



Reactie Amnesty International Nederland en Bits of Freedom op de “Oproep tot input AI-systemen voor risicobeoordelingen strafbare feiten, verbodsbepaling in EU-verordening 2024/1689(AI-verordening)”

02 April 2025

Inleiding

De Autoriteit Persoonsgegevens (hierna: AP) heeft een oproep uitgezet met een verzoek om input te leveren en enkele vragen te beantwoorden over de verbodsbepaling op AI-systemen voor ‘risicobeoordelingen strafbare feiten’ in de AI-verordening. Amnesty International Nederland en Bits of Freedom (hierna: Amnesty e.a.) maakt graag gebruik van deze kans.

Inhoud

1. Het verbod op ‘risicobeoordeling van strafbare feiten’
2. Risicoprofilering in criminaliteitsbestrijding valt onder de verbodsbepaling op ‘risicobeoordeling van strafbare feiten’
3. Discriminerende ‘risicobeoordeling van strafbare feiten’ is een reëel en geen hypothetisch probleem
4. De bescherming van mensenrechten moet leidend zijn bij de uitleg van de verboden
5. Discriminerende systemen zijn verboden ongeacht of ze onder de reikwijdte van de AI-verordening vallen
6. Registraties van onverdachte personen op politielijsten mogen geen aanleiding zijn voor ‘risicobeoordeling van strafbare feiten’
7. Handhavingsinstanties mogen geen registraties uitvoeren op basis van ‘risicobeoordeling van strafbare feiten’
8. ‘Objectieve en verifieerbare feiten’: redelijk vermoeden van schuld als drempel voor controles in de criminaliteitsbestrijding
9. Ook AI-systemen in het bestuursrecht kunnen onder verbodsbepaling D vallen
10. De AP moet waken voor misbruik van de uitzonderingsgronden
11. Betekenisvolle menselijk tussenkomst: “uitsluitend” moet niet misbruikt worden
12. Automation bias vormt risico voor mensenrechten
13. Concrete voorbeelden van (potentieel) discriminatoire toepassingen van risicoprofilering, respectievelijk “risicobeoordeling van strafbare feiten”

Uiteraard zijn Amnesty e.a. bereid om de bijdrage toe te lichten in een gesprek.

1. Het verbod op risicobeoordeling van strafbare feiten

In de AI-Verordening¹ is het verbod op ‘risicobeoordeling van strafbare feiten’ vastgelegd in Artikel 5, lid 1, onderdeel D (“verbod D”). Dat luidt:

“Het in de handel brengen, het voor dit specifieke doel in gebruik stellen of het gebruiken van een AI-systeem voor risicobeoordelingen van natuurlijke personen om het risico dat een natuurlijke persoon een strafbaar feit pleegt te beoordelen of te voorspellen, uitsluitend op basis van de profilering van een natuurlijke persoon of op basis van de beoordeling van diens persoonlijkheidseigenschappen en -kenmerken; dit verbod geldt niet voor AI-systemen die worden gebruikt ter ondersteuning van de menselijke beoordeling van de betrokkenheid van een persoon bij een criminele activiteit, die reeds op rechtstreeks met de criminele activiteit verband houdende objectieve en verifieerbare feiten is gebaseerd”

2. Risicoprofilering in criminaliteitsbestrijding valt onder de verbodsbepaling op ‘risicobeoordeling van strafbare feiten’

Amnesty e.a. menen dat geautomatiseerde risicoprofilering² bij criminaliteitsbestrijding onder de reikwijdte van het verbod op ‘risicobeoordeling van strafbare feiten’ in de AI-verordening moet vallen. De basis van risicoprofilering is dat aan de hand van ervaringen in het verleden en/of statistische analyse aannames worden gedaan over of correlaties worden vastgesteld tussen bepaalde kenmerken van personen/groepen en specifieke gedragingen. In het bijzonder wanneer risicoprofielen worden ingezet bij (politie)toezicht en controles, opsporing of handhaving inzake potentiële (strafrechtelijke) normovertreding als doel, gaat het om de inschatting van de *vermeende* kans dat een individu een wet gaat overtreden of een strafbaar feit gaat plegen. Kenmerkend voor risicoprofilering is dus dat dit door overheidsorganisaties, uitvoeringsorganisaties of private partijen proactief wordt ingezet, dus zonder concrete, geïndividualiseerde signaal of vermoeden dat iemand een wet gaat overtreden.

