; OECD 2014 (*supra* noot 1), p. 15.

19 Zie onder meer Verbond van Verzekeraars, *Grip op data*, april 2016 (online publiek).

20 Zie bijvoorbeeld de Veilig Rijden Verzekering van ANWB: www.anwb.nl/verzekeringen/autoverzekering/veiligrijden.

Voor de toetsing van deze big-datatoepassing aan het non-discriminatierecht is het zinvol om een onderscheid te maken tussen twee fasen in de inzet van big-datatechnologie. Ten eerste is er de fase waarbij een model wordt toegepast bij de besluitvorming doordat automatisch bepaalde besluiten worden genomen of voorgesteld. Zo kan een computer standaard een bepaalde beslissing (zoals wel/geen lening) nemen als aan bepaalde gedefinieerde voorwaarden is voldaan of op grond van die variabelen een bepaalde waarde (zoals een credit score) berekenen die de basis vormt voor verdere besluitvorming. In zijn meest eenvoudige vorm (als P, dan Q) ziet zo'n besluitvormingsmodel of algoritme er bijvoorbeeld als volgt uit: als een persoon in een bepaalde wijk woont, krijgt die persoon geen lening. Big-datatoepassingen werken op dezelfde manier, met dit verschil dat er met veel meer variabelen rekening wordt gehouden dan één kenmerk. Er wordt doorgaans gewerkt met profielen, die zijn opgebouwd uit een waaier van variabelen. Dankzij big-datatoepassingen worden die profielen steeds uitgebreider en kunnen de consequenties die aan bepaalde (combinaties van) variabelen moeten worden verbonden, steeds nauwkeuriger worden berekend.

Big-datasytemen worden doorgaans ook gebruikt in de fase voorafgaand aan de besluitvorming, te weten bij het ontwikkelen van de besluitvormingsmodellen met behulp van algoritmes.²¹ In deze zogeheten trainingsfase gaat het systeem in één of meerdere datasets (de trainingsdata) op zoek naar correlaties tussen bepaalde gegevens, bijvoorbeeld naar gegevens die correleren met het al dan niet terugbetalen van een lening. Als het systeem vaststelt dat er een verband bestaat tussen enerzijds de schermgrootte van een televisie die iemand koopt en anderzijds de tijd die het duurt dat een lening voor een televisie wordt terugbetaald, kan de variabele 'schermgrootte' worden toegevoegd aan het profiel dat een elektronikawinkel gebruikt om te bepalen onder welke voorwaarden iemand een televisie op krediet kan kopen. Zelf-lerende systemen kunnen aldus automatisch de modellen genereren op basis waarvan vervolgens beslissingen worden genomen of geadviseerd.

In beide fasen wordt data gebruikt, maar wel andere. Bij het opstellen van het model wordt beschikbare data over een groot aantal gevallen of situaties gebruikt die eerder verzameld zijn, de trainingsdata. In de besluitvormingsfase wordt dit model dan toegepast op de specifieke gegevens over het individu op wie het te nemen besluit betrekking heeft.

3 Big data en negatieve discriminatie: wat zijn de sluiptwegen?

3.1 De positieve en negatieve effecten van big-datatoepassingen

De toepassing van big-datatechnieken bij het nemen van besluiten over personen vormt vanuit het oogpunt van gelijke behandeling niet alleen een risico, maar ook een kans om discriminatie te verminderen en gelijkheid te bevorderen. Daarbij kan het verhelderend zijn om een onderscheid

21 Zie V. M.I. Jordan, T.M. Mitchell, 'Machine learning : Trends, perspectives, and prospects', *Science*, 2015, nr. 6245, p. 255-260; LeCun, Y., Bengio, Y. en Hinton, G., 'Deep learning', *Nature* 2015, p. 436-44; S. Das, A. Dey, A. Pal en N. Roy, 'Applications of Artificial Intelligence in Machine Learning : Review and Prospect', *International Journal of Computer Applications* 2015, nr. 9, p. 31-41; Custers 2017 (*supra* noot 5), p. 22-24.

te maken tussen drie verschillende mechanismes van discriminatie: vooroordelen, rationele discriminatie en niet-intentionele discriminatie.²²

Vooroordelen betreffen ongerechtvaardigde negatieve opinies en houdingen ten opzichte van een bepaalde groep. Aan het louter behoren tot een bepaalde groep wordt dan een negatieve beoordeling gekoppeld. Zulke menselijke vooroordelen kunnen leiden tot discriminatie. De computersystemen waarmee big data worden verwerkt hebben deze vooroordelen in beginsel niet, althans niet als die niet bewust of onbewust opgenomen zijn in het gebruikte besluitvormingsmodel. Toepassing van die systemen kan dus voorkomen dat beslissingen bewust of onbewust worden genomen op basis van vooroordelen.

Rationele discriminatie daarentegen vertrekt niet van ongerechtvaardigde stereotypen, maar van rationeel denken en correcte gegevens, waaraan een discriminerende conclusie verbonden wordt voor wie tot een bepaalde groep behoort. Daarbij wordt een gebrek aan kennis over een individu gecompenseerd met het gebruik van eerder opgedane kennis over de gemiddelde eigenschappen of resultaten in een groep waartoe het individu behoort. Op zich is dergelijke rationele discriminatie gebaseerd op objectieve gronden en daarom in principe niet verboden. Wat in principe wel verboden is, is om bepaalde zeer algemene kenmerken als ras of geslacht als grond voor een beslissing in aanmerking te nemen. Als jonge vrouwen gedurende een tijd niet beschikbaar zijn voor een werkgever vanwege een zwangerschap heeft dat financiële gevolgen voor die werkgever. Ook zijn er objectieve verschillen tussen mannen en vrouwen wat betreft levensduur en het voorkomen van bepaalde ziektes. Het niet aannemen van jonge vrouwen door een werkgever omwille van die financiële gevolgen of het differentiëren van verzekeringspremies volgens geslacht is dan een discriminerende handeling die niet gebaseerd is op een vooroordeel, maar op een rationele afweging en objectieve gronden. Het non-discriminatie-recht verbiedt zulk een discriminerende behandeling ondanks het objectieve onderscheid toch vanwege van de specifieke grond die als differentiërende factor gebruikt wordt.

Ook in dit opzicht kan de toepassing van big data discriminatie verminderen. Het verzamelen van big data kan er immers toe leiden dat er een veel completer beeld over het individu bestaat, waardoor beslissingen kunnen worden genomen op basis van gegevens die beter recht doen aan het individu. Zo kan een persoon die in het verleden geen lening kreeg omdat de kredietinstelling geen gegevens had waaruit bleek dat die persoon kredietwaardig was, die lening nu wel krijgen omdat meer bronnen worden meegenomen.²³

Bovendien kunnen big-datatoepassingen actief worden ingezet om materiële ongelijkheid te verminderen. *Georgia State University* heeft bijvoorbeeld het slagingspercentage van minder kansrijke groepen aanzienlijk weten te verhogen door de vorderingen van studenten nauwgezet bij te houden en extra begeleiding te bieden waar dat nodig was.²⁴

Tegelijkertijd vergroot de opkomst van big data het risico op ongelijke behandeling, simpelweg omdat die technieken meer differentiatie mogelijk maken. Zo was het voor een verzekeraar tot recent niet zinvol om onderscheid te maken in de hoogte van premies, omdat verzekeraars

22 Voor een uitgebreide bibliografie over deze vormen van discriminatie, zie A. Romei en S. Ruggieri, 'Discrimination Data Analysis: A Multi-disciplinary Bibliography', in B. Custers, T. Calders, B. Schermer, & T. Zarsky (eds.), *Discrimination and privacy in the information society: Data mining and profiling in large databases*, Berlijn/Heidelberg: Springer-Verlag 2013, p. 110-111.

23 Zie White House, Executive Office of the President, *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*, mei 2016, p. 11-12 (online publiek).

24 White House, 2016 (*supra* noot 23), p. 17.

niet of nauwelijks inzicht hadden in de risico's op het niveau van individuen. De opkomst van big data verandert dat²⁵ en zal tot gevolg hebben dat in de toekomst vaker onderscheid zal worden gemaakt tussen verzekeringsnemers, bijvoorbeeld op grond van allerlei gegevens die ze zelf aanleveren via apps (over rijgedrag) en wearables.

