

# Ter Visie - Algoritmische discriminatie

mr. M. Jovović<sup>1</sup>

## 1. Inleiding

Het algoritme<sup>2</sup> is in opkomst. ‘Slimme algoritmes’ (*machine learning algoritmes*) bepalen steeds vaker wie promotie krijgt of welk cv geselecteerd wordt in een sollicitatieprocedure. Dit snijvlak tussen Human Resources en informatietechnologie wordt ook wel HR-Analytics genoemd – een industrie die in Nederland nog in de kinderschoenen staat.<sup>3</sup>

De hoeveelheid gegevens die door HR-Analytics systemen kan worden gebruikt, is duizelingwekkend te noemen. Zo bestaat een algoritme voor werving en selectie dat – naast de gebruikelijke informatie zoals opleidingsniveau en studierichting – ‘duizenden bits aan informatie cruncht om zodoende ongeveer 300 grotere variabelen die betrekking hebben op een individu uit te rekenen: op welke sites brengt iemand zijn tijd door; het type taal, positief of negatief, dat hij of zij gebruikt om verschillende soorten technologie te omschrijven; de vaardigheden die iemand van zichzelf heeft opgegeven op LinkedIn; en de projecten waaraan iemand heeft gewerkt, en voor hoe lang.’<sup>4</sup> Ook in de opkomende<sup>5</sup> platformeconomie<sup>6</sup> worden veel in de

arbeidsverhouding van belang zijnde beslissingen genomen door geautomatiseerde datagedreven systemen: wie doet welke klus op welke plek op welk tijdstip en tegen welke beloning?

Het automatiseren van dit soort beslissingen draagt de belofte van toegenomen objectiviteit in het arbeidsproces in zich. Computers hebben immers geen vooroordelen.<sup>7</sup> In de woorden van oud-Wired hoofdredacteur Anderson: ‘*With enough data, the numbers speak for themselves.*’<sup>8</sup> Nu blijkt echter dat ook discriminatie verscholen kan zijn in de cijfers. Algoritmes kunnen namelijk tot niet objectief te rechtvaardigen resultaten komen, zowel doordat het algoritme de vooroordelen van de programmeurs of hun opdrachtgevers weerspiegelt,<sup>9</sup> als doordat het is getraind met historische oefendata die is voortgekomen uit een ongelijke historische situatie.<sup>10</sup>

Hoe past dit alles in het gelijkebehandelingsrecht? Zoals ik in deze bijdrage toelicht, zal het gelijkebehandelingsrecht in steeds meer arbeidsverhoudingen een rol gaan spelen, naarmate steeds meer alledaagse werkgeversbeslissingen worden genomen met behulp van *machine learning*. Voor de duidelijkheid: het gaat hier om méér arbeidsverhoudingen dan enkel die welke als arbeidsovereenkomst in de zin van art. 7:610 BW kunnen worden gekwalificeerd. Er is immers een veelvoud aan (Europese en nationale) regelgeving op grond waarvan ongeoor-

1. Marko Jovović is advocaat bij Stibbe.
2. ‘Algoritme’ is een zeer brede term die zowel betrekking kan hebben op recepten als op complexe zelflerende systemen die dieren op foto’s kunnen herkennen. Organisaties hebben vaak een veelvoud aan verschillende algoritmes in gebruik. In dit artikel bespreek ik uitsluitend de zogeheten ‘*supervised machine learning*’ algoritmes, zie: Bernard Marr, ‘What Is The Difference Between Deep Learning, Machine Learning and AI?’, *Forbes* 8 december 2016, online via: <https://www.forbes.com/sites/bernard-marr/2016/12/08/what-is-the-difference-between-deep-learning-machine-learning-and-ai/> (laatstelijk geraadpleegd op 3 december 2018).
3. K. Kuijpers, T. Muntz & T. Staal, ‘Privacy? Achterhaald’, *De Groene Amsterdammer* 31 oktober 2018 (nr. 2018-44-45).
4. P.T. Kim, ‘Data-Driven Discrimination at Work’, *58 Wm. & Mary L. Rev.*, p. 862. De betreffende onderneming lijkt niet meer actief te zijn; de website is in elk geval niet meer online vindbaar.
5. Ter Weel, Van der Werff, Bennaars, Scholte, Fijnje, Westerveld & Mertens, *De opkomst en groei van de kluseconomie in Nederland*, Amsterdam: SEO Economisch Onderzoek 2018.
6. Volgens het Rathenau-instituut: ‘*De technologische, economische en sociaal-culturele infrastructuur voor het faciliteren en organiseren van online sociaal en economisch verkeer tussen gebruikers en aanbieders, met (gebruikers)*

