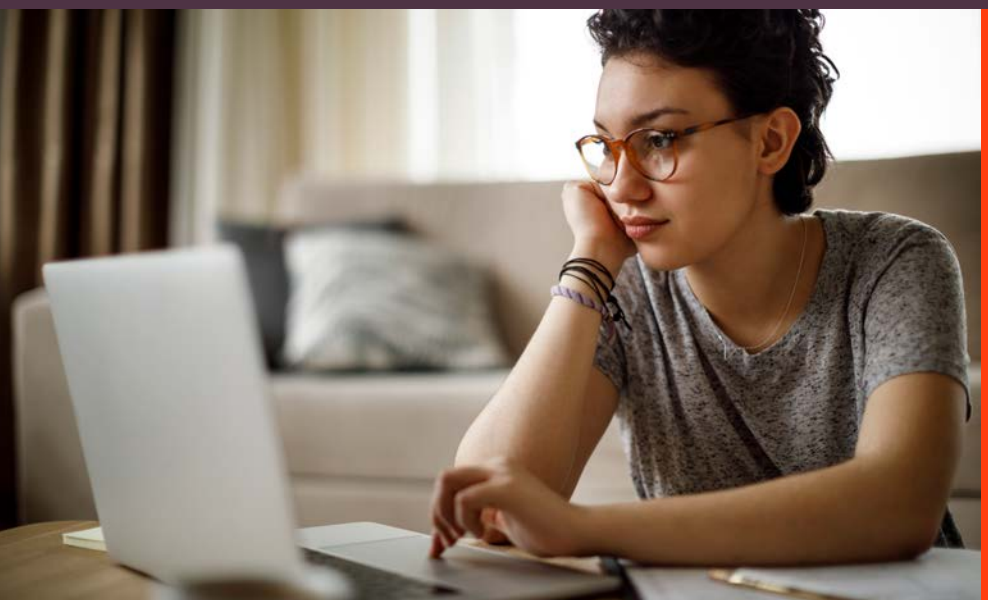
# Als computers je CV beoordelen, wie beoordeelt dan de computers?

Algoritmes en discriminatie
bij werving en selectie



COLLEGE VOOR
**DE RECHTEN VAN DE MENS**

September 2020

**College voor de Rechten van de Mens**

Kleinesingel 1-3
Postbus 16001
3500 DA Utrecht

**T** 030 888 38 88
Teksttelefoon: 030 888 38 29
**F** 030 888 38 83
**E** info@mensenrechten.nl
**W** www.mensenrechten.nl

Voor vragen kunt u een e-mail sturen en
op werkdagen bellen van 10.00 - 16.00 uur.

# Inhoudsopgave

# Samenvatting

Het College voor de Rechten van de Mens (afgekort: het College) verkent in dit onderzoek hoe de inzet van algoritmes bij werving en selectie kan leiden tot discriminatie. De risico's en oorzaken van deze vorm van discriminatie gaan vaak verscholen achter complexe computercode. Mede daarom ligt de nadruk bij dit onderzoek op het geven van voorbeelden om discriminatie bij werving en selectie door algoritmes inzichtelijker te maken.

In veel recruitmenttechnologie vormen algoritmes een centrale bouwsteen. Organisaties gebruiken deze technologie om gerichter en meer kandidaten voor vacatures te bereiken. Een voorbeeld is het zoekalgoritme van LinkedIn. Bij wervings- en selectieprocessen zal volgens deskundigen de rol van recruitmenttechnologie steeds groter worden.

In simpele woorden zijn algoritmes een soort instructieregels waarmee de computer op basis van ingevoerde informatie – zoals de selectiecriteria bij een vacature – besluiten kan nemen. Net als mensen kunnen algoritmes vooroordelen hebben (ook wel *bias* genoemd bij algoritmes) tegen bepaalde groepen. Een bekend voorbeeld hiervan is het selectiealgoritme van het Amerikaanse bedrijf Amazon dat op basis van kenmerken van werknemers uit het verleden een voorkeur had voor mannen bij bepaalde functies.

## Oorzaken van discriminatie door algoritmes

In dit literatuuronderzoek worden twee oorzaken van discriminatie door algoritmes beschreven: *bias* in het algoritme en *bias* in de data. *Bias* in het algoritme ontstaat bij het ontwerp van het algoritme en de keuze van variabelen. Variabelen die op het eerste gezicht niets met discriminatie te maken hebben, kunnen namelijk leiden tot discriminatie. Zo kan de variabele 'onafgebroken dienstjaren' een indicatie zijn van goed functioneren, maar ook van geslacht, omdat vrouwen vaker kortere aanstellingen hebben. Zelfs zonder expliciet informatie te verschaffen over iemands geslacht, ras, godsdienst of een ander beschermd kenmerk kan een algoritme bepaalde groepen dus discrimineren.

*Bias* in de data ontstaat, omdat data bestaande vooroordelen uit de maatschappij weergeven. Als deze data vervolgens worden gebruikt om een algoritme te trainen of aan te passen, dan kan dit weer leiden tot *bias* in het algoritme. Een voorbeeld is een selectiealgoritme bij een universiteit in het Verenigd Koninkrijk die een voorkeur had voor mannen zonder een migratieachtergrond. Het algoritme was getraind op basis van data uit een periode dat er weinig vrouwen en migranten mochten komen studeren.

Daarnaast kan een algoritme dat al in gebruik is op basis van gebruikersdata *bias* aanleren. Zo liet een algoritme vaker online advertenties voor technische vacatures aan mannen zien omdat mannen vaker dan vrouwen op deze advertenties klikten. Hierdoor hadden vrouwen minder kans om op de vacatures te klikken, terwijl er wellicht vrouwen waren die geschikt waren en interesse hadden voor deze vacatures.

Ook kan er *bias* in de data ontstaan, omdat bepaalde groepen te weinig hierin voorkomen. Mensen met een beperking en ouderen maken bijvoorbeeld minder gebruik van sociale media waardoor zij onvoldoende in bepaalde trainingsdata weergegeven worden.

## Verhoogd risico op discriminatie

Hoewel mensen ook kunnen discrimineren, brengt de aard van algoritmes vier risico's met zich mee die discriminatie verder kunnen verergeren.

1. Het kan vaak onduidelijk en niet transparant zijn hoe algoritmes tot besluiten komen. Algoritmes zijn hiervoor soms te complex. Daarnaast willen bedrijven niet altijd het algoritme openbaar maken vanwege commerciële belangen. Deze ondoorzichtigheid (ook wel de *black box* gemoemd) van algoritmes brengt mee dat sollicitanten vaak niet kunnen achterhalen waarom ze zijn afgewezen. Zo kunnen ze discriminatie moeilijk of niet aantonen.
2. Ten tweede associëren mensen computers en algoritmes veelal met rationale besluitvorming en foutloosheid (dit wordt *automation bias* genoemd). Hierdoor ontstaat het risico dat recruiters ongerechtvaardigd veel vertrouwen hebben in algoritmes zonder scherp te zijn op discriminatierisico's. Juist door de hoge tijdsdruk bij het wervings- en selectieproces is het risico op *automation bias* hoog.
3. Ten derde hebben algoritmes de potentie om *bias* – en daarmee discriminatie – te systematiseren. Algoritmes kunnen *bias* niet zomaar afleren en *biased* algoritmes kunnen ook makkelijk op grote schaal in gebruik genomen worden waardoor discriminatie zich verspreidt.
4. Tot slot kunnen algoritmes zeer veel complexe verbanden leggen waardoor de mogelijkheden om mensen te differentiëren (en dus ook te discrimineren) groter wordt.

## Discriminatie met algoritmes tegengaan

Ondanks de risico's voor werving en selectie, heeft het College ook oog voor de potentie van algoritmes om objectiever te zijn dan mensen en discriminatie tegen te gaan. Technieken om *bias* uit algoritmes te halen staan echter nog in de kinderschoenen en kennen nog steeds veel risico's. Verder onderzoek naar het verbeteren van algoritmes en bevorderen van gelijke behandeling in de ontwerpfase van een algoritme (*non-discrimination by design*) is essentieel om discriminatie door algoritmes te voorkomen.

Bovendien zullen technieken om *bias* te corrigeren vaak juist gebruik moeten maken van informatie over geslacht, ras of andere discriminatiegronden (zie box 2 op pagina 7). Het is de vraag of dit wenselijk is gezien het risico op directe discriminatie en omdat dit soort gegevens volgens de algemene verordening gegevensbescherming niet zomaar verwerkt mogen worden. Daarbij komt dat het lastig te beoordelen is wat de grens is tussen *bias* corrigeren in een algoritme en het geven van een voorkeursbehandeling waaraan strenge wettelijke eisen gesteld worden.

## Bewustwording en onderzoek

Het is essentieel dat algoritmeontwerpers en HR-professionals zich bewust zijn van de discriminatierisico's bij de inzet van algoritmes bij werving en selectie. Met de komst van algoritmes in het wervings- en selectieproces ontstaan nieuwe uitdagingen op het gebied van mensenrechten. Dit onderzoek laat zien dat niet enkel privacy of dataprotectie, maar ook discriminatie een risico is bij de inzet van algoritmes in het wervings- en selectieproces.

De vraag is echter hoe de eisen van het gelijkebehandelingsrecht zich verhouden tot algoritmes. Dit constateren ook de onderzoekers van het recente TNO-onderzoek *Digitale Arbeidsmarktdiscriminatie* dat uitgevoerd is in opdracht van de Inspectie Sociale Zaken en Werkgelegenheid (SZW).[1] In een vervolgonderzoek wil het College aan deze vraag aandacht besteden.

## Aanbevelingen

Op basis van dit onderzoek, doet het College de volgende aanbevelingen.

Voor de overheid:

- Vergroot het bewustzijn onder burgers over de risico's van discriminatie door algoritmes door middel van voorlichting.
- Bespreek met werkgevers de risico's van algoritmes bij werving en selectie.
- Geef voorlichting aan algoritmeontwerpers over hun wettelijke verplichtingen vanuit (onder andere) de gelijkebehandelingswetgeving.

Voor werkgevers:

- Informeer sollicitanten op een begrijpelijke manier voorafgaand aan de sollicitatieprocedure over de rol en de werking van algoritmes bij de selectie.
- Gebruik geen recruitmenttechnologie waarbij het niet duidelijk is op basis van welke afwegingen de software beslissingen maakt.

Voor algoritmeontwerpers:

- Informeer afnemers van software over de mogelijke risico's van de inzet van algoritmes bij werving en selectie.
- Maak op een begrijpelijke manier inzichtelijk op basis van welke variabelen sollicitanten beoordeeld worden.
- Voer met regelmaat validaties uit om te controleren of uw software niet leidt tot discriminatie.