Die nieuwe vormen van onderscheid kunnen problematisch worden vanuit het oogpunt van gelijke behandeling vanwege een groeiend risico op niet-intentionele discriminatie. Niet-intentionele discriminatie treedt op wanneer er geen intentie is om te discrimineren op basis van de verboden gronden, maar men bij gebrek aan inzicht in de gevolgen praktijken gebruikt die leiden tot discriminatie. Zoals in de volgende paragraaf uiteen zal worden gezet, kunnen de gebruikte gegevens of de daarop gebaseerde besluitvormingsmodellen een niet-gekende bias ten aanzien van bepaalde groepen bevatten. Het computersysteem past dan in de besluitvormingsfase het model wel neutraal en objectief toe zonder zich te laten leiden door vooroordelen, maar zulke vooroordelen kunnen wel niet-intentioneel tijdens de eerdere opbouw van het model in dit model zijn geslopen.

3.2 Fouten en aannames bij big-datapraktijken waardoor discriminatie binnensluipt

Keuzes bij het opstellen van een besluitvormingsmodel en de in dat verband gebruikte data-analyses kunnen op verschillende manieren leiden tot discriminatie. Meestal gaat het om vormen van niet-intentionele discriminatie, maar het technologisch ondersteund beslisproces kan ook gebruikt worden om een vorm van bewuste discriminatie te verhullen. We kunnen de sluiptwegen van mogelijke discriminatie, bewust of onbewust, in dit proces nagaan op drie niveaus, te weten op het niveau van het besluitvormingsmodel, het niveau van de representativiteit van de trainingsdata op basis waarvan het model wordt opgesteld, en het niveau van de kwaliteit van die trainingsdata, in de zin dat ze ongewenste bias kunnen bevatten.²⁶ De eerste sluiptweg betreft de vormgeving van het model, met name de keuze van variabelen waartussen we de relaties proberen te verklaren en eventuele vooronderstellingen over deze relaties. De tweede en derde sluiptweg hebben betrekking op de data op basis waarvan dit model uitgewerkt wordt. In de tweede sluiptweg is het de bedoeling om een bepaalde realiteit in het model te reproduceren en stelt zich de vraag of de gebruikte data daarvoor representatief zijn. In de derde sluiptweg stelt zich de vraag of de realiteit weerspiegeld in de data wel de gewenste is en al dan niet ongewijzigd gereproduceerd mag worden.

In het volgende gaan we dieper in op deze drie sluiptwegen of niveaus van mogelijke discriminatie in het big-dataproces.

25 Zie Rathenau Instituut, *Berekende risico's. Verzekeren in de datagedreven samenleving*, Den Haag 2015 (online publiek); Verbond van Verzekeraars 2016 (*supra* noot 19).

26 T. Calders & I. Pliobaitė, 'Why Unbiased Computational Processes Can Lead to Discriminative Decision Procedures', in: B. Custers, T. Calders, B. Schermer & T. Zarsky, *Discrimination and Privacy in the Information Society*, Berlijn/Heidelberg: Springer-Verlag 2013; S. Barocas & A.D. Selbst, 'Big Data's Disparate Impact', *California Law Review* 2016, p. 671-732.

3.3 Twee mogelijke fouten in het gekozen big-databesluitvormingsmodel (sluipweg 1)

Het model moet een relatie leggen tussen een probleem- of vraagstelling, uitgedrukt in één of meerdere te verklaren doelvariabelen (betaalt iemand zijn lening terug?), en een reeks predictoren of verklarende variabelen (in welke wijk woont iemand?). Hierbij is de vraag of het besluitvormingsmodel adequaat de onderliggende realiteit weerspiegelt. Een weinig adequate conceptuele vormgeving van het model, of het selecteren van de variabelen die met elkaar in verband worden gebracht, kan leiden tot discriminatie. Een eerste bron van discriminatie ligt bij de te verklaren doelvariabele, een tweede bij de predictoren.

Ten eerste, kan er discriminatie ontstaan door de omschrijving van de problemen waarvoor een oplossing wordt gezocht. Aan de *data miner* of data-analist komt de moeilijke taak toe om dit probleem te vertalen in een meer formele vraagstelling waar computers mee aan de slag kunnen, onder meer door het bepalen van de te modelleren doelvariabelen. Voor zo'n te verklaren doelvariabele moet ook een serie mogelijke waarden of 'labels' die deze variabele kan hebben, worden vastgelegd. Die labels kunnen een aantal discrete waarden of klassen zijn (bijvoorbeeld kredietwaardig of niet) of een te schatten waarde (bijvoorbeeld een risicoscore). Deze vertaalslag van een probleem naar concrete te modelleren variabelen en de waarden die die variabelen kunnen krijgen, kan resulteren in een beslissysteem dat negatief uitdraait voor bepaalde groepen. Definities van doelvariabelen – met andere woorden wat is een 'goed' resultaat of wat is een 'risico' – kunnen subjectief zijn of worden beïnvloed door andere factoren, zoals de noodzaak een probleem te vertalen in meetbare en in de database vindbare criteria.²⁷

Een voorbeeld vormen diploma's of academische graden. Deze kunnen wijzen op bepaalde competenties, maar diezelfde competenties kunnen ook op andere manieren worden verkregen, zoals ervaring of zelfstudie. Het meetbare criterium van een diploma is bijgevolg niet zaligmakend en kan andere sociale factoren weerspiegelen, zoals het hebben van de economische middelen om te betalen voor een academische studie. Bovendien kan het hebben van een diploma variëren tussen leeftijdsgroepen of geografische herkomst, afhankelijk van hoe toegankelijk het formele academische of hoger onderwijs was of is. Het hebben van een diploma kan gebruikt worden om competenties te meten, maar onder omstandigheden kan er dus een bias optreden met discriminerende effecten als resultaat. Een ander voorbeeld is het gebruik van gegevens over de lengte van aanstellingen als graadmeter voor het nemen van beslissingen over het aannemen van nieuw personeel. In dat geval wordt het 'goede resultaat' gedefinieerd als een werknemer met een langdurige aanstelling en wordt vervolgens onderzocht welke kenmerken voorspellen of een nieuwe werknemer al dan niet lang zal blijven. Wanneer de gegevens weerspiegelen dat vrouwen kortere aanstellingen hebben als gevolg van zwangerschappen of zorg voor kinderen, kan dit criterium leiden tot een onbedoelde maar indirecte discriminatie van vrouwen. Het probleem kan ook opduiken zonder dat big-datatechnieken worden gehanteerd, maar er is een niet geringe kans dat bij het gebruik van technologisch ondersteunde beslisprocessen de discriminatie heviger en minder zichtbaar wordt.²⁸

Ten tweede, kan er discriminatie ontstaan op het niveau van het model door de keuze van de input-variabelen of predictoren. Welke de voorspellende variabelen kunnen zijn, wordt aangereikt

²⁷ Barocas & Selbst 2013 (*supra* noot 26), p. 671-732.

²⁸ Barocas & Selbst 2013 (*supra* noot 26), p. 671-732.

via de beschikbare attributen in de dataset. In de klassieke statistische vraagstelling wordt op voorhand een model vastgelegd in een statistische hypothese, die vervolgens op de data getest wordt. Hier wordt dus een expliciete keuze gemaakt van relevante attributen door de data-analist. Bij datamining beschikt men over een grote reeks mogelijke variabelen aanwezig in de dataset. Welke variabelen relevant zijn en als predictor in het model worden opgenomen, wordt nu geselecteerd via het datamining-algoritme. Dit datamining-algoritme stelt dus tegelijk de hypothesen op en selecteert welke het meest geschikt zijn om als model te functioneren. Bij sommige algoritmes wordt het ondoorzichtig welke variabelen effectief gebruikt worden in een model of hoe het algoritme tot een bepaalde selectie van predictoren heeft besloten. Een slordige selectie van attributen te gebruiken als input-variabelen, ook wel 'feature-selectie', kan discriminerend werken voor een groep, met name als factoren die verantwoordelijk zijn voor statistische variatie binnen deze groepen hetzij ontbreken of te grofmazig worden binnengebracht. In beginsel houdt elke representatie door een aantal attributen of kenmerken een reductie in van een veel complexere realiteit. Het is immers onmogelijk om met alle factoren rekening te houden. Het terugbrengen van een verschijnsel tot een reeks kenmerken, zelfs een heel uitgebreide, is dus altijd een vorm van stereotypering. Een voorstelling kan daarom falen in het onderscheidingsvermogen en niet voldoende in detail gaan bij het onderscheiden van uitkomsten met betrekking tot bepaalde groepen. Toegeschreven kenmerken zijn dan niet meer dan 'coarse proxies' of grove benaderingen van bepaalde fenomenen en kunnen discriminatie in de hand werken.²⁹

De controversiële praktijk van *redlining* is gebaseerd op een dergelijke grove proxy.³⁰ Financiële instellingen gebruiken postcodes of geografische criteria als een gemakkelijk beschikbare indicatie voor kredietwaardigheid en weigeren leningen of andere diensten aan mensen in bepaalde wijken. De postcode wordt dan een indicator over het inkomen of het vermogen of het vermogen terug te betalen. Het is een goedkope, maar zeer grove indicator, zeker in vergelijking met een individueel zorgvuldig onderzoek naar de kredietwaardigheid van een persoon. In de literatuur worden nog meer voorbeelden gegeven van discriminerende praktijken gebaseerd op het gebruik van dergelijke proxies.³¹

Grove proxies zijn dus een vorm van rationele discriminatie, verborgen achter één of meerdere attributen die gecorreleerd zijn met de verboden grond. Bij data mining kunnen niet-intentioneel zulke met een verboden grond gecorreleerde attributen in het besluitvormingsmodel opgenomen worden en tot discriminatie leiden.

