

Artikel

Praktische handvatten voor empirisch toezicht op profileringsalgoritmes

Jurriaan Parie, Ylja Remmits, Brinn Hekkelman en Mark Kattenberg*

Bij toezicht op algoritmes en AI staat niet alleen het recht, maar ook de statistiek centraal. Aan de hand van een concreet profileringsalgoritme lichten we toe dat bestaande kaders, handreikingen en soft law onvoldoende houvast bieden voor het invullen van open normen uit het gelijkebehandelingsrecht. We illustreren dat beproefde methoden uit de empirische wetenschap handvatten bieden om deze normen in te vullen. Aan de hand van het praktijkvoorbeeld wordt een vragenlijst geïntroduceerd die toezichthouders ondersteunt in het stellen van gerichte vragen over indirect onderscheid door algoritmes en AI. Deze aanpak bouwt voort op wettelijke kaders en stelt toezichthouders in staat met beperkte middelen effectief toezicht te houden op algoritmegeïndreven besluitvormingsprocessen.

Inleiding

Binnen de democratische rechtstaat vervullen toezichthouders een belangrijke rol om rechtsbescherming en rechtszekerheid te bieden aan burgers, consumenten en organisaties. Om deze rol te vervullen met betrekking tot de verantwoorde inzet van algoritmes en artificiële intelligentie (AI) zijn nieuwe vaardigheden nodig. Kwantitatieve methoden spelen hierbij een steeds belangrijkere rol. Bij onderzoek naar een risicoprofileringsalgoritme in het proces ‘controle uitwonendenbeurs’ (CUB) van de Dienst Uitvoering Onderwijs (DUO)

is indirecte discriminatie bijvoorbeeld vastgesteld aan de hand van een datastudie waarin willekeurige steekproeven, statistische hypothesetoetsen en de migratieachtergrond van meer dan 300.000 studenten in de periode 2012-2023 retrospectief zijn geanalyseerd.¹ Voortbouwend op deze kwantitatieve studie heeft de controlerende macht – zowel de rechtspraak als de Tweede Kamer – de werking van het controleproces concreet kunnen normeren.²

Het is echter de taak van toezichthouders om dergelijke grondrechtenschendingen proactief te voorkomen. Met de almaar toenemende digitalisering en de beperkte mate waarmee organisaties risico's van AI-systemen weten te beheersen rijzen vragen.³ Hoe kan het dat er na de toeslagenaffaire ‘*bitter weinig is veranderd*’ om discriminerende algoritmes te voorkomen bij de Nederlandse overheid?⁴ Zal de AI-verordening het toezicht op algoritmes versterken door de ontwikkeling van concrete normen voor profileringsalgoritmes? En hoe kunnen toezichthouders met beperkte middelen hun taakstelling beter uitvoeren?

* J. Parie is wiskundige bij de Stichting Algorithm Audit. Y. Remmits is informatica bij de Stichting Algorithm Audit. Dr. B. Hekkelman is wiskundige bij het Centraal Planbureau. Dr. M. Kattenberg is econoom bij het Centraal Planbureau.

- 1 PwC, *Onderzoek misbruik uitwonendenbeurs*, 2024; Stichting Algorithm Audit (in opdracht van DUO), *Auditrapport Vooringenomenheid voorkomen*, 2024; en Stichting Algorithm Audit, *Addendum Vooringenomenheid voorkomen*, 2024.
- 2 Rb. Overijssel 29 oktober 2024, ECLI:NL:RBOVE:2024:5627; *Kamerstukken II* 2023/24, 24724, nr. 231, *Kamerstukken II* 2023/24, 24724, nr. 229.
- 3 Uit het onderzoek *Focus op AI bij de Rijksoverheid* van de Algemene Rekenkamer (2024) blijkt dat bij 35% van de AI-systemen die in gebruik zijn onbekend is wat de prestaties zijn. Onder meer deze uitkomst leidde tot 82 Kamervragen van de vaste commissie Digitale Zaken, *Kamerstukken II* 2024/25, Z16114 (Lijst van vragen inzake Publicatie Focus op AI bij de rijksoverheid (26643, nr. 1226).
- 4 Interview met Aleid Wolfsen, voorzitter van de Autoriteit Persoonsgegevens, “Bij zo’n beetje elke tegel die we lichten, ontdekken we discriminerende algoritmen bij de overheid”, zegt de Autoriteit Persoonsgegevens, *de Volkskrant* 2 juli 2024.

Dit artikel introduceert een vragenlijst aan de hand waarvan toezichthouders gericht toezicht kunnen houden op indirect onderscheid gemaakt door algoritmes en AI. Het bouwt hiermee voort op de oproep uit het Toetsingskader risicoprofilering van het College voor de Rechten van de Mens (CRM) om gebruik te maken van empirische methoden om indirect onderscheid te onderzoeken.⁵ Enerzijds is dit thema actueel – de inzet van algoritmes en AI-systemen zal alleen maar toenemen in de nabije toekomst. Anderzijds toont de DUO-casus aan dat reeds sinds decennia gevestigde methoden uit de empirische wetenschap hun weg slechts langzaam weten te vinden naar de beleids- en toezichtpraktijk.⁶

Aan de hand van het risicoprofileringsalgoritme van DUO wordt geïllustreerd hoe empirische methoden kunnen helpen om waardenspanningen, onder meer tussen effectiviteit en ongelijke behandeling, te beslechten. Achteraf bleek dat de profileringskenmerken uit het risicoprofiel (opleidingsniveau, leeftijd en afstand tot ouder(s)) sterk samenhangen met migratieachtergrond van studenten. Op basis van maatwerktabel van het Centraal Bureau voor de Statistiek (CBS) kon op populatieniveau exact worden vastgesteld dat in 2014 63,3% van de 16.023 mbo 1-2-studenten en 13,2% van de 104.814 wo-studenten een niet-Europese migratieachtergrond had. Omdat in het risicoprofiel een hogere risicoscore aan mbo-studenten werd toegekend, werd indirect dus ook een hogere risicoscore aan studenten met een migratieachtergrond toegekend. Aan de hand van deze kwantitatieve inzichten kon geïnformeerd worden afgewogen of het gemaakte onderscheid evenredig was.

In maart 2024 kwam de Minister van Onderwijs, Cultuur en Wetenschap (OCW) tot een oordeel: namens het kabinet bood minister Dijkgraaf excuses aan voor indirect onderscheid in het CUB-proces.⁷ In november 2024 werd een compensatieregeling voor meer dan 10.000 studenten aangekondigd ter waarde van ruim € 61 miljoen.⁸ Tegelijkertijd blijft er parlementaire steun voor de inzet van profilering door de Nederlandse overheid. In hetzelfde debat waarin de excuses werden aangeboden, stelde toenmalige minister Dijkgraaf dat: '(...) wij ons bewust zijn van de waarde van risicogericht toezicht. (...) Het is wel belangrijk dat risicogericht toezicht de juiste waarborgen kent.'⁹ Kwantitatieve methodes zoals besproken in dit artikel zullen hierbij een belangrijke rol spelen.

