



High-Tech versus High-Tech: hoe digitale discriminatie digitaal te detecteren

Claartje Thijs

Aanbevolen citeerwijze bij dit artikel

Claartje Thijs, 'High-Tech versus High-Tech: hoe digitale discriminatie digitaal te detecteren', *Beleidsonderzoek Online* mei 2019, DOI: 10.5553/BO/221335502019000005001

1 Inleiding

In veel landen, waaronder Nederland, is gelijke behandeling een fundamenteel grond- en mensenrecht en is het wettelijk verboden te discrimineren op kenmerken zoals geslacht, etniciteit, seksuele geaardheid en religie. Het is daarom van groot belang om het recht om niet gediscrimineerd te worden in de praktijk te beschermen en op naleving te toetsen. Dat is een voorname taak van het College voor de Rechten van de Mens (hierna 'het College'). Zoals op zijn website te lezen valt, beschermt, belicht en bevordert het College de mensenrechten in Nederland, onder andere door onderzoek, advies en voorlichting. Mensen die zich gediscrimineerd voelen, kunnen dit melden bij het College. Het College beoordeelt dan of er ook daadwerkelijk sprake is van verboden onderscheid.

Met de intrede van het grootschalige gebruik van digitale technologie zijn nieuwe (verschijnings)vormen van discriminatie ontstaan. Als voorbeeld hiervan gaat het in dit artikel specifiek over arbeidsmarktdiscriminatie. Daar waar een discriminerende vacaturetekst in een krant relatief eenvoudig als zodanig kan worden herkend door een potentiële sollicitant of controlerende instantie zoals het College, is het detecteren van een discriminerende vacature die zich aan het directe zicht onttrekt (bijvoorbeeld doordat deze door online platforms als Facebook en Google slechts aan een selecte deelverzameling van alle potentiële geïnteresseerden wordt aangeboden), een geheel nieuwe en ook pittige uitdaging. Het is immers vrijwel onmogelijk om erachter te komen wat je als gebruiker

van dergelijke platforms als gevolg van een verborgen vorm van wettelijk verboden onderscheiding zelf niet eens te zien krijgt.

De positieve keerzijde van deze technologische medaille is dat algoritmes en big data weliswaar nieuwe manieren van discriminatie in de hand werken, maar dat ze tevens kunnen worden ingezet om deze discriminatie te traceren en te bestrijden. Dit wordt mede geïllustreerd door een onderzoek dat het College in samenwerking met de Vrije Universiteit uitvoerde naar verboden leeftijdsonderscheid in vacatureteksten. In dit artikel lichten we beide zijden van de medaille toe.

2 Digitale technologie maakt nieuwe vormen van arbeidsmarktdiscriminatie mogelijk

De manier waarop vraag naar en aanbod van arbeid bij elkaar komen, is met de komst van het internet en sociale media drastisch veranderd. Dit brengt kansen met zich mee: werkgevers kunnen zich bij de werving veel directer op potentiële werknemers richten en zijn niet meer gebonden aan de doelgroep van bijvoorbeeld een dagblad. Het geeft echter ook risico's, omdat het tot uitsluiting van groepen kan leiden. Uitsluiting die voor de uitgesloten vaak niet eens zichtbaar is.

Door hun enorme aantallen gebruikers zijn online platforms zoals Facebook, Google en LinkedIn aantrekkelijk voor werkgevers om er hun vacatures op gerichte wijze op uit te zetten. De hulpmiddelen die bijvoorbeeld Facebook daarbij biedt (Speicher et al., 2018), zijn 'attribute based targeting', 'custom audience targeting' en 'look-alike audience targeting'. Bij attribute based targeting kan de adverteerder gebruikers selecteren die bepaalde eigenschappen (de zogeheten 'attributes') hebben en vervolgens alleen aan deze gebruikers de vacature (laten) tonen. Deze eigenschappen worden deels door de gebruikers zelf expliciet opgegeven, maar ook deels bepaald aan de hand van het online browse-gedrag. Bij custom audience targeting dient de adverteerder een lijst te geven met zogeheten PII's: persoonlijke identificeerbare informatie(eenheden), zoals postcodes en geboortedata. Facebook legt deze lijst naast zijn gebruikersbestand en toont de advertentie aan de gebruikers met de opgegeven PII's. Tot slot is look-alike audience targeting een variant van audience targeting waarbij Facebook de advertentie (al dan niet) laat zien aan een groep die een bepaalde mate van overeenkomst vertoont met een door de adverteerder samengestelde 'source audience', die doorgaans relatief bescheiden van omvang is in vergelijking met het door Facebook gegenereerde look-alike audience.