3.4 Twee onjuiste veronderstellingen over de geschiktheid van vertrek- of trainingsdata (sluipweg 2)

De relaties tussen de variabelen in het model worden vervolgens bepaald op basis van de beschikbare gegevens of trainingsdata. Die trainingsdata bevatten eerder gedane observaties of metingen van bepaalde variabelen over een grote reeks gevallen, die samen geacht worden een goede weergave van de onderliggende realiteit te zijn.

29 Calders & Pliobaitė 2013 (*supra* noot 26); Barocas & Selbst 2013 (*supra* noot 26), p. 671-732.

30 Barocas & Selbst 2013 (*supra* noot 26), p. 671-732.

31 Barocas & Selbst 2013 (*supra* noot 26), p. 671-732.

De vraag is natuurlijk of dit wel klopt en of de gekozen trainingsdata adequaat die realiteit weerspiegelen. Het uitwerken van een model op basis van data steunt op twee vooronderstellingen met betrekking tot die data. Ten eerste dat de kenmerken van de populatie waarop het model gebouwd wordt, dezelfde blijven bij de populatie waarop het model in de toekomst zal worden toegepast. Ten tweede dat de initiële gegevens of trainingsdata een betrouwbare weergave van deze populatie geven.³² Wanneer deze vooronderstellingen niet blijken te kloppen, kunnen onbedoelde discriminerende effecten optreden door een model gesteund op een specifieke realiteit op een populatie buiten deze specifieke context toe te passen. Bekijken we achtereenvolgens beide aannames kritisch.

Ten eerste, de vooronderstelling, dat populaties niet veranderen. Deze houdt in onze ogen geen stand wanneer relevante factoren veranderen ten opzichte van de trainingsdata. Denk bijvoorbeeld aan veranderende economische (bijvoorbeeld inkomensniveaus en werkgelegenheidsgraad in tijden van crisis versus betere economische tijden) of andere omstandigheden (bijvoorbeeld veranderingen in de beschikbaarheid van technologie, zoals smartphones). Deze veranderingen worden niet weerspiegeld in het model, waardoor het een steeds meer van de realiteit afwijkende beschrijving vormt. Modellen zijn dus maar zolang houdbaar als de data representatief blijven en het gebruik van verouderde modellen kan leiden tot een bias ten aanzien van bepaalde groepen.³³

Een voorbeeld daarvan is dat de veranderingen in communicatietechnologie, van vaste telefoons naar mobiele telefoons en communicatie via het internet, verschillend lopen naargelang leeftijd of bevolkingsgroep. Hierdoor wordt het bijvoorbeeld voor opiniepeilers moeilijker om een representatieve selectie uit de bevolking te maken en te bevragen. Zij moeten daarom selectiemethodes aanpassen om toch tot een betrouwbare opiniepeiling te komen.³⁴

Ten tweede, is er de vooronderstelling dat de initiële gegevens een betrouwbare weergave van deze populatie geven. Dit beeld kan evenwel vertekend zijn als de genomen steekproef een inaccuraat weerspiegeling is van de onderliggende werkelijkheid. Bij een inaccuraat steekproef kunnen bepaalde delen van de gekozen populatie onder- of oververtegenwoordigd zijn in de gegevens, wat de resultaten van het beslismodel negatief kan beïnvloeden.³⁵

Dit laatste kan het onbedoelde gevolg zijn van de gekozen wijze van verzamelen van gegevens. Zo wordt bij bepaalde crisisinterventies, bijvoorbeeld bij natuurrampen, gebruikgemaakt van Twitter-data. Hierdoor kan een vertekend beeld ontstaan over een populatie omdat teveel wordt uitgegaan van mensen met internet of smartphones. De kansarme bevolking die niet over deze instrumenten beschikt, komt dan niet in beeld.³⁶ Een ander voorbeeld is dat van *Street Bump* in Boston, een programma dat de wegkwaliteit meet via gegevens over plotselinge veranderingen gesignaleerd door middel van versnellingsmeters in smartphones. Dergelijke ogenschijnlijk slimme initiatieven geven een onvolledig beeld vanwege de ondervertegenwoordig-

32 Calders & Pliobaitë 2013 (*supra* noot 26).

33 Calders & Pliobaitë 2013 (*supra* noot 26).

34 A. Mohorko, E., de Leeuw, & J. Hox, 'Coverage Bias in European Telephone Surveys: Developments of Landline and Mobile Phone Coverage across Countries and over Time', *Survey Methods: Insights from the Field*, 2013 (online publiek).

35 Calders & Pliobaitë 2013 (*supra* noot 26).

36 G. Vega-Gorgojo et al., *Case Study Reports*, BYTE Deliverable D3.2, 2015.

ging van arme buurten, waar minder auto's en smartphones aanwezig zijn.³⁷ Soms is een bias in de data het rechtstreekse gevolg van bestaande discriminerende praktijken, die aan de basis van de dataverzameling liggen. Bijvoorbeeld raciale profilering door de politie kan leiden tot een oververtegenwoordiging van bepaalde bevolkingsgroepen in strafzaken en veiligheidsstatistieken.

De tweede vooronderstelling is ook niet vervuld wanneer de attributen of mogelijke voorspellende variabelen in de dataset weinig geschikt zijn om de populatie of een deel ervan op een relevante wijze te beschrijven. Dit hebben we reeds in de vorige sluiptegenwoordiging besproken.

3.5 De data bevat ongewenste informatie (sluiptegenwoordiging 3)

De trainingsdata voor het model kan ook een weerspiegeling zijn van praktijken waarvan we nu afstand nemen, zoals vroegere discriminerende praktijken of vooroordelen die aanwezig blijven in de dagelijkse praktijk. Die praktijken worden gereproduceerd als data ongewijzigd worden gebruikt als basis voor een besluitvormingsmodel. In dit geval vormt de data wel een adequate representatie van de vroegere realiteit, maar willen we daar bewust van afwijken. De ongewenste realiteit kan, zoals besproken in de vorige paragraaf, weerspiegeld zijn in de steekproef waardoor bepaalde groepen over- of onder-gerepresenteerd zijn. Te denken valt aan de praktijk van etnische profilering door de politie, die zorgt voor een over-representatie van bepaalde bevolkingsgroepen in data met betrekking tot criminaliteit.

Maar de ongewenste discriminatie kan ook ingang vinden via de *labeling* van data (het benoemen van de uitkomsten). In voorkomend geval zit het probleem niet in de methode van dataverzameling, maar schuilt de discriminatie in de gebruikte labels (de uitkomsten voor de doelvariabele). Er is geen probleem wanneer de labels een weergave zijn van objectieve metingen en waarnemingen. Zelfs wanneer er een bias is in de verzameling van de gegevens kunnen de labels correct zijn. Dat geldt bijvoorbeeld voor de uitkomst van een alcoholtest, ook al werd die uitgevoerd op grond van etnische profilering bij het verzamelen van gegevens. Maar wanneer het plakken van labels op een uitkomst steunt op een subjectieve menselijke beoordeling, is het duidelijk dat vooroordelen kunnen binnensluipen. Zo zullen gegevens over aanstellingen en toelatingen uit het verleden ook iets vertellen over de toentertijd gehanteerde vooroordelen van hen die betrokken waren bij deze procedures.³⁸

Dit onderscheid tussen verschillende bronnen van partijdigheid of bias in de data is belangrijk omdat elk een andere aanpak vereist om ze te corrigeren. Over- of ondervertegenwoordiging van bepaalde populaties kan worden gecorrigeerd door het gebruik van wegingsfactoren. De impact van een ondervertegenwoordigde groep wordt dan vergroot door haar gewicht naar verhouding te vergroten en vice versa. Vooroordelen binnengeslopen in de labels zelf zijn minder eenvoudig vast te stellen en te corrigeren.³⁹

37 K. Crawford, 'Think Again: Big Data', *Foreign Policy*, 10 mei 2013 (online publiek); Barocas & Selbst 2013 (*supra* noot 26), p. 671-732.