*data als brandstof*.’ Zie: K. Frenken, A. van Waes, M. Sminck & R. van Est, *Eerlijk delen – Waarborgen van publieke belangen in de deeleconomie en de kluseconomie*, Den Haag: Rathenau Instituut 2017, p. 22-26.

7. Zie ook J.M. Van Slooten, ‘Ter Visie – Platformarbeid: nog een reden tot rethinking van het arbeidsrecht’, *TAO* 2017, nr. 2, p. 3-6: ‘[G]een (of althans minder) discriminatie op achternaam of geslacht. Zolang je de klus goed klaart, is er niemand die zeurt.’
8. Chris Anderson, ‘The End of Theory: The Data Deluge Makes the Scientific Method Obsolete’, *Wired* 23 juni 2008, online via: <https://www.wired.com/2008/06/pb-theory/> (laatstelijk geraadpleegd op 3 september 2018).
9. Denk hierbij aan het algoritme dat witte mannen herkent, maar zwarte vrouwen niet: T. Simonite, ‘Photo algorithms ID white men fine – black women, not so much’, *Wired* 2 juni 2018. Online via: <https://www.wired.com/story/photo-algorithms-id-white-men-fine-black-women-not-so-much/> (laatstelijk geraadpleegd op 27 november 2018).
10. M.J. Vetzto, J.H. Gerards & R. Nehmelman, *Algoritmes en grondrechten*, Den Haag: Boom Juridisch 2018, p. 49.

loofd onderscheid is verboden.<sup>11</sup> Ook lopen de personele werkingssferen van deze regelingen nogal uiteen.<sup>12</sup> In deze bijdrage hanteer ik de begrippen ‘werkgever’ en ‘werknemer’.

## 2. Twee arbeidsrechtelijk relevante eigenaardigheden van algoritmes

### 2.1. Veel data, veel factoren, veel potentieel (indirect) onderscheid

*Machine learning* systemen kunnen omgaan met zeer grote aantallen variabelen. Deze variabelen kunnen in beginsel betrekking hebben op alle eigenschappen waarover data voorhanden is. Zowel de aard van de gegevens als de wijze van analyse daarvan, hebben implicaties voor het gelijkebehandelingsrecht. Telkens moet worden bezien of sprake is van directe of indirecte discriminatie. Ik werk dit hieronder uit.

Modellen die rechtstreeks rekening houden met factoren als godsdienst, levensovertuiging, politieke gezindheid, ras, geslacht, nationaliteit, hetero- of homoseksuele gerichtheid of burgerlijke staat,<sup>13</sup> chronische ziekte of handicap<sup>14</sup> of leeftijd,<sup>15</sup> riskeren op te lopen tegen het verbod op direct onderscheid.<sup>16</sup> Er zal dan immers sprake zijn van een verschil in behandeling op grond van een kenmerk dat in het kader van het gelijkebehandelingsrecht gezien wordt als een ‘verboden variabele’. Zelfs als de afwijking voor de leden van de betrokken groep positief uitpakt, is sprake van onderscheid dat niet altijd zal zijn toegestaan. Het gaat dan in feite om algoritmische positieve actie, een rechtvaardigingsgrond die op grond van de Nederlandse wetgeving niet in alle gevallen is toegestaan<sup>17</sup> en welke door

het Europees Hof van Justitie beperkt wordt uitgelegd.<sup>18</sup>

Wanneer het model niet uitdrukkelijk rekening houdt met verboden variabelen, kan toch sprake zijn van indirecte discriminatie. Dat kan gebeuren doordat ogenschijnlijk neutrale factoren de neiging hebben discriminatoir uit te pakken wanneer er eenzijdigheden verscholen zitten in de data. Zo ontwikkelde het sollicitatiealgoritme van Amazon een voorkeur voor mannelijk taalgebruik als ‘executed’ en ‘captured’ om de verdeling tussen mannen en vrouwen in de trainingsdata, ontstaan doordat Amazon in het verleden minder vrouwen dan mannen had aangenomen, te kunnen verklaren. Later weigerde het algoritme de sollicitaties van vrouwen op basis van deze voorkeuren.<sup>19</sup> Dit is een typisch voorbeeld van een ogenschijnlijk neutrale bepaling, maatstaf of handelwijze die een categorie beschermde personen in vergelijking met andere personen bijzonder treft.<sup>20</sup>

