



# Van datingapp tot schadeverzekering: het algoritme bepaalt onze kansen

**Datum:** 5 oktober 2023

**Locatie:** Datamiddag bij het Verbond van Verzekeraars, Den Haag

**Spreker:** Sven Stevenson, programmadirecteur Directie Coördinatie Algoritmes (DCA), Autoriteit Persoonsgegevens (AP)

Beste data scientists,

Dank allereerst aan het Verbond van Verzekeraars voor de uitnodiging om met jullie in gesprek te gaan op deze datamiddag. Bijeenkomsten zoals deze zijn cruciaal om kennis te delen in een snel veranderende wereld waarin de mogelijkheden van algoritmes en AI eindeloos lijken. Maar, net zo eindeloos lijken de risico's voor mensen die aan algoritmes en AI worden onderworpen. Dit brengt een grote verantwoordelijkheid met zich mee voor bedrijven en organisaties. Het draait om het vinden van evenwicht tussen innovatie en het beschermen van mensen. Een middag als vandaag laat zien dat de zoektocht naar dat evenwicht daadwerkelijk gaande is.

Het is belangrijk te weten dat bedrijven die zoektocht niet in hun eentje hoeven aan te gaan. Sommige van de ethische vragen die opdoemen bij de inzet van algoritmes en AI zijn bij uitstek maatschappelijke vragen die ook een maatschappelijk antwoord vergen.

Laat ik maar beginnen met een voorbeeld.

Wie van u gebruikt weleens een datingapp? Nee, geen handen graag – dat antwoord is privé. Maar, als u een trouwe gebruiker bent van datingapps, kent u misschien Breeze, een Nederlandse app die een paar jaar geleden werd gelanceerd. Breeze baseert zich net als andere datingapps op het 'swipe-and-match'-concept, maar verschilt doordat er daarna geen chat maar een 'echte' date volgt.

Volgens Breeze zijn er via haar app inmiddels al meer dan 100.000 dates georganiseerd. Maar geeft het algoritme van Breeze iedereen wel een gelijke kans? Breeze had reden om daar zelf aan te twijfelen. Op basis van meldingen en eigen onderzoek vermoedde Breeze dat haar algoritme gemiddeld een lagere matchingskans berekende voor gebruikers met een donkere huidskleur en/of niet-Nederlandse achtergrond. En dit terwijl het algoritme niet direct gebruikmaakte van gegevens die te linken zijn aan etnische afkomst of huidskleur. De app vroeg deze gegevens niet op.

Het algoritme van Breeze is zelflerend, en het bedrijf gaf aan niet precies te weten hoe het algoritme werkt. Wel is duidelijk dat het likegedrag van gebruikers invloed heeft op de matchingskans.

En hiermee komen we bij de vermoedelijke oorzaak van de ongelijke behandeling. Verreweg de meeste gebruikers van de app waren, in de periode dat deze problemen aan het licht kwamen, van Nederlandse afkomst en met een lichte huidskleur. En omdat veel mensen bij het zoeken naar een partner of date vaak een voorkeur hebben voor personen met een vergelijkbare etnische achtergrond, maakt dit zelflerende algoritme deze voorkeur tot de norm. We kunnen dit eigenlijk zien als een vorm van 'social scoring'. En de



voorkeuren van de meerderheid – personen van Nederlandse afkomst met een lichte huidskleur – zorgen daarmee automatisch voor slechtere kansen voor de rest.

Met deze moeilijke situatie is Breeze eind vorig jaar naar het College voor de Rechten van de Mens gestapt. Breeze wou maatregelen treffen om te zorgen dat het algoritme, via de likes van haar gebruikers, niet discrimineert. Maar zou het toegestaan zijn om maatregelen te treffen om eventueel onderscheid op grond van etnische achtergrond te compenseren? Twee maanden geleden heeft het College voor de Rechten van de Mens hier uitspraak over gedaan en het oordeel was duidelijk: het is niet alleen toegestaan om de discriminatie door het algoritme te voorkomen of te compenseren, het is zelfs verplicht.<sup>i</sup>

Er is veel belangrijk aan deze uitspraak, maar graag wil ik uw aandacht vragen voor de openheid en de pro-activiteit waarin dit heeft plaatsgevonden. Dit keer geen gedupeerde die zich heeft gemeld bij een publiek orgaan, maar de ondernemer in kwestie die twijfelt over de ethische werking van zijn product. En dit keer niet ver nadat het op grote schaal is misgegaan met ontelbare gedupeerden, maar vroegtijdig in de beginfase van de inzet van een nieuwe toepassing. Iets om toe te juichen.