38 Calders & Pliobaitė 2013 (*supra* noot 26).

39 Calders & Pliobaitė 2013 (*supra* noot 26).

4 Big data in het licht van de non-discriminatiewetgeving

4.1 Algemeen

Het non-discriminatierecht kan worden onderverdeeld in enerzijds het algemene gelijkheidsbeginsel, zoals opgenomen in diverse mensenrechtencatalogi, en anderzijds de specifieke wettelijke regels op het gebied van gelijke behandeling. Omdat de laatstgenoemde regels een concrete invulling vormen van het algemene gelijkheidsbeginsel is het in de praktijk nuttig eerst te toetsen of die regels van toepassing zijn, en zo ja, of aan die regels is voldaan, alvorens na te gaan of het algemene gelijkheidsbeginsel is geschonden. Daarom zullen hierna eerst de gelijke-behandelingswetten worden besproken.

4.2 Het toepassingsbereik van de AWGB en andere gelijke-behandelingswetten

Bij de beantwoording van de vraag of een bepaalde big-datapraktijk in overeenstemming is met het non-discriminatierecht moet ten eerste worden beoordeeld of de specifieke wetgeving op het gebied van gelijke behandeling van toepassing is. Als dat zo is, moet de big-datatoepassing immers worden getoetst aan het vrij restrictieve regime van die wettelijke regels. Het gaat in dit verband om de Algemene Wet Gelijke Behandeling (AWGB), de Wet gelijke behandeling op grond van leeftijd (WGBL), de Wet gelijke behandeling op grond van handicap of chronische ziekte (WGBH/CZ), de Wet gelijke behandeling van mannen en vrouwen (WGB), de Wet onderscheid arbeidsduur (WOA) en de Wet onderscheid bepaalde en onbepaalde tijd (WOBOT).

Deze wetten implementeren diverse Europese richtlijnen.⁴⁰ Daarom moet bij de uitleg van de genoemde wetten rekening worden gehouden met de rechtspraak van het Hof van Justitie over die richtlijnen.

Het toepassingsbereik van de genoemde wetten hangt af van zowel de grond (bijvoorbeeld ras of leeftijd) als het gebied (bijvoorbeeld arbeid of de levering van goederen/diensten). Zo zal big data in de private sector vaak worden gebruikt in het kader van de levering van goederen en diensten, zoals verzekeringen en kredieten.⁴¹ Op dat terrein is de AWGB van toepassing, maar gelden de overige gelijkebehandelingswetten niet. Big data kunnen ook in arbeidsrechtelijke verhoudingen een rol spelen. In dat geval zijn alle genoemde wetten van toepassing.

Verder is van belang dat de wettelijke regels niet van toepassing zijn als de gegevensverwerking niet leidt tot een verschil in behandeling van personen. Een verzekeringsbedrijf kan bijvoorbeeld gegevens over geslacht verwerken in modellen om haar risicoblootstelling op bedrijfsniveau in te schatten of om gemiddelde verzekeringspremies te berekenen. De regels

40 Richtlijn 2000/43/EG van de Raad van 29 juni 2000 houdende toepassing van het beginsel van gelijke behandeling van personen ongeacht ras of etnische afstamming; Richtlijn 2000/78/EG van de Raad van 27 november 2000 tot instelling van een algemeen kader voor gelijke behandeling in arbeid en beroep; Richtlijn 2004/113/EG van de Raad van 13 december 2004 houdende toepassing van het beginsel van gelijke behandeling van mannen en vrouwen bij de toegang tot en het aanbod van goederen en diensten; Richtlijn 2006/54/EG van het Europees Parlement en de Raad van 5 juli 2006 betreffende de toepassing van het beginsel van gelijke kansen en gelijke behandeling van mannen en vrouwen in arbeid en beroep.

41 McKinsey 2011 (*supra* noot 1), p. 64.

inzake gelijke behandeling worden pas van toepassing wanneer er sprake is van differentiatie op individueel niveau.⁴²

4.3 Directe discriminatie bij big data door gebruik van verboden gronden

Als een van deze gelijkebehandelingswetten van toepassing is, moet ten tweede worden beoordeeld of er sprake is van directe discriminatie. Van directe discriminatie is sprake wanneer iemand op basis van een van de ‘verboden gronden’ die onder de betreffende wet vallen, ongunstiger wordt behandeld dan een ander in een vergelijkbare situatie wordt, is of zou worden behandeld.⁴³ Bij een big-datatoepassing zal directe discriminatie zich kunnen voordoen als een van die verboden gronden is opgenomen in het model dat wordt gebruikt in de fase van besluitvorming, bijvoorbeeld als ras is opgenomen in het profiel op basis waarvan het algoritme bepaalt of iemand in aanmerking komt voor een lening. Het bijzondere aan big-datatoepassingen is echter dat het altijd gaat om een veelheid aan factoren. Een beslissing zal dus nooit worden gebaseerd enkel op één specifiek kenmerk, maar op basis van een veelheid van variabelen, waaronder mogelijk een verboden grond. Dat brengt mee dat het bij big-datatoepassingen voor een slachtoffer van discriminatie bijzonder lastig zal zijn om – zonder kennisneming van het model – directe discriminatie aan te tonen. Daarvoor moet immers niet alleen komen vast te staan dat een negatieve beslissing is genomen, maar ook dat die beslissing anders zou zijn uitgevallen voor iemand die zich, afgezien van de verboden grond, in eenzelfde situatie bevindt.⁴⁴ Strikt genomen zou dat betekenen dat een voorbeeld moet worden gevonden van een persoon van wie alle in het profiel opgenomen kenmerken identiek zijn, met uitzondering van de verboden grond.

Daarbij moet wel worden bedacht dat van het bedrijf of de organisatie die het beslismodel toepast, kan worden verlangd te bewijzen dat de verboden grond *niet* in het beslismodel voorkomt zodra er een vermoeden van discriminatie is. De gelijke-behandelingswetten schrijven namelijk voor dat de verweerder moet aantonen dat het beginsel van gelijke behandeling niet werd geschonden als de klager feiten kan aanvoeren die discriminatie doen vermoeden.⁴⁵ Als een bedrijf of organisatie geen inzage kan of wil geven in de werking van het model, bijvoorbeeld vanwege de concurrentiegevoeligheid daarvan, bestaat voor dat bedrijf of die organisatie dus het risico dat directe discriminatie wordt aangenomen, ook als er geen verboden gronden in het beslismodel staan. Aldus kan een ondoorzichtig beleid ertoe leiden dat directe discriminatie wordt aangenomen.⁴⁶

42 Zie bijvoorbeeld Europese Commissie, Richtsnoeren betreffende de toepassing van Richtlijn 2004/113/EG van de Raad op verzekeringen, in het licht van het arrest van het Hof van Justitie van de Europese Unie in zaak C-236/09 (*Test-Achats*), 13 januari 2012.

43 Zie bijvoorbeeld de definitie van het begrip direct onderscheid in artikel 1 lid 1 sub b AWGB.

44 Zie o.m. European Union Agency for Fundamental Rights, *Handbook on European non-discrimination law*, Luxemburg: Imprimerie Centrale 2010, paragraaf 2.2.2; HvJ EU 13 januari 2004, C-256/01, ECLI:EU:C:2004:18 (*Allonby*).

45 Zie bijvoorbeeld artikel 10 lid 1 AWGB; zie ook FRA 2010 (*supra* noot 44), paragraaf 5.3; vgl. HR 10 december 1982, NJ 1983/687; EHRM 13 november 2007, 57325/00, ECLI:CE:ECHR:2007:1113JUD005732500 (*D.H./Tsjechië*).

46 HvJ EU 10 oktober 1989, C-109/88, ECLI:EU:C:1989:383.

4.4 Mogelijke rechtvaardiging van directe discriminatie

Als er sprake is van directe discriminatie moet vervolgens worden beoordeeld of het onderscheid valt onder een van de wettelijke rechtvaardigingsgronden. Omdat die rechtvaardigingsgronden beperkt zijn, is directe discriminatie in het algemeen moeilijk te rechtvaardigen, zelfs als er een objectieve, door zorgvuldige big-data-analyses ondersteunde basis bestaat voor het onderscheid. Zo voorziet de AWGB in een uitzondering op het verbod van ongelijke behandeling gebaseerd op geslacht wanneer het geslacht bepalend is of als het de bescherming van de vrouw betreft (artikel 2 lid 2 sub a AWGB). Die uitzonderingsbepaling is limitatief uitgewerkt in artikel 1 van het Besluit gelijke behandeling.⁴⁷ Het bepalen van de hoogte van de premie van een autoverzekering valt bijvoorbeeld niet onder die uitzonderingen. De AWGB staat dus niet toe dat geslacht wordt opgenomen in een model voor het nemen van besluiten over de hoogte van de premie van een autoverzekering. Een door big data gevoede analyse die objectief aantoont dat vrouwen echt minder vaak ongelukken veroorzaken, rechtvaardigt dat onderscheid niet.