5 College voor de Rechten van de Mens, *Toetsingskader risicoprofilering. Normen tegen discriminatie op grond van ras en nationaliteit*, 2025, in het bijzonder p. 25-30.

6 De Wet op het CBS stelt sinds 2003 overheidsorganisaties in staat om in het kader van statistisch onderzoek specifieke populatiestatistieken op te vragen. Het trekken van willekeurige steekproeven en uitvoeren van statistische hypothesetoetsen om effecten, bijvoorbeeld van medicijnen, te meten is sinds de jaren vijftig van de vorige eeuw een beproefde methode.

7 *Kamerstukken II* 2023/24, 24724, nr. 220.

8 *Kamerstukken II* 2023/24, 24724, nr. 243.

9 *Kamerstukken II* 2023/24, D07566 (Beslisnota's bij Kabinetsreactie onderzoek naar controleproces uitwonendenbeurs).

De DUO-casus staat niet op zichzelf. Ondoordachte risicoprofilering is exemplarisch voor het bredere gebruik van algoritmes bij de overheid. 'Bij zo'n beetje elke tegel die we lichten, ontdekken we discriminerende algoritmes', aldus Aleid Wolfsen, voorzitter van de Autoriteit Persoonsgegevens (AP).¹⁰ Ook in de private sector is discriminatie door middel van profilering een risico. In februari 2025 rapporteerde Argos dat Rabobank gebruikmaakte van discriminerende kenmerken in transactie-monitoringssystemen en oordeelde het CRM dat Meta indirect verboden onderscheid maakt op grond van geslacht bij het tonen van advertenties voor vacatures.¹¹ Ook in deze casus hebben datastudies een belangrijke rol gespeeld om ongelijke behandeling aan te tonen.

Nu profileringsalgoritmes tot de digitale haarvaten van de maatschappij zijn doorgedrongen is het zaak dat toezichthouders burgers, consumenten en organisaties rechtsbescherming bieden tegen discriminatie. Daartoe worden in dit artikel de kaderstellende normen uit het non-discriminatierecht geanalyseerd. Daarna worden aan de hand van de DUO-casus empirische methoden uitgewerkt die behulpzaam zijn om profileringsalgoritmes aan de wettelijke kaders te toetsen. Vervolgens introduceren we een vragenlijst aan de hand waarvan toezichthouders gericht vragen kunnen stellen aan algoritmeontwikkelaars en gebruikersverantwoordelijken over indirect onderscheid door profileringsalgoritmes. Het artikel sluit af met een conclusie.

De open juridische norm: wie vult deze in?

Onderscheid is van alle tijden. Zowel in de klassieke oudheid als in de huidige maatschappij wordt onderscheid gemaakt voor onder meer toegang tot onderwijs, arbeid en huisvesting. Vandaag de dag zijn sommige vormen van onderscheid echter verboden. De Grondwet, Europese verdragen en de Algemene wet gelijke behandeling verbieden direct onderscheid op basis van beschermde gronden, zoals etniciteit en nationaliteit.¹² Verboden onderscheid kan ook indirect voorkomen via

10 Interview Aleid Wolfsen, *de Volkskrant* 2 juli 2024.

11 'Rabobank discrimineerde bij klantcontroles', NPO Radio 1, 15 februari 2025; CRM oordeel 2025-17, 18 februari 2025, 'Meta Platforms Ireland Ltd. maakt verboden onderscheid op grond van geslacht bij het tonen van advertenties voor vacatures aan gebruikers van Facebook in Nederland'; Onderzoek dat aanleiding vormde tot klacht: 'New evidence of Facebook's sexist algorithm', NGO Global Witness, 12 juni 2023.

12 Het discriminatieverbod volgt onder andere uit verdragen van de Europese Unie (EU) en het Europees Verdrag voor de Rechten van de Mens (EVRM). In Nederland is het discriminatieverbod verankerd in art. 1 Grondwet en uitgewerkt in de Awgb. Art. 1 Awgb vermeldt de volgende beschermde gronden: 'godsdienst, levensovertuiging, politieke gezindheid, ras, geslacht, nationaliteit, hetero- of homoseksuele gerichtheid of burgerlijke staat'. Beschermde gronden vermeld in EU-verdragen verschillen per context. Leeftijd is bijvoorbeeld wel een beschermde grond in Richtlijn 2000/78/EG tot instelling van een algemeen kader voor gelijke behandeling in arbeid en beroep (art. 1), maar niet in Richtlijn 2000/43/EG houdende toepassing van het beginsel van gelijke behandeling van personen ongeacht ras of etnische afstamming (art. 2).

zogenoemde *proxykenmerken*. In dat geval vindt onderscheid niet direct plaats op basis van een beschermde grond, maar leiden ogenschijnlijk neutrale selectiecriteria tot onevenredige benadeling van deze groepen. Voorbeelden van proxy's voor de beschermde gronden etniciteit en nationaliteit zijn postcode, hoogte van het inkomen, kenteken, familielid in het buitenland en laaggeletterdheid.¹³ Omdat correlaties in een statistisch model per definitie voorkomen, hebben profileringskenmerken altijd in meerdere of mindere mate een proxykarakter. Het maken van onderscheid op basis van proxykenmerken is niet per se verboden. Het gebruik moet echter goed gemotiveerd kunnen worden. Gelijkebehandelingswetgeving biedt hiervoor kaders. In deze paragraaf volgt een analyse hoe de cijfermatige wereld van algoritmes en AI zich verhoudt tot de letter van de wet, inclusief ontwikkelde *soft law*-kaders.

Wetgeving: te algemeen voor verantwoorde profileringsalgoritmes

Onder andere de Algemene verordening gegevensbescherming (AVG),¹⁴ de Algemene wet bestuursrecht (Awb) en de Algemene wet gelijke behandeling (Awgb) reguleren de inzet van profileringsalgoritmes. De concreetste normen voor gelijke behandeling komen uit het non-discriminatierecht. De Awgb en verdragen van de Europese Unie en de Raad van Europa stellen dat er mogelijk sprake is van verboden indirect onderscheid (discriminatie) wanneer een ogenschijnlijk neutrale bepaling, maatstaf of handelwijze personen van een beschermde groep in vergelijking met andere personen bijzonder treft. Er is geen sprake van discriminatie als het gemaakte onderscheid objectief gerechtvaardigd kan worden. Deze rechtvaardigingstoets bestaat uit drie onderdelen:¹⁵

- geschikt: zijn de profileringskenmerken voldoende relevant en objectief om op een effectieve(re) wijze bij te dragen aan de verwezenlijking van het legitiem nagestreefde doel?
- noodzakelijk: zijn er andere, minder ingrijpende, manieren voorhanden om het gewenste doel te bereiken?
- proportioneel: weegt het legitiem nagestreefde doel voldoende zwaar om het onderscheid te rechtvaardigen?