Voor werkzoekenden:

- Maak melding indien u denkt dat u gediscrimineerd bent door een wervings- en selectiealgoritme. Dit kan ook als u alleen een vermoeden heeft. U kunt hiervoor terecht bij een van de discriminatiemeldpunten (kijk op www.discriminatie.nl voor meer informatie). Ook kunt u bij het College terecht voor een verzoek om een oordeel, het melden van discriminatie of vragen.

1 F. Grommé, S. Emmert, N. Wiezer, C. Thijs, *Digitale arbeidsmarktdiscriminatie: Inzicht in de risico's op arbeidsmarktdiscriminatie door de inzet van recruitment technologieën in werving en selectie.* Den Haag: TNO 2019, p. 60-62.

# 1. Inleiding

Door de inzet van recruitmenttechnologie bij het vervullen van vacatures kan (onbewust) discriminatie optreden. Technologie krijgt bij wervings- en selectieprocessen een steeds grotere rol en deskundigen verwachten dat in de toekomst dit soort technologie steeds vaker zal worden gebruikt.[2]

Als nationale toezichthouder maakt het College voor de Rechten voor de Mens (afgekort: het College) zich zorgen over de gevolgen van digitalisering voor de bescherming van mensenrechten. Daarom heeft het College hiervoor aandacht in zijn nieuwe strategische programma Digitalisering en Mensenrechten (box 1). Dit onderzoek, dat de oorzaken van discriminatie door (computer)algoritmes bij werving en selectie en de rol van het gelijkebehandelingsrecht in kaart brengt, is een uitwerking van dit programma.

**Box 1: Strategisch programma Digitalisering en Mensenrechten**

In 2020 is het College het strategische programma Digitalisering en Mensenrechten gestart. De samenleving digitaliseert in hoog tempo en het College wil niet dat mensen buiten de boot vallen. Twee mensenrechten staan centraal in dit programma: non-discriminatie en rechtsbescherming. Voor rechtsbescherming ligt de focus op de inzet van geautomatiseerde procedures en daaruit voorvloeiende beslissingen in de publieke sector. Bij non-discriminatie ligt de focus met name op digitalisering bij werving en selectie. Ook reageert het College op actualiteiten rondom digitalisering, zoals het gebruik van apps voor het contactonderzoek bij coronapatiënten.

Recruitmenttechnologie helpt organisaties om gerichter en meer kandidaten voor vacatures te bereiken. Denk bijvoorbeeld aan een wervingscampagne op sociale media of software die automatisch informatie uit cv's analyseert en op basis van selectiecriteria kandidaten rangschikt.[3]

Een centrale bouwsteen voor dit soort technologie zijn algoritmes. Dit zijn een soort instructieregels waarmee de computer op basis van ingevoerde informatie – zoals de selectiecriteria bij een vacature – besluiten kan nemen. Soms leidt de inzet van algoritmes tot ongewenste situaties waarbij (semi-) automatisch besluiten over mensen worden genomen waarbij sprake is van een aantasting van de mensenrechten.[4] Vaak wordt daarbij in eerste instantie gedacht aan het recht op privacy of dataprotectie.[5]

## Uitdagingen op het gebied van gelijke behandeling

Door de inzet van algoritmes bij werving en selectie ontstaan er echter ook uitdagingen op het gebied van discriminatie en gelijke behandeling. Algoritmes kunnen namelijk vooringenomenheden van mensen herhalen, verbreden en zelfs verdiepen. Een mens is vooringenomen als deze bewust of onbewust denkt dat bijvoorbeeld vrouwen niet technisch zijn. Bij een algoritme noemt men vooringenomenheid ook wel *bias*.

*Bias* in algoritmes kan op verschillende manieren ontstaan: zo kan een algoritme kandidaten voor technische functies selecteren op basis van de kenmerken van technische werknemers uit het verleden. Als deze werknemers allemaal mannen zijn, kan het algoritme onterecht mannen met technische vaardigheden gaan associëren en hebben vrouwen dus minder kans op de baan.[6]

## Kansen

Andersom hebben zorgvuldig ontworpen en gecontroleerde algoritmes de potentie om vooringenomenheid bij mensen te corrigeren, herhaling te voorkomen en zo diversiteit blijvend te bevorderen. Inzet van algoritmes en bestrijden van discriminatie kan dus beide kanten op werken.

## Niet zichtbare discriminatie

Discriminatie door algoritmes is vaak onzichtbaar. Voor sollicitanten is het bijvoorbeeld lastig te achterhalen waarom ze zijn afgewezen door een algoritme.[7] In het algemeen zijn de voorbeelden van discriminatie door algoritmes bij werving en selectie schaars. Het gebrek

2 Grommé e.a. 2019 (supra noot 1), p. 43, 52. D. Das, R. de Jong, L. Kool, *Werken op waarde geschat – Grenzen aan digitale monitoring op de werkvloer door middel van data en AI,* Den Haag: Rathenau Instituut 2020, p. 54.

3 Zie voor een overzicht van recruitmenttechnologieën Grommé e.a. 2019 (supra noot 1) hoofdstuk 5. Zie voor een indruk van de werking van bepaalde recruitmentsoftware ook D.V. van Ulden, *De geautomatiseerde sollicitatieprocedure* (masterscriptie Amsterdam) 2015, p. 6-13.

4 Sommige auteurs stellen daarom de vraag of er nieuwe (grond)rechten nodig zijn voor het digitale tijdperk. B. Custers, 'Nieuwe digitale (grond)rechten', *NJB* 2019/2775.

5 H. Lammerant, P. Blok, P. de Hert, 'Big data besluitvormingsprocessen en sluipwegen van discriminatie', *NTM/NJCM-bull* 2018/1, par 1.

6 Grommé e.a. 2019 (supra noot 1), p. 16.

7 W.L. Roozendaal, 'Big data op de werkvloer', TRA 2018/66.

aan zichtbaarheid betekent echter niet dat discriminatie in deze context niet voorkomt. Daarom is het extra relevant om te begrijpen wat de oorzaken van discriminatie door algoritmes zijn.

In het recente TNO-onderzoek *Digitale Arbeidsmarktdiscriminatie* – waarop het huidige onderzoek voortbouwt – bevelen de onderzoekers aan dat het beschrijven van duidelijke voorbeelden van digitale discriminatie van belang is om de bewustwording te versterken.[8] Het doel van dit onderzoek is om met behulp van voorbeelden inzichtelijk te maken hoe de inzet van algoritmes bij werving en selectie kan leiden tot discriminatie. Zo kunnen werkgevers, beleidsmakers en werkzoekenden zich met dit onderzoek meer bewust worden van de discriminatierisico's van algoritmes.

Deze publicatie is gebaseerd op literatuuronderzoek. Het bestaat naast deze inleiding uit vijf hoofdstukken en een conclusie. Na een korte uitweiding over de betekenis van de begrippen discriminatie en algoritmes (hoofdstuk 2), zullen twee oorzaken van discriminatie door algoritmes worden beschreven (hoofdstuk 3): *bias* in het algoritme (3.1.) en *bias* in de data (3.2.).[9] Vervolgens wordt ingegaan op een aantal discriminatieversterkende effecten die kunnen ontstaan door algoritmes (hoofdstuk 4). Daarna wordt het potentieel van algoritmes om discriminatie tegen te gaan bekeken (hoofdstuk 5). Tot slot vat het zesde hoofdstuk de belangrijkste conclusies samen en worden er een aantal aanbevelingen gedaan.

8 Grommé e.a. 2019 (supra noot 1), p. 60.

9 In de literatuur worden veelal vijf oorzaken beschreven hoe de inzet van algoritmes kan leiden tot discriminatie in navolging van het grondleggende werk van S. Barocas, A.D. Selbst, 'Big Data's Disparate Impact', *California Law Review* (104), 2016, afl. 3, p. 671. Maar ook wordt er een onderverdeling in twee oorzaken voor: White House, *Big Data: A Report on Algorithmic Systems Opportunity, and Civil Rights,* Executive Office of the President: 2016, p. 6. Of drie oorzaken: H. Lammerant, P. de Hert, P.H. Blok, 'Big data en gelijke behandeling'. In: P.H. Blok (red.), Big data & het recht: *Een overzicht van het juridisch kader voor big data toepassingen in de private sector,* Sdu uitgeverij 2017: Monografieën Recht en Informatietechnologie 10, p. 127.

# 2. Wat is discriminatie en wat zijn algoritmes?

## Discriminatie

Discriminatie is het anders behandelen, achterstellen of uitsluiten op basis van (persoonlijke) kenmerken. Deze kenmerken, zoals ras, geslacht of leeftijd, worden ook wel discriminatiegronden genoemd. Er zijn twaalf discriminatiegronden bij arbeid die door de wet verboden zijn (box 2). De wet gebruikt de term 'verboden onderscheid' in plaats van discriminatie, maar in dit paper wordt de gangbare term discriminatie gebruikt.

Discriminatie kan zowel direct als indirect zijn.
Als iemand op basis van deze twaalf verboden op een andere wijze wordt behandeld dan een ander in een vergelijkbare situatie, dan spreken we van directe discriminatie. Hiervan is bijvoorbeeld sprake als iemand de baan niet krijgt vanwege een discriminatiegrond zoals ras, leeftijd of geslacht.[10]

**Box 2: De twaalf verboden discriminatiegronden bij arbeid**

- godsdienst
- levensovertuiging
- politeke gezindheid
- ras
- geslacht
- nationaliteit
- hetero- of homoseksuele gerichtheid
- burgerlijke staat
- handicap of chronische ziekte
- leeftijd
- arbeidsduur
- soort contract

Bij indirect onderscheid is sprake van een ogenschijnlijk neutrale bepaling, maatstaf of handelswijze die niet rechtstreeks verwijst naar een van de discriminatiegronden in de wet, maar waardoor personen van een bepaalde beschermde groep (zoals vrouwen, ouderen of mensen met een beperking) wel in het bijzonder worden getroffen. Ter illustratie: een uitzendbureau dat alleen maar studenten werft maakt weliswaar geen direct onderscheid op leeftijd, maar de meeste studenten zijn wel jonger dan dertig jaar waardoor ouderen praktisch uitgesloten worden.[11]

10 Zie bijvoorbeeld College oordeel 2018-129, 2020-12 en 2019-27.

11 Zie bijvoorbeeld College oordeel 2019-37.

Zowel directe als indirecte discriminatie is in principe verboden, tenzij er voldaan kan worden aan een wettelijke rechtvaardigingsgrond.