En hoe duidelijk ook dat het hier echt het algoritme zelf is waarbinnen de problemen ontstaan. Ik citeer graag de woorden van Breeze data-analist Thomas Crul over wat voor hem de belangrijkste les is: “Ik had zelf een redelijk laconieke houding ten opzichte van ethische problemen binnen het maken van algoritmes. Ik dacht: wij stoppen geen informatie over afkomst in het systeem, dus er kan ook geen verschil te zien zijn en dus geen probleem ontstaan. Maar zo simpel is het dus niet...”<sup>iii</sup>

Hoewel het antwoord van het College voor de Rechten van de Mens een duidelijke richting geeft, betekent dat niet dat er voor Breeze nu een makkelijke oplossing voorhanden is. Etniciteit is immers een bijzonder persoonsgegeven. Hoe dan ook, de opgave is nu bekend.

.. En dan toch: als we het meenemen van de voorkeuren van andere gebruikers in het voorschotelen van profielen als een vorm van social scoring kunnen zien, moeten we dat eigenlijk wel willen? Is de simpele oplossing dan niet gewoon te verbieden dat het likegedrag van andere gebruikers medebepalend is of ik wel of niet het profiel van andere gebruikers als mogelijke match te zien krijg? Soms kan de oplossing ook zijn om te besluiten dat een algoritme iets niet mag doen. Een gesprek dat we al snel op het niveau van de samenleving als geheel moeten voeren, want dit raakt niet alleen Breeze, maar ook soortgelijke apps als Bumble en Tinder. Grote bedrijven dus, met soms een miljardenomzet.

Beste mensen, met een voorbeeld als dit in het achterhoofd zult u begrijpen dat wij bij de Autoriteit Persoonsgegevens het afgelopen jaar met een uitdagende nieuwe taak op pad zijn gestuurd. In het regeerakkoord uit december 2021 werd besloten het takenpakket van de AP uit te breiden met de rol van algoritmetoezichthouder die mede bijdraagt aan het versterken van de samenwerking en samenhang tussen de diverse digitale toezichthouders. Een tweede pijler van de AP dus, naast onze voor u bekende werkzaamheden als toezichthouder op de AVG.

Dit jaar zijn wij met deze nieuwe taak van coördinerend algoritmetoezichthouder aan de slag gegaan. Concreet is daartoe binnen de AP de Directie Coördinatie Algoritmes (DCA) opgericht. We zijn begonnen met een klein budget, maar zetten de eerste stappen. Belangrijk voor ons is de missie die wij voor onszelf zien. Door te wijzen op ontwikkelingen, risico's en oplossingsrichtingen, wil de DCA bijdragen aan een goede bescherming van publieke waarden en grondrechten bij de inzet van algoritmes en AI. Om welke algoritmes en AI gaat het dan? Daarover zijn we redelijk agnostisch. Publieke waarden en grondrechten



kunnen immers net zo in het geding zijn bij een simpel statisch algoritme met enkele variabelen, als bij een complex zelflerend algoritme.

Waar we op dit moment specifiek naar kijken zijn drie risico's voor publieke waarden en grondrechten: 1) discriminatie, 2) willekeur en 3) gebrek aan transparantie. Laat ik daar kort verder bij stilstaan.

Ons eerste aandachtspunt is discriminatie.

Bewuste of onbewuste discriminatie is, helaas, een risico van algoritmes waar vrijwel iedereen in Nederland maar al te bekend mee is.