Het kan hierbij gaan om een bewuste strategie van de werkgever om, door daarvoor in de plaats een op het eerste gezicht neutraal kenmerk te gebruiken, toch rekening te kunnen houden met een verboden kenmerk.<sup>21</sup> Het algoritme kan echter ook rekening houden met beschermde eigenschappen door verschillende – op zich neutrale en te goeder trouw in het model geïncorporeerde factoren – op een specifieke wijze te wegen. Een algoritme dat rekening houdt met postcode, inkomen en opleidingsniveau kan indirect onderscheid maken op grond van etniciteit, zonder dat de variabele ‘eticiteit’ ooit door de programmeurs in het model is opgenomen.<sup>22</sup> Dit gebeurt doordat *machine learning* algoritmes worden getraind met behulp van historische gegevens. Het doel van de training is het algoritme zo in te stellen dat het in het overgrote merendeel van de historische gevallen op basis van de historische ‘input’ de daarbij behorende historische ‘output’ geeft. Daarna kan het algoritme worden toegepast op nieuwe situaties om voorspellingen te maken. Hierbij is van belang dat *machine learning* algorit-

11. Zie voor een beknopt overzicht M.J. Vetzo, J.H. Gerrards en R. Nehmelman, *Algoritmes en grondrechten*, Den Haag: Boom Juridisch 2018, p. 88-92.

12. Aan de verhoudingen en overlappingsen tussen de werkingssferen van al deze bepalingen en regelingen op het gebied van het gelijkebehandelingsrecht zou een heel artikel gewijd kunnen worden, maar vanwege de beperkte ruimte laat ik verdere bespiegelingen daarover achterwege.

13. Art. 1 Awgb.

14. Art. 1 Wgbh/cz.

15. Art. 1 Wgbl.

16. In art. 21 Handvest worden zelfs nog meer antidiscriminatiegronden genoemd: geslacht, ras, kleur, etnische of sociale afkomst, genetische kenmerken, taal, godsdienst of overtuiging, politieke of andere denkbeelden, het behoren tot een nationale minderheid, vermogen, geboorte, een handicap, leeftijd of seksuele gerichtheid.

17. Art. 7 lid 1 Richtlijn 2000/78/EG maakt positieve actie mogelijk voor alle discriminatiegronden, maar de Nederlandse wetgever heeft positieve actie enkel mogelijk willen maken voor onderscheid dat tot doel heeft vrouwen of personen behorende tot een bepaalde etnische of culturele minderheidsgroep een bevoorrechte positie toe te kennen (zie bijvoorbeeld art. 7:646 lid 4 BW, art. 2 lid 3 Awgb of art. 5 lid 1 Wgbhmv).

18. Zie bijv. HvJ EG 17 oktober 1995, ECLI:EU:C:1995:322 (*Kalanke*), r.o. 18.

19. Zie L. Moerel, ‘Algoritmes kunnen discriminatie juist beperken, maar alleen met goede data’, *Het Financieel Dagblad* 11 november 2018.

20. Vgl. de definitie van indirect onderscheid uit art. 1 lid 1 aanhef en onder b Awgb.

21. Zo wordt het gebruik van gegevens met betrekking tot woonplaats als wijze om via een omweg rekening te houden met etniciteit of ras in de Amerikaanse literatuur ‘redlining’ genoemd, zie: A. Hiller, ‘Spatial analysis of historical redlining’, *Journal of Housing Research* 14(1), p. 137-168 en G. Squires, ‘Racial profiling, insurance style: insurance redlining and the uneven development of metropolitan areas’, *Journal of Urban Affairs* 25(4), p. 391-410. Ook wordt dit fenomeen wel eens met de term ‘masking’ aangeduid.