## Algoritmes

Een algoritme is een reeks instructies, in een bepaalde volgorde uit te voeren.[12] Zelfs voor de computer bestonden algoritmes al. Zo kan een rekensom of een recept om een maaltijd op tafel te krijgen ook beschouwd worden als een algoritme. Veelal worden algoritmes gebruikt om door middel van de computer processen te automatiseren. In dit onderzoek hebben we het dan ook over computeralgoritmes.

Een alledaags voorbeeld van een computeralgoritme is het algoritme dat zorgt voor jouw persoonlijke aanbevelingen op websites. Zo komen de filmaanbevelingen op Netflix of YouTube tot stand door een algoritme.[13] Op basis van jouw kijkgedrag (inputdata) rangschikt en beveelt het algoritme films aan die jij waarschijnlijk leuk zal vinden (outputdata) doordat het algoritme scores toekent aan bijvoorbeeld hetzelfde genre of dezelfde acteur.

Algoritmes kunnen ook scores toekennen aan mensen, dit doen datingsites bijvoorbeeld. De bekende datingapp Tinder geeft gebruikers een score op basis van hoeveel *likes* (*swipes* naar rechts) een gebruiker van andere gebruikers krijgt.[14] Ook is deze score hoger als de like afkomstig is van een gebruiker die zelf een hoge score heeft. Daarnaast kent Tinder scores toe aan onder andere je biografie, locatie en leeftijd. Tinder laat gebruikers vervolgens profielen zien die passen bij dezelfde score die zijzelf hebben gekregen.[15]

Op dezelfde manier kunnen bij werving en selectie kandidaten worden gescoord op basis van geschiktheid voor een baan, bijvoorbeeld het aantal jaren werkervaring of de genoten opleiding.

In de literatuur zijn er veel definities te vinden van een algoritme en sommige auteurs vragen zich af of het überhaupt nuttig (of mogelijk) is om een algoritme te definiëren.[16] De definitie van een algoritme is heel algemeen en algoritmes kennen vele verschijningsvormen.[17] Voor de doeleinden van dit onderzoek is het met name van belang om te begrijpen wat een algoritme allemaal kan. Er zijn grofweg vier kerntaken die een algoritme kan uitvoeren: prioriteren (TomTom selecteert de snelste route), classificeren (advertenties bieden iemand babyvoeding aan omdat die persoon geclassificeerd is als jonge moeder), associëren (Amazon die productaanbevelingen doet op basis van eerdere aankopen) en filteren (Facebook en Twitter laten je eerst *posts* zien die bij jou passen).[18]

## Zelflerende algoritmes

Veel algoritmes hebben een simpele 'als-dan' structuur. Bijvoorbeeld *als* de temperatuur in de huiskamer zakt *dan* stuurt de thermostaat een seintje zodat de verwarming aan gaat.[19] Deze 'simpele' algoritmes kunnen worden onderscheiden van zelflerende algoritmes die op basis van eerder behaalde resultaten en trainingsdata autonoom verbanden kunnen leggen en beslisregels kunnen maken. Een voorbeeld is een spamfilter. Die is voorafgaand 'getraind' op basis van data – bijvoorbeeld met zinnen als 'magische pil tegen gewichtstoename' – om spam te herkennen en te onderscheiden van niet-spam. Door feedback van gebruikers kunnen zelflerende algoritmes continu bijleren en nieuwe beslisregels maken.

Elke keer als een gebruiker een email in de spamfolder zet of terughaalt leert het algoritme van deze informatie om spamberichten beter te herkennen.[20] Bij het aannemen of afwijzen van een sollicitant door een HR-medewerker kan dit hetzelfde werken. Het algoritme leert steeds beter kandidaten te herkennen die de medewerker zelf ook zou hebben gekozen.

12 Dikke van Dale woordenboek van de Nederlandse taal (online).

13 S. van den Braak, S. Choenni, 'Voorspellen met big-data-modellen: over de valkuilen voor beleidsmakers'. In: Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC), *De toekomst verkennen en voorspellen,* Justitiële verkenningen 2019/45, p. 22.

14 S. Blauw, 'Wat is een algoritme', *de Correspondent* 2 juli 2019.

15 K. Tiffany, 'The Tinder algorithm, explained', *Vox* (online, bijgewerkt 18 maart 2019).

16 Y. Gurevich, 'What is an algorithm? (revised)', Microsoft Research 2017.

17 Blauw 2019 (supra noot 14).

18 H. Fry, *Hello World: Being Human in the Age of Algorithms,* New York/London: W.W. Norton & Company 2018.

19 C. Adriaansz, 'Betekenisvolle transparantie voor algoritmische besluitvorming', Computerrecht, 2020/43, par. 2.

20 Uitleg gebaseerd op: V. Frissen, M. van Eck, T. Drouen, *Toezicht op het gebruik van algoritmen door de overheid,* Hooghiemstra en Partners 2019, p. 9, 10.

Een zelflerend algoritme kan op basis van grote hoeveelheden data zeer veel verbanden leggen die vaak zo complex zijn dat mensen ze niet snel zelf zullen kunnen observeren. In de consumentendata van een elektronicazaak zou een algoritme bijvoorbeeld kunnen ontdekken dat er een verband is tussen de schermgrootte van een televisie die iemand koopt en hoelang iemand nodig heeft om de consumentenlening voor die televisie af te betalen.[21] Kortom, de mogelijkheden en verschijningsvormen van algoritmes zijn groot.

# 3. Oorzaken van discriminatie door algoritmes bij werving en selectie

Arbeidsmarktdiscriminatie is een hardnekkig en wijdverbreid probleem en algoritmes spelen daar in toenemende mate een rol in. Verschillende onderzoeken tonen aan dat mensen met een niet-westerse achtergrond vaker worden uitgesloten bij het zoeken naar een baan dan mensen zonder migratieachtergrond.[22] Een kwart van de werkende Nederlanders ervaart wel eens discriminatie op de werkvloer.[23] Meer dan een kwart van de discriminatiemeldingen bij antidiscriminatievoorzieningen in de afgelopen vijf jaar had betrekking op de arbeidsmarkt, waarvan 43 tot 51 procent ging over werving en selectie.[24] Ook bij het College heeft arbeidsmarktdiscriminatie, met name bij werving en selectie, het grootste aandeel in de vragen, meldingen en (verzoeken om) oordelen.[25]

Mensen discrimineren niet altijd met opzet. Vaak spelen onbewuste vooroordelen over bepaalde groepen een rol. Doordat HR-professionals besluiten moeten nemen op basis van weinig informatie en onder hoge tijdsdruk is er een risico dat (onbewuste) vooroordelen het wervings- en selectieproces insluipen. Van alle vormen van arbeidsmarktdiscriminatie lijkt het risico op discriminatie in het wervings- en selectieproces daarom het grootst.[26]

Doordat het wervings- en selectieproces vaak grotendeels achter de schermen gebeurt, is dit niet goed zichtbaar. Ten opzichte van andere vormen van discriminatie twijfelen mensen vaak of er sprake is van discriminatie bij het zoeken naar werk (13 procent).[27] Ook is het de gangbare 'cultuur' dat men zich erbij neerlegt.[28]

Bij algoritmes spelen onbewuste vooroordelen – *bias* – eveneens een rol. Zoals verder zal worden uitgewerkt in dit hoofdstuk, ontstaat *bias* in algoritmes mede omdat algoritmes onbewuste vooroordelen in de samenleving reproduceren.[29] Net als discriminatie door onbewuste vooroordelen is discriminatie door algoritmes vaak slecht zichtbaar. In het algemeen zijn er weinig voorbeelden van discriminatie door algoritmes bij werving en selectie, maar net als bij discriminatie door mensen betekent dit niet dat discriminatie door algoritmes niet voorkomt.

Om meer inzicht te verschaffen over hoe de inzet van algoritmes bij werving en selectie tot discriminatie kan leiden, zullen in dit hoofdstuk aan de hand van een aantal voorbeelden twee oorzaken van discriminatie door algoritmes worden beschreven: *bias* in het algoritme (3.1.) en *bias* in de data (3.2.).[30] Deze twee oorzaken zijn echter niet helemaal van elkaar te scheiden, zo kan door *bias* in de data ook *bias* in het algoritme ontstaan.

21 *Voorbeeld van* Lammerant, de Hert en Blok 2017 (supra noot 9), p. 122.

22 L. Thijssen, M. Coenders, B. Lancee, 'Etnische discriminatie op de Nederlandse arbeidsmarkt: Verschillen tussen etnische groepen en de rol van beschikbare informatie over sollicitanten', *Mens & maatschappij* (94) 2019, nr. 2.

23 I. Andriessen, J. Hoegen Dijkhof, A. van der Torre, E. van den Berg, I. Pulles, J. Iedema, M. de Voogd-Hamelink, *Ervaren discriminatie in Nederland II*, Den Haag: Sociaal en Cultureel Planbureau 2020, p. 77.

24 Discriminatie.nl, *Monitor arbeidsdiscriminatie 2015-2019,* Rotterdam: 2020, p. 43.

25 College voor de Rechten van de Mens, *Monitor Discriminatiezaken 2019,* 2020.

26 College voor de Rechten van de Mens, *Interventies om discriminatie bij de werving en selectie tegen te gaan*, 2019, p. 11.

27 Andriessen e.a. 2020 (supra noot 23), p. 72.

28 A.H. Pranger, P.C. Vas Nunes, 'Kroniek gelijke behandeling bij arbeid', *Arbeidsrecht* 2017/52, par. 6.

29 Barocas en Selbst 2016 (supra noot 9), p. 673.

30 Hiermee wordt het onderscheid gevolgd van: White House 2016 (supra noot 9), p. 6. en M. Vetzo, J. Gerards, R. Nehmelman, *Algoritmes en grondrechten,* Den Haag: Boom Juridisch 2018, p. 142.

## 3.1 Discriminatie vanwege *bias* in het algoritme

Wat is de ideale werknemer? Die vraag is niet makkelijk precies te beantwoorden (box 3).[31] Ontwerpers van algoritmes staan voor de lastige uitdaging om een antwoord te formuleren dat gemeten kan worden door het algoritme: denk aan verkoopcijfers, diploma's of bijvoorbeeld het aantal ononderbroken dienstjaren bij vorige werkgevers. Dit zijn resultaten, ofwel variabelen, waar het algoritme naar kan zoeken bij sollicitanten. Maar niet alles wat telbaar is, telt mee.