Al deze drie door Facebook aangeboden targeting-opties kunnen aanleiding geven tot onwettige uitsluiting. Zo toonde de Amerikaanse

non-profitorganisatie ProPublica (die sinds de oprichting in 2007 in samenwerking met diverse nieuwsorganisaties onderzoeksjournalistiek bedrijft) in december 2017 aan dat Facebook adverteerders de mogelijkheid bood vacatures slechts uit te zetten onder een specifieke leeftijdsgroep, zoals mensen tussen de 25 en 36 jaar (Angwin et al., 2017). Uit het onderzoek bleek zelfs dat organisaties zoals Verizon, Amazon, Goldman Sachs, Target, maar ook Facebook zelf, hier daadwerkelijk gebruik van maakten en zich dus schuldig maakten aan directe leeftijdsdiscriminatie bij de werving. Facebook verdedigde zich met het argument dat, indien op verantwoorde wijze gebruikt, leeftijdstargeting helpt om mensen uit iedere leeftijdscategorie aan passend werk te helpen. Facebook verschaft gereedschappen, maar het is aan de adverteerders om deze op verantwoorde wijze in te zetten bij de werving. Deze verschillende standpunten illustreren op treffende wijze het spanningsveld tussen de voor- en nadelen van gericht adverteren.

Een ander voorbeeld betreft (vrijwel) directe discriminatie naar etniciteit (Speicher et al., 2018) die ontstond doordat Facebook adverteerders toestond het attribute 'etnische affiniteit' te gebruiken bij hun selectie. Deze eigenschap betrof weliswaar niet de etniciteit van de Facebook-gebruiker zelf, maar wel de interesse van deze gebruiker in etnische gemeenschappen. Op basis hiervan konden werkgevers relatief eenvoudig een doelgroep bepalen die onwettig onderscheidend was. Na de mediareel die hieruit ontstond, verwijderde Facebook het attribute meteen als mogelijke eigenschap om op te selecteren. Het verwijderen van de direct onderscheidende attributes uit de lijst van selectieparameters lijkt een eenvoudige oplossing voor de problemen. Het blijkt voor adverteerders echter nog altijd mogelijk om sterk discriminerende selecties toe te passen door een slim gekozen combinatie van individuele, op zichzelf onschuldig lijkende kenmerken te gebruiken (Speicher et al., 2018).

Tot slot is duidelijk dat ook look-alike audience targeting risico's van ongewenste uitsluiting met zich meebrengt. Zo kan arbeidsdiscriminatie in een lokaal personeelsbestand van relatief kleine omvang (de source audience) doorwerken in de uitgebreidere, minder lokale look-alike audience waaronder de nieuwe vacatures worden uitgezet. Dit mogelijke indirecte onderscheid is eigenlijk voor niemand zichtbaar. In de samenstelling van de nieuwe uitgebreide doelgroep kan zowel opzettelijke als onbedoelde discriminatie bij het samenstellen van de source audience (ook nu weer opzettelijk of onbedoeld) optreden (Speicher et al., 2018). Ook hier blijkt dus dat het voordeel voor de werkgever om een vacature gericht uit te zetten onder een grotere populatie (hetgeen dus leidt tot een efficiëntere en goedkopere procedure) een verhoogd risico op arbeidsdiscriminatie kan betekenen.

Bovenstaande voorbeelden hebben voornamelijk betrekking op Facebook. Facebook is echter een innovatieve marktleider in de ontwikkeling van gericht adverteren, en sociale netwerken als Twitter, Pinterest, LinkedIn en YouTube zijn geneigd om de innovaties over te nemen. De omvang van de problemen is daarmee in potentie groot.

3 Digitale technologie maakt de detectie van nieuwe arbeidsmarktdiscriminatie mogelijk

Om goed te kunnen onderzoeken in welke mate groepen bij de werving via digitale platforms als LinkedIn, Google en Facebook worden uitgesloten als gevolg van gerichte advertenties, zouden de gebruikte selectiemechanismen bekend moeten zijn, en zouden de profielen van gebruikers van deze platformen bekeken moeten kunnen worden om te zien welke advertenties erop worden getoond. Dit is problematisch door de enorme omvang van de data alsook door hun beperkte publieke toegankelijkheid. Hierdoor ontstaat de noodzaak om dergelijk onderzoek uit te voeren met behulp van dezelfde technologieën als die hierboven zijn beschreven om mensen op onwettige gronden uit te sluiten: namelijk door algoritmes toe te passen op big data. Dergelijke data kunnen door gebruikers vrijwillig beschikbaar zijn gesteld, maar kunnen ook zijn verkregen door voldoende grote aantallen gesimuleerde profielen aan te maken op de betreffende digitale platforms.