Laat ik deze keer het voorbeeld noemen van het project 'analytics uitkeringsfraude', een risicomodel voor het selecteren van personen die mogelijk bijstandsfraude zouden plegen. Dit algoritme werd gebruikt door de gemeente Rotterdam. Dankzij een verkenning van de lokale rekenkamer in 2020 is best wat bekend over dit model.<sup>iii</sup>

Wat de onderzoekers direct opviel is dat er in het model variabelen zaten die sterk correleerden met etniciteit. Denk aan 'geslacht', 'inburgeringsbehoefte' en 'taalbeheersing'. We noemen dit 'proxy discriminatie', en we weten helaas dat dit ook maar al te vaak bewust kan gebeuren. Een voorbeeld uit de financiële sector is de 'redlining' in de Verenigde Staten in het midden van de 20<sup>e</sup> eeuw. Er werd bewust gebruikgemaakt van prohibitieve prijsdifferentiatie langs de grenzen van postcodegebieden om indirect te zorgen dat er geen leningen en verzekeringen werden verstrekt in zwarte wijken.

De gemeente Rotterdam verklaarde richting de Rekenkamer dat zij niet wist of een variabele zoals 'inburgeringsbehoefte' gecorreleerd zou kunnen zijn met etniciteit en daarmee discriminerend. Máár... de gevolgen van onbewuste discriminatie zijn voor personen net zo groot. Hoe dan ook: de stormachtige opkomst van algoritmes in alle facetten van onze samenleving vergroot de kans dat dit soort proxy discriminatie kan optreden.

Ons tweede aandachtspunt is willekeur.

Als data scientists zult u het met mij eens zijn dat dit op het eerste gezicht een tegenstrijdigheid lijkt. Immers, als we een algoritme in principe ergens niet van kunnen betichten, dan is dat willekeur. Het algoritme is een (mathematisch of logisch) model, dat precies handelt zoals opgedragen in de code. Maar, een algoritme raakt ook aan statistiek en de vraag is wat iemand met die statistische probabiliteit vervolgens doet. Veel algoritmes worden gebruikt om tot waarschijnlijkheden, rangschikkingen of profilering te komen. Als u de Backstreet Boys leuk vindt, wilt u misschien ook naar NSYNC luisteren... Als u een studie rechten summa cum laude heeft afgerond, bent u waarschijnlijk een van de betere kandidaten voor een juridische vacature... Of... als uw leerlingen een veel lagere Cito-score behalen dan die in een andere klas, bent u misschien wel een slechtere docent...

Maar: er zijn onzekerheidsmarges in het spel. Als daar géén rekening mee wordt gehouden, dan dreigt willekeur. Amerikaanse docenten die werden ontslagen vanwege een beoordelingsmodel waarbij de toetsresultaten van hun leerlingen leidend waren kunnen daarover meepraten. De steekproef (n) van enkele tientallen leerlingen per jaar is te klein om hier zekerheid aan te verbinden. Dat heeft geleid tot schrijnende voorbeelden.<sup>iv</sup> In het algemeen geldt dat algoritmes vaak bijzonder slecht uitpakken als jij “de



uitzondering op de regel” bent. Als voor onzekerheid geen ruimte is in de wijze waarop met het algoritme wordt omgegaan, dreigt dus willekeur.<sup>v</sup>

Ons derde aandachtspunt is gebrek aan transparantie.

Stel, u bent die “uitzondering op de regel” en uw aanvraag voor een arbeidsongeschiktheidsverzekering is onterecht afgewezen en dat is in belangrijke mate beïnvloed door een algoritme... Of, u bent onterecht beschuldigd van fraude en dat had indirect te maken met discriminatie op etniciteit. Om daar achter te komen, moet u weten dat er een algoritme in het spel was, en wat de rol van dit algoritme was. Maar ook toezichthouders moeten dit kunnen weten en idealiter ook het bredere publiek. De informatiebehoefte en mate waarin verantwoording afgelegd kan en mag worden zal daarbij verschillen van doelgroep tot doelgroep. Maar duidelijk is dat, om de risico's van algoritmes aan te pakken, we eerst moeten weten wáár we moeten kijken. Transparantie, los van de precieze vorm, is dus een voorwaarde om publieke belangen en grondrechten te waarborgen.

Vervolg vraag is: hoe pakken wij onze missie op?

Zoals gezegd, we zijn met deze nieuwe coördinerende algoritmetaak net uit de startblokken gekomen. De werkzaamheden van de DCA zijn samen te vatten in drie type activiteiten: 1) vroegtijdige signalering van ontwikkelingen en risico's in het gebruik van AI, 2) het samenbrengen van toezichthouders die met algoritmes en AI geconfronteerd worden in hun toezichttaak en 3) het werken aan beleid, uitleg en regelgeving op het terrein van algoritmes en AI.