Bovenstaande normen beschrijven *wat* de algemene wettelijke kaders zijn; *hoe* deze open normen ingevuld moeten worden voor profilering is afhankelijk van de context. In de DUO-casus is het niet direct duidelijk hoe de geschiktheid, noodzakelijkheid en proportionaliteit

van de drie profileringskenmerken getoetst kunnen worden.

Ook de AI-verordening¹⁶ lijkt niet in deze behoefte te gaan voorzien. De productveiligheidswetgeving ziet toe op de verantwoorde ontwikkeling en verantwoord gebruik van AI-systemen door publieke en private organisaties, waarmee de veiligheid, gezondheid en grondrechten van burgers van de Europese Unie (EU) worden beschermd. Specifieke toepassingen van AI-systemen die zijn aangemerkt als 'hoog risico' moeten volgens de AI-verordening voldoen aan geharmoniseerde normen – ook wel bekend als CE-markering.¹⁷ De Europese Commissie heeft standaardiseringsorganisatie CEN-CENELEC dan ook verzocht tien normen te ontwikkelen die de eisen in de AI-verordening nader uitwerken, onder meer voor een risicomanagement- en kwaliteitsmanagementsysteem.¹⁸ Onder experts bestaan echter zorgen of onder de aanzienlijke tijdsdruk nuttige normen voor specifieke toepassingen van AI-systemen, zoals profilering, uit dit standaardiseringsproces zullen voortkomen.¹⁹ Vooralsnog zijn de geharmoniseerde normen procedureel van aard en zullen zij algemene normen, zoals uitlegbaarheid, menselijke tussenkomst en non-discriminatie, niet nader invullen. Daarnaast maken veel Nederlandse profileringsalgoritmes, zoals het DUO-algoritme, geen gebruik van een AI-systeem waardoor eisen uit de AI-verordening niet voor dit type algoritme van toepassing zullen zijn.²⁰ Hierdoor is het noodzakelijk dat de nationale wetgever en toezichthouders initiatieven van eigen bodem blijven verkennen om open juridische normen voor de verantwoorde inzet van profileringsalgoritmes te concretiseren.

Soft law-kaders: te soft voor verantwoorde profileringsalgoritmes

Dat er onduidelijkheden zijn hoe profileringsalgoritmes genormeerd moeten worden aan de hand van algemene wettelijke kaders is door toezichthouders en beleidsmakers niet onopgemerkt gebleven. Om nadere invulling te geven aan *hoe* open normen ingevuld moeten worden, zijn door verschillende instituties *soft law*-kaders ontwikkeld. Hoe bruikbaar deze kaders in de algoritmische praktijk zijn, wordt toegelicht aan de hand van de DUO-casus.

Het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK) heeft verschillende handreikingen, instrumenten en kaders die de laatste jaren zijn ontwik-

13 CRM, *Toetsingskader risicoprofilering*, 2025.

14 Verordening (EU) 2016/679 van het Europees Parlement en de Raad van 27 april 2016 betreffende de bescherming van natuurlijke personen in verband met de verwerking van persoonsgegevens en betreffende het vrije verkeer van die gegevens en tot intrekking van Richtlijn 95/46/EG (algemene verordening gegevensbescherming) (PbEU 2016, L 199/1-88).

15 De objectieve rechtvaardigingstoets geldt alleen als een legitiem doel wordt nagestreefd. In het geval van DUO wordt hieraan voldaan, namelijk naleving van de Wet studiefinanciering 2000 (WSF 2000).

16 Verordening (EU) 2024/1689 van het Europees Parlement en de Raad van 13 juni 2024 tot vaststelling van geharmoniseerde regels betreffende de artificiële intelligentie (verordening artificiële intelligentie) (PbEU L 2024/1689).

17 Onder meer art. 1 lid 2 en art. 40 AI-verordening.

18 Europese Commissie, Draft standardisation request amending implementing decision C(2023)3215 on a standardisation request in support of Union policy on artificial intelligence, 2023.

19 De auteurs gelieerd aan Algorithm Audit zijn via Stichting Koninklijk Nederlands Normalisatie Instituut (NEN) actief in verschillende werkgroepen van CEN-CENELEC.

20 Stichting Algorithm Audit, Richtlijnen voor AI-verordening implementatie – Definitie van een AI-systeem, <https://algorithmaudit.eu>, februari 2025.

keld samengebracht in het *Algoritmekader*.²¹ Hierin wordt voor een ‘toelichting over het uitvoeren van de rechtvaardigheidstoets’ verwezen naar het *Toetsingskader risicoprofilering* van het CRM (hierna: *Toetsingskader*).²² Het Algoritmekader en het Toetsingskader waren op het moment dat de vermoedelijke overselectie van studenten met een migratieachtergrond bij de controle op de uitwonendenbeurs aan het licht kwam in de zomer van 2023 nog niet gepubliceerd. De onderzoekers hebben destijds andere kaders geraadpleegd die weinig houvast boden om de specifieke profileringskenmerken te toetsen aan de geldende normen uit het gelijkebehandelingsrecht.²³ Wanneer het vermoeden van ongelijke behandeling is ontstaan, hoe moet dan de geschiktheid, noodzakelijkheid en proportionaliteit van de profileringskenmerken getoetst worden?

Bij een vermoeden van ongelijke behandeling is de eerste stap onderzoeken of er sprake is van indirect onderscheid. In de DUO-casus betekent dit nagaan of studenten met een migratieachtergrond via proxykenmerken in onevenredige mate zijn getroffen. Het Toetsingskader specificeert dat ‘met ‘onevenredige mate’ wordt bedoeld dat door de profilering de kans op nadere controle groter is vanwege een kenmerk vallend onder ‘ras’ of nationaliteit dan de gemiddelde kans om gecontroleerd te worden bij aselechte steekproefcontroles. De verwijzing naar gemiddelde kansen en aselechte steekproefcontroles veronderstelt een kwantitatieve methodologie om deze evenredigheid te onderzoeken.

