

Inbreng College voor de Rechten van de Mens op verzoek van de commissie voor Digitale Zaken, 13 mei 2022

- ❖ *Drie aandachtspunten naar aanleiding van de aanbevelingen uit het WRR-rapport ‘Opgave AI. De nieuwe systeemtechnologie’:*

Aandachtspunt 1: Het verbod op (indirecte) discriminatie wordt nog onvoldoende meegenomen bij de ontwikkeling, inzet en controle van AI-systemen

De inzet van AI kan zowel tot direct als indirect verboden onderscheid (discriminatie) leiden. Beide vormen van onderscheid zijn in strijd met de gelijkebehandelingswetgeving en het recht op gelijke behandeling. Direct verboden onderscheid is het rechtstreeks benadelen van personen op basis van een beschermde grond zonder dat hiervoor een rechtvaardiging bestaat. Denk hierbij aan een algoritme dat wordt ingezet om sociale welzijnsfraude te signaleren en dat hiervoor het kenmerk ‘etniciteit’ als selectiecriteria gebruikt. Bij *indirect* verboden onderscheid is er sprake van een op het eerste gezicht neutraal criterium dat niet rechtstreeks verwijst naar een discriminatiegrond, maar wel een bepaalde groep in het bijzonder treft. Denk bijvoorbeeld aan een AI-recruitmentsysteem dat de factor onafgebroken jaren werkervaring van een sollicitant gebruikt als criterium voor een geschikte kandidaat. Op het eerste gezicht lijkt dit onschuldig omdat dit iets kan zeggen over de kwaliteit van de sollicitant, maar tegelijkertijd treft dit vrouwen vaker dan mannen omdat zij vanwege een zwangerschap of zorgtaak voor kinderen vaker dan mannen een gat op hun cv hebben.¹ Het risico op indirecte discriminatie neemt nog verder toe bij ‘zelflerende algoritmes’, waarbij het nog moeilijker wordt te achterhalen waarom een systeem een bepaalde beslissing neemt.

Het College constateert dat ontwerpers en gebruikers van AI-systemen onvoldoende op de hoogte zijn van de gelijkebehandelingswetgeving, in het bijzonder bij indirect verboden onderscheid. Gebruikers en ontwikkelaars denken nog te vaak dat een algoritme niet discrimineert zolang het niet rechtstreeks selecteert op bijvoorbeeld geslacht, afkomst of persoonsgegevens. Het niet verwerken van persoonsgegevens, is geen garantie dat de inzet een AI-systeem niet alsnog leidt tot verboden onderscheid.

Aandachtspunt 2: Door gebrekkige transparantie en uitlegbaarheid van AI vindt discriminatie onopgemerkt plaats

Als een AI-systeem discriminatie veroorzaakt, is dit niet direct duidelijk. De sollicitant die afgewezen is door een AI-systeem of de burger die extra streng onderzocht wordt op fraude heeft zelf niet altijd de informatie dat AI is gebruikt. En weet de burger dit wel, dan beschikt deze niet over de gegevens (zoals output-data) of expertise om te beoordelen of AI discrimineert. Mensen hebben daarnaast de neiging besluiten van computers voor waar aan te nemen (*automation bias*). Ook de ambtenaar, recruiter of andere gebruiker van AI wordt hierdoor gehinderd.

Door gebrekkige transparantie, uitlegbaarheid en complexiteit van AI (de bekende *black box*) kan discriminatie lang verborgen voortduren. Er is nog weinig Nederlandse casuïstiek over discriminatie door AI. Dit kan betekenen dat het niet plaats vindt, maar een plausibeler verklaring is dat discriminatie verhuuld is achter complexe algoritmes die alleen door een kleine groep experts te begrijpen is. Discrimineert een geautomatiseerd systeem dan zijn de gevolgen snel groot, omdat het juist vaak gebruikt wordt op grote schaal. Denk hierbij aan de toeslagenaffaire, waarbij de inzet van fraudesystemen bijdroeg aan tienduizenden slachtoffers. De risico's van AI vragen om controle en kenbaarheid. Het College beveelt de commissie aan om garanties te vragen ten aanzien van de

¹ Andere voorbeelden: een recruitmentsysteem dat maximaal aantal jaren werkervaring meeweegt (grond leeftijd), of een gezichts- en spraakherkenningsysteem dat minder goed werkt voor mensen met een beperking (handicap/chronische ziekte) of mensen met een donkere huidskleur (ras) als gevolg van onder-representatie in trainingsdata.

transparantie over gebruik van AI, en uitlegbaarheid van de werking AI-systemen. Het is tevens belangrijk de capaciteit van maatschappelijke organisaties te ondersteunen die AI onderzoeken.²

Aandachtspunt 3: De rechtsbescherming loopt gevaar bij de toepassing van AI

In aansluiting op het vorige punt: bij de inzet van AI worden burgers veelal alleen met de uitkomst van het proces geconfronteerd. Je moet weten hoe AI werkt om je ertegen te verweren. Daar komt bij dat deze systemen behalve voor ambtenaren en burgers ook voor de rechtelijke macht lastig te doorgronden kunnen zijn.

