Datum	2 april 2025
Behandeld door	Ambtelijk secretaris
E-mail	cpa@amsterdam.nl

Onderwerp Bias in algoritmes

Advies van de Commissie Persoonsgegevens Amsterdam (CPA) over bias in algoritmes. Het agendapunt is behandeld op de openbare vergadering van 30 januari 2025.

Geachte heer 5.1, 2, e

Tijdens de vergadering van de CPA van 30 januari 2025 was u aanwezig bij het agendapunt "Dialoog: Hoe de gemeente omgaat met bias in algoritmes", tezamen met 5.1, 2, e .

Bijgaand treft u aan het advies van de CPA.

A. Samenvatting advies

De Commissie geeft u de volgende **adviezen** ter overweging mee:

- maak expliciet hoe het afwegingskader rekening houdt met de impact op kwetsbare burgers in de samenleving. Integreer een toetsingsmechanisme dat het risico op discriminatie en achterstelling beoordeelt;
- breid de bias-analyse uit met verschillende bias-analysemethoden;
- hanteer een duidelijk onderscheid tussen bias en eerlijkheid, waarbij bias wordt gedefinieerd als een technisch en statistisch concept;
- hanteer een passende verwerkingsgrondslag onder de AVG en de AI-verordening voor de verwerking van persoonsgegevens in een bias-analyse, zoals een beroep op een wettelijke verplichting of het algemeen belang;
- gebruik afkomstgerelateerde kenmerken niet in algoritmische risicoprofilering, tenzij hier een goed onderbouwde en noodzakelijke reden voor bestaat en een grondige ethische en (privacy)juridische toetsing is uitgevoerd;
- beoordeel algoritmes te allen tijde op uitlegbaarheid.

De Commissie:

- wenst te vernemen op welke wijze de gemeente Amsterdam gevolg heeft gegeven aan haar adviezen.

B. Aanleiding en omschrijving

Het Plan van aanpak Algoritmen[1] bevat verbeterinitiatieven om het gebruik van algoritmes binnen de gemeente Amsterdam eerlijker en transparanter te maken. Dit plan, eerder besproken met de

[1] https://openresearch.amsterdam/nl/page/107246/plan-van-aanpak-algoritmen

Commissie Persoonsgegevens Amsterdam (CPA) en gepresenteerd op het CPA-symposium, resulteert in een geactualiseerd beheerskader voor het verantwoord gebruik van algoritmes en een diepgaande analyse van de algoritmes die op dit moment door de gemeente worden gebruikt.

Een belangrijk onderdeel van dit beheerskader is de bias-analyse, die momenteel als onduidelijk en praktisch onwerkbaar wordt ervaren. Daarom wordt door de gemeente gewerkt aan een verbeterde versie die begrijpelijker en toepasbaar is, zodat de gemeente beter kan beoordelen of burgers met een bepaald algoritme "eerlijk" behandeld worden. Hiervoor zijn uitgangspunten geformuleerd die worden uitgewerkt in een praktisch instrument en moeten bepaalde keuzes worden gemaakt, omdat een bias-analyse zelden een eenduidig goed of fout kent. Ook is het vrijwel onmogelijk om bias volledig uit te sluiten.

Om tot een robuust en praktisch instrument te komen, is advies gevraagd van zowel interne als externe experts, waaronder Stichting Algorithm Audit, de Hogeschool van Amsterdam en andere kennisinstellingen en overheidsinstanties. De feedback van deze partijen is grotendeels verwerkt in de huidige conceptversie van het document 'Achtergrond bias-analyse - uitgangspunten en keuzes'. Na afronding zal het document ter goedkeuring worden voorgelegd aan het DT van het I-domein, de Wethouder ICT en Digitale Stad en uiteindelijk de gemeenteraad.

Verzoek aan de CPA

De gemeente verzoekt de CPA om feedback, advies en/of suggesties op het document 'Achtergrond bias-analyse - uitgangspunten en keuzes' in het algemeen en in het bijzonder over de keuzes die beschreven zijn.

C. Advies

De CPA dankt de bij dit adviesverzoek betrokken medewerkers van de gemeente voor hun toelichting en de aangeleverde stukken.