De volgende stap is het uitvoeren van de rechtvaardigingstoets. Deze begint met het onderzoeken van de *geschiktheid* van profileringskenmerken. Het Toetsingskader licht toe dat ‘de essentie van deze toets is na te gaan of het gebruik van risicoprofilering daadwerkelijk bijdraagt aan het vaststellen van meer normovertredingen dan wanneer enkel aselechte controle (steekproeven) zouden worden gehanteerd’. Het kader schrijft echter niet voor hoe de ‘effectiviteit van elk van de afzonderlijke profielfactoren’ vastgesteld moet worden. Het Toetsingskader stelt wel dat dit ‘in feite een empirische toets’ betreft. In de DUO-casus betroffen de effectiviteit van aselechte steekproeven uit 2014 en 2017 respectievelijk 3,6 en 3,8%. De effectiviteit van het CUB-proces – waarvan het risicoprofiel een onderdeel was – bracht een effectiviteit van 38,9 en 35,3% voort.²⁴ Door de verwevenheid tussen toepassing van het risicoprofileringsalgoritme en handmatige doorselectie in het controleproces kon in de DUO-casus alleen de effectiviteit van het gehele controleproces vastgesteld worden. Hoe de geschiktheid van afzonderlijke profileringskenmerken empirisch getoetst is, bespreken we in de paragraaf *Empirische methoden*.

Wanneer is aangetoond dat het risicoprofiel de effectiviteit van controles verhoogt, is de volgende stap in de rechtvaardigingstoets het nagaan van *noodzakelijkheid*. Er moet worden nagegaan of er geen minder bezwaarlijke alternatieven zijn om het legitiem nagestreefde doel te bereiken. Het Toetsingskader stelt dat

‘het vergelijken van alternatieven een kwestie [is] van meten en vergelijken van effectiviteit, maar vergt ook een afweging: een alternatief handavingsprofiel dat minder of geen (indirect) onderscheid maakt geldt ook als redelijk alternatief als het marginaal meer kost en of minder effectief is dan het aanvankelijke risicoprofiel’.

Afgezet tegen de aselechte steekproef is het risicoprofiel effectief, maar DUO heeft tijdens de ontwikkeling van het algoritme onvoldoende onderzocht welke alternatieven voorhanden waren.²⁵ Hiermee is de noodzaak van inzet van het risicoprofiel onvoldoende onderbouwd. Het Toetsingskader stelt dat als de profileringsmethode geschikt en noodzakelijk wordt geacht de *proportionaliteit* van het risicoprofiel moet worden gewogen: ‘[zijn] de behaalde doelen van de risicoprofilering proportioneel ten opzichte van de negatieve effecten die deze vorm van risicoprofilering toebrengt?’. Hiertoe dienen alle ‘positieve en negatieve effecten van het risicoprofiel zorgvuldig in kaart te worden gebracht’. Er is geen zuivere oplossing om deze afweging tussen effectiviteit en inbreuk op gelijkheidsrechten te beslechten. Aan de hand van de DUO-casus worden in de paragraaf hierna data-analysetechnieken uitgewerkt die deze proportionaliteitsafweging informeren.

Empirische methoden

Digitale informatiesystemen stellen organisaties in staat om op schaal empirische methoden toe te passen tijdens ontwikkeling en gebruik van algoritmes.²⁶ Het CUB-proces van DUO kon zorgvuldig onderzocht worden omdat data over willekeurige steekproeven en de migratieachtergrond van studenten retrospectief voor de periode 2012-2022 onderzocht konden worden. De in dit onderzoek toegepaste empirische methoden worden in deze paragraaf nader besproken. Daarnaast wordt toegelicht hoe de methoden de rechtvaardigingstoets kunnen ondersteunen.

Datakwaliteit en de willekeurige steekproef

Als empirische methoden gebruikt worden om de rechtvaardigheidstoets te informeren moeten de beschikbare data kunnen worden vertrouwd. In ieder geval de volgende vragen moeten positief worden beantwoord: Zijn de data representatief voor de bijbehorende populatie? Zijn datapunten objectief en verifieerbaar? Kunnen

21 Ministerie van BZK, Algoritmekader v2.1, 2025.

22 CRM, *Toetsingskader risicoprofilering*, 2025.

23 ‘Studenten met migratieachtergrond opvallend vaak beschuldigd van fraude, minister wil systeem grondig nagaan’, NOS, 21 juni 2023; Ministerie van BZK, Impact Assessment Mensenrechten en Algoritmes (IAMA), 2021; en Audit Dienst Rijk, Onderzoekskader Algoritmes, 2023.

24 Zie de rapporten uit 2024 van PwC en Stichting Algorithm Audit.

25 Zie de rapporten uit 2024 van PwC en Stichting Algorithm Audit.

26 J. Kleinberg, J. Ludwig, S. Mullainathan & C.R. Sunstein, ‘Discrimination in the age of algorithms’, *Journal of Legal Analysis* 2018, p. 113-174.

handmatig toegekende labels worden vertrouwd? Vertrouwen in de datakwaliteit valt of staat met documentatiestandaarden. Hierin worden onder meer de betekenis, betrouwbaarheid en actualiteit van de data bijgehouden. Organisaties gelieerd aan het Ministerie van Justitie en Veiligheid (JenV) hebben bijvoorbeeld het JenV Afsprakenstelsel Gegevens & Algoritmes (JAGA) ingericht, waarin beleid en standaardisatie is opgenomen over het verwerken van gegevens.²⁷

Tijdens onderzoek naar de DUO-casus waren de beschikbare data over uitwonende studenten betrouwbaar. De profileringskenmerken onderwijsvorm, leeftijd en afstand tot ouders(s) konden allemaal eenduidig en objectief worden vastgesteld.²⁸ De data over de uitkomsten van het CUB-proces waren echter niet representatief, omdat jonge mbo-studenten onevenredig vaak zijn gecontroleerd en deze groep niet representatief is voor de gehele populatie uitwonende studenten. Door dit zogenoemde *vergrootglaseffect* zijn de data over de uitkomsten van het controleproces daarom ook niet representatief. De woonsituatie van jonge mbo-studenten is doorgaans anders dan de woonsituatie van wo-studenten.²⁹ Er is daardoor onvoldoende informatie beschikbaar over de effectiviteit van het CUB-proces voor bijvoorbeeld hbo- en wo-studenten.