Het probleem met variabelen is dat ze nooit een volledig beeld van een sollicitant geven. Iemand kan bijvoorbeeld geen diploma hebben, maar alsnog vaardigheden hebben verkregen door zelfstudie of ervaring. Het risico is ook dat digitale instrumenten alleen meten wat makkelijk te meten is, terwijl kwaliteiten die lastig te vangen zijn in variabelen – zoals collegialiteit – minder belangrijk worden.[32]

### Een statistisch verband is niet altijd een oorzakelijk verband

Een variabele kan daarnaast voor meerdere uitleg vatbaar zijn, en daarmee soms onjuist.[33] Door de komst van zelflerende algoritmes kunnen (subtiele) verbanden – correlaties – in data worden ontdekt tussen schijnbaar onbelangrijke kenmerken en de kwaliteiten van een goede werknemer. Zo bleek uit een onderzoek dat nieuwe werknemers die een andere webbrowser op hun computer installeren langer in dienst blijven dan werknemers die een al geïnstalleerde browser gebruiken.[34] Ook bleek uit een ander onderzoek een statistisch verband tussen het liken van krulfriet op Facebook en hogere intelligentie.[35] Dergelijke opmerkelijke verbanden kunnen meegenomen worden in de afweging die het algoritme maakt. Dit kan heel ver gaan.

**Box 3: Hunkemöller's algoritme bij video-interviews**

Sollicitanten bij het bedrijf Honkemöller worden niet beoordeeld op basis van een cv, maar op basis van een videofragment van anderhalve minuut dat de sollicitant moet opsturen. Om de video's te analyseren gebruikt het bedrijf algoritmes die getraind zijn om goede werknemers te herkennen op basis van honderd succesvolle medewerkers die al bij Hunkemöller werken.

De lichaamstaal van een sollicitant speelt hierbij een rol. Zo is de stand van een wenkbrauw volgens het bedrijf een indicator van extravert zijn. Het analyseren van videofragmenten is niet zonder risico's. Zo liet een onderzoek zien dat algoritmes moeite hebben met het registreren van emoties van vrouwen met een donkere huidskleur.

Bogen en Rieke 2018 (supra noot 61), p. 37.J. Buolamwini, T. Gebru, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification', *Proceedings of Machine Learning Research* (81) 2018. B. van de Haterd, 'Algoritme van de maan: hoe video-analyse bepaalt of je bij Hunkemöller past', *Werf&* (online, bijgewerkt op 15 februari 2019).

Het Amerikaanse bedrijf Gild, dat andere bedrijven helpt met het rekruteren van programmeurs, gebruikt algoritmes om op het internet open-source code en sociale media te analyseren om zo goede programmeurs op te sporen. Volgens de *chief scientist* van het bedrijf is een interesse voor een specifieke Japanse animatie (manga) website een sterke voorspeller voor een goede programmeur.[36] Interesse voor deze website zorgt dus voor een betere kans op een aanbod, ondanks dat er geen duidelijke oorzaak is waarom liefhebbers van Japanse animatie betere programmeurs zouden zijn. Waarschijnlijk leer je niet veel over programmeren door tijd door te brengen op deze website.

Twee gegevens die statistisch verbonden zijn, hoeven dus niet in oorzakelijk verband te staan met elkaar. 'Op dagen dat er veel zomerse kleding wordt gedragen, wordt er veel ijs verkocht' zou de redenering van een algoritme op basis van data kunnen zijn, terwijl een mens meteen zou begrijpen dat mensen meer ijs kopen, omdat het warm is en niet omdat ze zomerse kleding aan hebben.

31 Barocas en Selbst 2016 (supra noot 9), p. 679.

32 Das, de Jong en Kool 2020 (supra noot 2), p. 72.

33 P.T. Kim, 'Data-driven discrimination at work', *William & Mary Law Review* (58) 2017, p. 922.

34 J. Pinsker, 'People Who Use Firefox or Chrome Are Better Employees', *The Atlantic* (online, 16 maart 2015).

35 M. Kosinski, D. Stillwell, T. Graepel, 'Private traits and attributes are predictable from digital records of human behavior', *Proceedings of the National Academy of Sciences of the United States of America* (PNAS) (110) 2013 (nr. 15), p. 5804.

36 D. Peck, 'They're Watching You at Work', *The Atlantic* (online, december 2013)

## Schijnbaar neutrale variabelen

Variabelen die op het eerste gezicht niets met discriminatie te maken lijken te hebben kunnen ook indirect leiden tot discriminatie van bepaalde beschermde groepen. Een veel gebruikt voorbeeld in de literatuur van een dergelijke schijnbaar neutrale variabele is een postcode. Postcodes zeggen namelijk niet alleen iets over iemands postadres, maar ook of je in een wijk woont waar veel mensen van een bepaalde etniciteit wonen. Dit gegeven kan gebruikt worden om een bepaalde groep uit te sluiten.[37]

Een ander voorbeeld: de variabele 'onafgebroken dienstjaren' kan een indicatie zijn van loyaliteit en goed functioneren, maar ook geslacht. Immers hebben vrouwen vaak meer korte aanstellingen vanwege zwangerschappen of kinderzorg.[38] Zelfs 'zelden laat zijn' zou een discriminerende variabele kunnen zijn, omdat dit geen volledig beeld geeft van de achtergrond van een sollicitant. Zo wonen mensen met een migratieachtergrond misschien vaker verder van werk waardoor ze meer in de file staan of vertraging hebben met het openbaar vervoer.[39] Kortom, met dezelfde schijnbaar neutrale variabelen die relevant lijken voor de selectie van personeel, kunnen algoritmes onderscheid maken tussen beschermde groepen met als gevolg: indirecte discriminatie.[40]

Ook als er niet per definitie sprake is van discriminatie, kunnen deze schijnbaar neutrale variabelen bijdragen aan ongelijkheid op de arbeidsmarkt. Zo zegt het hebben van een diploma iets over iemands sociale status. Leerlingen uit gezinnen met een lage sociaaleconomische status verlaten vaker vroegtijdig hun school.[41] Daarom zal een algoritme wat 'het hebben van een diploma' als variabele heeft, zonder nadere toevoeging over mogelijk relevante vervangende ervaring, minder mensen uit lagere sociaaleconomische milieus selecteren. Ook was toegang tot hoger onderwijs vroeger minder vanzelfsprekend, waardoor ouderen minder kans maken.[42]

Sommige algoritmes kunnen op basis van schijnbaar neutrale data afleiden of iemand tot een bepaalde groep of klasse behoort en die informatie gebruiken bij de ranking van sollicitanten. Dit kan dus nadelig uitpakken voor beschermde groepen, zelfs zonder dat iemand specifiek informatie heeft gegeven over bijvoorbeeld zijn of haar ras of godsdienst.[43] Zo blijkt uit een onderzoek van de Autoriteit Persoonsgegevens dat Facebook de seksuele geaardheid van gebruikers kan afleiden door de websites die zij bezoeken om op basis hiervan gerichte advertenties te tonen.[44] Ook op basis van vriendschappen kan Facebook seksuele voorkeuren voorspellen terwijl gebruikers hierover geen expliciete informatie hebben gegeven.[45] Een ander onderzoek laat zien dat Facebook ras (wat Facebook overigens 'etnische affiniteit' noemt) kan afleiden uit het like- en klikgedrag van gebruikers; ook deze informatie gebruikt Facebook voor gericht adverteren.[46]

Wat bij Facebook kan, is ook mogelijk bij recruitmentwebsites. Onderzoek naar de websites Indeed, Monster en Careerbuilder laat zien dat vrouwen minder kans hebben op een baan (vooral bij technische beroepen en beroepen die met mannen geassocieerd worden, zoals vrachtwagenchauffeur) terwijl kandidaten geen informatie hebben gegeven over hun gender. De onderzoekers kunnen niet zonder twijfel uitleggen waarom vrouwen minder kans hebben op deze banen, maar ze denken dat dit te maken kan hebben met een verborgen neutrale variabel die samenhangt met geslacht. Ook het klikgedrag van bevooroordeelde HR-professionals die minder vaak op vrouwen klikken voor dit soort vacatures kan een van de oorzaken zijn dat het algoritme vrouwen benadeeld (zie 3.2.).[47]

Vergelijkbare problemen kunnen ontstaan bij software die de cv's, sollicitatiebrieven of andere informatie van kandidaten analyseert. *Bias* kan bijvoorbeeld ontstaan tegen kandidaten met een ander dialect.[48] Dit is relevant

37 F.J. Zuiderveen Borgesius, 'Strengthening legal protection against discrimination by algorithms and artificial intelligence, *The International Journal of Human Rights* 2020, p. 6.

38 Lammerant, Blok en de Hert 2018 (supra noot 5), par 3.3.

39 F. Zuiderveen Borgesius, *Discrimination, artificial intelligence, and algorithmic decision-making,* Strasbourg: Council of Europe 2018, p. 10.

40 Barocas en Selbst 2016 (supra noot 9), p. 691

41 T. Traag, R.K.W. van der Velden, 'Early school-leaving in the Netherlands: The role of family resources, school composition and background characteristics in early school-leaving in lower secondary education', *Irish Educational Studies* (30) 2011 (nr. 30), p. 54.

42 Lammerant, de Hert en Blok 2017 (supra noot 9), p. 127.

43 Grommé e.a. 2019 (supra noot 1), p. 20.

44 Autoriteit Persoonsgegevens, *Onderzoek naar het verwerken van persoonsgegevens van betrokkenen in Nederland door het Facebook-concern,* 2017. p. 155.

45 C. Jernigan, B.F.T. Mistree, 'Gaydar: Facebook friendships expose sexual orientation', *First Monday* (14) 2009 (nr. 10).

46 J. Angwin, T. Parris, 'Facebook lets Advertisers Exclude Users by Race', *ProPublica* (online, bijgewerkt 28 oktober 2016).

47 L. Chen, A. Hannák, R. Ma, C. Wilson, 'Investigating the Impact of Gender on Rank in Resume Search Engines', (Proceedings of the 2018 CHI Conference on Human Factors in Computing).

48 A. Sutton, T. Lansdall-Welfare, N. Cristianini, 'Biased embeddings from wild data: Measuring, understanding and removing' (International Symposium on Intelligent Data Analysis 2018).

voor software die spraak en taalgebruik als variabele analyseert. Bedrijven als L'Oréal, Accenture, Danone, Deloitte en Heineken gebruiken bijvoorbeeld de software Seedlink om onbewuste patronen in taalgebruik van kandidaten te analyseren om zo te testen of iemand past in de bedrijfscultuur.[49] Het is mogelijk dat bij dit type software ook (niet opzettelijk) patronen worden ontdekt waardoor mensen uit bepaalde beschermde groepen minder kans hebben op een baan.

Het gebruiken van schijnbaar neutrale variabelen is dus niet zondermeer zonder risico. Ook het Amerikaanse bedrijf Amazon (zie ook 3.2.) nam minder vrouwen in dienst omdat het algoritme een voorkeur had voor 'mannelijke woorden'.[50] Zelflerende algoritmes kunnen (onbedoeld) variabelen die verband hebben met een door de gelijkebehandelingswetgeving beschermde discriminatiegrond meenemen bij het wervings- en selectieproces. Dit verband is een belangrijk aandachtspunt, omdat het hier gaat over bindende wettelijke normen.