Afwegingskader

De gemeente onderscheidt drie domeinen voor het afwegingskader dat in het instrument wordt gebruikt om te beoordelen of burgers met een bepaald algoritme eerlijk behandeld worden: organisatorisch, juridisch en technisch. Het beheerskader moet het mogelijk maken een algoritme te analyseren binnen een bepaalde context. Het dient als handreiking voor de gemeente om zelf weloverwogen afwegingen te kunnen maken die leiden tot een beargumenteerd en transparant oordeel ten aanzien van het gebruik van een algoritme.

De CPA constateert enkele aandachtspunten bij het huidige afwegingskader en de bias-analyse binnen het beheerskader. Zo maakt het afwegingskader onderscheid tussen de genoemde organisatorische, juridische en technische aspecten, maar ontbreekt een expliciete koppeling met de werkelijke gevolgen voor kwetsbare groepen. Dit is juist van belang voor mensen die al een verhoogd risico lopen op discriminatie, achterstelling of schending van mensenrechten. Daarnaast constateert de CPA onvoldoende diversiteit in de bias-analysemethoden. Er bestaan verschillende vormen van bias-analyse, maar het huidige afwegingskader reflecteert hier nog beperkt op. Dit kan leiden tot een onvolledig beeld van mogelijke bias in algoritmes. Ook wordt in het document eerlijkheid in verband gebracht met bias, terwijl dit twee verschillende concepten zijn. Bias is een meetbaar en statistisch

concept, terwijl eerlijkheid een normatief en contextafhankelijk vraagstuk is. De verwarring tussen deze begrippen kan leiden tot onduidelijkheid in de toepassing van het beheerskader.

De CPA adviseert om de maatschappelijke invalshoek te versterken, bijvoorbeeld door expliciet te maken hoe het afwegingskader rekening houdt met de impact op kwetsbare groepen in de samenleving. De gemeente kan daarbij een toetsingsmechanisme integreren dat het risico op discriminatie en achterstelling beoordeelt. Daarnaast adviseert de CPA om de bias-analyse uit te breiden met verschillende bias-analysemethoden. Afhankelijk van het type algoritme en de doelgroep kunnen verschillende methoden worden toegepast. Tot slot adviseert de CPA om een duidelijk onderscheid tussen bias en eerlijkheid te hanteren. Definieer bias hierbij als een technisch en statistisch concept en behandel eerlijkheid separaat, bijvoorbeeld binnen een ethisch afwegingskader waarbij wordt gekeken naar de mogelijke effecten op mensen in hun werkelijke belevingswereld.

De verwerking van persoonsgegevens bij een bias-analyse

De verwerking van persoonsgegevens vormt een uitdaging bij het uitvoeren van een effectieve bias-analyse. Dit komt voort uit de behoefte aan voldoende data voor een betrouwbare bias-analyse. Een grondige bias-analyse vereist namelijk een grote hoeveelheid (bijzondere) persoonsgegevens om patronen en mogelijke discriminatie op te kunnen sporen. Dit staat op gespannen voet met de beperkingen die voortvloeien uit de AVG ten aanzien van het gebruik van persoonsgegevens.

De CPA benadrukt dat de AVG niet voorschrijft dat zo min mogelijk gegevens moeten worden gebruikt, maar dat er niet méér gegevens mogen worden verwerkt dan nodig is voor het vooraf vastgestelde doel (artikel 5 lid 1 sub c AVG). Het doel van de bias-analyse moet hierbij onderscheiden worden van het doel van het algoritme. Wanneer de gemeente zich kan beroepen op een grondslag uit de AVG, dan kan de verwerking van persoonsgegevens voor de bias-analyse gerechtvaardigd zijn, mits de persoonsgegevens noodzakelijk zijn voor het doel. De CPA verwijst in dit verband naar het rapport van het College voor de Rechten van de Mens, waarin uitzonderingsgronden van artikel 9 AVG (inzake de verwerking van bijzondere persoonsgegevens) worden erkend bij bias-analyses.[2]