Net als ieder ander die nadenkt over het beter beheersen van algoritmes en AI is het in ons werk op dit moment pionieren. En één van de eerste resultaten van onze pioniersactiviteiten is de Rapportage Algoritmerisico's Nederland die wij in juli 2023 voor het eerst gepubliceerd hebben. Vanaf nu zullen wij deze rapportage twee keer per jaar publiceren, zodat we continu de vinger aan de pols houden wat betreft de belangrijkste algoritmerisico's en de mate waarin we daar als samenleving grip op krijgen.

Maar wat zijn nu de belangrijkste risico's? Zijn die meetbaar en is het niet appels en peren vergelijken? Een belangrijke vraag, die tegelijkertijd al deels het huidige antwoord geeft. We hebben in ons eerste rapport dan ook aangegeven dat we op dit moment twee voornaamste risico's zien.

Het eerste risico is de snelheid waarmee nieuwe AI-innovaties, zoals intelligente chatbots gebaseerd op large language models (LLMs) onze samenleving binnendringen. En dat er zoals ik eerder heb geschetst weinig bewustzijn is over hoe deze modellen werken en wat daarvan de gevolgen en risico's kunnen zijn. Naast veel andere risico's is er ook hier de dreiging van willekeur. Want generatieve AI genereert output zonder dat er inzicht is in de kans dat die min of meer samenhangende zinnen ook daadwerkelijk 'waar' zijn.

Het tweede risico is het gebrekkig inzicht in bestaande algoritmes. Als we onvoldoende weten wat voor algoritmes met hoog risico waar gebruikt worden, dan is het ook moeilijk zicht krijgen op de mate van beheersing van bijkomende risico's. Ontwikkelingen zoals registers juichen we dan ook toe, en de AI Act die nu in onderhandeling is gaat hopelijk helpen om te komen tot duidelijke spelregels over hoogrisico-algoritmes.



Een belangrijke boodschap die wij afgeven is dat politiek, bestuur en organisaties voortgang dienen te blijven maken met de inspanning voor beleid, implementatie en intern en extern toezicht dat ziet op algoritmes. Het vraagt ook dat organisaties meer personeel inzetten en verder werk maken van scholing om algoritmerisico's te beheersen.

We roepen ook op tot kritische zelfreflectie. We vinden dat organisaties die twijfelen aan de adequaatheid van hun risicobeheersing er verstandig aan doen om voorlopig terughoudend te zijn met de inzet van algoritmes.

Een boodschap die overigens ook door anderen wordt uitgedragen. In juni 2023 publiceerde de Data & Trust Alliance, een consortium van bedrijfsleven en consultancy, een kort document over de urgentie van AI-governance.<sup>vi</sup> Het document stelt dat iedere partij moet beginnen met het vaststellen van de volwassenheid van de eigen organisatie om vast te stellen wat de sterke en zwakke punten zijn in de mate waarin de risico's van AI gemanaged worden. Belangrijk is de kanttekening die hierbij wordt geplaatst: zo'n volwassenheidsbeoordeling "klinkt heel basaal, maar de meeste organisaties doen dat niet of kunnen dat niet."

Aan het einde van mijn verhaal wil ik graag de sprong maken naar de verzekeringssector.

Data en verzekeren hoort bij elkaar als boter en brood. Maar dat geldt evengoed voor het spanningsveld tussen enerzijds premiedifferentiatie die volgt uit schadestatistieken en risicomodellen en anderzijds de beginselen van gelijkheid en solidariteit. De link met AI en algoritmes is hier dat (potentiële) klanten een steeds verder gepersonaliseerd aanbod krijgen. Dat kan de inclusiviteit vergroten – voor onverzekerden ontstaat nieuw aanbod – maar kan ook verliezers opleveren. In Trendzicht 2023 waarschuwt de AFM expliciet dat "bepaalde consumenten op basis van het gehanteerde algoritme [kunnen] worden gediscrimineerd of uitgesloten van de markt".<sup>vii</sup>

Met dat in het achterhoofd ziet het ethisch kader dat is opgesteld door het Verbond van Verzekeraars er vanaf de buitenkant uit als een goede basis. De normen liggen er. En op basis van de verhalen die wij lezen begrijpen wij dat in ieder geval sommige verzekeraars met interne algoritmeregisters zijn gaan werken. Ook doet de Stichting toetsing verzekeraars als onafhankelijk keuringsinstituut een toetsing van de naleving van deze zelfregulering.