Het verbod zoals omschreven in Artikel 5, lid 1, onderdeel D (“verbod D”) geldt expliciet niet “voor AI-systemen die worden gebruikt ter ondersteuning van de menselijke beoordeling van de betrokkenheid van een persoon bij een criminele activiteit, die reeds op rechtstreeks met de criminele activiteit verband houdende objectieve en verifieerbare feiten is gebaseerd”. Het duidelijk definiëren van wat verstaan wordt onder ‘objectieve en verifieerbare feiten’ is hier van uiterst belang. In de volgende paragrafen gaan Amnesty e.a. hierop in.

3. Discriminerende ‘risicobeoordeling van strafbare feiten’ is een reëel en geen hypothetisch probleem

Bij de vragen die gesteld worden in de oproep van de AP over de verbodsbepaling over

¹ https://eur-lex.europa.eu/legal-content/NL/TXT/PDF/?uri=OJ:L_202401689

² Onder ‘geautomatiseerde risicoprofilering’ verstaat Amnesty International de inschatting, evaluatie of berekening van de kans op normovertreding van mensen of groepen (soms wordt gesproken van ‘voorspellen’ of ‘risicotaxatie’). Het kan gaan om min of meer gestructureerde of vastgelegde vormen van selectie, ondersteund door algoritmen. Vervolgens wordt op basis van deze inschatting een selectiebeslissing genomen, bijvoorbeeld over (extra) controle of handhaving, het inzetten van een aanpak of interventie, of het bepalen van de reactie op de normovertreding. Kenmerkend voor risicoprofilering is dat dit door overheidsorganisaties, uitvoeringsorganisaties of private partijen proactief wordt ingezet, dus zonder concrete, geïndividualiseerde signaal of vermoeden dat iemand een wet gaat overtreden. Risicoprofilering is gebaseerd op één of meerdere kenmerken (criteria, indicatoren), op basis waarvan een inschatting wordt gemaakt van een risico op normovertreding. De kenmerken hangen volgens de ambtenaar of organisatie die het risicoprofiel hanteert samen met een hoger risico op overtreding van een norm, regel of wet. Beslissingen om personen te controleren op basis van risicoprofilering zijn dus gebaseerd op statistische kansberekening, niet op feitelijk individueel gedrag dat wijst op fraude of criminaliteit.

Zie voor een uitgebreide bespreking, Amnesty International, “Etnisch profileren is overheidsbreed probleem”, 2024, <https://www.amnesty.nl/actueel/het-kabinet-moet-burgers-beschermen-tegen-etnisch-profileren>

‘risicobeoordeling van strafbare feiten’, wordt steeds gekozen voor de formulering “denkbeeldige AI-systemen”. Dit wekt de indruk dat ‘risicobeoordeling van strafbare feiten’ systemen een toekomstig of speculatief risico vormen. De praktijk laat echter zien dat discriminerende ‘risicobeoordeling van strafbare feiten’ een reëel en geen hypothetisch probleem is in Nederland.

In de afgelopen jaren hebben de schandalen over discriminerende risicoprofielen van overheidsorganisaties zich opgestapeld, onder meer bij de politie, de Koninklijke Marechaussee (KMar), de Belastingdienst, gemeenten, Dienst Uitvoering Onderwijs (DUO), de Immigratie en Naturalisatiedienst (IND) en de ministeries van Sociale Zaken en Werkgelegenheid (SZW) en Buitenlandse Zaken (BZ). Deze voorbeelden zijn door Amnesty uitgebreid gedocumenteerd in een in maart 2024 gepubliceerd rapport.¹ De Parlementaire enquêtecommissie Fraudebeleid en Dienstverlening publiceerde in maart 2024 het rapport ‘Blind voor mens en recht’ over fraudebestrijding in de sociale zekerheid. De commissie concludeerde dat als gevolg van het risicogericht toezicht, waarvan risicoprofilering de basis vormt, ‘bepaalde groepen mensen vaker in beeld zijn gekomen bij uitvoeringsorganisaties als mogelijke fraudeurs. Daarbij zijn de grondrechten op de eerbiediging van de persoonlijke levenssfeer en gelijke behandeling geschonden’.³ Ook in de private sector is discriminatie door risicoprofilering een veelvoorkomend probleem.⁴