Bij leeftijdsdiscriminatie ligt dat anders, omdat in dat kader een ruimere rechtvaardigingsgrond beschikbaar is. Die luidt hetzelfde als de rechtvaardigingsgrond voor indirecte discriminatie, die hierna zal worden besproken: onderscheid op basis van leeftijd mag als het objectief gerechtvaardigd wordt door een legitiem doel en de middelen voor het bereiken van dat doel passend en noodzakelijk zijn (artikel 7 lid 1 sub b WGBL). Uit rechtspraak van het Hof van Justitie over de Europese richtlijn waarvan de WGBL de implementatie vormt, blijkt dat een leeftijdsvereiste bij fysiek zware beroepen als brandweerman, piloot of politieman op die grond in principe gerechtvaardigd kan zijn.⁴⁸ De rechtvaardiging moet wel met objectieve gegevens worden verantwoord, net als het niet voorhanden zijn van minder ingrijpende maatregelen (bijvoorbeeld fysieke tests). Big-data-analyses die aantonen dat er daadwerkelijk een correlatie bestaat tussen de geschiktheid voor een beroep en leeftijd kunnen dan wel behulpzaam zijn, mits die analyses zelf geen bias bevatten.

4.5 Indirecte big data discriminatie (benadeling ondanks ogenschijnlijke neutraliteit)

Als er geen directe discriminatie is, moet worden getoetst of er sprake is van indirecte discriminatie. Indirecte discriminatie doet zich voor wanneer een ogenschijnlijk neutrale bepaling, maatstaf of handelwijze personen met een van de specifieke kenmerken in vergelijking met andere personen bijzonder benadeelt.⁴⁹ Daarvan is sprake als personen met een 'verboden kenmerk' in overwegende of disproportionele mate deel uitmaken van de groep van benadeelde personen. Daarbij wordt zowel gelet op de mate van oververtegenwoordiging van die groep als op de

47 Besluit van 18 augustus 1994, houdende nadere omschrijving van gevallen waarin het geslacht bepalend is, van gevallen waarin het de bescherming van de vrouw betreft, van gevallen waarin uiterlijke kenmerken die samenhangen met het ras van een persoon bepalend zijn en van gevallen waarin de nationaliteit bepalend is, *Stb.* 1994, 657, zoals laatstelijk gewijzigd bij *Stb.* 2012, 565.

48 HvJ EU 12 januari 2010, C-229/08, ECLI:EU:C:2010:3 (*Wolf*); HvJ EU 13 september 2011, C-447/09, ECLI:EU:C:2011:573 (*Prigge e.a.*); HvJ EU 13 november 2014, C-416/13, ECLI:EU:C:2014:2371 (*Vital Pérez*).

49 Zie bijvoorbeeld artikel 1 lid 1 sub c AWGB.

continuïteit van een verschil in behandeling.⁵⁰ Bij een big-datatoepassing kan indirecte discriminatie optreden als er weliswaar geen verboden gronden in het beslismodel zijn opgenomen, maar er een correlatie bestaat tussen de wel in het model gebruikte variabelen en de verboden gronden. Zo heeft onderzoek uitgewezen dat er een sterk verband bestaat tussen de *likes* die iemand geeft op facebook en onder meer het ras (95% nauwkeurigheid), het geslacht (93% nauwkeurigheid) en de seksuele voorkeur (88% nauwkeurigheid) van die persoon.⁵¹ Een onderscheid op basis van iemands *likes* kan dus meebrengen dat indirect onderscheid wordt gemaakt op basis van ras, geslacht en seksuele voorkeur.

Ook voor indirecte discriminatie geldt dat in beginsel het slachtoffer de discriminatie moet bewijzen, maar dat die bewijslast omdraait als het slachtoffer een vermoeden van discriminatie hard kan maken.⁵² Dat vermoeden kan bijvoorbeeld worden gebaseerd op statistische gegevens die aantonen dat het beslismodel dat een bedrijf of organisatie hanteert, resulteert in een disproportionele uitsluiting van groepen met een van de beschermde kenmerken, zoals een bepaald geslacht of ras.

4.6 Mogelijke rechtvaardiging voor indirecte discriminatie: objectief, legitiem en passend en noodzakelijk

Als een beslismodel personen met een van de specifieke kenmerken bijzonder benadeelt, zal de organisatie die het model toepast een objectieve rechtvaardiging moeten kunnen aanvoeren. Voor indirecte discriminatie bestaat, anders dan voor de meeste vormen van directe discriminatie, geen gedetailleerd uitgewerkte uitzonderingsbepaling, maar uitsluitend een algemene rechtvaardigingsgrond. Die houdt in dat het onderscheid objectief gerechtvaardigd moet zijn door een legitiem doel en dat de middelen voor het bereiken van dat doel passend en noodzakelijk moeten zijn.⁵³

De vereiste objectiviteit van de rechtvaardiging brengt mee dat er bewijs moet worden geleverd van het verband tussen de grond waarop het onderscheid wordt gemaakt en het doel. Een elektriciteitsbedrijf kan er bijvoorbeeld voor kiezen in een bepaalde wijk beveiligingsmaatregelen te treffen om beschadiging en illegale stroomaftap te voorkomen. Als dan blijkt dat in die wijk een bepaalde etnische groep is oververtegenwoordigd en dat de maatregelen die groep dus bijzonder benadelen, moet het bedrijf kunnen aantonen dat zich in de betreffende wijk daadwerkelijk beschadiging en illegale aftap hebben voorgedaan en dat het risico bestaat dat dergelijke gedragingen voortduren.⁵⁴

Daarnaast moet worden aangetoond dat de getroffen maatregel een passend en noodzakelijk middel vormt om die legitieme doelstelling te bereiken. ‘Passend’ wil zeggen dat de maatregel geschikt is om het nagestreefde legitieme doel te bereiken. Geschikt betekent niet dat doelstellin-

50 Zie FRA 2010 (*supra* noot 44), paragrafen 2.3.2 en 5.3; A. Ellis & P. Watson, *EU Anti-discrimination Law*, Oxford: Oxford University Press, 2012, p. 148-155; J.H. Gerards & A.W. Heringa, *Wetgeving gelijke behandeling*, Deventer: Kluwer, 2003, paragraaf 2.2.3.

51 M. Kosinski e.a., ‘Private traits and attributes are predictable from digital records of human behavior’, *PNAS* 2013, p. 5802-5805.

52 Zie onder meer artikel 10 lid 1 AWGB.

53 Zie onder meer artikel 2 lid 1 AWGB.

54 Zie HvJ EU 16 juli 2015, C-83/14, ECLI:EU:C:2015:480, r.o. 111-117 (*CHEZ*).

gen volledig bereikt moeten worden. Als een door een energiebedrijf getroffen beveiligingsmaatregel een merkbaar verschil in schade of illegale stroomaftap oplevert, is dat voldoende om als geschikt te worden beschouwd.⁵⁵ 'Noodzakelijk' impliceert een toets aan het subsidiariteits- en het proportionaliteitsbeginsel.⁵⁶ Het subsidiariteitsbeginsel vereist dat de legitieme doelstelling niet met andere passende, maar minder ingrijpende maatregelen kan worden bereikt.⁵⁷ Het proportionaliteitsbeginsel houdt in dat de veroorzaakte nadelen niet onevenredig mogen zijn aan de nagestreefde doelstellingen en dat de maatregel geen excessieve inbreuk mag maken op de legitieme belangen van de betrokkenen.⁵⁸

Omdat bij een vermoeden van indirecte discriminatie de wederpartij moet bewijzen dat niet in strijd met de gelijke-behandelingswet is gehandeld,⁵⁹ impliceert het voorgaande ten eerste dat de organisatie die het model gebruikt, bij een vermoeden van indirecte discriminatie inzicht zal moeten verschaffen in de werking van het model en de totstandkoming daarvan. Ten tweede is het de vraag of het feit dat het beslismodel is gebaseerd op correlaties die big-data-analyses hebben gevonden, meebrengt dat de daardoor geboden rechtvaardiging voldoende objectief is. Dat is niet zonder meer het geval. Zoals hiervoor besproken in paragraaf 3, kan namelijk via verschillende wegen een bias in het model zijn geslopen. Van de toepasser van een big-datatechniek kan worden verlangd dat hij aantoonde dat hij redelijke maatregelen heeft genomen om een dergelijke bias te voorkomen.