### Vooroordelen van algoritmeontwerpers

Onbewuste vooroordelen van algoritmeontwerpers kunnen bij de selectie en formulering van variabelen ook in het algoritme sijpelen.[51] Een vrouw in het Verenigd Koninkrijk kon bijvoorbeeld niet met haar pasje de vrouwenkleedkamer bij de sportschool in omdat ze de titel 'Dr.' (Doctor) gebruikte. Het systeem herkende dit als een exclusieve indicator voor het mannelijke geslacht.[52] Een vergelijkbaar probleem kan voorkomen bij wervings- en selectiealgoritmes als iemand zijn titel op zijn cv zet.

Dikwijls wordt de zorg geuit dat veel software wordt gemaakt in *Silicon Valley*, een omgeving die wordt gekenmerkt door mannen met een specifieke achtergrond. Hierdoor kunnen geslachtseigen en sociale normen ingebed raken in de algoritmes die daar gemaakt worden.[53] Ook kan een HR-medewerker bij een organisatie waar veel mannen werken (eventueel onbewust) doelvariabelen kiezen waardoor meer mannen worden aangenomen, net als bij de sportschool hierboven.

Een organisatie zoekt vaak naar sollicitanten die passen bij de bedrijfscultuur (ook wel werven op culture-fit) zoals bij het eerdergenoemde voorbeeld van Seedlink. Selectiealgoritmes kunnen daardoor een digitale variant van de similar-to-me-*bias* nabootsen en versterken.[54]

## 3.2 Discriminatie vanwege *bias* in de data

Het gebruik van een algoritme kan tot discriminerende resultaten leiden als het algoritme is getraind op basis van *biased* data. Grofweg gebeurt dit in twee gevallen, namelijk als de data bestaande of oude discriminerende vooroordelen bevatten die worden gereproduceerd door het algoritme of omdat de verzamelde data niet representatief zijn voor de doelgroep.[55]

### Vooroordelen reproduceren

Het eerste geval van *biased* data deed zich voor bij een algoritme dat gebruikt werd voor de selectie van nieuwe geneeskundestudenten in het Verenigd Koninkrijk in de jaren tachtig.[56] Het algoritme discrimineerde vrouwen en mensen met een migratieachtergrond, omdat de data waren verzameld in een periode dat er weinig vrouwen en migranten mochten komen studeren. Het algoritme introduceerde dus geen nieuwe discriminatie, maar het reproduceerde wel de vooroordelen tegen vrouwen en migranten uit het verleden.

Een recenter voorbeeld is een selectiealgoritme van het bedrijf Amazon in de Verenigde Staten. Gebaseerd op de werknemers bij Amazon van de afgelopen tien jaar, leerde het algoritme dat mannen de voorkeur hadden bij technische functies. CV's met een referentie naar het woord 'vrouw', zoals 'voorzitter vrouwenschaakclub', belandden daarom onderop de stapel.[57]

### Data verzamelen over personeel

Door de komst van *HR-Analytics* (box 4) is het mogelijk om steeds meer en meer gedetailleerde data te verzamelen over de huidige werknemers van een bedrijf. Deze ontwikkeling kan ook gevolgen hebben voor het wervings- en selectieproces. Dat kan bijvoorbeeld als deze data worden gebruikt om een profiel op te stellen van 'succesvolle werknemers' die algoritmes kunnen gebruiken om nieuwe werknemers te rangschikken. *HR-Analytics* kan ertoe leiden dat meer *biased* data

49 P. Boerman, 'Selectie 2.0: Hoe Seedlink je success voorspelt aan de hand van je taalgebruik', *Werf&* (online, bijgewerkt 4 april 2017).

50 J. Dastin, 'Amazon scraps secret AI recruiting tool that showed bias against women', *Reuters* (online, bijgewerkt 10 oktober 2018).

51 White house 2016 (supra noot 9), p. 6.

52 J. Fleig, 'Doctor locked out of women's changing room because gym automatically registered everyone with Dr title as male', *Mirror* (online, bijgewerkt 19 maart 2015).

53 N.C. van Oostrom-Streep, 'Over de ethiek van de toekomst en achterhaalde concepten', *WPNR* 2017/7154, p. 566.

54 Das, de Jong en Kool 2020 (supra noot 2), p. 39.

55 Zuiderveen Borgesius 2018 (supra noot 39), p. 11.

56 S. Lowry, G. Macpherson, 'A blot on the profession', *British Medical Journal* (296) 1988 (nr. 6623).

57 Dastin 2018 (supra noot 50).

het selectieproces in sluipen. Een manager met vooroordelen over mensen met een migratieachtergrond kan bijvoorbeeld het werk van zijn personeel met deze achtergrond kritischer beoordelen of minder ondersteuning geven. Dit is vervolgens terug te zien in de data door negatieve functioneringsbeoordelingen en prestaties. Een latere manager, die deze vooroordelen niet deelt, zal in beginsel geen reden hebben om aan het algoritme te twijfelen, waardoor er nog steeds minder mensen met een migratieachtergrond in dienst worden genomen – terwijl hij dit zelf zonder het algoritme misschien wel had gedaan.[58]

**Box 4: HR-Analytics**
Steeds meer organisaties maken gebruik van *HR-analytics*: technologieën om met behulp van data en algoritmes de productiviteit van werknemers te beoordelen. Hierdoor kunnen promoties of beoordelingsgesprekken op basis van data plaatsvinden.

Er zijn bij *HR-analytics* zorgen over privacy en discriminatie. Deze instrumenten kunnen gevoelige gegevens verzamelen zoals e-mails, bewegingspatronen of gezichtsuitdrukking. De verwachting is dat het gebruik van *HR-analytics* sterk zal toenemen.

Das, de Jong en Kool 2020 (supra noot 2).

## Aanpassingen op basis van zoek- en klikgedrag

Als een algoritme al in gebruik is, kan een algoritme vanwege nieuwe gebruikersdata (bijvoorbeeld het zoek- en klikgedrag van gebruikers) ook *bias* vertonen. Algoritmes kunnen namelijk bijleren op basis van deze nieuwe informatie. Zo suggereerde LinkedIn's algoritme mannennamen terwijl er vrouwen gezocht werden omdat deze namen vaker werden gezocht door gebruikers (box 5).[59]

Uit onderzoek in de Verenigde Staten bleek dat het opzoeken van typische Afro-Amerikaanse namen op Google vaker leidt tot Google-advertenties die iets te maken hebben met arrestaties en strafbladen (bijvoorbeeld een advertentie met de vraag: 'Heeft deze persoon een strafblad?'). De onderzoeker denkt dat dit mogelijk komt omdat mensen die op Google zochten op Afro-Amerikaanse namen daarna vaak klikten op advertenties die te maken hadden met arrestaties en strafbladen of daarna direct zochten op de trefwoorden 'arrestatie' of 'strafblad'. Hierdoor leerde het Google-algoritme waarschijnlijk dat mensen die informatie zoeken over Afro-Amerikaanse namen, waarschijnlijk ook meer informatie willen over arrestaties en strafbladen. Zodoende zouden advertenties over deze onderwerpen sneller voorkomen bij de zoekopdrachten naar Afro-Amerikaanse namen.[60]

De vooroordelen die bepaalde groepen over zichzelf hebben kan ook *bias* in de data veroorzaken. Als een aantal vrouwen bijvoorbeeld minder snel klikken op managementposities omdat ze denken dat ze als vrouw minder kans maken, dan kan het algoritme niet alleen deze vrouwen naar verloop van tijd minder van dit type vacatures laten zien, maar ook andere vrouwen die niet deze verwachtingen over zichzelf hebben.[61]

**Box 5: 'Bedoelde u Stephen Williams?'**
In 2016 kwam LinkedIn in de Verenigde Staten in opspraak. Als iemand op een vrouwennaam zocht op LinkedIn, suggereerde de zoekmachine soms een mannelijke naam die daarop leek. Zo vroeg LinkedIn 'bedoelde u Stephen Williams?' als er gezocht werd naar 'Stephanie Williams'.

Ook bij andere vrouwennamen kwam dit voor: Danielle werd bijvoorbeeld Daniel en Alexa werd Alex. Volgens LinkedIn kwamen deze suggesties voort uit de zoekdata van gebruikers; deze namen werden vaker gezocht. LinkedIn heeft de zoekfunctie inmiddels aangepast.

M. Day, 'LinkedIn changes search algorithm to remove female-to-male name prompts', Seattle Times 8 september 2016.

Andersom heeft de feedback van HR-professionals ook impact op het algoritme. Zo is het mogelijk dat een website als LinkedIn, minder vaak vrouwen of mensen van een bepaald ras aanbeveelt omdat deze groepen vaker worden 'weggeklikt' vanwege discriminerende vooroordelen bij werkgevers.[62] Het gevolg hiervan kan zijn dat sollicitanten uit een beschermde groep de vacature niet eens te zien krijgen, waardoor ze bij voorbaat

58 Barocas en Selbst 2016 (supra noot 9), p. 687.

59 Barocas en Selbst 2016 (supra noot 9), p. 682.

60 L. Sweeney, 'Discrimination in Online Ad Delivery', *Queue* (11) 2013 (3), p. 14

61 M. Bogen, A. Rieke, Help Wanted: *An Examination of Hiring Algorithms, Equity, and Bias,* Washington DC: Upturn 2018, p. 21.

62 Barocas en Selbst 2016 (supra noot 9), p. 683.

al minder kans maken. Een onderzoek toont bijvoorbeeld aan dat mannen vaker online advertenties zagen voor technische banen en goed betaalde banen dan vrouwen.[63]

### Niet representatieve data

Het tweede geval, niet-representatieve data, doet zich voor als data worden verzameld op een manier waarop bepaalde groepen worden uitgesloten of oververtegenwoordigd zijn. Om natuurrampen te bestrijden wordt bijvoorbeeld gebruik gemaakt van Twitterdata voor crisisinterventies. Het probleem hiervan is echter dat veel mensen geen internet of smartphones hebben waardoor interventies de bevolking slechter bereiken.[64]

Hetzelfde kan gebeuren bij recruitmenttechnologie waarbij gebruik wordt gemaakt van data op sociale media. Veel jonge mensen zitten op sociale media, terwijl ouderen en mensen met een beperking hierop vaak minder actief zijn. Over ouderen en mensen met een beperking zijn dus minder data aanwezig. Dit heeft weer invloed op het 'trainen' van het algoritme waardoor ouderen of mensen met beperking vanwege gebrekkige data anders door het algoritme beoordeeld kunnen worden.[65] Dit heeft dus ook gevolgen voor ouderen en mensen met een beperking die wel handig zijn met een smartphone of een computer, omdat het algoritme minder data heeft verzameld over de groep waartoe zij behoren.[66]

### Verouderde data

Een ander probleem is dat trainingsdata niet meer representatief zijn voor geschikte kandidaten omdat de omstandigheden zijn veranderd. Denk bijvoorbeeld aan de komst van nieuwe technologie (zoals smartphones) of verminderde werkgelegenheid vanwege een pandemie. Dat kan invloed hebben op de dataverzameling.