4. De bescherming van mensenrechten moet leidend zijn bij de uitleg van de verboden

De vele schandalen in Nederland over discriminerende risicoprofilering in de afgelopen jaren laten zien dat het gebruik van risicoprofilering respectievelijk ‘risicobeoordeling van strafbare feiten’, inherente mensenrechtenrisico’s met zich meebrengt,⁵ inclusief, zoals in overweging 31 van de AI-verordening wordt benadrukt, “*discriminerende resultaten en de uitsluiting van bepaalde groepen[...]. Deze systemen kunnen een schending inhouden van het recht op waardigheid en non-discriminatie en van waarden van gelijkheid en rechtvaardigheid.*”⁶ Overweging 28 over verboden praktijken, waaronder ‘risicobeoordeling van strafbare feiten’, luidt: “*Dergelijke praktijken zijn bijzonder schadelijk en abusief en moeten worden verboden omdat zij in strijd zijn met de waarden van de Unie, namelijk eerbied voor de menselijke waardigheid, vrijheid, gelijkheid, democratie en de rechtsstaat, en met de grondrechten van de Unie die zijn vastgelegd in het Handvest, waaronder het recht op non-discriminatie, gegevensbescherming en privacy, en de rechten van het kind.*”

De bescherming van mensenrechten moet centraal staan bij de normuitleg van de verschillende termen in deze verbodsbepaling en in de AI-verordening in het algemeen. Het doel van deze verordening is namelijk dat AI-systemen mensenrechten respecteren.⁷ In dit licht moedigen Amnesty e.a. de AP aan om een heldere uitleg te ontwikkelen van de definitie van een AI-systeem een ‘risicobeoordeling van strafbare feiten’ systeem dat sterke mensenrechtenbescherming biedt. Het risicoprofiel van DUO dat jarenlang studenten discrimineerde op grond van ras en sociaaleconomische status, was zogenaamd ‘rule-based’. Het is voor rule-based risicoprofielen nog onduidelijk of deze systemen onder de reikwijdte van de AI-verordening vallen. De exacte interpretatie van de verschillende termen in de verordening of

³ Tweede Kamer der Staten-Generaal, “Blind Voor Mens En Recht. Rapport van de Parlementaire Onderzoekscommissie over Fraudebeleid.”, 2024.

⁴ Oordeel: ING discrimineert klanten op basis van hun afkomst bij transactiecontroles <https://pilp.nu/ing-discrimineert-klanten-op-basis-van-hun-afkomst-bij-transactiecontroles/> ; <https://www.mensenrechten.nl/actueel/nieuws/2024/07/25/ing-discrimineert-twee-klanten-door-betalingen-te-blokkeren-en-controlleren-vanwege-niet-nederlands-klinkende-naam>

⁵ Amnesty International, “Etnisch profileren is overheidsbreed probleem”, 2024, <https://www.amnesty.nl/actueel/het-kabinet-moet-burgers-beschermen-tegen-etnisch-profileren>

⁶ Recital 31 AI-verordening, https://eur-lex.europa.eu/legal-content/NL/TXT/PDF/?uri=OJ:L_202401689

⁷ <https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/nieuwe-technologieen-data-en-ethiek/artificiele-intelligentie-ai/ai-verordening/>

in Artikel 5 lid 1, onderdeel D (“verbod D”) moet de handhaving van discriminerende systemen niet in de weg staan.

De lat om te kunnen beargumenteren dat behandeling op basis van een risicoprofiel, respectievelijk ‘risicobeoordeling van strafbare feiten’ systeem, proportioneel is, moet dan ook hoog worden gelegd. Richtlijnen over de implementatie van Artikel 5, lid 1, onderdeel D moeten duidelijk maken dat voor het aantonen van de rechtmatigheid van een AI-systeem, de bewijslast bij de gebruiker van het AI-systeem ligt en niet bij burgers die hieraan worden onderworpen.

5. Discriminerende systemen zijn verboden ongeacht of ze onder de reikwijdte van de AI-verordening vallen

Amnesty e.a. wijzen de AP erop dat wanneer een AI-systeem of algoritme discriminatie als gevolg heeft, dit systeem verboden is en onmiddellijk uit de lucht gehaald moet worden, ongeacht of het onder de reikwijdte van de AI-verordening valt. Burgers die onderworpen worden aan risicovolle systemen, met inbegrip van risicoprofilering respectievelijk ‘risicobeoordeling van strafbare feiten’, moeten worden beschermd tegen de potentiële discriminerende gevolgen hiervan. Nu de AP de enige aangewezen toezichthouder is met handhavende bevoegdheden over AI-systemen, dringen Amnesty e.a. erop aan om proactief onderzoek uit te voeren en zo nodig overgaan tot handhaving. De grote aandacht die op de implementatie van de AI-verordening ligt, inclusief de onderhavige verbodsbepaling, zou kunnen leiden tot blinde vlekken op algoritmische systemen die niet aan de definities van AI- of ‘risicobeoordeling van strafbare feiten’ systemen voldoen, maar wel discrimineren. Amnesty e.a. maken zich hier zorgen over. De implementatie van de AI-verordening mag niet leiden tot minder daadkrachtig toezicht op discriminerende algoritmische systemen in brede zin.