En wees gerust: als DCA binnen de AP zullen wij straks met interesse kennismaken van deze evaluatie.

We hopen op een zo transparant mogelijk overzicht van toetsingsuitkomsten. Want, dit is uiteindelijk wel de stap die genomen moet worden. Welke acties zetten verzekeraars precies om de doelstellingen van het ethisch kader na te leven. Wat kan daarover verteld worden. En wat is de voortgang? Voor ons is het nu lastig te beoordelen. Transparantie is ook hier uiteindelijk de basis voor het vertrouwen.

En dan tot slot de vraag: welke rol kan een data scientist hier individueel spelen? Ik wil u twee boodschappen meegeven.

Allereerst, onderschat niet uw impact. Een algoritme of AI-toepassing die u mede ontwikkelt, of waarvan u bijdraagt aan de implementatie in uw organisatie, raakt al snel miljoenen bestaande of potentiële klanten. Ook als het gaat om fraudebestrijding. Denk aan hen die de uitzondering op de regel zijn. Hoe pakt het algoritme voor hen uit? Wees ook alert als het gaat om een algoritme dat medewerkers binnen de



organisatie slechts 'adviseert'. Met een 'human in the loop' zijn zorgen over algoritmes niet automatisch weggenomen. In de wetenschap omtrent algoritmes en AI is 'overreliance' inmiddels een bekend concept: de menselijke natuur is dat we structureel vaker een foutief advies van een algoritmisch model toch overnemen, dan dat we een juist advies onterecht corrigeren.

Daarnaast, overschat niet het automatische zelfcorrigerend vermogen van uw organisatie. Risicovolle inzet van algoritmes en AI wordt niet altijd uit zichzelf ondervangen. En dit is niet bedoeld als kritiek, maar een erkenning van de realiteit die dus ook wordt uitgesproken namens grote internationale organisaties. Zoals we zagen in het voorbeeld van de datingapp, de risico's van bijvoorbeeld discriminatie zitten soms diep verstopt in het algoritme.

Het zijn mede het werk en de expertise van statistici en data scientists die kunnen helpen deze problemen boven water te krijgen. En niet alleen in de ontwikkelingsfase, maar zeker ook gedurende inzet. Het monitoren van uitkomsten en het algoritme waar nodig corrigeren is essentieel. Dit geldt voor statische algoritmes. Maar het geldt ook voor zelflerende algoritmes waarin discriminatoire of willekeurige patronen en voorkeuren steeds dieper kunnen inslijten.

Ik kom tot een afronding. Mijn verhaal begon met de boodschap dat de zoektocht naar een evenwicht tussen algoritmische innovatie en het adequaat beschermen van mensen in volle gang is. Ik denk dat we met zijn allen kunnen concluderen dat dit geen makkelijke opgave is. Laten we daarom vooral in openheid het gesprek aangaan en de inhoud induiken. Ik kijk uit naar de rest van deze middag.

Dank u wel voor uw aandacht.

---

<sup>i</sup> <https://www.mensenrechten.nl/actueel/nieuws/2023/09/06/dating-app-breeze-mag-en-moet-algoritme-aanpassen-om-discriminatie-te-voorkomen>

<sup>ii</sup> <https://mtsprout.nl/groei/breeze-algoritme-discriminatie>

<sup>iii</sup> Zie Rekenkamer (2020), "Gekleurde technologie: verkenning ethisch gebruik algoritmes".

<sup>iv</sup> Zie Cathy O'Neil (2016), "Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy".

<sup>v</sup> Ook bij veel AI toepassingen wordt volledig voorbij gegaan aan het bepalen inzicht in onzekerheidsmarges. Dat is een groot verschil met 'ouderwetse' statistische methoden.

<sup>vi</sup> <https://www.ibm.com/downloads/cas/MV9EXNV8>

<sup>vii</sup> Zie AFM (2022), "Trendzicht 2023".