22. Zie I. Žliobaitė en B. Custers, ‘Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models’, *Artificial Intelligence and Law* (24), p. 186. Het rekenvoorbeeld op pagina 191 van het artikel is bijzonder verhelderend.

mes over het algemeen geen onderscheid maken tussen correlaties en oorzakelijke verbanden: iedere correlatie weegt mee in het systeem. Als dus personen met een bepaalde etniciteit over- of ondervertegenwoordigd zijn in de trainingsdata, zal het algoritme al snel andere factoren vinden die daarmee correleren, om zodoende alsnog de verdeling in de trainingsdata te kunnen reproduceren. Denk hierbij aan het hierboven besproken voorbeeld van Amazon.

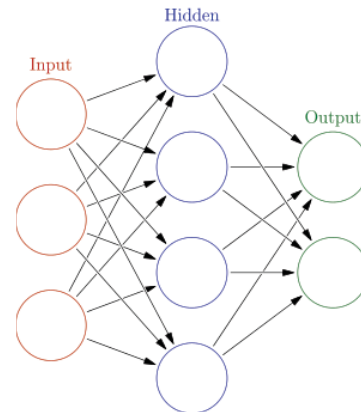
In de (met name buitenlandse) literatuur gaan stemmen op om, enigszins contra-intuïtief, het algoritme juist dit soort vooroordelen 'af te leren' door het juist wel rekening te laten houden met 'verboden variabelen'.<sup>23</sup> Hier zullen werkgevers echter aanlopen tegen het hiervoor gesignaleerde probleem dat positieve discriminatie niet in alle gevallen is toegestaan.<sup>24</sup> Dit probleem zou kunnen worden ondervangen door, in plaats van telkens te beoordelen of een specifieke variabele discriminatoir is, het effect van het gehele algoritme in ogenschouw te nemen. In die benadering wordt het gehele algoritme aangemerkt als 'ogenschijnlijk neutrale bepaling, maatstaf of handelwijze'. De individuele variabelen zijn niet meer dan onderdelen daarvan, met weliswaar juridische relevantie maar zonder zelfstandige, doorslaggevende juridische betekenis (te vergelijken met een enkel woord in een cao). Indien het de ontwerpers van het algoritme daadwerkelijk is gelukt de historische ongelijkheid te elimineren door rekening te houden met 'verboden variabelen', zal dit toegestaan zijn in de door mij voorgestelde benadering. Zo wordt recht gedaan aan de specifieke aard van *machine learning* systemen: het 'slimme' gedrag is een eigenschap van het geheel, van de wegenen en van de dwarsverbanden tussen de informatie. Dit sluit niet alleen aan bij de wijze waarop menselijke oordelen worden getoetst, maar het zou ook een manier kunnen zijn om de 'black box'-problematiek, die hieronder aan bod komt, enigszins het hoofd te bieden.

23. Zie bijvoorbeeld P.T. Kim, 'Data-Driven Discrimination at Work', 58 *Wm. & Mary L. Rev.*, p. 857 e.v., I. Žliobaitė en B. Custers, 'Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models', *Artificial Intelligence and Law* (24), p. 183-201, B.A. Williams, C.F. Brooks & Y. Shmargad, 'How Algorithms Discriminate Based on Data they Lack: Challenges, Solutions and Policy Implications', *Journal of Information Policy* 2018-8, p. 78-115, L. Moerel, 'Algoritmes kunnen discriminatie juist beperken, maar alleen met goede data', *Het Financieel Dagblad* 11 november 2018.

24. Een ander probleem hangt samen met het privacyrecht: de meeste 'verboden variabelen' zijn bijzondere persoonsgegevens in de zin van art. 9 lid 1 Algemene Verordening Gegevensbescherming en kunnen zodoende – met name in arbeidsverhoudingen – moeilijk worden verwerkt, zie: WP29, Guidelines on consent under Regulation 2016/679, 28 november 2017, p. 7.

## 2.2. De 'black box'

De term 'black box' is de laatste tijd bijna een eigen leven gaan leiden. Kort en goed gezegd komt het erop neer dat slimme algoritmes in data<sup>25</sup> (meestal in heel grote hoeveelheden daarvan)<sup>26</sup> op zoek gaan naar correlaties tussen gegevens. Voordat dit kan gebeuren, worden de gegevens verrijkt, gecomprimeerd of op andere wijze voorbereid. Meestal is de informatie daarna voor mensen onbegrijpelijk geworden. Daarna gaat het zelflerende algoritme met de gegevens aan de slag:



Zeer versimpeld beeld van een neuraal netwerk.