Daarnaast moet de kwaliteit van de labels – in de DUO-casus de uitkomsten van het controleproces ('rechtmatig' of 'onrechtmatig' ontvangen uitwonendenbeurs) – vertrouwd kunnen worden. Het is denkbaar dat inspecteurs door menselijke vooringenomenheid of ambigue werkinstructies bepaalde studenten vaker het oordeel 'onrechtmatig' toebedelen. De verzamelde labels zijn in dat geval vooringenomen en mogelijk incorrect.³⁰ Verder is het belangrijk dat succesvolle bezwaarprocedures in de data gecorrigeerd worden. In het geval van de DUO-casus is de betrouwbaarheid van de labels en eventuele vooringenomenheid hierin niet als zodanig onderzocht, omdat ongelijke behandeling aan de hand van een datastudie naar alle verschillende stappen uit het CUB-proces doel van het onderzoek was.³¹

Het trekken van een willekeurige steekproef – ook wel een *aselecte steekproef* – gecombineerd met een zorgvuldig ontworpen dataverzamelingsproces is daarom een *best practice*.³² Bij een willekeurige steekproef worden betrokkenen geselecteerd voor controle zonder dat daarbij een profileringsmethode is toegepast om personen of organisaties te selecteren voor controle. Bij voor-

keur wordt een profileringsmethode ontworpen enkel op basis van data afkomstig van een willekeurige steekproef. Wanneer dit niet mogelijk is, moet de profileringsmethode geëvalueerd worden gebaseerd op de uitkomsten van een parallel aan het selectieproces getrokken willekeurige steekproef. In de DUO-casus was het al toegepaste CUB-risicoprofiel niet gebaseerd op een willekeurige steekproef. Het gebruikte risicoprofiel is vervolgens wel geëvalueerd aan de hand van 387 en 293 willekeurig geselecteerde studenten voor controle in 2014 en 2017.³³ Op basis van deze willekeurige steekproeven zijn aannames in het risicoprofiel getoetst.

Het toetsen van aannames

Als de data vertrouwd worden, kunnen empirische methoden worden ingezet ter ondersteuning van de rechtmatigheidstoets. Op basis van de willekeurige steekproef kan zowel de geschiktheid van profileringskenmerken uit een risicoprofiel, de noodzakelijkheid van het gehele model, als de proportionaliteit worden onderzocht. De methoden worden toegepast op de DUO-casus.

Aan de hand van door DUO getrokken willekeurige steekproeven kan statistisch worden onderzocht of er een verband bestaat tussen profileringskenmerken uit het risicoprofiel en onrechtmatig gebruik van de uitwonendenbeurs.³⁴ Door middel van een hypothesetoets kunnen aannames uit het profiel worden getoetst: is het inderdaad zo dat jongere studenten vaker onrechtmatig gebruikmaken van de uitwonendenbeurs dan oudere studenten? Hoe zit dit voor oudere studenten die ver bij hun ouders vandaan wonen? Na het toepassen van de statistische hypothesetoets (Z-toets) op de aselecte steekproef bleek alleen het kenmerk 'afstand tot ouder(s)' voorspellende waarde te hebben.³⁵ De andere kenmerken worden daarom vanuit deze eenvoudige statistische toets, los van mogelijke kwalitatieve bezwaren, ongeschikt geacht.

Om de geschiktheid van profileringskenmerken te onderzoeken kunnen ook geavanceerdere statistische methoden worden toegepast. Dit is nodig wanneer verschillende profileringskenmerken met elkaar samenhangen. Dit kan het geval zijn wanneer er voor individuele profileringskenmerken geen verband is tussen het kenmerk en de uitkomst, terwijl dit verband er binnen het gebruikte risicoprofiel wel is. Het risicoprofiel zou dan alsnog niet-geschikte kenmerken gebruiken.³⁶ Dit kan ondervangen worden door de individuele

27 Justitiële Informatiedienst, JenV Gegevensstandaard Referentiegegevens, www.justid.nl, 2023.

28 Onderwijsgegevens (mbo 1-2, mbo 3-4, hbo of wo) van studenten worden door DUO opgehaald bij onderwijsinstellingen, leeftijd wordt bijgehouden via de Basis Registratie Personen (BRP) en afstand tot ouder(s) wordt intern bepaald aan de hand van de afstand tussen postcodes.

29 Kenniscentrum Studentenhuysvesting, Landelijke Monitor Studentenhuysvesting 2014.

30 Een zogeheten ruisonderzoek kan inzicht bieden in de mate waarin menselijke beoordelingen van elkaar verschillen, zie D. Kahneman, O. Sibony & C.R. Sunstein, *Ruis* [vert.], Nieuw Amsterdam 201. Zie ook 'Bias experiment invloed labels op beslissing', *Kamerstukken II 2023/24, 2024D17779*.

31 Zie de rapporten uit 2024 van PwC en Stichting Algorithm Audit.

32 Rapportage AI- & Algoritmerisico's Nederland, Editie 3, 2024.

33 Zie de rapporten uit 2024 van PwC en Stichting Algorithm Audit.

34 Richtlijnen voor de grootte van de steekproef kunnen worden gevonden in Stichting Algorithm Audit, 'Grootte willekeurige steekproef', 2024.

35 Zie de rapporten uit 2024 van PwC en Stichting Algorithm Audit. Merk op dat 'afstand tot ouders' alleen in de analyse van de steekproef uit 2014 voorspellende waarde heeft en niet in de steekproef van 2017. Dit houdt onder andere verband met verschillen in de onderliggende populatie. In 2014 bestond de populatie uit mbo-, hbo- en wo-studenten. In 2017 bestond de steekproef, naar aanleiding van het afschaffen van de uitwonendenbeurs, enkel uit mbo-studenten. De voorspellende waarde van het kenmerk 'afstand tot ouder(s)' hangt dus af van het specifieke doel waarvoor de profileringsmethode wordt toegepast en moet per situatie worden onderzocht en onderbouwd.

36 Het tegenovergestelde kan ook waar zijn, waardoor aan het risicoprofiel relevante variabelen onthouden worden en het minder effectief is.

relatie tussen een kenmerk en de uitkomst te onderzoeken via het toetsen van conditionele statistische significantie.³⁷

De willekeurige steekproef is ook relevant om de noodzakelijkheid van een profileringsmethode na te gaan. De effectiviteit van geschikt geachte profileringskenmerken moet worden afgezet tegen de effectiviteit van de willekeurige steekproef. Zoals eerder vermeld was de effectiviteit van de door DUO getrokken willekeurige steekproeven uit 2014 en 2017 respectievelijk 3,6 en 3,8%. De effectiviteit van een risicoprofiel moet tegen deze cijfers worden afgezet. Deze uitkomsten informeren de proportionaliteitsafweging. Deze procedure dient herhaald te worden voor ieder geschikt kenmerk dat aan de profileringsmethode wordt toegevoegd.