Zo ontwikkelde ProPublica (Angwin et al., 2017) software die indirect inzicht geeft in de gemaakte selecties: vrijwilligers kunnen een applicatie installeren die bijhoudt welke advertenties worden getoond op hun Facebook-pagina. Deze zogeheten ‘Facebook Political Ad Collector’ werd in eerste instantie ontwikkeld om de gerichte verspreiding van politieke advertenties te onderzoeken, en geeft de volgende uitleg over zichzelf:

When users log into Facebook, the tool will collect the ads displayed on the user’s news feed and guess which ones are political. The tool will also show users political ads that others are seeing, but that weren’t targeted at them.

The tool does not collect any personally identifiable information, and will not learn which ads are shown to which user. The political ads that are collected will be contributed to a public database.

Als bijvangst werd inzicht verkregen in hoe adverteerders bepaalde categorieën werkzoekenden direct of indirect uitsloten. Dit leidde onder andere tot de constatering van discriminatie op grond van leeftijd zoals beschreven in de vorige paragraaf. ProPublica slaagde er ook in om zonder problemen op Google en LinkedIn te adverteren binnen de leeftijdsgroep tot 40 jaar. Naar aanleiding van het

onderzoek heeft LinkedIn zijn werkwijze aangepast (Angwin et al., 2017).

Speicher et al. (2018) betogen dat noch de intentie van de adverteerder noch het proces dat tot de selectie leidt, eenvoudig te bestuderen is. Bovendien bieden beide geen goede maatstaf om discriminatie te kwantificeren. Zoals reeds opgemerkt kan ook een combinatie van onschuldige attributes zoals aangereikt door Facebook immers leiden tot een discriminerende selectie door de adverteerder. In dat geval is noch de adverteerder noch het proces dat Facebook aanbiedt onomstotelijk als tekortschietend aan te wijzen. Speicher et al. stellen daarom voor om de uiteindelijke samenstelling van de groep van mensen die benaderd wordt bij de targeting als maatstaf te nemen om discriminatie te kwantificeren. Zelfs indien gecompliceerde algoritmes worden ingezet om deze groep te definiëren, en ongeacht de al dan niet discriminerende intentie van de adverteerder, is het bestuderen van de uiteindelijke samenstelling van die groep (relatief) eenvoudig.

In Speicher et al. (2018) worden de drie hulpmiddelen die Facebook aanreikt om gericht te adverteren ('attribute based targeting', 'custom audience targeting', en 'look-alike audience targeting') kwantitatief onderzocht op discriminatie. Door gebruik te maken van publiek beschikbare data (zoals in de VS de namen, adressen en telefoonnummers van alle geregistreerde stemmers) konden, soms met behulp van bedrijven die uit publieke data andere data kunnen voorspellen, binnen Facebook custom audiences worden samengesteld die met onevenredig grote waarschijnlijkheid een bepaald gevoelig attribute (zoals geslacht, etniciteit of leeftijd) bevatten. Hieruit blijkt dat custom audience targeting het adverteerders niet erg moeilijk maakt om op oneigenlijke gronden mensen uit te sluiten. Het voorbeeld uit de vorige paragraaf over het attribute 'ethnic affinity' liet al zien dat ook attribute based targeting directe discriminatie mogelijk maakt. Bovendien laat slim geautomatiseerd gebruik van de Facebook Ad Interface (het programma van Facebook waarmee de advertentiedoelgroep kan worden bepaald) zien dat de combinatie van onschuldige attributen indirecte discriminatie door adverteerders mogelijk maakt, zelfs op kenmerken als geloof en seksuele gerichtheid. Tot slot wordt aangetoond dat een source audience met een omvang van ongeveer 10.000 mensen, waarin bewust een bias is aangebracht, door Facebook middels look-alike audience targeting uitgebreid kan worden tot een omvang van ruim vier miljoen, en dat de bias hierin aanwezig blijft.