Bron: Wikimedia Commons.<sup>27</sup>

Heel kort door de bocht gezegd, leidt dit alles – mede in combinatie met het feit dat *machine learning* algoritmes doorgaans bedrijfsgeheim zijn<sup>28</sup> – ertoe

25. Informatie die digitaal is opgeslagen, soms een persoonsgegeven in de zin van art. 4 lid 1 AVG, maar niet altijd (meteorologische gegevens zijn immers net zo goed data die een rol kunnen spelen in een model).

26. M.J. Vetzto, J.H. Gerards en R. Nehmelman, *Algoritmes en grondrechten*, Den Haag: Boom Juridisch 2018, p. 14-24.

27. Op deze afbeelding is een zeer versimpeld neuraal netwerk te zien. Neurale netwerken zijn één van de typen algoritme die bij *machine learning* worden ingezet. Elk rode bolletje staat voor een bepaalde inputvariabele, bijvoorbeeld (in de context van een papierclipfabriek): 'aantal gewerkte uren in de afgelopen 7 dagen', 'gemiddeld aantal geproduceerde papierclips per 24 uur' en 'gemiddeld ijzerdraadverbruik per papierclip'. De blauwe en groene bolletjes bevatten elk een gewogen gemiddelde van de waarden van alle bolletjes in de vorige laag. Het algoritme wordt 'getraind' door de wegenen op basis van heel veel trainingsdata zo in te stellen dat het algoritme in staat is de historische output te reproduceren op basis van de historische input. Het is wiskundig gezien zeer moeilijk te reconstrueren wat de invloed is geweest van een specifieke inputvariabele – deze heeft immers op een groot aantal plekken in het algoritme een rol gespeeld.

Bron: [https://en.wikipedia.org/wiki/Artificial\\_neural\\_network#/media/File:Colored\\_neural\\_network.svg](https://en.wikipedia.org/wiki/Artificial_neural_network#/media/File:Colored_neural_network.svg).

28. Zie ook A.M. Helstone & T. Barkhuysen, 'Ter Visie – Bescherming van bedrijfsgeheimen', *TAO* 2018, nr. 1, p. 3-6.

dat ook deskundigen niet altijd in staat zijn exact te duiden waarom het *machine learning* algoritme heeft gedaan wat het heeft gedaan: het is een black box.<sup>29</sup> Naar verwachting zal de 'black box'-problematiek in procedures omtrent discriminatie van belang zijn voor de stelplicht en bewijslast van zowel de gediscrimineerde partij als de discriminerende partij.

Op grond van (bijvoorbeeld) art. 10 Awgb verschuift de bewijslast, zodra degene die meent dat in zijn nadeel onderscheid is gemaakt in rechte feiten stelt en bewijst die onderscheid doen vermoeden. Dat kan door aan te tonen dat bepaalde personen in vergelijking met andere personen bijzonder zijn benadeeld. Dat kan op 'enigerlei wijze, ook op basis van statistische gegevens'.<sup>30</sup>

De werknemer zal hierin eerder slagen, naarmate het door de werkgever gehanteerde systeem minder doorzichtig is. Ik illustreer dit aan de hand van twee oudere arresten van het Europees Hof van Justitie, waarin het Hof de verschuiving van de bewijslast ontwikkelde. In de *Enderby*-zaak oordeelde het Europees Hof van Justitie dat de bewijslast kan worden verlegd wanneer dat noodzakelijk is om de kennelijk gediscrimineerde werknemers niet elk doeltreffend middel te ontnemen om het beginsel van gelijke beloning te doen eerbiedigen.<sup>31</sup> En in het *Danfoss*-arrest overwoog het Europees Hof van Justitie dat de werkgever<sup>32</sup> die een stelsel van beloning toepast dat wordt gekenmerkt door een volstrekt gebrek aan doorzichtigheid, moet bewijzen dat zijn beloningspraktijk niet discriminerend is zodra de werknemer met betrekking tot een relatief groot aantal loontrekkenden aantoonde, dat de gemiddelde beloning voor vrouwelijke werknemers lager is dan die voor mannelijke werknemers.<sup>33</sup> Laat nu juist van *machine learning* systemen bekend zijn dat zij in het algemeen bijzonder ondoorzichtig zijn – zelfs experts slagen er vaak niet in te doorgronden