Hoewel de AI-verordening het in artikel 10 lid 5 mogelijk maakt om bijzondere persoonsgegevens te verwerken indien dit strikt noodzakelijk is voor het opsporen van bias, ontslaat dit de gemeente niet van de verplichting om een verwerkingsgrondslag te hebben onder artikel 6 AVG. De gemeente kan zich hierbij mogelijk beroepen op artikel 6 lid 1 sub c AVG (wettelijke verplichting), omdat de AI-verordening een wettelijke verplichting kent tot het uitvoeren van een bias analyse (artikel 10 lid 2 AI-verordening). Een andere mogelijkheid biedt artikel 6 lid 1 sub e AVG (taak van algemeen belang). Bias-analyse draagt onder meer bij aan de bescherming van grondrechten, waaronder het discriminatieverbod van artikel 1 van de Grondwet. Dit is bij uitstek een taak van algemeen belang dat aan de overheid – waar de gemeente Amsterdam deel van uitmaakt – is toebedeeld. De CPA adviseert de gemeente om, waar mogelijk, in dit verband contact op te nemen met de Autoriteit Persoonsgegevens om, de AVG en AI-verordening indachtig, af te stemmen welke grondslagen en uitzonderingsmogelijkheden toepasbaar zijn voor bias-analyse binnen de gemeentelijke context.

Gebruik van afkomstgerelateerde kenmerken

Het gebruik van afkomst-gerelateerde kenmerken in algoritmische risicoprofilering kan

[2] College voor de Rechten van de Mens, Toetsingskader risicoprofilering, integrale versie (januari 2025), p. 36.

discriminatierisico's met zich meebrengen. De gemeente neigt naar een standpunt in het beleidskader waarbij deze kenmerken standaard niet worden gebruikt, tenzij hier een sterke rechtvaardiging voor bestaat. In dit verband is in een AP-onderzoek naar het gebruik van een algoritme voor fraudeaanpak door DUO geconcludeerd dat selectiecriteria zonder objectieve rechtvaardiging discriminerend en onrechtmatig kunnen zijn.

De CPA adviseert de gemeente om het gebruik van afkomstgerelateerde kenmerken niet te gebruiken, tenzij er een goed onderbouwde en noodzakelijke reden voor bestaat. Volgens de CPA is het veelal niet noodzakelijk om dergelijke gevoelige kenmerken te verwerken om een model te bereiken met vrijwel gelijke accuratesse. Indien de gemeente dringende overwegingen heeft om afkomstgerelateerde kenmerken te gebruiken, dan dient een grondige ethische en (privacy)juridische toetsing te worden uitgevoerd, in lijn met de Quality Assurance Acceptatiecriteria van de gemeente. In deze toetsing moet ook worden geanalyseerd of alternatieve modellen zonder afkomst-gerelateerde kenmerken vergelijkbare accuratesse kunnen bereiken. Ook moeten de bevindingen uit het AP-onderzoek naar fraudeaanpak door DUO worden meegenomen. Hiermee kan de gemeente het risico op discriminatie minimaliseren en blijft risicoprofilering juridisch en ethisch verantwoord.

Machine learning algoritmes

De gemeente heeft ervoor gekozen geen machine learning algoritmes te gebruiken voor profileringsdoeleinden om uitlegbaarheid te kunnen waarborgen. De gemeente vindt het belangrijk om goed uit te kunnen leggen aan burgers waarom iemand een bepaald profiel is toegewezen.

De CPA merkt hierbij op dat ook algoritmes die niet onder de categorie machine learning vallen, niet per se transparant zijn. Transparantie neemt voornamelijk af naarmate de complexiteit van het algoritme toeneemt. Dit geldt ook voor algoritmes die niet onder de categorie machine learning vallen. De CPA adviseert daarom om algoritmes te allen tijde te beoordelen op uitlegbaarheid. Ook benadrukt de CPA het belang dat algoritmes niet alleen transparant moeten zijn voor burgers, maar ook voor de ambtenaren die met de algoritmes werken.

Tot slot

Als laatste punt brengt de Commissie in herinnering dat de rol van de CPA adviserend is, en niet instemmend. Het verzoek aan de Commissie om 'in te stemmen' of 'positief te adviseren' over het voornemen past niet bij de rol en taak van de CPA als extern adviesorgaan.

Verder verwijst de CPA u naar bijgevoegde notulen van de vergadering over het betreffende agendapunt.