5 Big data in het licht van de non-discriminatiegrondrechten (Grondwet en EVRM)

5.1 Algemeen

Als een bepaalde praktijk buiten het toepassingsbereik van de specifieke gelijke-behandelingswetten valt, kan die worden getoetst aan het grondrechtelijk verbod op discriminatie. Dit verbod volgt onder meer uit artikel 14 van het Europees Verdrag voor de Rechten van de Mens (EVRM), het twaalfde protocol bij het EVRM en de artikelen 20 en 21 van het Handvest van de Grondrechten van de Europese Unie. Het wordt door het Hof van Justitie bovendien gezien als algemeen beginsel van Unierecht.⁶⁰ Deze grondrechten en beginselen schrijven voor dat allen gelijk zijn voor de wet en dat eenieder aanspraak heeft op gelijke behandeling. Het vereist dat vergelijkbare situaties niet verschillend en verschillende situaties niet gelijk worden behandeld.

55 Zie onder meer conclusie Advocaat-Generaal HvJ EU 12 maart 2015, C-83/14, ECLI:EU:C:2015:170 (*CHEZ*), paragraaf 123; en conclusie Advocaat-Generaal HvJ EU conclusie Advocaat-Generaal HvJ EU 20 september 2012, C-394/11, ECLI:EU:C:2012:585, par. 108 (*Belov*).

56 Zie HR 18 december 2015, ECLI:NL:HR:2015:3628.

57 Zie onder meer *Chez*-arrest (*supra* noot 54), r.o. 120.

58 Zie onder meer *Chez*-arrest (*supra* noot 54), r.o. 123 en HvJ EU 12 oktober 2010, C-499/08, EU:C:2010:600, r.o. 32 en 47 (*Ingeniorforeninger i Danmark*).

59 Zie bijvoorbeeld HR 18 december 2015, ECLI:NL:HR:2015:3628 over de bewijslast in het kader van de WGB.

60 Zie HvJ EU 22 november 2005, C-144/04, ECLI:EU:C:2005:709, r.o. 75, 77 (*Mangold*); HvJ EU 19 januari 2010, C-555/07, ECLI:EU:C:2010:21, r.o. 21, 50 en 51 (*Swedex*); HR 13 juli 2012, ECLI:NL:HR:2012:BW3367; zie over de bronnen van non-discriminatie recht in het kader van de EU: Ellis & Watson 2012 (*supra* noot 45), hoofdstuk 1 en 3.

Grondrechten zijn primair van toepassing in de verticale verhouding tussen de overheid en de burgers, maar kunnen op verschillende manieren doorwerken in (horizontale) verhoudingen tussen burgers en bedrijven onderling.⁶¹ Ten eerste houdt het grondrecht voor de overheid veelal niet alleen een negatieve verplichting in zich te onthouden van schending van het recht, maar ook een positieve verplichting om die rechten te waarborgen. Dat kan meebrengen dat de overheid maatregelen moet nemen om die rechten ook in horizontale verhoudingen te beschermen. De hiervoor behandelde wetgeving op het gebied van gelijke behandeling is daarvan een voorbeeld. Ten tweede kunnen grondrechten horizontale werking hebben, wat wil zeggen dat zij ook in horizontale verhouding worden toegepast. Met betrekking tot het gelijkheidsbeginsel gebeurt in het Nederlandse recht voornamelijk op indirecte wijze, te weten in die zin dat het algemene privaatrechtelijke begrippen inkleurt. Zo heeft de rechtbank in de Agfa-Schoolleerling-zaak de eisen van goed werkgeverschap ingevuld met 'het in het algemeen rechtsbewustzijn levend beginsel dat werknemers recht hebben op een billijke beloning, hetgeen onder meer meebrengt dat gelijke arbeid in gelijke omstandigheden op gelijke wijze gehonoreerd moet worden, behoudens een objectieve rechtvaardigingsgrond'. In cassatie heeft de Hoge Raad de klachten tegen dat oordeel verworpen.⁶²

5.2 Het verbod op onderscheid dat niet is gebaseerd op redelijke en objectieve gronden

Net zoals de hoger besproken wetgeving, is ook het grondrechtelijk verbod op discriminatie niet absoluut. Het verbiedt niet alle vormen van onderscheid, maar alleen onderscheid dat niet is gebaseerd op redelijke en objectieve gronden. Wat betreft het EVRM vereist een objectieve en redelijke rechtvaardiging dat het verschil in behandeling een legitiem doel dient en dat er 'a reasonable relationship of proportionality' bestaat tussen dat doel en een toegepaste maatregel.⁶³ De eisen die worden gesteld aan de rechtvaardiging zijn afhankelijk van het soort onderscheid dat wordt gemaakt. De lat ligt hoog voor onderscheid op gronden die nauw of onlosmakelijk met een persoon verbonden zijn, zoals geslacht en ras. Deze kenmerken worden specifiek genoemd in de hiervoor genoemde artikelen van het EVRM en het protocol en zijn vanuit het oogpunt van discriminatie 'verdacht'. Rechtvaardiging voor een onderscheid op deze specifieke gronden vereist volgens het EHRM 'very weighty reasons'.⁶⁴

De vereiste objectiviteit van de rechtvaardiging brengt mee dat vooroordelen en ongefundeerde generalisaties niet worden aanvaard als rechtvaardiging.⁶⁵ Daarnaast weegt het EHRM mee

61 Zie onder meer R. Nehmelman en C.W. Noorlander, *Horizontale werking van grondrechten, Over een leerstuk in ontwikkeling*, Deventer: Kluwer 2013; J. Gerards, *EVRM. Algemene beginselen*, Den Haag: Sdu uitgeverij 2011, paragraaf 5.3.

62 HR 8 april 1994, ECLI:NL:HR:1994:ZC1322, NJ 1994, 704; zie ook HR 24 maart 2017, ECLI:NL:HR:2017:492.

63 Zie onder meer EHRM 24 juli 2003, 40016/98, ECLI:CE:ECHR:2003:0724JUD004001698, r.o. 37 (*Karner/Oostenrijk*); EHRM 2 maart 2010, 13102/02, ECLI:CE:ECHR:2010:0302JUD001310202, r.o. 91 (*Kozak/Polen*); zie ook D.J. Harris e.a., *Law of the European Convention on Human Rights*, Oxford: Oxford University Press 2014, p. 792 e.v.

64 Zie de in noot 63 aangehaalde rechtspraak en literatuur.

65 Zie bijvoorbeeld EHRM 16 december 2003, 64927/01, ECLI:CE:ECHR:2003:1216JUD006492701, r.o. 42 (*Palau-Martinez/France*); EHRM 10 maart 2011, 2700/10, ECLI:CE:ECHR:2011:0310JUD000270010, r.o. 68 (*Kiyutin/Rusland*); EHRM 22 maart 2012, 30078/06, ECLI:CE:ECHR:2012:0322JUD003007806, r.o. 141 en 143 (*Konstantin Markin/Rusland*); EHRM 13 oktober 2015, 33631/06, ECLI:CE:ECHR:2015:1013JUD003363106, r.o. 75 (*Vrountou/Cyprus*); EHRM 24 mei 2016, 38590/10, ECLI:CE:ECHR:2016:0524JUD003859010, r.o. 126 (*Biao/Denmark*).

of een behandeling al dan niet is gebaseerd op een categorische uitsluiting van een bepaalde groep en of er ruimte is voor een individuele beoordeling. Zo oordeelde het EHRM dat een weigering van een verblijfsvergunning op basis van iemands HIV-status niet door de beugel kon, onder meer omdat geen rekening werd gehouden met overige gezondheidsgegevens en gegevens over familiebanden van de betrokkene.⁶⁶

Voor de invulling van de rechtvaardigingsgrond onder het Unierecht kan worden verwezen naar hiervoor besproken rechtvaardigingsgrond voor indirecte discriminatie. Die is namelijk gebaseerd op de rechtspraak met betrekking tot het algemene gelijkheidsbeginsel als grondprincipe van het Unierecht.⁶⁷