6. Registraties van onverdachte personen op politielijsten mogen geen aanleiding zijn voor ‘risicobeoordeling van strafbare feiten’

Amnesty e.a. zien dat het veiligheidsrisico-denken leidt tot grootschalige informatieverzameling met onvoldoende aandacht voor mensenrechten. De brede controlebevoegdheden van de politie en de opdracht aan agenten om zo uitgebreid mogelijk te registreren staat haaks op de fundamentele rechten en vrijheden van mensen. De afgelopen jaren analyseerde Amnesty verschillende vormen van gegevensverwerkingen door de politie en de NCTV, onder meer in de rapporten *Ongecontroleerde Macht* en *In beeld bij de politie*.⁸ Uit deze en andere onderzoeken, zoals recent het onderzoek van de Nationale ombudsman, *Blind vertrouwen? Een onderzoek naar CTER-registraties en de impact ervan op het leven van burgers*, blijkt dat de overheid op grote schaal gegevens verzamelt zonder de noodzakelijke mensenrechten-waarborgen, wettelijke kaders en onafhankelijk toezicht.⁹

Deze registraties kunnen, ondanks dat ze in een meerderheid van de gevallen niet onderzoekswaardig worden geacht¹⁰, grote gevolgen hebben voor het leven van onschuldige burgers. Zo krijgt een politieagent CTER-codes in beeld bij een ID-controle, net als allerlei andere omschrijvingen en registraties van niet-strafbare gedragingen. Ook bij een scan van een kenteken komen registraties naar boven, die vervolgens aanleiding kunnen geven voor een controle.¹¹

⁸ Amnesty International, ‘Politie schendt rechten van demonstranten’, 1 juni 2023, <https://www.amnesty.nl/actueel/politie-schendt-rechten-van-demonstranten>; Amnesty International, ‘Camerasurveillance door politie moet aan banden worden gelegd’, 16 oktober 2024, <https://www.amnesty.nl/actueel/camera-surveillance-door-politie-moet-aan-banden-worden-gelegd>

⁹ Amnesty International, “Toezicht Gezocht: Amnesty Roept Op Tot Controle Op Politiedatabanken,” 2025, <https://www.amnesty.nl/actueel/amnesty-international-wil-beter-toezicht-op-politiedatabanken-in-nederland>.

¹⁰ Ombudsman, *Blind vertrouwen? Een onderzoek naar CTER-registraties en de impact ervan op het leven van burgers*, 12 november 2024, p. 32.

¹¹ Amnesty International, ‘Politie schendt rechten van demonstranten’, 1 juni 2023,

Bij de implementatie van het verbod op ‘risicobeoordeling van strafbare feiten’ vastgelegd in Artikel 5, lid 1, onderdeel D (“verbod D”), mogen dergelijke registraties geen onderdeel vormen van zogenaamde ‘objectieve en verifieerbare feiten’ die aanleiding zijn om over te gaan tot ‘risicobeoordeling van strafbare feiten’, respectievelijk risicoprofilering. Dit zou onschuldige burgers, inclusief vreedzame demonstranten, potentieel blootstellen aan willekeur, machtsmisbruik en discriminatie door politie en andere handhavingsinstanties.

7. Handhavingsinstanties mogen geen registraties uitvoeren op basis van ‘risicobeoordeling van strafbare feiten’

Eveneens mogen de politie en andere handhavingsinstanties geen registraties van onschuldige burgers (al dan niet uitsluitend) uitvoeren op basis van ‘risicobeoordeling van strafbare feiten’, respectievelijk risicoprofilering. In het bijzonder wanneer risicoprofielen worden ingezet bij (politie)toezicht en controles, opsporing of handhaving inzake potentiële (strafrechtelijke) normovertreding als doel, gaat het om de inschatting van de *vermeende* kans dat een individu een wet gaat overtreden of een strafbaar feit gaat plegen. De inschatting die gemaakt wordt door het risicoprofiel is dus geen bewijs dat iemand zich daadwerkelijk schuldig heeft gemaakt aan een strafbaar feit. De drempel om mensen te registreren op verdachten- of signalementlijsten moet hoog liggen en gepaard gaan met mensenrechtenwaarborgen, wettelijke kaders en onafhankelijk toezicht. Zie ook paragraaf 8 over redelijk vermoeden van schuld.