Hoogachtend,

Mevrouw H. Janssen
Voorzitter Commissie Persoonsgegevens Amsterdam

Uittreksel van het verslag van het agendapunt "Dialoog: Hoe de gemeente omgaat met bias in algoritmes" van de vergadering op 30 januari 2025

De omgang van de gemeente met bias in algoritmes
Dit is de derde keer dat het Plan van Aanpak Algoritmes van de gemeente in de CPA passeert. Eerder gebeurde dat in een reguliere vergadering en op het symposium. Deze keer gaat het om de omgang met bias. Er is vooraf bij de vergaderstukken een document over bias-analyse toegestuurd, De projectleider stelt voor in de bespreking nu vooral in te gaan op de richtingen voor het maken van keuzes rond enkele problemen, zoals die in hoofdstuk 5 worden aangegeven.

Het algemene uitgangspunt is dat de gemeente een instrument wil hebben om te beoordelen of burgers met een bepaald algoritme eerlijk behandeld worden. Analoog aan de nationale overheid worden drie domeinen voor de kaders gehanteerd: organisatorisch, juridisch en technisch. Het gaat erom de problemen en de oplossingen met elkaar te verbinden in een afwegingskader. Op technologisch gebied zijn er veel mogelijkheden, maar ook belemmeringen vanuit de andere twee domeinen. Verder is het zo dat het toetsen van bias lastiger wordt naarmate er meer kenmerken in het algoritme verwerkt worden. Er moeten dus keuzes gemaakt worden en 100% uitsluiting van bias is zo goed als onmogelijk.

De CPA vraagt welk belang al deze onderscheiden hebben voor de werkelijke wereld, met name voor de groepen die sowieso al een het risico lopen te maken te krijgen met discriminatie, achterstelling en schending van hun mensenrechten. Er zijn verschillende vormen van biasanalyse en daarvan vindt men weinig terug in het voorliggende stuk. De CPA heeft ook moeite met het betrekken van eerlijkheid bij het benoemen van bias. Bias is vooral een afweging op statistische gronden; eerlijkheid is een heel ander vraagstuk. Het organisatorische domein ziet men vooral rond de vraag op welk niveau de ethische dialoog gevoerd wordt.

De bedoeling van het beheerskader gaat vooral om het analyseren binnen een context. Het doel is niet oplossingen aan te leveren voor alle problemen, maar een handreiking te bieden om binnen de gemeentelijke organisatie zelf aan de slag te gaan en afwegingen te maken en zo te komen tot beargumenteerde en transparante oordelen. Het nu aangeleverde stuk gaat onderdeel worden van een totaal document voor de gemeente waarin alle aspecten aan de orde komen.

Een volgend probleem speelt rond de verwerking van persoonsgegevens. Voor een goede biasanalyse is een grote hoeveelheid persoonsgegevens nodig en dat staat op gespannen voet met de AVG die juist voorschrijft het gebruik van persoonsgegevens zo veel mogelijk te beperken.

De CPA merkt op dat de AVG het niet heeft over zo min mogelijk persoonsgegevens gebruiken, maar om niet meer gegevens te gebruiken dan voor het gestelde doel nodig is. Het doel van het algoritme zelf is daarbij iets anders dan het doel van de biasanalyse. Als de grondslag in orde is, kan men dus meer gegevens gebruiken. De CPA verwijst in dit verband ook naar uitspraken van het College voor de Rechten van de Mens over uitzonderingsgronden van artikel 9, onder meer bij bias-analyses. Het advies is daar dan contact over op te nemen met de AP. Er zijn knelpunten met de AVG, maar daar hoeft men niet per se een negatieve uitstraling aan te verbinden. Ook de AVG gaat over de bescherming van grondrechten. Het is niet altijd even eenvoudig, maar er blijkt meer mogelijk dan velen denken.

De projectleider zegt dat de gedachte was een algemene waarschuwing te laten uitgaan om zo veel mogelijk af te zien van het gebruik van (bijzondere) persoonsgegevens om te voorkomen dat men voor elk algoritme een uitgebreid traject zou moeten doorlopen aan de hand van zijn team dat een beperkte capaciteit heeft voor individuele begeleiding. Het lijkt hem wel nuttig te helpen bij het vinden van uitzonderingsgronden. Hij houdt zich aanbevolen voor tekstuele opmerkingen en aanvullingen van de kant van de CPA om die met de nodige snelheid in het document te verwerken. Het is de bedoeling het document op 3 maart met de wethouder te bespreken.