5.3 Relevante Europese rechtspraak: Jyske Finans, Test-Aankoop, e.a.

Met betrekking tot het gebruik van big data en data mining stelt zich in de eerste plaats de vraag of de resultaten van zulk een analyse kunnen worden gebruikt ter rechtvaardiging van een bepaald onderscheid op basis van 'verdachte kenmerken' zoals geslacht of ras. De rechtspraak van het Hof van Justitie met betrekking tot de kwalificatie als directe of indirecte discriminatie is niet eenduidig. Aan de grondslag daarvan liggen verschillende theorieën van wat directe discriminatie is, zoals tot uiting komt in de verschillende conclusies van de advocaten-generaal in de Achbita- en de Bougnaoui-zaken.⁶⁸ Het Hof heeft in zijn arresten in deze en volgende zaken deze discussie niet op een eenduidige wijze beslecht. Om onze vraag te beantwoorden kunnen we echter abstraheren van deze discussie. Deze kwalificatie is weliswaar van belang om te bepalen welke rechtvaardigingsgronden voorhanden zijn, maar het al dan niet gebruik maken van big data om tot een bepaalde behandeling te besluiten speelt in deze discussie geen rol. Aan de basis van de beoordelingen als directe of indirecte discriminatie, of buiten het terrein van de anti-discriminatierichtlijnen als inbreuken op het gelijkheidsbeginsel, ligt de vraag of een verschillende behandeling objectief te rechtvaardigen is en dus berust op een objectief en redelijk criterium. Voor de beoordeling van op big data gebaseerde besluitvorming is het wel relevant wat als objectief beschouwd wordt. De vraag is dus of statistische verschillen kunnen gelden als een voldoende objectief criterium om twee situaties te onderscheiden.

Het Hof van Justitie heeft in haar rechtspraak nog geen antwoord op die vraag gegeven, maar in verschillende conclusies van de advocaten-generaal zijn er wel aanzetten tot een antwoord gegeven.

In de recente zaak Jyske Finans kwam deze vraag zijdelings aan bod in de conclusie van de Advocaat-Generaal bij de behandeling van het begrip 'objectieve rechtvaardiging' in de context van indirecte discriminatie. De Advocaat-Generaal wijst erop dat, wanneer de regelgeving ter bestrijding van witwassen in richtlijn 2005/60 'indirect is gebaseerd op stereotypen van personen of transacties („het opstellen van profielen“)', er bij de toepassing gebruik moet gemaakt worden

⁶⁶ EHRM 10 maart 2011, 2700/10, ECLI:CE:ECHR:2011:0310JUD000270010, r.o. 73 en 74 (*Kiyutin/Rusland*).

⁶⁷ Vergelijk met HvJ EU 16 december 2008, C-127/07, ECLI:EU:C:2008:728, (*Société Arcelor Atlantique et Lorraine e.a.*), r.o. 47 en de daar geciteerde rechtspraak.

⁶⁸ Conclusie Advocaat-Generaal HvJ EU, C-157/15, ECLI:EU:C:2016:382 (*Achbita*); Conclusie Advocaat-Generaal HvJ EU, C-188/15, ECLI:EU:C:2016:553 (*Bougnaoui*). Voor een duidelijkere uiteenzetting van de onderliggende opvattingen over directe discriminatie, zie de conclusie van Advocaat-Generaal Sharpston HvJ EU, C-73/08, ECLI:EU:C:2009:396 (*Bressol*).

van een geïndividualiseerde beoordeling van het risico.⁶⁹ Met andere woorden, statistische verschillen mogen niet ter rechtvaardiging van een verschillende behandeling op basis van een beschermd groepskenmerk dienen. Het verschil moet gebaseerd zijn op een concreet aan de persoon gebonden risico. De kredietinstelling is dan wel in een aantal gevallen verplicht tot verscherpte klantenonderzoeksmaatregelen, maar ze kan dit niet zomaar verruimen tot een brede groep. Het Hof kwam tot zijn arrest zonder op deze discussie in te moeten gaan.

In de beroemde *Test-Aankoop*-zaak speelde eveneens de kwestie van het gebruik van statistische verschillen ter rechtvaardiging van een verschillende behandeling.⁷⁰ In dat arrest werd artikel 5(2) van Richtlijn 2004/113/EG ongeldig verklaard wegens strijd met het algemene gelijkheidsbeginsel in het licht van het doel van die richtlijn. Die bepaling bood lidstaten de ruimte in verzekeringsovereenkomsten 'proportionele verschillen in premies voor individuele personen toe te staan in de gevallen waarin sekse een bepalende factor is bij de beoordeling van het risico op basis van relevante en nauwkeurige actuariële en statistische gegevens'. Anders gesteld, deze bepaling liet directe discriminatie op grond van geslacht toe wanneer het ging om rationele discriminatie op basis van actuariële en statistische gegevens. Dat is vergelijkbaar met discriminatie op basis van big-data-analyses. Het Hof in Luxemburg beschouwde de bepaling in strijd met de doelstelling van gelijke behandeling van de richtlijn en onverenigbaar met de artikelen 21 en 23 van het Handvest. Het Hof wees erop dat de richtlijn uitging van de 'vergelijkbaarheid van de respectieve situaties van vrouwen en mannen', terwijl de gewraakte bepaling veronderstelde dat er wel een relevant verschil was. Het nam hiermee het eerste argument van de Advocaat-Generaal over, dat vooral draaide om een incoherente motivering in de richtlijn. De conclusie van de Advocaat-Generaal Kokott gaat dieper in op de toepassing van het gelijkheidsbeginsel en op welke gegevens mogen gebruikt worden om te stellen dat er sprake is van een verschil in situatie.⁷¹ Zij wijst erop dat het betwiste artikel een verschillende behandeling toelaat op basis van louter statistische verschillen en niet enkel wanneer sprake is van duidelijke biologische verschillen. Dit is volgens haar niet aanvaardbaar. Statistische gegevens kunnen niet zomaar dienen om beschermde gronden als differentiërende factor aan te nemen. Dit kan naar haar oordeel enkel gemotiveerd worden als met zekerheid kan worden vastgesteld dat het onderscheid uitsluitend is gebaseerd op 'objectieve criteria die niets te maken hebben met discriminatie op grond van geslacht'.

De hier besproken conclusies tonen dat verschillende advocaten-generaal, zowel in een casus rond directe (*Test-Aankoop*) en rond indirecte (*Jyske Finans*) discriminatie en los van de discussie over de kwalificatie daarvan, statistische gegevens niet zomaar aanvaarden om een verschillende behandeling te rechtvaardigen met een beschermde grond als proxy. Deze afwijzing lijkt vrij strikt bij gronden als geslacht, ras en etnische afstamming. In deze gevallen moet een verschillende behandeling gemotiveerd worden met aan de persoon verbonden feiten en gegevens, en niet louter met feiten verbonden aan de beschermde grond of daar heel nauw mee samenhangende eigenschappen. Advocaat-Generaal Kokott laat wel meer speelruimte bij leeftijd, wat ook steun vindt in de rechtspraak.⁷²

69 Conclusie Advocaat-Generaal HvJ EU, C-668/15, ECLI:EU:C:2016:914.

70 HvJ EU 1 maart 2011, C-236/09, ECLI:EU:C:2011:100.

71 Conclusie Advocaat-Generaal Kokott EU HvJ 30 september 2010, C-236/09, ECLI:EU:C:2010:564.

72 Zie de in voetnoot 48 aangehaalde rechtspraak.

Dit laatste laat vermoeden dat het gebruik van statistische gegevens voor het onderscheiden van situaties niet even streng wordt beoordeeld voor alle beschermde gronden. Van strengheid schijnt vooral sprake te zijn bij gronden zoals geslacht, ras en etnische afstamming.

Zo gaat alvast de Europese Commissie ervan uit dat de Test-Aankoop-uitspraak geen impact heeft op de voorgestelde richtlijn betreffende de toepassing van het beginsel van gelijke behandeling van personen ongeacht godsdienst of overtuiging, handicap, leeftijd of seksuele geaardheid.⁷³ Deze bevat een gelijkaardig artikel dat een verschillende behandeling toelaat op basis van leeftijd of handicap wanneer dit gemotiveerd kan worden in de risicobeoordeling ‘op basis van relevante en accurate actuariële of statistische gegevens’. Ook het Hof stelde zich in de eerder vermelde rechtspraak over leeftijdsdiscriminatie soepeler op en wees het gebruik van leeftijds grenzen als proxy voor fysieke geschiktheid niet af. Dergelijk gebruik als proxy is analoog aan een statistische benadering en beschouwt een leeftijdsgroep als een vergelijkbare groep wat fysieke geschiktheid betreft.