Misschien wel het meest geavanceerde onderzoek naar discriminatie als gevolg van gericht adverteren is dat van Datta et al. (2015). Hierin is Google het onderwerp van studie, en met name (de invloed van) de Google Ad Settings. Deze settings zijn bedoeld om de gebruikers

informatie en controle over advertentie-settings te verschaffen. Met behulp van een geavanceerd geautomatiseerd gereedschap genaamd AdFisher (vernoemd naar R.A. Fisher, auteur van het boek *The Design of Experiments* uit 1935) wordt de interactie tussen deze Ad Settings, het browse-gedrag van de gebruiker en de getoonde advertenties gepoogd bloot te leggen. Het bijzondere aan AdFisher is dat het vrijwel volledig geautomatiseerd werkt met gesimuleerde gebruikersprofielen die, afgezien van het te bestuderen aspect, zo min mogelijk van elkaar verschillen om de invloed van ruis te minimaliseren. Door de helft van deze profielen een bepaald browse-gedrag voor te schrijven, of in de helft van de profielen bepaalde Ad Settings te wijzigen, kan worden onderzocht wat het effect hiervan is in vergelijkingen met de andere helft. Opmerkelijk is dat zelfs dit onderzoek geautomatiseerd plaatsvindt in de zin dat zelflerende algoritmes worden gebruikt om zowel te detecteren in welk opzicht er verschillen zijn als dat de statistische analyse erdoor wordt uitgevoerd.

De belangrijkste conclusies zijn dat Ad Settings de gebruiker weliswaar enigszins invloed geven op welke advertenties worden getoond, maar dat verschillend browse-gedrag hierop ook van invloed is zonder dat de Ad Settings zichtbaar veranderen. Hierdoor ontstaat een gebrek aan transparantie over de werking van de Ad Settings dat kan resulteren in discriminatie. Expliciet genoemd wordt een voorbeeld waarin aan mannen meer dan aan vrouwen vacatures worden getoond voor hoogbetaald werk (als resultaat van het gestructureerd wijzigen van het geslacht in de Ad Settings).

4 Digitale technologie kan advertentieteksten grootschalig onderzoeken op discriminatie

Mensen die een vacature tegenkomen waarin bijvoorbeeld een bepaalde leeftijdsgroep wordt uitgesloten, kunnen daarover een klacht indienen bij het College. Het College beoordeelt dan of er echt sprake is van verboden onderscheid. Het aantal keer dat mensen dit doen, is vanzelfsprekend geen goede indicatie van de mate waarin er discriminatie optreedt. Ook de mate van perceptie van discriminatie is hiervoor geen goede graadmeter, al is het maar omdat de wetgeving rondom leeftijdsdiscriminatie complex is en niet alle werkzoekenden daarvan op de hoogte zijn.

Om toch inzicht te krijgen in de mate waarin leeftijdsonderscheid in vacatureteksten voorkomt, is dus zowel een andere manier van dataverzameling als een andere manier van data-analyse nodig dan voorheen. Daarbij zou moeten worden gewerkt met een steekproef van advertentieteksten die handmatig zouden moeten worden gecategoriseerd. Omdat vacatures tegenwoordig vrijwel allemaal ook online verschijnen, zijn ze gelukkig ook geautomatiseerd te verzamelen en te bestuderen. Dit vergt wel een digitaliseringsslag.

Het College is samen met de Vrije Universiteit ingesprongen op deze mogelijkheid. Bij hun onderzoek hebben ze gebruikgemaakt van een vacaturedatabase met daarin 95% van alle online verschenen vacatureteksten. Deze database is van Textkernel, een bedrijf dat gespecialiseerd is in geautomatiseerde oplossingen voor het samenbrengen van vraag en aanbod op de arbeidsmarkt. Zij onderhouden en beheren een vacaturedatabase (genaamd Jobfeed) die geautomatiseerd het internet afzoekt naar nieuwe vacatures. Het dagelijks verzamelen van actuele vacatures op het internet werkt middels 'spidering', digitale spinnen die het web aflopen op zoek naar informatie. Hierbij worden miljoenen websites binnen het Nederlandse web continu op vacatures doorzocht ('brede spiders'), maar worden ook vacatures van specifieke websites gehaald ('specifieke spiders'). Onder die laatste categorie vallen jobboards, kranten, sites van intermediairs en sites van grote werkgevers (Fokkens et al., 2018). Het onderzoek had betrekking op 1,8 miljoen unieke vacatures die in 2017 online zijn verschenen.