8. “Objectieve en verifieerbare feiten”: redelijk vermoeden van schuld als drempel voor controles in de criminaliteitsbestrijding

In een rechtsstaat mogen burgers verwachten dat de overheid hen vanuit vertrouwen bejegt en hen behandelt als burgers met vrijheden en rechten.¹² Burgers moeten erop kunnen rekenen dat overheden zorgvuldig en transparant handelen.¹³ Vanwege de inherente risico’s van risicoprofilering, zouden overheidsorganisaties hiermee zeer terughoudend moeten zijn.

Bij criminaliteitsbestrijding is een redelijk vermoeden van schuld aan een strafbaar feit de drempel om een burger te mogen controleren. Dit vermoeden moet objectief, concreet en geïndividualiseerd zijn. Om rassendiscriminatie en etnisch profileren tegen te gaan, moeten beslissingen om burgers te controleren volgens de ECRI gebaseerd zijn op individueel gedrag en/of ‘accumulated intelligence’.¹⁴

In tegendeel constateren Amnesty e.a. dat de Nederlandse wetgever de redelijk vermoeden van schuldstandaard in de afgelopen jaren steeds meer loslaat in de aanpak van terrorisme, ondermijning en andere vormen van criminaliteit.¹⁵ Voor gemeenten en samenwerkingsverbanden van (een groot aantal) overheidsorganisaties zijn ‘signalen’ steeds vaker voldoende om gegevens over personen uit te wisselen. Maatregelen die sinds 9/11 zijn genomen om terrorisme te voorkomen richten zich voor een aanzienlijk deel op het vroegtijdig signaleren van mogelijke radicalisering aan de hand van gedrag dat niet strafbaar is.¹⁶ Overweging 42 in de AI-verordening benoemt het redelijk vermoeden van schuld expliciet.¹⁷

<https://www.amnesty.nl/actueel/politie-schendt-rechten-van-demonstranten>

¹² Nationale Ombudsman, “Behoorlijkheidswijzer,” 2019,

<https://www.nationaleombudsman.nl/system/files/infomateriaal/No%20Behoorlijkheidswijzer%20juni%202019.pdf>.

¹³ Nationale Ombudsman, “Behoorlijkheidswijzer,” 2019.

¹⁴ European Commission against Racism and Intolerance, “ECRI, General Policy Recommendation 11, on Combating Racism and Racial Discrimination in Policing,” 2007, <https://rm.coe.int/ecri-general-policy-recommendation-no-11-on-combating-racism-and-racial/16808b5adf>.

¹⁵ Amnesty International, “Etnisch Profileren Is Overheidsbreed Probleem.”

¹⁶ Amnesty International, “Etnisch Profileren is Overheidsbreed Probleem.”

¹⁷ Recital 42 AI-verordening, https://eur-lex.europa.eu/legal-content/NL/TXT/PDF/?uri=OJ:L_202401689

Amnesty e.a. hebben herhaaldelijk zorgen geuit over potentieel indringende controles gericht op personen die niet verdacht worden van strafbare feiten. Amnesty e.a. roepen de AP op om de redelijk vermoeden van schuld-standaard aan te houden en beschermen bij de beoordeling van 'risicobeoordeling van strafbare feiten', respectievelijk risicoprofilering. Mensen mogen niet blootgesteld worden aan potentieel discriminerende risicoprofilering als ze niet concreet en individueel verdacht worden van een strafbaar feit. De richtlijnen door de AP zouden dus ook moeten specificeren dat 'signalen' of vermoedens van strafbare feiten, die de drempel van verdenking niet behalen, niet gekwalificeerd kunnen worden als een "objectieve en verifieerbare" menselijke beoordeling.

9. Ook AI-systemen in het bestuursrecht kunnen onder verbodsbepaling D vallen

De erosie van het redelijk vermoeden van schuld als drempel voor controles door handhavingsinstanties, gaat bovendien gepaard met een verschuiving van een aanpak binnen het strafrechtelijk kader naar een aanpak binnen het bestuursrechtelijk kader, met minder strikte waarborgen voor de rechten van burgers.