Populatiestatistieken

Een van de meest complexe aspecten om de effecten van het risicoprofiel in kaart te brengen is het vaststellen van het proxykarakter van de profileringskenmerken. Een zuivere methode om het proxykarakter vast te stellen is een datastudie aan de hand van populatiestatistieken. Als de mate van indirect onderscheid is vastgesteld, moet dit tijdens de proportionaliteitsafweging worden afgezet tegen de noodzaak en geschiktheid om op basis van dit kenmerk onderscheid te maken. Voor de DUO-casus wordt toegelicht hoe het proxykarakter van de profileringskenmerken is vastgesteld.

Allereerst moet worden bepaald ten aanzien van welke beschermde grond het proxykarakter bepaald wordt. Het Europese non-discriminatierecht stelt dat 'ras' en etniciteit te allen tijden beschermde gronden zijn.³⁸ Dit is niet altijd het geval voor leeftijd.³⁹ Veel overheidsorganisaties, maar bijvoorbeeld ook banken, hebben gegevens tot hun beschikking (zoals nationaliteit) aan de hand waarvan het proxykarakter van profileringskenmerken vastgesteld kan worden. Deze bijzondere persoonsgegevens mogen echter niet altijd zomaar verwerkt worden volgens de AVG.⁴⁰ De AVG zelf biedt enkele uitzonderingen om deze data alsnog te mogen verwerken voor onderzoek naar mogelijke vooringenomenheid en ook de AI-verordening biedt hier een uitzondering voor.⁴¹ Wanneer deze gegevens niet beschikbaar zijn of ze niet verwerkt kunnen worden voor dit doel, kunnen overheidsorganisaties een maatwerkverzoek bij het CBS indienen.⁴² Aan de hand van statistie-

ken over groepssamenstelling kan zo het proxykarakter van profileringskenmerken worden bepaald.⁴³

Tijdens het DUO-onderzoek is gebruikgemaakt van CBS-aggregatiestatistieken om het proxykarakter voor de profileringskenmerken onderwijsvorm, leeftijd en afstand tot ouder(s) voor de beschermde grond migratieachtergrond vast te stellen. De resultaten zijn weergegeven in figuur 1. De rode lijn geeft het proxykarakter per profileringskenmerk weer. Hieruit volgt dat onderwijsvorm een sterk proxykarakter heeft: 13,2% van de wo-studenten heeft een niet-Europese migratieachtergrond, vergeleken met 63,3% van de mbo 1-2-studenten. Voor leeftijd geldt dat oudere studenten vaker een niet-Europese migratieachtergrond hebben. Hoe verder studenten bij hun ouder(s) vandaan wonen, des te vaker zij van Nederlandse herkomst zijn. Merk op dat er voor ieder profileringskenmerk in meerdere of mindere mate altijd sprake is van een proxykarakter. Een uitgebreide analyse van deze aggregatiestatistieken kan worden gevonden in het rapport *Addendum Vooringenomenheid voorkomen*.⁴⁴

Nu het proxykarakter van de profileringskenmerken bekend is, kan de proportionaliteit van het risicoprofiel worden gewogen. Aan de hand van de statistische hypothesetoets weten we dat de kenmerken 'onderwijsvorm' en 'leeftijd' ongeschikt zijn. Deze kenmerken hoeven in deze stap dus niet te worden overwogen. Hoe het kenmerk 'afstand tot ouder(s)' gewogen moet worden ten aanzien van het proxykarakter is een waardengedreven afweging die transparant en in samenspraak met belanghebbenden moet worden beslecht.⁴⁵

Als de beschermde grond op persoonsniveau beschikbaar is, dan kan het effect van proxykarakters op de uitkomst van de profileringsmethode (wie wordt wel of niet geselecteerd voor controle) gemonitord worden. In deze stap kan vertekening door proxykenmerken het meest direct worden vastgesteld. Dit is relevant omdat een risicoprofiel gebaseerd op meerdere criteria ervoor kan zorgen dat het proxykarakter van een individueel kenmerk in samenhang met andere proxykenmerken een ander effect heeft dan wanneer het profiel enkel uit één kenmerk bestaat. Aan de hand van de uitkomsten kan ook een verband worden gelegd tussen de verwachte effectiviteit van de profileringsmethode en de mate waarin de selectie afwijkt van de representatieve steekproef. Dit verband kan bijvoorbeeld aan de hand van *bijgestuurde selectie* vastgesteld worden en helpt om de proportionaliteitsafweging verder te informeren. Deze techniek wordt toegelicht in de paragraaf *Bijsturen op kenmerken*.

37 Zo kunnen twee versies van het risicoprofiel gemaakt worden, een model met en een model zonder het betreffende kenmerk, en kan vervolgens getest worden of er een verschil is in beide voorspellingen.

38 'Ras' is een juridische verzamelterm voor persoonskenmerken als huidskleur, etniciteit en nationale of etnische herkomst, zie CRM, *Toetsingskader risicoprofilering*, 2025; zie ook de beschrijving van het discriminatieverbod in noot 14.

39 Zie de beschrijving van het discriminatieverbod in noot 14.

40 Art. 9 AVG.

41 M. van Bekkum, 'Using sensitive data to de-bias AI systems: Article 10 (5) of the EU AI act', *Computer Law & Security Review* 2025, nr. 106115; art. 10 AI-verordening.