Deze zijn niet handmatig te onderzoeken op leeftijdsonderscheid. In plaats daarvan is gebruikgemaakt van een algoritme dat kwalificeert welke tekst onderscheid maakt naar leeftijd, hierbij rekening houdend met de context. Zo levert 'jong van geest' geen onderscheid op, maar een tekst waarin gevraagd wordt om een 'jong persoon' wel. Het algoritme is getest op betrouwbaarheid. Het algoritme bleek weinig vacatures onterecht als verboden onderscheidend aan te merken. Het kwam vaker voor dat een vacature waarin sprake was van verboden onderscheid, door het algoritme werd gemist. In totaal werden ruim 63.000 van de 1,8 miljoen vacatures (3,5 procent) geclassificeerd als verboden leeftijdsdiscriminatie (Fokkens et al., 2018; College, 2018a). Hierin werden voornamelijk oudere werkzoekenden uitgesloten. Uitsplitsing naar bedrijfsniveau liet zien dat twaalf bedrijven verantwoordelijk waren voor meer dan 40% van de onderscheid makende vacatures. Het College is met deze organisaties in gesprek gegaan over hoe zij in de toekomst leeftijdsdiscriminatie bij werving en selectie kunnen voorkomen. Het is de bedoeling het onderzoek in 2020 te herhalen. Daarbij zal worden gekeken of deze twaalf bedrijven hun vacatures voortaan anders opstellen. In tegenstelling tot het behandelen van individuele klachten over discriminatie kan op deze manier een groot deel van het probleem bij vacatures worden aangepakt.

In aanvullend onderzoek heeft het College gekeken wat voor een effect verboden onderscheid op grond van leeftijd in een vacaturetekst had op de intenties van oudere werkzoekenden om te solliciteren. Dit effect werd onderzocht door middel van vignetten: werkzoekende ouderen werd gevraagd verschillende vacatures te beoordelen, waarbij in de vacature leeftijdsonderscheid een van de variërende condities was. Dit onderzoek liet zien dat vacatureteksten waarin ouderen

worden uitgesloten, niet alleen verboden kunnen zijn, maar dat ze bovendien echt ontmoedigen om te solliciteren (College, 2018b).

5 Conclusies

Grote online platforms zoals Google, LinkedIn en Facebook bieden adverteerders (zoals werkgevers) diverse mogelijkheden om hun advertenties (zoals vacatures) gericht uit te zetten onder bepaalde doelgroepen. Dit levert vaak efficiëntie en daarmee kostenbesparing op. Het gevaar is echter dat de targeting, soms direct soms indirect, resulteert in discriminatie (zoals arbeidsmarktdiscriminatie), mede doordat de digitale technologie achter deze mogelijkheden vaak niet transparant is, vooral als er gecompliceerde algoritmes worden gebruikt die grote hoeveelheden data manipuleren.

Hoopgevend is dat de detectie van dergelijke vormen van discriminatie, en uiteindelijk ook de bestrijding ervan, gebruik kan maken van dezelfde digitale technologie. Deze is in staat om grootschalige simulaties los te laten op de black box van targetingmethodes en de resulterende grote hoeveelheden data met geavanceerde algoritmes te analyseren op discriminatie. En HR-analytics kan, mits juist ingezet, ook een deel van de menselijke bias wegnemen (Hoffman et al., 2017). Daarnaast kunnen ook grote hoeveelheden advertenties (zoals vacatures) worden onderzocht op uitsluitende ingrediënten.

Samengevat levert de toegenomen digitalisering bij de werving en selectie risico's op van discriminatie en uitsluiting die minder zichtbaar zijn voor de individuele werkzoekende. Tegelijkertijd kan deze uitsluiting op grotere schaal worden onderzocht en bestreden dankzij diezelfde digitalisering.

Literatuur

Angwin, J., Scheiber, N., & Tobin, A. (2017). *Dozens of companies are using Facebook to exclude older workers from job ads*. ProPublica. Retrieved from <https://www.propublica.org/article/facebook-ads-age-discrimination-targeting>

College voor de Rechten van de Mens. (2018a). *Gezocht: jonge hond*. Utrecht: College voor de Rechten van de Mens.

College voor de Rechten van de Mens. (2018b). *Om mij wordt niet gevraagd*. Utrecht: College voor de Rechten van de Mens.

Datta, A., Tschantz, M.C., & Datta, A. (2015). Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies*, (1), 92-112.

Fokkens, A., Beukeboom, C.J., & Maks, I. (2018). *Leeftijdscriminatie in vacatureteksten: Een geautomatiseerde inhoudsanalyse naar verboden leeftijd-gerelateerd taalgebruik in vacatureteksten*. Rapport in opdracht van het College voor de Rechten van de Mens.

Hoffman, M., Kahn, L.B., & Li, D. (2017). Discretion in hiring. *The Quarterly Journal of Economics*, 133(2), 765-800.

Speicher, T., Ali, M., Venkatadri, G., Ribeiro, F., Arvanitakis, G., Benevenuto, F., Gummadi, K.P., Loiseau, P., & Mislove, A. (2018). Potential for discrimination in online targeted advertising. *Proceedings of Machine Learning Research*, (81), 1-15.