De Europese Commissie (hierna: EC) heeft in haar guidance¹⁸ over de verbodsbepalingen in de AI-verordening, als voorbeeld van verboden toepassing een omschrijving gebruikt van een AI-systeem dat ingezet wordt om belastingfraude op te sporen, dat lijkt op het risicoclassificatiemodel dat ontwikkeld werd door de Dienst Toeslagen.¹⁹ Dit laat zien dat verboden AI-systemen niet enkel in het strafrecht voorkomen, maar ook systemen die in het bestuursrecht onder deze verbodsbepaling kunnen vallen. Overigens valt volgens de EC een dergelijke AI-toepassing zowel onder verbodsbepaling C ("social scoring") als onder verbodsbepaling D ("risicobeoordeling van strafbare feiten").

10. De AP moet waken voor misbruik van uitzonderingsgronden

Amnesty e.a. en andere organisaties hebben zorgen geuit over de uitzonderingsgronden in de AI-verordening, bijvoorbeeld voor migratieautoriteiten en nationale veiligheid.

Deze uitzonderingsposities zijn zorgwekkend, omdat juist het gebruik van algoritmen in deze contexten vergaande gevolgen kunnen hebben voor mensenlevens en mensenrechten, zoals het recht op vrijheid, het recht op een eerlijk proces, de onschuldpresumptie en het recht op gelijke behandeling. Dit blijkt onder meer uit Amnesty's rapporten over het gebruik van algoritmen voor criminaliteitsbestrijding en de gevolgen hiervan voor verschillende mensenrechten.²⁰ De potentiële omvang en ernst van de impact van geautomatiseerde risicoprofilering in de context van migratie, opsporing, rechtshandhaving en nationale veiligheid vereisen een hoge norm van bescherming.

Het gebruik van algoritmen in deze contexten vereist dus niet minder waarborgen, maar juist meer. Amnesty e.a. roepen de AP op om bij de implementatie van de AI-verordening en in haar nieuwe rol als coördinator grondrechtentoezicht om te waken voor misbruik van deze uitzonderingsgronden en om de bescherming van grond- en mensenrechten voorop te stellen.

11. Betekenisvolle menselijk tussenkomst: "uitsluitend" moet niet misbruikt worden

¹⁸ ANNEX to the Communication to the Commission, Approval of the content of the draft "Communication from the Commission Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)"

¹⁹ Amnesty International, "Xenophobic Machines" 2021.

²⁰ Amnesty International, "Trapped in the Matrix: Secrecy, Stigma and Bias in the Met's Gang Database," 2018, <https://www.amnesty.org.uk/files/reports/Trapped%20in%20the%20Matrix%20Amnesty%20report.pdf>; Amnesty International, "We Sense Trouble," 2020.

Artikel 5, lid 1, onderdeel D ('Verbod D') verbiedt het gebruik van 'risicobeoordeling van strafbare feiten' wanneer deze plaatsvindt "*uitsluitend op basis van de profilering van een natuurlijke persoon of op basis van de beoordeling van diens persoonlijkheidseigenschappen en -kenmerken; dit verbod geldt niet voor AI-systemen die worden gebruikt ter ondersteuning van de menselijke beoordeling...*".

Dit gedeelte van het verbod overlapt met de definitie van geautomatiseerde besluitvorming zoals vastgelegd in art. 22 AVG. Hierbij merken Amnesty e.a. op dat elke risicoprofilering, met of zonder menselijk tussenkomst, en ook al leidt het niet tot overgang tot controle, een persoon in aanmerkelijke mate treft. De besluitvorming kan niet uitsluitend als geautomatiseerd worden geclassificeerd als er sprake is van betekenisvolle menselijke tussenkomst, maar dit mag niet slechts een "symbolisch gebaar" zijn.²¹ De vraag of de menselijke tussenkomst al dan niet betekenisvol is, is onder andere afhankelijk van hoeveel tijd wordt besteed aan de afhandeling van een individuele casus, bijvoorbeeld hoeveel tijd besteed wordt aan het verzamelen en afwegen van informatie.²² Bovendien heeft het HvJ-EU in 2023 geoordeeld dat het vaststellen van een risicoscore kan worden beschouwd als geautomatiseerde besluitvorming als het daaropvolgende besluit "hoofdzakelijk" afhangt van de risicoscore.²³

Amnesty e.a. maken zich zorgen over mogelijk misbruik van "uitsluitend", waarbij een slechts symbolische menselijk tussenkomst gebruikt wordt om het verbod op 'risicobeoordeling van strafbare feiten' te omzeilen. Amnesty e.a. vragen bijzondere waakzaamheid van de AP op dit punt en roept de AP op om concreet af te bakenen wat wordt verstaan onder "*ter ondersteuning van de menselijke beoordeling*". Wanneer een menselijk besluit "hoofdzakelijk" afhangt van een 'risicobeoordeling van strafbare feiten', respectievelijk risicoprofilering, moet dit in lijn met het SCHUFA-arrest alsnog worden aangemerkt als verboden praktijk onder de onderhavige verbodsbepaling.