Het bestaande gelijkebehandelingsrecht werkt ‘klachtafhankelijk’: burgers moeten zelf aannemelijk maken dat zij ongelijk zijn behandeld. Die bewijspositie is bij intermenselijk onderscheid soms al lastig, maar bij AI is dit nog ingewikkelder. Het klachtafhankelijke gelijkebehandelingsrecht heeft dus een aanvulling nodig in de vorm van (AI-)toezicht.

Uit onderzoek blijkt dat gemeenten die algoritmes inzetten op dit moment geen speciale voorzieningen treffen voor rechtsbescherming.³ Zij vertrouwen op de standaard procedures van bezwaar en beroep of de klachtfase. Het afhandelen van bezwaar of klachten loopt echter vaak gescheiden van de collega’s met kennis over kunstmatige intelligentie. Hierdoor ontstaat ook het risico dat feedback op het systeem onvoldoende aandacht krijgt.

❖ *Eventuele lacunes of ongewenste overlap in wet- en regelgeving bij de toepassing van AI:*

Voorkom verminderde effectiviteit van het gelijkebehandelingsrecht door AI

Het College adviseert de commissie op grond van bovengenoemde aandachtspunten om na te denken over een effectievere naleving van het gelijkebehandelingsrecht en de mensenrechten bij de inzet van AI. In het bijzonder:

- Verplicht transparantie over het gebruik en de werking van gehanteerde algoritmes.
- Deskundig toezicht op AI-systemen bij de overheid als aanvulling op het individuele klachtrecht is noodzakelijk om burgers te beschermen tegen discriminatie.
- Maak wetgeving die eisen stelt aan de gebruiker voor en tijdens de ingebruikneming van het AI-systeem. Een voorbeeld hiervan is de vergewisplicht voor werkgevers, zoals opgenomen in het Wetsvoorstel *toezicht gelijke kansen werving en selectie*.⁴
- Kijk naar nieuw juridisch instrumentarium binnen de bestaande wetgeving gelijke behandeling, zoals het omkeren van de bewijslast als een organisatie geen transparantie kan verschaffen over het AI-systeem.⁵

Regelgeving zit het herkennen van (indirecte) discriminatie door AI soms in de weg

Zoals gezegd kan AI ongemerkt indirecte discriminatie veroorzaken. Indirecte discriminatie is te signaleren door na te gaan of een AI-systeem voor bepaalde groepen met een door beschermd persoonskenmerk (zoals etniciteit, of geslacht) nadeliger uitpakt dan voor andere groepen.⁶ Technieken om onevenredige effecten te meten en te corrigeren dienen daarom in principe gebruik te maken van (bijzondere) persoonsgegevens. Op het moment lijkt er echter geen verwerkingsgrond aanwezig in de AVG ten behoeve van het biasvrij maken van AI-systemen. Hiertoe wordt wel de mogelijkheid geschapen in artikel 10 lid 5 van de voorgestelde Europese AI verordening, maar tot die tijd bestaat hier frictie tussen de bescherming van persoonsgegevens en het discriminatievrij maken van AI-systemen.

² Aansluitend bij Aanbeveling 5 uit het WRR-rapport ‘Opgave AI’. Zie in dit kader ook het WRR-working paper ‘AI van replek gediend? Een verkenning van tegenmacht vanuit maatschappelijke organisaties’ dat gelijktijdig verscheen.

³ W. Choi, M. van Eck en H. Hukshorn 2021, Hoe gemeenten besluiten over algoritmen & mensenrechten, Onderzoek voor het College voor de Rechten van de Mens.

⁴ College voor de Rechten van de Mens 2022, Notitie uitwerking van normen bij Wet toezicht gelijke kansen bij werving en selectie. <https://mensenrechten.nl/nl/publicatie/6261292d20f34f5d836210e7>

⁵ Equinet 2020, Meeting the new challenges to equality and non-discrimination from increased digitalization and the use of artificial intelligence, p. 70.

⁶ Ook is deze informatie belangrijk om te beoordelen of een AI-systeem binnen de wettelijke grenzen van voorkeursbeleid (‘positieve discriminatie’) blijft.