Hierdoor kan de trainingsdata zijn verouderd met mogelijk discriminerende gevolgen ten aanzien van bepaalde groepen.[67] Stel dat een algoritme mensen die op het moment van solliciteren geen baan hebben minder hoge scores geeft. Bij een plotse golf van werkloosheid kunnen opeens veel jonge flexwerkers hun baan verliezen en hebben opeens veel meer jongeren bij dit algoritme een kleinere kans op een baan, terwijl ze misschien wel goede kandidaten zijn.

Representatie kan daarnaast ook worden beïnvloed omdat kwetsbare groepen zich bewust zijn van mogelijke discriminatie bij het zoeken naar werk. Werkzoekenden uit deze groepen die het gevoel hebben dat ze minder kans hebben op een baan zijn misschien minder gemotiveerd om te solliciteren en te investeren in een opleiding. Dit heeft vervolgens een zichzelf versterkend effect op de trainingsdata, omdat werkzoekenden met deze achtergrond minder vaak in de data voorkomen.[68]

## 4. Verhoogd risico op discriminatie door algoritmes

Nu de oorzaken van discriminatie door algoritmes zijn besproken gaat het in dit deel over hoe algoritmes zelf discriminatie kunnen versterken. Want hoewel mensen ook kunnen discrimineren en *bias* (mede) ontstaat doordat bestaande ongelijkheid in de samenleving door het algoritme wordt gereproduceerd, hebben algoritmes een aantal eigenschappen waardoor de effecten van discriminatie erger worden.[69]

Dit is ook een bevinding in het recente rapport van de speciaal VN-rapporteur voor hedendaagse vormen van racisme, rassendiscriminatie, vreemdelingenhaat en aanverwante onverdraagzaamheid, E. Tendayi Achiume. Zij concludeert dat nieuwe digitale technologieën bestaande ongelijkheden kunnen verergeren.[70]
De risico's van algoritmes om discriminatie te verergeren kwamen ook naar voren in het literatuuronderzoek, deze risico's zullen hier kort worden samengevat.

63 A. Lambrecht, C. Tucker, 'Algorithmic *Bias*? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads', *Management Science* (65) 2019 (7). A. Datta, M.C. Tschantz, A. Datta, 'Automated Experiments on Ad Privacy Settings', *Proceedings on Privacy Enhancing Technologies* 2015.

64 Lammerant, Blok en de Hert 2018 (supra noot 5), par. 3.4.

65 Grommé e.a. 2019 (supra noot 1), p. 21.

66 Barocas en Selbst 2016 (supra noot 9), p. 685.

67 Lammerant, Blok en de Hert 2018 (supra noot 5), par. 3.4.

68 Kim 2017 (supra noot 33), p. 882.

69 Institut Monataigne, *Algorithmes : contrôle des biais S.V.P.,* Parijs: 2020, p. 23. Conseil supérieur de l'égalité professionelle entre les femmes et les hommes, *Égalité entre les femmes et les hommes dans les procedures RH: Le réflexe égalité à chaque étape*, 2019, p. 202.

70 *Racial Discrimination and Emerging Digital Technologies:* A Human Rights Analysis. Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance tot he United Nations Human Rights Council, 2020, A/HRC/44/57. Par. 4.

**Box 6: Platformeconomie en discriminatie**
Platformdiensten, zoals Uber en Airbnb, kunnen ook te maken krijgen met discriminerende input, en vervolgens, als niet wordt ingegrepen, discriminatie in de hand werken. Vaak krijgen de aanbieders op dit soort platforms beoordelingen (*ratings*) van klanten. Onderzoek laat zien dat platformwerkers met een donkere huidskleur en vrouwen slechtere of minder beoordelingen krijgen dan (witte) mannen. Algoritmes die de beoordelingsdata in het zoeksysteem ongefilterd of onbewerkt verder verwerken kunnen discriminerende data op deze platformen vervolgens versterken.

S. Burri, S. Heeger-Hertter, 'Discriminatie in de platformeconomie juridisch bestrijden: geen eenvoudige zaak', *Ars Aequi* 2019. M. Kullmann, 'Platformwerk, besluitvorming door algoritmen en bewijs van algoritmische discriminatie', *Ondernemingsrecht* 2019/8.

## Black box

Ten eerste kunnen algoritmes leiden tot meer discriminatie omdat algoritmes niet transparant of inzichtelijk kunnen zijn. Als discriminatie immers niet ontdekt wordt, kan er weinig aan gedaan worden om het te corrigeren of te verminderen. Vooral zelflerende algoritmes die zichzelf aanpassen kunnen zeer complex worden. Hierdoor kan achteraf niet meer achterhaald worden wat er gebeurd is. Dit geldt zelfs voor de ontwerpers van het algoritme. De belangrijkste reden is dat de code niet meer door mensen te interpreteren is.

Bovendien zijn zelflerende algoritmes vaak geheim en wordt de werking vanwege commerciële redenen niet openbaar gemaakt.[71] Door deze ondoorzichtigheid worden algoritmes ook wel omschreven als een *black box*. Niet-zelflerende algoritmes, met meer eenvoudige als-dan-beslisbomen, zullen veelal inzichtelijker zijn.[72] Voor deze algoritmes is er echter niet altijd een voorkeur, omdat ze soms makkelijker te misleiden zijn.

Zo kan een sollicitant met kwade bedoelingen een trefwoord als 'Oxford' bijvoorbeeld verwerken met een witte tekstkleur in de kop- of voettekst van een CV. Hoewel het trefwoord dan voor het menselijk oog niet zichtbaar is, zal een simpel als-dan-algoritme alle woorden in het document scannen, inclusief de kop- en voetteksten en dus ook het woord 'Oxford' detecteren. Het algoritme zal hierdoor de sollicitant hoger waarderen, ook al heeft hij of zij niet aan Oxford gestudeerd.[73]

De *black box* maakt het verhullen van opzettelijke discriminatie makkelijker. Door variabelen en trainingsdata aan te passen, zoals eerder beschreven, kunnen bepaalde groepen opzettelijk buiten het wervings- en selectieproces vallen.[74] Verboden gronden hoeven ook niet in het algoritme of de data voor te komen om te zorgen dat beschermde groepen bewust worden buitengesloten zoals we eerder zagen.[75] Het probleem is dus niet alleen dat algoritmes discriminatie kunnen veroorzaken, maar ook bewuste discriminatie (blijvend) kunnen verhullen.[76] Bewuste discriminatie kan door algoritmes schuil gaan achter een façade van ogenschijnlijk neutrale data-analyse.

Maar nog belangrijker dan het verhullen van discriminatie is dat sollicitanten vanwege de *black box* niet kunnen achterhalen waarom ze zijn afgewezen. Hierdoor kan discriminatie moeilijker te ontdekken zijn. Het is traditioneel al lastig om discriminatoire bedoelingen te achterhalen. Meestal wordt er gebruik gemaakt van een schijnbaar neutrale maatregel die een beschermde groep harder raakt.

Door het gebruik van algoritmes zal dit nog lastiger zijn, omdat er nog een 'vertaalslag' overheen is gegaan. Ook tast de *black box* de effectiviteit van rechtsmiddelen aan om discriminatie aan te kaarten, en waar nodig te veroordelen. Het is dus moeilijker om je recht te halen.[77] Dit is schrijnend omdat discriminatie strafrechtelijk vervolgbaar is en omdat de *black box* potentieel de toepassing van de gelijkebehandelingswetgeving op algoritmes bemoeilijkt.[78]

71 Adriaansz 2020 (supra noot 19), par. 4.

72 Van den Braak en Choenni 2019 (supra noot 13), p. 24.

73 Das, de Jong en Kool 2020 (supra noot 2), p. 35.

74 Barocas en Selbst 2016 (supra noot 9), p. 692.

75 Vetzo, Gerards en Nehmelman 2018 (supra noot 30), p. 145.

76 Kim 2017 (supra noot 33), p. 884.

77 M. Vetzo, J.H. Gerards, 'Algoritme-gedreven technologieën en grondrechten', *Computerrecht* 2019/3, par. 3.2.2. Comittee of experts on internet intermediaries (MSI-NET), *Algorithms and Human Rights: Study on the human rights dimensions of automated data processing techniques and possible regulatory implications*, Strasbourg: Council of Europe 2018, p. 24.

78 Art. 137g Sr en 429quater Sr.

## Automation bias

Ten tweede versterken *biased* algoritmes discriminatie omdat mensen onterecht denken dat juist de meest objectieve uitkomst wordt bereikt door wiskundige berekeningen.[79] Hoewel een algoritme wiskundig gezien objectief is, zijn de keuzes voor variabelen en data menselijke keuzes. Computers worden veelal geassocieerd met rationaliteit en foutloosheid en mensen proberen mogelijk hun verantwoordelijkheid te minimaliseren door het volgen van computers. Dit fenomeen wordt in de literatuur ook wel *automation bias* genoemd.[80]

Uit gesprekken met HR-professionals ontstaat de indruk dat organisaties vertrouwen hebben in recruitmenttechnologie en zich niet altijd bewust zijn van de risico's op discriminatie.[81] Een recruiter zegt in de krant zelfs een 'bijna blind' vertrouwen te hebben in selectiealgoritmes.[82] Vaak gebeurt werven en selectie onder hoge tijdsdruk waardoor het risico op discriminatie groot is.[83] Juist bij werving en selectie zal het risico op *automation bias* dus reëel zijn.