12. Automation bias vormt risico voor mensenrechten

Overheidsorganisaties die menen onder de uitzondering op Art. 1, lid 1, onderdeel D te vallen ("*ter ondersteuning van de menselijke beoordeling*"), dienen zich bewust te zijn dat er bij het gebruik van algoritmen voor ondersteuning van beslissingen het risico bestaat op 'automation bias'. Dat wil zeggen dat ambtenaren zich te veel laten leiden door de uitkomst van het algoritme, in dit geval de risicoscores en categorisering. Dit staat op zeer gespannen voet met het zorgvuldigheidsbeginsel. Automatisering kan dehumaniserend zijn en een aantasting van de menselijke waardigheid, wanneer het minder ruimte laat voor het meewegen van persoonlijke omstandigheden (de 'menselijke maat').²⁴ Het principe dat er altijd een mens betrokken is bij de besluitvorming over mensen ('human in the loop' of menselijk tussenkomst) is dus niet vanzelfsprekend een waarborg voor zorgvuldige beslissingen over wie en wie niet te controleren. De menselijke betrokkenheid moet betekenisvol zijn (zie ook paragraaf 11).

13. Concrete voorbeelden van (potentieel) discriminatoire toepassingen van risicoprofilering, respectievelijk "risicobeoordeling van strafbare feiten"

a- Toepassingen door de politie en in de strafrechtketen

In 2020 publiceerde Amnesty het rapport *We sense trouble* over een experiment met 'predictive policing'.²⁵ Amnesty stelde vast dat de werkwijze van de politie in dit project neerkwam op etnisch

²¹ Article 29 Data protection Working Party, "Guidelines on ADM and Profiling under GDPR," 2018.

²² Gerechtshof Amsterdam, 4 april 2023, ECLI:NL:GHAMS:2023:793 (Uber II), r.o. 3.24.

²³ CJEU, C-634/21, 7 december 2023, SCHUFA Holding (Scoring).

²⁴ UN Special Rapporteur on Extreme Poverty and Human Rights, "Report of the Special Rapporteur on Extreme Poverty and Human Rights Philip Alston, A/74/50," 2019, https://www.ohchr.org/Documents/Issues/Poverty/A_74_48037_AdvanceUneditedVersion.docx.

²⁵ Amnesty International, "We Sense Trouble."

profileren. Het algoritmisch risicomodel had tot doel om vast te stellen of personen in geselecteerde voertuigen van Oost-Europese afkomst waren. De politie veronderstelde op basis van politieregistraties dat deze groep oververtegenwoordigd zou zijn bij winkeldiefstallen en zakkenrollerij in een winkelcentrum. Dit is een voorbeeld van de manier waarop etnisch profileren is ingebed in 'informatiegestuurd politiewerk'. De politie zette het experiment in 2020 stop.

De EC heeft in haar guidance²⁶ over de verbodsbepalingen in de AI-verordening, als voorbeeld van toepassingen die buiten de reikwijdte van het verbod op 'risicobeoordeling van strafbare feiten' vallen, meerdere voorbeelden genoemd van AI-systemen die het risico op recidive beoordelen bij de rechtelijke beslissing op voorlopige hechtenis of eerdere vrijlating van gevangenen. Amnesty e.a. wijzen de AP erop dat ongeacht of zulke toepassingen onder de reikwijdte van Art 5, lid 1, onderdeel D vallen, leiden ze tot hoge risico's op discriminatie.

In het kader van informatiegestuurd politiewerk gebruikte de politie verschillende risicotaxatieinstrumenten die tot doel hebben de kans op recidive in te schatten van bij de politie bekende personen.²⁷ In 2023 besloot de politie de inzet van het risicotaxatie-instrument RTI-Geweld stop te zetten, na een kritisch nieuwsbericht. De politie gebruikte RTI-Geweld sinds 2014. Tot 2017 kregen personen met een specifieke migratieachtergrond een hogere risicoscore. Deze criteria werd in 2017 verwijderd.