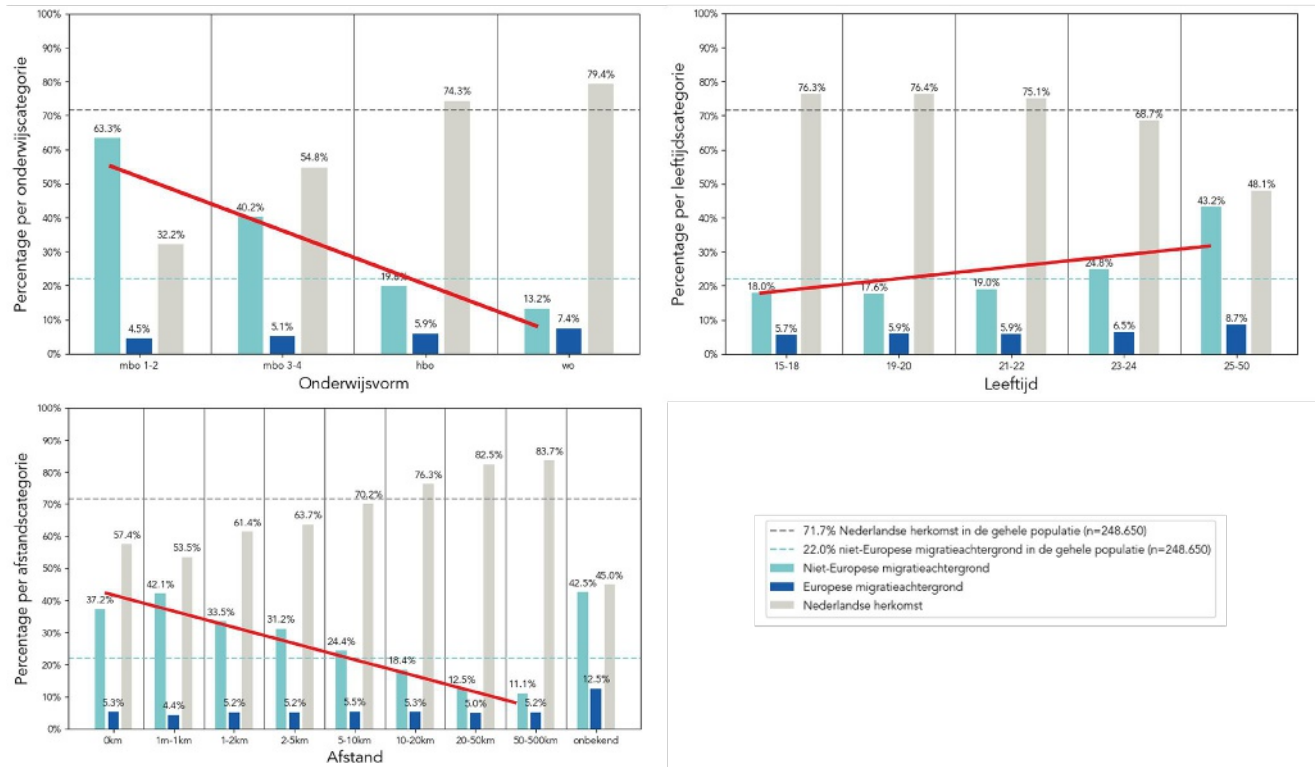
42 Art. 41-42 Wet op het Centraal Bureau voor de Statistiek. Zie ook CBS, *Maatwerk en microdata*, 2025.

43 Door middel van door het CBS opgestelde protocollen kunnen data op een veilige wijze worden verwerkt. Zo worden data via een beveiligde omgeving geanalyseerd en worden alleen aggregatiestatistieken gedeeld en dus geen privacygevoelige gegevens van individuele personen. Resultaten over groepen kleiner dan tien personen worden niet gepubliceerd.

44 Zie de rapporten uit 2024 van PwC en Stichting Algorithm Audit.

45 A. Meuwese, J. Parie & A. Voogt, 'Hoe "algorudentie" kan bijdragen aan de verantwoorde inzet van machine learning algoritmes', *NJB* 2024, afl. 10.

Figuur 1 Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst per onderwijs-, leeftijd en afstandscategorie in de uitwonendenbeurspopulatie 2014 (n=248.650) (bron: Stichting Algorithm Audit, Addendum Vooringenomenheid voorkomen, 2024).



Hoe toezichthouders inzage kunnen krijgen in de gemaakte keuzes ten aanzien van het proxykarakter van profileringsmethoden beschrijven we in *Vragenlijst voor toezichthouders*.

Demografische groepen niet beschikbaar

Ook als populatiestatistieken niet geanalyseerd kunnen worden, bijvoorbeeld omdat interne of CBS-data niet beschikbaar zijn, zijn er manieren om indirect onderscheid te onderzoeken. Naar deze methoden wordt doorgaans verwezen als *unsupervised learning of anomaly detection*. Hiervoor kunnen beproefde statistische methoden zoals clustering worden ingezet.

In het kader van wetenschappelijk onderzoek is clustering op de DUO-casus toegepast, om te onderzoeken welke mate van vooringenomenheid gedetecteerd had kunnen worden als CBS- aggregatiestatistieken over de migratieachtergrond van studenten niet beschikbaar waren geweest. Op basis van profileringskenmerken (onderwijsvorm, leeftijd, afstand tot ouder(s)) zijn studenten gegroepeerd (in clusters) die vaker dan gemiddeld door het profileringsalgoritme als 'hoog risico' zijn aangemerkt. Het meest benadeelde cluster bestaat uit mbo-studenten die gemiddeld genomen relatief dicht bij hun ouder(s) wonen.⁴⁶ Zonder toegang tot de migratieachtergrond van studenten had deze uitkomst als een eerste signaal kunnen fungeren voor domeinexperts om een vermoeden van ongelijke behandeling in het CUB-proces nader te onderzoeken.

Bijsturen op kenmerken

Door de uitkomsten van een profileringsalgoritme zorgvuldig te monitoren kan ongelijke behandeling worden gemitigeerd door middel van bijsturen.⁴⁷ Bijsturen houdt in dat de beheerder van het algoritme de selectie aanpast zodat het aandeel personen met een bepaald kenmerk op een gewenste hoeveelheid uitkomt.⁴⁸ Bijsturen biedt grip op onverwacht en ongewenst onderscheid door proxykenmerken. Door de willekeurige steekproef als referentiepunt te nemen kan ongelijke behandeling worden voorkomen. Een voordeel van bijsturen is dat het verlies in verwachte effectiviteit van het risicoprofiel minimaal is.⁴⁹ Wel is hiervoor toegang tot beschermde kenmerken, zoals migratieachtergrond, vereist. Bijsturen is onderdeel van het monitoren van een profileringsalgoritme tijdens gebruik. Na de toekenning van risicoscores wordt een extra stap geïntroduceerd waarin de selectie wordt bijgestuurd op basis van een specifiek kenmerk. In de DUO-casus zou dit betekenen dat de studenten met een migratieachtergrond met de laagste risicoscores in de 'hoogrisicocategorie' vervangen worden door evenveel studenten zonder migratieachtergrond met de hoogste risicoscores die niet al in deze categorie

46 F. Holstege e.a., 'Auditing a Dutch public sector risk profiling algorithm using an unsupervised bias detection tool', 2025, <https://arxiv.org/abs/2502.01713>.

47 Zie J. Kleinberg, J. Ludwig, S. Mullainathan & A. Rambachan, 'Algorithmic Fairness', *AEA Papers and Proceedings* 2018, 108, p. 22-27 of B. Heekelman, M.A.C. Kattenberg & B.J. Scheer, *Rechtvaardige Algoritmes*, CPB juli 2023, Document voor een toepassing op Nederlandse data en de term 'bijgestuurde selectiemethode'.

48 Het is mogelijk om selecties bij te sturen op meerdere kenmerken, voor een uitwerking hiervan zie: B. Heekelman, M.A.C. Kattenberg & B.J. Scheer, *The Costs of Affirmative Action: Evidence from a Medical School Lottery*, CPB Discussion Paper 2024.