## Systematiseren van discriminatie

Ten derde kunnen algoritmes discriminatie verder inbedden en systematiseren in de samenleving. Algoritmes zijn in beginsel consequent. Besluiten genomen op basis van vooropgestelde parameters in een algoritme waarin *bias* voorkomt zal deze parameters in beginsel op alle besluiten toepassen.[84] Mensen daarentegen – hoewel niet vrij van vooroordelen – hebben keuzevrijheid en intuïtie om zich bewust te worden van vooroordelen en deze af te leren.[85]

Ook kan een *biased* algoritme op grote schaal in gebruik genomen worden en daarmee bredere effecten hebben. Zo kan een algoritme, getraind op grond van data uit een bepaalde regio waar veel racisme voorkomt, gebruikt worden in een omgeving waar racisme minder vaak plaatsvindt. De *bias* die het algoritme vertoont kan dan geëxporteerd worden naar een andere regio.[86]

Bovendien hebben zelflerende algoritmes met *bias* eerder de neiging om *bias* te bevestigen en te versterken dan te ontkrachten. Een *biased* algoritme kan bijvoorbeeld mannen een hogere ranking geven bij een sollicitatie. Hierdoor worden er waarschijnlijk ook meer mannen aangenomen. Een zelflerend algoritme dat leert van de keuzes van recruiters zal vervolgens mannen een nog hogere ranking geven, waardoor er een vicieuze cirkel ontstaat (*feedback loops*).[87]

Daarnaast zullen werkgevers minder in contact komen met werknemers uit minderheidsgroepen waardoor zij minder gelegenheid hebben om de prestaties van deze minderheidsgroepen te observeren en minder redenen hebben om te twijfelen aan het algoritme.[88] Met andere woorden: het inzetten van algoritmes creëert en versterkt een alternatieve realiteit.

## Meer mogelijkheden om onderscheid te maken

Ten vierde vergroten algoritmes, in samenhang bezien met nieuwe technieken om data te verzamelen, het risico op ongelijke behandeling omdat deze techniek zeer veel complexe verbanden kan leggen. Hierdoor is het risico groter dat groepen geclassificeerd kunnen worden op basis van allerlei neutrale gegevens die onder de traditionele discriminatiegronden vallen zoals of iemand rijk of arm is.[89]

Ook kunnen algoritmes meer verbanden ontdekken tussen schijnbaar neutrale informatie en discriminatiegronden, waardoor algoritmes een niet-gekende *bias* ten aanzien van bepaalde groepen kunnen bevatten.[90] Kortom, er zijn meer mogelijkheden om mensen te differentiëren en dus ook om te discrimineren.

De bovenstaande risico's laten zien dat voorzichtigheid geboden is bij de inzet van algoritmes. Ondanks deze risico's worden algoritmes ook ingezet om discriminatie tegen te gaan. Hierop zal in het volgende hoofdstuk worden ingegaan.

79 D. Mijnheer, 'De zoektocht naar een "eerlijk" gebruik van kunstmatige intelligentie', *Tijdschrift voor compliance* 2020.

80 Zuiderveen Borgesius 2018 (supra noot 39), p. 8.

81 Das, de Jong en Kool 2020 (supra noot 2), p. 20. Grommé e.a. 2019 (supra noot 1), p. 40. G. Zwenne, 'Wat heb ik aan transparantie van een algoritme', SC 2016/19.

82 C. Don, 'Algoritme geeft werkzoekende sollicitatie-advies', *NRC* 27 juli 2017.

83 Juist door technologie in te zetten kunnen bedrijven voldoen aan normen om binnen een korte tijd te reageren op alle sollicitanten zoals opgesteld in de sollicitatiecode stelt van Ulden 2015 (supra noot 3), p. 22.

84 MSI-NET 2018 (supra noot 77), p. 26. L. Edwards, M. Veale, 'Slave to the algorithm? Why a "right to an explanation" is probably not the remedy you are looking for', *Duke Law & Technology Review* (16) 2017 (nr. 18), p. 26.

85 M. Kullmann, 'Discriminating job applicants through algorithmic decision-making', *Ars Aequi 2019*, p. 49.

86 Institut Monataigne 2020 (supra noot 69), p. 23.

87 I. Smeets, '510 staandehoudingen en de zelfversterkende feedback-loop', ionica.nl 12 mei 2020.

88 Kim 2017 (supra noot 33), p. 882, 895, 896.

89 Zuiderveen Borgesius 2018 (supra noot 39), p. 35.

90 Lammerant, Blok en de Hert 2018 (supra noot 5), par. 3.1.

# 5. Discriminatie met algoritmes tegengaan

De risico's van algoritmes zijn dominant in het publieke debat.[91] Desondanks kunnen algoritmes ook tegen discriminatie ingezet worden. Zo heeft het College in 2018 en 2020 voor een onderzoek naar discriminerende vacatureteksten op leeftijd een algoritme van de Vrije Universiteit ingezet om meer dan 4,9 miljoen online vacatures te doorzoeken.[92] Daarnaast is het goed om niet te vergeten dat werving en selectie zonder technologie ook niet neutraal hoeft te zijn. Ook mensen kunnen bedoeld en onbedoeld discrimineren.

Verschillende onderzoeken tonen aan dat bepaalde groepen (zoals vrouwen, etnische minderheden, religieuze groepen, mensen met een beperking, homoseksuelen en minder aantrekkelijke personen) structureel worden benadeeld bij werving en selectie en minder positieve reacties ontvangen naar aanleiding van hun sollicitatie.[93]

Het is dan ook maar de vraag of mensen het wel beter doen dan algoritmes bij werving en selectie. Sterker nog, in de voorgaande secties werd beschreven dat algoritmes vaak *bias* vertonen *omdat* ze zijn getraind op data of feedback die aangetast zijn door de vooroordelen die in de samenleving leven.[94]

## Streven naar objectieve algoritmes

Algoritmes hebben de potentie om objectiever te zijn dan mensen en kunnen worden ingezet om (onbewuste) discriminatie door mensen bloot te leggen. Eeuwenoude menselijke vooroordelen bij werving en selectie kunnen juist worden gecorrigeerd door de inzet van algoritmes, stellen voorstanders van het gebruik van algoritmes.[95] Volgens het bedrijf Unilever zijn sinds dit jaar vijftig procent van alle managers vrouw – mede – vanwege het gebruik van selectiealgoritmes (box 7).[96]

**Box 7: Unilever's middelen om gendergelijkheid te bereiken**

Unilever maakt gebruik van gendergelijke interviewvragen en streeft ernaar om hun personeelsbestand divers en inclusief te krijgen. Het concern houdt het aantal benoemingen van vrouwen als senior leaders bij. Daarnaast maakt Unilever gebruik van algoritmes om gendergelijkheid te bereiken. Zo gebruikt het bedrijf videointerviewsoftware en moeten sollicitanten spelletjes spelen die kandidaten op gewenste persoonlijkheidskenmerken testen.

Het gebruik van videointerviewsoftware (zie box 3) en spelletjes is echter niet zonder risico. Zo is er weinig onderzoek gedaan naar de validiteit van online games (Grommé e.a. 2019 (supra noot 2), p. 40).

Een zelflerend algoritme kan subtiele kenmerken van geschikte kandidaten herkennen waardoor een beter beeld van een individu ontstaat. In plaats van de typische *high-potentials* krijgen kandidaten die HR-medewerkers volgens oude methoden niet zouden overwegen juist wel een kans.[97]

## Bias uit algoritmes halen

Om discriminatie door algoritmes te voorkomen, proberen algoritmeontwerpers *bias* uit algoritmes te halen. LinkedIn heeft bijvoorbeeld in 2018 zijn algoritme aangepast zodat zoekresultaten een correcte genderbalans hebben.[98] Het bedrijf Pymetrics, dat online games (assessments) aanbiedt aan onder andere Unilever, heeft een (tegen)algoritme ontwikkeld om *bias* in een algoritme te detecteren.[99]

91 Institut Montaigne 2020 (supra noot 69), p. 22.

92 A. Fokkens & C.J. Beukeboom, *'Leeftijdsdiscriminatie in vacatureteksten: Een herhaalde geautomatiseerde inhoudsanalyse naar verboden leeftijd-gerelateerd taalgebruik in vacatureteksten uit 2017 en 2019'*. Rapport in opdracht van het College voor de Rechten van de Mens, 2020 & College voor de Rechten van de Mens, *'Gezocht: Jonge Hond' – Onderzoek naar de omvang en het effect van leeftijdsdiscriminatie in vacatureteksten,* 2018.

93 Zie inleiding van dit hoofdstuk.

94 Zuiderveen Borgesius 2018 (supra noot 39), p. 21. Zo kwam Google in opspraak omdat de zoekmachine bij de zoekopdracht 'drie zwarte tieners' portretfoto's van zwarte gevangenen lieten zien. De zoekopdracht 'drie witte tieners' leverde echter lachende witte tieners op. Volgens Google kwam dit verschil voort uit het klik- en zoekgedrag van de gebruikers van Google. Het ligt volgens Google dus niet aan algoritme, maar aan de samenleving. A. Allen, 'The "three black teenagers" search shows it is society, not Google, that is racist' *the Guardian* 10 juni 2016.

95 J. Kleinberg, J. Ludwig, S. Mullainathan, C.R. Sunstein. 'Discrimination in the age of algorithms' Journal of Legal Analysis (10) 2018.

96 Unilever, 'Unilever bereikt wereldwijd gelijk percentage vrouwen en mannen in managementfuncties', unilever.nl 3 maart 2020.

97 White House 2016 (supra noot 9), p. 16.

98 S. Cem Geyik, K. Kenthapadi, 'Building Representative Talent Search at LinkedIn', engineering.linkedin.com 10 oktober 2018.

99 K. Johnson, 'Pymetrics open-sources Audit AI, an algorithm *bias* detection tool', *Venturebeat* (online, bijgewerkt 31 maart 2018).

Ook grote bedrijven als Microsoft en Facebook werken aan dit soort oplossingen.[100]

Daarnaast zijn er ook technieken die de trainingsdata van algoritmes analyseren om te corrigeren op *bias*.[101] Verder onderzoek naar het verbeteren van algoritmes en bevorderen van gelijke behandeling in de ontwerpfase (non-discrimination by design) is essentieel om discriminatie door algoritmes te voorkomen.[102]

## Voorzichtigheid is geboden

Ondanks de potentie van algoritmes of (tegen)algoritmes om op technische wijze *bias* te corrigeren, moeten organisaties ook zelf alert blijven bij het gebruik van algoritmes. De technieken om *bias* te corrigeren zijn nog volop in ontwikkeling en er zijn ook aan algoritmes inherente discriminatieverergerende risico's die aan het gebruik ervan kleven (zie 4.). Er is nog steeds discussie over *good practices* (goede methoden) en ontwikkelaars van recruitmenttechnologie focussen met name op het corrigeren van discriminatie op grond van geslacht of ras. Ook is er weinig transparantie over hoe bedrijven *bias* aanpakken.[103]

Daarbij komt ook het risico dat correctie van *bias* in een algoritme tegen een bepaalde groep kan leiden tot meer *bias* tegen een *andere* beschermde groep. Hierdoor verplaatst het discriminatieprobleem zich, in plaats van dat het weggenomen wordt.[104]

Ook het simpelweg weglaten van beschermde gronden in een algoritme is niet genoeg om discriminatie tegen te gaan. Zoals eerder beschreven, kan andere informatie verband houden met specifieke karakteristieken van bepaalde beschermde groepen waardoor er alsnog discriminatie kan plaatsvinden.