De technische bias-analyse levert ook als de (samengestelde) data van perfecte kwaliteit zijn, geen eenduidige uitkomsten op. Het kan bijvoorbeeld een cijfer opleveren, maar dan nog is er een beoordeling nodig aan de hand van juridische en ethische waarden nodig om een oordeel te vormen over de toepasbaarheid. Dat kan

men in hoge abstracte termen doen, maar de bedoeling is juist hier te komen met praktische handreikingen die aansluiten bij de informatiebehoefte van lijndirecties.
Amsterdam is voor zover bekend de enige gemeente die serieus bezig is met deze vragen rond bias-analyses en het werken aan zulke handreikingen aan lijndirecties en ze op die manier een handelingsperspectief te bieden. De bedoeling is deze aanpak openbaar te maken zodat andere gemeenten die kunnen overnemen.

De CPA adviseert meer onderscheid te maken tussen bias-analyses als een technische operatie aan de ene kant en anderzijds de mogelijke effecten op mensen in hun werkelijke belevingswereld. Wellicht komt men dan tot de conclusie dat bij bepaalde projecten geen nader onderzoek of extra maatregelen nodig zijn.

De projectleider benadrukt nogmaals dat dit document beoogt het hele speelveld van gegevensverwerking te bestrijken, maar dat bij het gebruik van de handreikingen in de praktijk de context expliciet in ogenschouw moet worden genomen. Rond AI-projecten moet nog veel uitgezocht worden en dat gaat ook gebeuren.
Een volgend punt is dat er verschillende meningen bestaan over het gebruik van afkomst-gerelateerde kenmerken in risicoprofilering via een algoritme. De richting die men in de handreikingen wil inslaan is dat men zulke kenmerken niet moet gebruiken tenzij er een met argumenten onderbouwde reden is om dit juist wel te doen. Daarbij neemt men ook de rapportages rond DUO ter harte.

De CPA tekent daarbij aan dat het vaak ook niet nodig blijkt te zijn dergelijke kenmerken te verwerken om te komen tot een model met vrijwel gelijke accuratesse.

Er bestaan verschillende opvattingen over het volledig automatiseren van processen waarin profilering een rol speelt. Mede op grond van ervaringen van DUO is ervoor gekozen willekeurige selectie naast algoritme - profilering-gedreven en signaal-gedreven selectie uit te voeren. Processen waarin hoger risico algoritmes gebruikt worden niet volledig te automatiseren maar ook (willekeurige) steekproeven te nemen om de eerlijkheid van het algoritme te testen. In de uitwerking wordt nader bepaald voor welk risico dit geldt en hoe de steekproeven ingericht dienen te worden qua scope, timing en proces.

De CPA denkt dat een reflectie op de vijf voorwaarden die de AP stelt om te voorkomen dat men in verkeerd vaarwater belandt, een verrijking van het document zouden opleveren.

De projectleider neemt die suggestie ter harte.

De CPA is benieuwd wat er gebeurt als men met dit kader tot de conclusie komt dat men moet stoppen met een bepaald algoritme.

De projectleider stelt dat het kader vooral bedoeld is om vooraf tot een goede afweging te komen of men al dan niet een bepaald algoritme zal invoeren. Over stoppen is wel nagedacht, maar daar is nog geen pasklare oplossing voor.

Tot slot merkt hij op dat er verschillende meningen bestaan over het gebruik van AI met zelflerende algoritmes bij profileringsprocessen. Vooralsnog is gekozen geen zelflerende algoritmes voor profileringsdoeleinden te gebruiken omdat men het belangrijk vindt te kunnen blijven uitleggen waarom iemand een bepaald profiel krijgt toebedeeld.

De CPA merkt op dat ook niet-zelflerende algoritmes niet transparant kunnen zijn. Over het algemeen neemt de transparantie af naarmate het instrument meer complex is. Bovendien is het belangrijk dat het instrument transparant is zowel voor de (individuele) burger als voor de ambtenaar die ermee moet werken.

De projectleider zegt toe dat dit in het Plan van Aanpak zal worden uitgewerkt. Hij bedankt voor dit in zijn ogen waardevolle gesprek en vraagt nogmaals aantekeningen en tekstsuggesties van CPA-leden los van het advies zo spoedig mogelijk aan hem toe te sturen zodat hij die tijdig kan verwerken in het document voor het gesprek met de wethouder op 3 maart.

De voorzitter bedankt hem voor deze dialoog en zegt een advies van de CPA toe.