6 De noodzaak van *non-discrimination by design*

Voorgaande rechtspraakanalyse toont aan dat de kwestie rond het gebruik van statistische gegevens voor het onderscheiden van situaties nog niet is uitgekristalliseerd. Dergelijk gebruik is op zich ongepast bij gronden als geslacht, ras en etnische afstamming en moet aangevuld worden met bijkomende motiveringsplichten. Bij andere beschermde gronden is statistisch ondersteunde besluitvorming minder uit den boze, maar blijft waakzaamheid en zorgvuldige motivering een plicht en rust een bewijslast hierover op de gebruiker van de big-datatoepassing. Dit laatste is mogelijk één van de belangrijkste *strongholds* van het non-discriminatie *acquis* opgebouwd in Europa. De zorgplicht strekt zich uit over alle gevolgen van de gemaakte keuze inzake het besluitvormingsproces. Ook al worden de regels inzake gelijke behandeling pas van toepassing bij differentiatie op individueel niveau in de besluitvormingsfase, hetgeen in de daaraan voorafgaande fase waarin het model wordt opgebouwd gebeurt, is van grote invloed op de beoordeling van de verenigbaarheid met die regels. In het algemeen kan gesteld worden dat besluitvorming op basis van statistische modellen toegelaten is, behalve als die aanleiding geeft tot een vorm van rationele discriminatie op basis van één van de verboden gronden of daar nauw mee samenhangende proxies. Bijgevolg kan worden geconcludeerd dat het juridisch kader de gebruiker van big data verplicht ervoor zorg te dragen dat zijn big-datatoepassing niet leidt tot een discriminerende behandeling en dat hij voldoende inzichtelijk en controleerbaar moet maken dat aan die verplichting is voldaan. Ook wanneer de rechtvaardiging soepeler wordt beoordeeld, zoals bij leeftijd, dan nog is de gebruiker van big data verplicht dit te verantwoorden wanneer iemand een ongelijke behandeling op die basis aannemelijk maakt.

Het is daarom aan te bevelen om hieraan reeds bij het ontwerp van zulk een systeem aandacht te besteden en de nodige maatregelen te treffen om eventuele discriminatie te kunnen opsporen en te kunnen voorkomen. Deze behoefte heeft de laatste jaren geleid tot vernieuwend onderzoek naar de oorzaken van eventuele niet-intentionele discriminatie via data mining, zoals

73 Europese Commissie, ‘Richtsnoeren betreffende de toepassing van Richtlijn 2004/113/EG van de Raad op verzekeringen, in het licht van het arrest van het Hof van Justitie van de Europese Unie in zaak C-236/09 (Test-Aankoop)’, *PbEU* 2012, C 11/01.

hierboven in de sluiptwegen besproken, en naar technische maatregelen. Verscheidene statistische testmethodes zijn ontwikkeld om na te gaan of een discriminerende bias aanwezig is in oude datasets of op basis daarvan ontwikkelde modellen. Dit maakt het mogelijk om big-datasystemen te auditen of bij gebruik te monitoren. Eveneens is een reeks methodes ontwikkeld om discriminatie te voorkomen of tegen te gaan bij datamining en de daarmee ontwikkelde modellen. *Pre-processing*-technieken verwijderen of compenseren de discriminatie in de training dataset, waarna die gebruikt kan worden met standaard datamining-technieken. *In-processing* technieken passen de datamining-algoritmes aan opdat ze de aanwezige bias compenseren. *Post-processing* technieken corrigeren de ontwikkelde modellen om de discriminerende bias te verwijderen.⁷⁴ Om dat te kunnen doen, zullen de technieken vaak gebruik moeten maken van gegevens over de 'verboden' gronden, zoals ras en godsdienst.⁷⁵ In deze fase van de data-analyse kan het gebruik van gegevens over die kenmerken dus juist bijdragen aan het voorkomen van discriminatie. Sommige van deze technieken kunnen genuanceerd gebruikt worden. Alle verschil in resultaat tussen groepen verwijderen kan neerkomen op positieve discriminatie, wat niet per definitie de gewenste doelstelling is. Daarom kan de correctie beperkt worden tot het verwijderen van het verschil in resultaat tussen groepen dat niet verklaarbaar is op basis van bepaalde objectieve factoren.⁷⁶

Dit soort technische maatregelen is nog volop in ontwikkeling en staat, ook al zijn er al bruikbare methoden voorhanden, nog in de kinderschoenen. Dit valt te vergelijken met de situatie toen de eerste *privacy by design*-principes werden geformuleerd en enkel een reeks *privacy-enhancing techniques* (PET) beschikbaar waren.⁷⁷ Ondertussen is *privacy by design* verder ontwikkeld tot uitgewerkte ontwerpmethodologieën met precieze technische doelstellingen, ontwerppatronen en -methodologieën, auditing- en assessment-procedures.⁷⁸ In de komende jaren kan *non-discrimination by design* zich ook op deze manier ontwikkelen en gebruikmaken van de *lessons learned* met *privacy by design*.

7 Conclusie

In dit artikel hebben we besproken hoe op big data gebaseerde besluitvorming zowel discriminatie kan tegengaan als veroorzaken. Discriminatie wordt tegengegaan door het bewust of onbewust baseren van beslissingen op vooroordelen in de besluitvormingsfase te voorkomen. Eveneens maakt big data het mogelijk om een veel completer beeld van een individu in die besluitvorming

74 F. Kamiran, T. Calders en M. Pechenizkiy, 'Techniques for Discrimination-Free Predictive Models' in: B. Custers, T. Calders, B. Schermer, & T. Zarsky (eds.), *Discrimination and privacy in the information society: Data mining and profiling in large databases*, Berlijn/Heidelberg: Springer-Verlag 2013.

75 I. Zliobaitė & B. Custers, 'Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models', *Artificial Intelligence Law* 2016/24, p. 183-201.

76 F. Kamiran & I. Zliobaitė, 'Explainable and Non-explainable Discrimination in Classification', in: B. Custers, T. Calders, B. Schermer & T. Zarsky, *Discrimination and Privacy in the Information Society*, Berlijn/Heidelberg: Springer-Verlag 2013.

77 A. Cavoukian, 'Privacy by Design: The 7 Foundational Principles', Information & Privacy Commissioner of Ontario, Canada, 2009, <https://www.privacybydesign.ca/index.php/about-pbd/7-foundational-principles/>

78 Privacy by Design is uitgegroeid tot een uitgebreid onderzoeksveld. Voor een goede introductie en stand van zaken, zie de ENISA rapporten: G. Danezis et al, *Privacy and Data Protection by Design - from policy to engineering*, ENISA, december 2014; G. D' Acquistio et al, *Privacy by design in big data*, ENISA, december 2015. Beide zijn online beschikbaar op <https://www.enisa.europa.eu/topics/data-protection/privacy-by-design>.

te betrekken en zo de noodzaak te verminderen om terug te vallen op rationele discriminatie via grove benaderingen. Anderzijds blijkt dat via verschillende sluiptwegen in de daaraan voorafgaande fase waarin het beslissingsmodel wordt opgebouwd toch weer onbedoelde bias in dit model terecht kunnen komen.

Het juridisch kader inzake gelijke behandeling is in principe in staat om met op big data gebaseerde besluitvorming om te gaan. Zowel directe als indirecte discriminatie op basis van de beschermde gronden is in principe niet toegelaten, tenzij het objectief gerechtvaardigd kan worden met een in het specifieke geval beschikbare rechtvaardigingsgrond. Voor 'verdachte gronden' kan zo'n objectieve rechtvaardiging niet op basis van statistische modellen gebeuren waarbij die grond als grove proxy functioneert. Eveneens is het aan de gebruiker van big data om, wanneer een slachtoffer discriminatie aannemelijk maakt, aan te tonen dat er geen sprake is van discriminatie op basis van de verboden gronden.

Dit impliceert dat wat er in de voorafgaande fase van de modelopbouw gebeurt, mee betrokken wordt in de beoordeling van de geïndividualiseerde toepassing van dit model tijdens de eigenlijke besluitvormingsfase op haar verenigbaarheid met het non-discriminatierecht. De gebruiker van big data heeft er daarom alle belang bij om tijdig niet-intentionele discriminatie op te sporen en zijn systeem voor zichzelf en voor mogelijke slachtoffers inzichtelijk te kunnen maken. Big data in overeenstemming brengen met de beginselen van non-discriminatie vergt dus in de eerste plaats de ontwikkeling van een technisch kader met bijbehorende standaarden die deze beginselen in de werking van de technologie integreert en controleerbaar maakt. Net zoals in het gegevensbeschermingsrecht *privacy by design* zich eerst als technologisch dispositief heeft ontwikkeld en vervolgens als juridische verplichting in dit recht is geïntegreerd, lijkt een gelijkaardige ontwikkeling van *non-discrimination by design* aangewezen.