In het geval van de Amsterdamse Top400 werden niet-strafrechtelijke factoren gebruikt als indicatoren voor criminaliteit ("u bent minstens 3 keer van basisschool veranderd"), evenals louter vermoedens van betrokkenheid bij criminaliteit, zonder feitelijk bewijs.²⁸

De politie gebruikt(e) verschillende risicotaxatie-instrumenten waaronder ProKid, RTI-Geweld en Preselect Recidive die op basis van een algoritme de kans op recidive van bij de politie bekende (verdachte) personen berekenen. Hiervoor worden onder meer gegevens gebruikt over strafbare feiten gepleegd in het verleden. De kans op indirecte discriminatie is hierbij reëel.²⁹

De Raad voor Strafrechtstoepassing en Jeugdbescherming (RSJ) waarschuwde in 2021 voor de methodologische, praktische en ethische beperkingen wanneer dit soort instrumenten worden gebruikt voor strafrechtelijke (selectie)beslissingen.³⁰ De risico-inschatting door deze instrumenten is gebaseerd op 'statische' gegevens, ook wel 'historische risicofactoren' genoemd, zoals delictgeschiedenis, leeftijd waarop iemand voor het eerst een delict pleegde, leeftijd en geslacht. De Raad waarschuwt voor het gebruik van statische gegevens omdat deze 'niet ongedaan gemaakt [kunnen] worden waardoor het risico bestaat dat desbetreffende personen blijvend negatief worden beoordeeld'. Een ander probleem, door de RSJ niet benoemd, met het gebruik van delictgeschiedenis als risico-indicator is dat verschillen in aanhouding, vervolging of berechting tussen personen niet kunnen worden beschouwd als een objectieve indicatie van verschillen in gedrag, omdat deze verschillen (mede) het gevolg zijn van etnisch profileren en andere manieren waarop personen ongelijk worden behandeld in de strafrechtstketen. Verschillen in delictgeschiedenis reflecteren ook prioriteiten voor bepaalde vormen van criminaliteit en

²⁶ ANNEX to the Communication to the Commission, Approval of the content of the draft "Communication from the Commission Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)"

²⁷ Amnesty International, "Etnisch Profileren Is Overheidsbreed Probleem."

²⁸ Amnesty International, "Etnisch Profileren Is Overheidsbreed Probleem."

²⁹ Amnesty International, "Etnisch Profileren Is Overheidsbreed Probleem."

³⁰ Raad voor Strafrechtstoepassing en Jeugdbescherming, "Risicotaxatie in de Strafrechtstoepassing," 2021, <https://www.rsj.nl/documenten/rapporten/2021/11/18/advies-risicotaxatie-in-de-strafrechtstoepassing>.

wijken van de politie en andere actoren in de strafrechtsketen. Hierdoor bestaat er een risico op indirecte discriminatie op grond van ras.

Voor overige voorbeelden refereren we naar hier geciteerde recente Amnesty rapporten.

b- Toepassingen in de sociale zekerheid

De Dienst Uitvoering Onderwijs (hierna: DUO), een uitvoeringsorganisatie die werkt onder verantwoordelijkheid van het ministerie van Onderwijs, Cultuur en Wetenschap (OCW), gebruikte tussen 2012 en juni 2023 een algoritmisch risicoprofileringsysteem om studenten te selecteren voor een huisbezoek om mogelijk misbruik van de uitwonendenbeurs op te sporen. Een recent Amnesty rapport laat zien dat de DUO-werkwijze studenten discrimineerde op grond van zowel ras als economische en sociale status.³¹

De Belastingdienst en de dienst Toeslagen hanteerden in de afgelopen jaren verschillende discriminerende wijzen om te zoeken naar potentiële fraudeurs, al dan niet met behulp van (geautomatiseerde) risicoprofielen. Sinds 2013 gebruikte de Belastingdienst onder meer het 'risicoclassificatiemodel' om te proberen fraude met kinderopvangtoeslagen te voorspellen. Dit was een zelflerend risicoprofiel dat het risico op onjuistheden van aanvragers bepaalde.

In het geval van de onlangs door Lighthouse Reports en partners onderzochte Zweedse algoritmen voor risicoscores in de bijstand, werden mensen die door algoritmen als hoog risico werden aangemerkt, automatisch onderworpen aan onderzoeken door fraudecontroleurs binnen de bijstandsinstelling, waarbij van meet af aan werd uitgegaan van "criminele bedoelingen".³²

Voor overige voorbeelden refereren we naar hier geciteerde recente Amnesty rapporten.

³¹ Amnesty International, "Roekeloze Risicoprofilering: Studenten in Nederland Geschaad Door Discriminerende Fraudeaanpak", 2024, <https://www.amnesty.nl/content/uploads/2024/11/Amnesty-2024-rapport-Roekeloze-risicoprofilering-DUO.pdf?x99298>.

³² <https://www.lighthousereports.com/investigation/swedens-suspicion-machine/>