waarom zulke systemen hebben gedaan wat ze hebben gedaan.<sup>34</sup>

De werknemer zal met een beroep op art. 15 AVG en art. 843a Rv naar verwachting voldoende informatie kunnen verkrijgen om – op zijn minst – aan te kunnen tonen dat sprake is van een stelsel dat wordt gekenmerkt door een volstrekt gebrek aan doorzichtigheid. De werknemer zal ook op grond van het *Meister*-arrest een beroep kunnen doen op de werkgever om gegevens omtrent de loonpolitiek te openbaren. Weigert de werkgever, dan geldt dat de weigering informatie te verstrekken volgens het Hof van Justitie een omstandigheid is die in aanmerking dient te worden genomen door de rechter bij de beoordeling of indirecte discriminatie is aangetoond.<sup>35</sup> Slaagt de werknemer erin het vermoeden van ongeoorloofde ongelijke behandeling te vestigen, dan rust de bewijslast vervolgens op de werkgever.

Rust de bewijslast eenmaal op de werkgever, dan moet hij aantonen dat hij niet heeft gediscrimineerd.<sup>36</sup> Dat zal een behoorlijke opgave zijn, waarbij de 'black box'-problematiek in de weg zal kunnen zitten. Weliswaar zal hetgeen de werkgever met de informatie heeft gedaan en welke bewerkingen hij hierop heeft losgelaten, enkel relevant zijn wanneer de individuele variabelen worden aangemerkt als 'ogenschijnlijk neutrale bepaling, maatstaf of handelwijze'. De werkgever zal echter in alle gevallen moeten aantonen dat hij een legitiem doel heeft, en dat hij heeft gehandeld in overeenstemming met de beginselen van proportionaliteit en subsidiariteit. Hoe kan hij hierin slagen tegen de achtergrond van de 'black box'-problematiek?

Het is goed kort stil te staan bij de mogelijke concrete invulling van de criteria 'legitiem doel' en 'noodzakelijkheid en passendheid van het middel' in de specifieke context van algoritmische ongelijke behandeling. Ik meen dat het, in gevallen van indirect onderscheid, met het legitieme doel niet snel

29. Zie dit illustratieve artikel in *Nature*: D. Castelvecchi, 'The black box of AI', *Nature* 6 oktober 2018 en (voor een bredere beschouwing) F. Pasquale, *The Black Box Society, The Secret Algorithms That Control Money and Information*, Cambridge, Massachusetts: Harvard University Press 2015.

30. Zie paragraaf 15, preambule Richtlijn 2000/43/EG en paragraaf 15, preambule Richtlijn 2000/78/EG. Enkele belangwekkende uitspraken omtrent mogelijk aan te voeren statistische gegevens zijn HvJ EG 9 februari 1999, ECLI:EU:C:1999:60 (*Seymour*), r.o. 60-61, HvJ EG 10 maart 2005 ECLI:EU:C:2005:141 (*Nikoloudi*).

31. Vgl. bijv. HvJ EG 27 oktober 1993, ECLI:EU:C:1993:859 (*Enderby*).

32. Het werkgeversbegrip van de Europese antidiscriminatie-richtlijnen (o.m. Richtlijn 2000/17/EG (Kader-richtlijn gelijke behandeling), Richtlijn 2000/43/EG (Richtlijn rassendiscriminatie), Richtlijn 2004/113/EG (Richtlijn gelijke behandeling mannen en vrouwen) en Richtlijn 2010/41/EU (de Richtlijn gelijke behandeling zelfstandige mannen en vrouwen) is autonoom gedefinieerd, zodat ook partijen die geen werkgever zijn in de zin van art. 7:610 BW.