Bovendien zullen technieken om *bias* te corrigeren paradoxaal genoeg wellicht juist gebruik moeten maken van discriminatiegronden.[105] LinkedIn gebruikt bijvoorbeeld '*gender buckets*' om in het eerdergenoemde voorbeeld een genderbalans in zoekresultaten te krijgen.

Ook geven deze gronden meer context aan de situatie waarin mensen zich bevinden. Stel dat etnische minderheden die in militaire dienst zijn geweest beter presteren op werk. Deze afweging wordt niet meegenomen als een algoritme een negatief statistisch verband legt tussen werkprestaties en militaire dienst in het algemeen. Als een werkgever op dit algoritme zou vertrouwen krijgen etnische minderheden dus juist minder kans dan ze op basis van hun eigen prestaties zouden verdienen, omdat het algoritme geen onderscheid maakt naar beschermde groepen.[106]

De ironie zou dan dus zijn dat om gelijk behandeld te worden door algoritmes, mensen uit beschermde groepen in sommige omstandigheden op hun sollicitatieformulieren juist hun ras of geloofsovertuiging zouden moeten opgeven. De vraag is of dit wenselijk is gezien het risico op directe discriminatie en omdat dit soort gegevens volgens de algemene verordening gegevensbescherming niet zomaar verwerkt mogen worden.[107]

Ook moeten HR-professionals voorzichtig zijn met het gebruik van deze technieken. Het resultaat kan immers zijn dat er overgecompenseerd wordt voor beschermde groepen wat in feite zou leiden tot voorkeursbeleid. Hieraan worden strenge voorwaarden gesteld in de gelijkebehandelingswetgeving.

100 K. Wiggers, 'Microsoft is developing a tool to help engineers catch *bias* in algorithms', *Venturebeat* (online, bijgewerkt 25 mei 2018). D. Gershgorn, 'Facebook says it has a tool to detect *bias* in its artificial intelligence', *Quartz* (online, bijgewerkt 3 mei 2018).

101 F. Kamiran, T. Calders, M. Pechenizkiy, 'Techniques for Discrimination-Free Predictive Models', In: B. Custers, T. Calders, B. Schermer, T. Zarsky (red.), *Discrimination and Privacy in the Information Society: Data Mining and Profiling in Large Databases,* Heidelberg/New York/Dordrecht/London: Springer 2013.

102 Lammerant, Blok en de Hert 2018 (supra noot 5), par. 6.

103 Bogen en Rieke 2018 (supra noot 61), p. 38.

104 A.G. King, M. Mrkonich, '"Big Data" and the Risk of Employment Discrimination', *Oklahoma Law Review* (68) 2016, p. 580.

105 I. Zliobaite, B. Custers, 'Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models', *Artificial Intelligence and Law* (24) 2016.

106 Kim 2017 (supra noot 33), p. 878.

107 Hierbij dient opgemerkt te worden dat verwerking van bijzondere persoonsgegevens mogelijk is als dit 'noodzakelijk is met het oog op uitvoering van verplichtingen of uitoefening van specifieke rechten van de potentiele werkgever of de sollicitant op het gebied van arbeidsrecht' (artikel 9 lid 2 sub b AVG). Hierbij kan gedacht worden aan bijzondere gegevens die nodig zijn voor het voeren van voorkeursbeleid. Grommé e.a. 2019 (supra noot 1), p. 83.

# 6. Conclusie

Het doel van dit onderzoek is om te verhelderen hoe de inzet van algoritmes bij werving en selectie kan leiden tot discriminatie. Ter illustratie zijn hiervoor verschillende voorbeelden van discriminatie door algoritmes gebruikt die betrekking hadden op zowel werving en selectie als andere situaties.

In het algemeen zijn de voorbeelden van discriminatie door algoritmes bij werving en selectie schaars. Dit betekent niet dat discriminatie door algoritmes bij werving en selectie niet voorkomt. Het gebrek aan zichtbaarheid betekent dat het extra relevant is om te begrijpen wat de oorzaken van discriminatie door algoritmes zijn. Hierdoor draagt dit onderzoek bij aan het algemene bewustzijn over de discriminatierisico's van algoritmes onder werkgevers, beleidsmakers en werkzoekenden.

## De duivel zit in de details

Het onderzoek laat zien dat details een belangrijke rol spelen bij het ontstaan van discriminatie door algoritmes. Keuzes bij het ontwerp van het algoritme voor bepaalde variabelen kunnen leiden tot *bias*, zelfs als deze variabelen op het eerste gezicht neutraal lijken. Zo kunnen postcodes een indicatie zijn van een migratieachtergrond als deze hoort bij een woonwijk waar veel migranten wonen. Daarnaast kan de data die gebruikt worden om een algoritme te trainen ongemerkt *bias* bevatten. Het klikgedrag van bevooroordeelde sollicitanten en recruiters op websites kan er bijvoorbeeld toe leiden dat vrouwen minder online advertenties te zien krijgen voor technische functies.

## Risico's en kansen

Uit het onderzoek blijkt dat de aard van algoritmes risico's met zich meebrengen waardoor discriminatie versterkt wordt. Centraal hierbij staat dat algoritmes vaak ondoorzichtig zijn (de *black box* problematiek) waardoor het moeilijk is om discriminatie door een algoritme te achterhalen. Ook hebben mensen vanwege *automation bias* de neiging om de uitkomsten van algoritmes zonder twijfel te volgen.

Ondanks de risico's, hebben algoritmes ook de potentie om discriminatie tegen te gaan en objectiever te zijn dan mensen. Technieken om *bias* uit algoritmes te halen staan echter nog in de kinderschoenen en er zijn nog steeds risico's.

Verder onderzoek naar het verbeteren van algoritmes en bevorderen van gelijke behandeling in de ontwerpfase (*non-discrimination by design*) is essentieel om discriminatie door algoritmes te voorkomen.

Bovendien zullen technieken om *bias* te corrigeren vaak juist gebruik moeten maken van discriminatiegronden. Het is de vraag of dit wenselijk is gezien het risico op directe discriminatie en omdat dit soort gegevens volgens de algemene verordening gegevensbescherming niet zomaar verwerkt mogen worden. Bij de inzet van algoritmes bij werving en selectie is dus voorzichtigheid geboden.

## Vervolgstappen van het College

Met de komst van algoritmes in het wervings- en selectieproces ontstaan nieuwe uitdagingen op het gebied van mensenrechten. Dit onderzoek heeft nader geïllustreerd dat niet enkel privacy, maar ook discriminatie een risico is bij de inzet van algoritmes in het wervings- en selectieproces. In het kader van het strategische programma Digitalisering & Mensenrechten (box 1) van het College, zal het College hier aandacht aan geven.

Om het bewustzijn van de risico's op discriminatie door algoritmes bij werving en selectie te vergroten zal het College op basis van dit onderzoek in gesprek gaan met de relevante bewindspersonen, beleidsmakers, werkgevers- en werknemersorganisaties en toezichthouders zoals de Autoriteit Persoonsgegevens en de Inspectie Sociale Zaken en Werkgelegenheid (SZW). Daarnaast zal het College in zijn publieksvoorlichting meer aandacht besteden aan discriminatie door algoritmes om de bewustwording onder het algemene publiek te vergroten.

Verder zal het College een vervolgonderzoek uitbrengen waarop ingegaan zal worden op de rol van de gelijkebehandelingswetgeving bij de regulering van algoritmes. Hierbij bouwt het College voort op de aanbevelingen in het recente TNO-onderzoek *Digitale Arbeidsmarktdiscriminatie* dat uitgevoerd is in opdracht van de Inspectie SZW.[108] De onderzoekers van het TNO-onderzoek constateren dat er weinig rechtspraak en onderzoek is over hoe de eisen van het gelijkebehandelingsrecht zich verhouden tot (gedeeltelijk) geautomatiseerde algoritmes.

108 Grommé e.a. 2019 (supra noot 1), p. 60-62.

# Aanbevelingen

## Bewustwording

Het is essentieel dat er meer bewustzijn ontstaat over de discriminatierisico's van de inzet van algoritmes bij het wervings- en selectieproces. Dit onderzoek draagt bij aan de maatschappelijke discussie hierover. In dit licht beveelt het College het volgende aan:

- **Voor de overheid:** Vergroot het bewustzijn onder burgers over de risico's van discriminatie door algoritmes door middel van voorlichting.
- **Voor de overheid:** Bespreek met werkgevers de risico's van algoritmes bij werving en selectie.
- **Voor algoritmeontwerpers:** Informeer afnemers van software over de mogelijke risico's van de inzet van algoritmes bij werving en selectie.

## Transparantie

Discriminatie voorkomen bij werving en selectie vraagt om transparantie en uitleg waarom bepaalde afwegingen zijn gemaakt. Echter, juist algoritmes worden gekenmerkt door niet-transparante procedures waardoor discriminatie vaak niet goed zichtbaar is. Daarom moet het volgende onderstreept worden:

- **Voor algoritmeontwerpers:** Maak op een begrijpelijke manier inzichtelijk op basis van welke variabelen sollicitanten beoordeeld worden.
- **Voor werkgevers:** Informeer sollicitanten op een begrijpelijke manier voorafgaand aan de sollicitatieprocedure over de rol en de werking van algoritmes bij de selectie.
- **Voor werkzoekenden:** Maak melding indien u denkt dat u gediscrimineerd bent door een wervings- en selectiealgoritme. Dit kan ook als u alleen een vermoeden heeft. U kunt hiervoor terecht bij een van de discriminatiemeldpunten (kijk op www.discriminatie.nl voor meer informatie). Ook kunt u bij het College terecht voor een verzoek om een oordeel, het melden van discriminatie of vragen.

## Preventie

Discriminatie is bij wet verboden. Schijnbaar neutrale variabelen of data kunnen alsnog indirect leiden tot *bias* in een algoritme. Daarbij komt dat technieken om *bias* uit algoritmes te halen nog steeds in de kinderschoenen staan. Gezien deze omstandigheden beveelt het College het volgende aan:

- **Voor de overheid:** Geef voorlichting aan algoritme-ontwerpers over hun wettelijke verplichtingen vanuit (onder andere) de gelijkebehandelingswetgeving.
- **Voor algoritmeontwerpers:** Voer met regelmaat validaties uit om te controleren of uw software niet leidt tot discriminatie.
- **Voor werkgevers:** Gebruik geen recruitment-technologie waarbij het niet duidelijk is op basis van welke afwegingen de software beslissingen maakt.