49 Zie Kleinberg e.a. 2018, p. 22-27.

zaten. Hierdoor daalt het aandeel studenten met een migratieachtergrond in de ‘hoogrisicocategorie’, terwijl de verwachte effectiviteit op basis van risicoscores zo hoog mogelijk blijft. Op deze manier wordt bijgestuurd om een selectie te verkrijgen die effectiever is dan een aselechte steekproef (omdat een profileringsalgoritme wordt gebruikt), maar die qua samenstelling sterk lijkt op de steekproef (omdat daarnaar bijgestuurd wordt). De uitdaging zit vooral in het beslissen wanneer en in welke mate bijsturing moet plaatsvinden. Dit is een normatieve keuze die relateert aan het rechtvaardigen van onderscheid. Zo draagt bijsturen bij aan de proportionaliteitsafweging, omdat de relatie tussen effectiviteit en ongelijke behandeling wordt gekwantificeerd.

Vragenlijst voor toezichthouders

Toezichthouders – zowel intern als extern – hebben een belangrijke waarborgfunctie om ervoor te zorgen dat profileringsalgoritmes verantwoord worden ingezet, in het bijzonder wanneer deze technologie wordt ingezet voor risicoselectie. Voortbouwend op de geldende juridische kaders voor gelijke behandeling, onder meer beschreven in het Toetsingskader risicoprofilering van het College van de Rechten van de Mens, ondersteunt onderstaande vragenlijst toezichthouders bij het stellen van gerichte vragen over de verantwoorde inzet van profileringsmethoden. De vragenlijst is empirisch van aard en richt zich op het voorkomen van onwenselijk indirect onderscheid. De opgehaalde antwoorden informeren de objectieve rechtvaardigingstoets uit het non-discriminatie recht aan de hand waarvan verboden indirect onderscheid vastgesteld kan worden. De vragen relateren aan verschillende fases van de algoritmelevenscyclus en helpen te voorkomen dat burgers, consumenten en organisaties worden gediscrimineerd. We moedigen toezichthouders aan om de vragenlijst voor te leggen aan publieke en private organisaties.

Organisatieverantwoordelijkheden

1. Welke algoritmes en AI-systemen worden gebruikt binnen de organisatie? Let er hierbij op dat algoritmes en AI-systemen niet altijd als zodanig herkend worden.

Probleemanalyse

2. Wordt met het proces waarin de profileringsmethode een rol speelt een legitiem doel nagestreefd?⁵⁰
3. Welk probleem wordt met de inzet van de profileringsmethode opgelost?
4. Zijn kwetsbare groepen in de populatie geïdentificeerd? Is nagegaan wat de nadelige

gevolgen voor deze groepen zijn? Zo nee, waarom niet?

Datakwaliteit

5. Is er een willekeurige steekproef gebruikt voor het verzamelen van ontwikkeldata? Zo nee, zijn de gebruikte ontwikkeldata representatief voor de te onderzoeken populatie?
6. In hoeverre zijn door medewerkers toegekende labels betrouwbaar?
7. Zijn profileringskenmerken op voorhand uitgesloten voor de profileringsmethode, omdat ze bekend staan als proxykenmerk voor een beschermde grond, wettelijk zijn verboden, of omdat ze subjectief of veranderlijk zijn?⁵¹
8. Is het proxykarakter van profileringskenmerken vastgesteld door analyse van populatiestatistieken? Zo nee, waarom niet?
9. Houden geselecteerde profileringskenmerken inhoudelijk verband met het legitiem nagestreefde doel? Zijn de profileringskenmerken objectief en verifieerbaar?

Profileringsalgoritme

10. Is er een willekeurige steekproef getrokken, waarmee de uitkomsten van het profileringsalgoritme vergeleken kunnen worden? Zo nee, waarom niet?
11. Zijn aannames in de profileringsmethode getoetst op geschiktheid aan de hand van een statistische hypothesetoets? Zo nee, waarom niet?
12. Is de noodzakelijkheid van het gebruikte profileringsalgoritme bepaald door alternatieven te overwegen? Zo nee, waarom niet?
13. Is gelijke behandeling van kwetsbare groepen onderzocht door analyse van populatiestatistieken? Zo nee, zijn andere analyses uitgevoerd om mogelijke ongelijke behandeling te identificeren?
14. Is de effectiviteit van het profileringsalgoritme bepaald? Is bijsturing van ongelijke behandeling onderzocht en het effect daarvan op de verwachte effectiviteit van het algoritme gekwantificeerd? Zo nee, waarom niet?
15. Hoe is de proportionaliteitsafweging tussen gelijke behandeling en effectiviteit gemaakt?

Gebruik

16. Welke deel van de uitgevoerde controles is geselecteerd door middel van toepassing van de profileringsmethode? Welke deel van de geselecteerde groep is willekeurig geselecteerd?

50 Voorbeelden: beschermen openbare veiligheid, preventie van criminaliteit, handhaven immigratiebeleid, fraudebestrijding enz.

51 Enkele voorbeelden van proxykenmerken voor beschermde grond ras of nationaliteit zijn postcode, hoogte van het inkomen, kenteken, familielid in het buitenland, laaggeletterdheid (CRM, *Toetsingskader risicoprofilering*, 2025); Stichting Algorithm Audit Publieke standaard profileringsalgoritme, 2024.

17. Wordt de samenstelling van de geselecteerde groep voor controle gemonitord? Is bijsturen van de selectie overwogen? Zo nee, waarom niet?

Slotbeschouwing

Het normeren van indirect onderscheid is complex. Uit onze analyse volgt dat het gelijkebehandelingsrecht open normen formuleert die gebruikers te weinig houvast bieden om profileringsmethoden in de praktijk verantwoord in te zetten. Ook ontwikkelde *soft law*-kaders zijn hiervoor beperkt bruikbaar. Dit artikel beschrijft empirische methoden om te bepalen of indirect onderscheid gerechtvaardigd kan worden op basis van geschiktheid, noodzakelijkheid en proportionaliteit. Gebruik van de willekeurige steekproef, hypothesetoetsen, populatiestatistieken en het bijsturen op kenmerken staan hierbij centraal. Moderne digitale infrastructuur maakt het mogelijk om deze gegevens op populatieniveau en retrospectief te analyseren. Op basis van praktijkervaring hebben we in dit artikel een vragenlijst opgesteld die toezichthouders helpt relevante informatie op te vragen over onderscheid door profileringsalgoritmes. Zo kunnen toezichthouders hun kennispositie verbeteren, zonder daarbij zelf direct te hoeven normeren. Een vraag kan gemakkelijk worden gesteld. Nu onderzoekers voorbeelden hebben uitgewerkt, is het aan toezichthouders om gebruik van de methoden op te schalen.