33. HvJ EEG 17 oktober 1989, ECLI:EU:C:1989:383 (*Danfoss*), r.o. 11.

34. D. Castelvecchi, 'The black box of AI', *Nature* 6 oktober 2018.

35. HvJ EU 19 april 2012, ECLI:EU:C:2012:217, (*Meister*), r.o. 37-44.

36. Dat zou kunnen door het verschil in behandeling te verklaren onder verwijzing naar objectieve factoren, zie bijvoorbeeld het working paper van Stanford-onderzoekers C. Cook, R. Diamond, Jonathan Hall, J.A. List & P. Oyer, 'The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers', online via: <https://web.stanford.edu/~diamondr/UberPayGap.pdf> (laatstelijk geraadpleegd op 3 december 2018). Het beloningsverschil van \$ 1,28 tussen mannen en vrouwen bleek na analyse van 1.000.000 ritten volkomen verklaarbaar doordat mannen gemiddeld vaker reden in gebieden waar de prijzen hoger lagen, doordat zij gemiddeld sneller reden, vaker ritten aannamen die dichterbij waren en doordat ze vaker langere ritten reden. De vraag is hoe veel verder de werkgever komt met deze analyse: de resultaten zijn weer nieuwe 'ogenschijnlijk neutrale factoren', die weer gerechtvaardigd moeten worden met een legitiem doel en een passend en noodzakelijk middel.



fout zal gaan. Een doel zal legitiem zijn wanneer het beantwoordt aan een echte behoefte van de onderneming,<sup>37</sup> of wanneer het een noodzakelijke doelstelling van sociaal beleid vormt.<sup>38</sup>

Op het eerste gezicht lijkt er weinig mis met het bevorderen van een doelmatig aannamebeleid (bij automatische cv-screenings) of het verbeteren van de kwaliteit van de dienstverlening (bij vervoers-apps die aan automatische matching doen). Het zal vaak gaan om maatregelen met als doel een bepaalde kernactiviteit van de onderneming te bevorderen. Wel moeten werkgevers al vroeg in het ontwerpproces nadenken over de vraag hoe het systeem moet worden ontworpen en waarom. Telkens zal men zich moeten afvragen of *dit* systeem in *deze vorm* daadwerkelijk beantwoordt aan een echte behoefte van de onderneming en anders of dit systeem in deze vorm het behalen van een noodzakelijke doelstelling van sociaal beleid bevordert.

Zoals gezegd, wordt het echter lastiger bij de beoordeling of het middel passend en noodzakelijk is. Hiervoor moet worden onderzocht of het doel niet met andere, niet- of minder discriminatoire middelen had kunnen worden bereikt.<sup>39</sup> Eerst zal hiervoor moeten worden bepaald wat exact het middel is: het gehele *machine learning* algoritme of juist het gebruik van een specifieke variabele. Bij beide opties speelt een gelijksoortig probleem. In beide gevallen zal namelijk moeten worden vastgesteld waarom een bepaalde technische oplossing is gekozen, en of een andere technische oplossing minder discriminatoire effecten zou hebben gehad. Ook in dit kader zullen werkgevers zich moeten afvragen waarom zij *dit* systeem in *deze vorm* inzetten. Beoordeeld zal moeten worden wat de concrete voordelen zijn van (bijvoorbeeld) het gebruik van het ene rekenmodel ten opzichte van het andere of van het weglaten of juist wel modelleren van een bepaalde variabele. Het zal volgens mij lonen het ontwikkelproces nauwgezet te documenteren om zo later, onder verwijzing naar objectieve criteria,<sup>40</sup> deugdelijk uit te kunnen leggen welke keuzes zijn gemaakt en waarom.

Bewijslast is bewijsrisico, en hierboven heb ik al betoogd dat de kans dat de werkgever de bewijslast zal krijgen zal toenemen naarmate de complexiteit van het systeem toeneemt. Werkgevers die niet op tijd stilstaan bij de fundamentele keuzes die bij de ontwikkeling van hun zelflerende systemen worden gemaakt, zullen hier te zijner tijd op kunnen worden aangesproken. Hierbij dreigt voor hen een reële en effectieve schadevergoeding.<sup>41</sup>

37. HvJ EG 13 mei 1986, ECLI:EU:C:1986:204 (*Bilka*).

38. HvJ EG 13 juni 1989, ECLI:EU:C:1989:328 (*Rinner-Kühn*).

39. Zie bijvoorbeeld HvJ EG 9 februari 1999, ECLI:EU:C:1999:60 (*Seymour*).

40. HvJ EG 5 maart 2009, ECLI:EU:C:2009:128 (*Age Concern England*), r.o. 51.

41. HvJ EG 22 april 1997, ECLI:EU:C:1997:208 (*Draehmpaehl*), r.o. 26.