



ALGORITMEBELEID KAN NIEUW TOESLAGENSCHANDAAL NIET VOORKOMEN

MENSEN IN NEDERLAND NOG STEEDS ONVOLDOENDE
BESCHERMD TEGEN DISCRIMINERENDE RISICOMODELLEN

BRIEFING

AMNESTY
INTERNATIONAL 

INHOUD

SAMENVATTING	3
1. AANLEIDING: DISCRIMINERENDE ALGORITMEN	4
2. AMNESTY'S RAPPORT <i>XENOFOBE MACHINES</i> (2021)	7
3. ALGORITMEGEBRUIK IN LIJN MET MENSENRECHTEN	9
3.1 EEN VERPLICHTE EN BINDEnde MENSENRECHTENTOETS	9
3.2 VERBOD OP AFKOMST ALS CRITERIUM IN RISICOPROFIELEN	10
3.3 VERBOD OP ZELFLERENDE ALGORITMEN IN DE UITVOERENDE FASE	12
3.4 VERBOD OP BLACK BOX-ALGORITMEN	13
4. TRANSPARANTIE: EEN VERPLICHT EN BINDEND ALGORITMEREGER	14
4.1 HET ALGORITMEREGER MOET VERPLICHT WORDEN	14
4.2 HET ALGORITMEREGER MOET COMPLEET ZIJN	15
4.3 REGISTRATIE OOK VOOR RECHTSHANDHAVING EN NATIONALE VEILIGHEID	16
5. TOEZICHT: EEN ALGORITMEWAAKHOND MET MANDAAT EN MIDDELEN	16
6. AANBEVELINGEN	18

SAMENVATTING

Naar aanleiding van het toeslagenschandaal is er bij de regering en andere overheidsinstellingen meer aandacht gekomen voor de risico's van het gebruik van algoritmen in de publieke sector. Vooral waar algoritmen door overheden worden gebruikt voor het identificeren of opsporen van potentiële normovertreding in de context van criminaliteitsbestrijding, sociale bescherming en migratie is het risico op (indirecte) discriminatie groot. Sinds het toeslagenschandaal aan het licht kwam zijn er meer onthullingen geweest over overheidsinstellingen die discriminerende algoritmen gebruik(t)en om burgers te selecteren voor controle. Dit toont niet alleen aan dat het gebruik van algoritmen in de publieke sector zeer risicovol is, maar ook dat de regering en uitvoerende overheidsinstellingen zich onvoldoende inspanssen om deze risico's in te dammen.

De regering heeft sinds het toeslagenschandaal diverse acties ondernomen met als doel om te waarborgen dat algoritmen met respect voor mensenrechten worden ontwikkeld en gebruikt. Hoewel hiermee de eerste stappen in de goede richting zijn gezet, zijn de genomen en toegezegde maatregelen onvoldoende om burgers te beschermen tegen discriminatie en etnisch profileren door de overheid. De regering laat de deur open voor het gebruik van afkomst als criterium in risicomodellen, er is geen verbod op zelflerende en black box-algoritmen, de mensenrechtentoets en het algoritmeregister hebben geen verplichtend of bindend karakter, en de nieuwe algoritmetoezichthouder heeft onvoldoende mandaat en middelen. Met het huidige algoritmebeleid kan een nieuw toeslagenschandaal niet worden voorkomen.

De overheid mag burgers niet discrimineren en heeft daarnaast de plicht om burgers effectief te beschermen tegen discriminatie. Amnesty roept de regering op om:

- Onmiddellijk een bindende en verplichte mensenrechtentoets in te voeren voor algoritmische besluitvormingssystemen¹ in de publieke sector.
- Vast te leggen en af te dwingen bij overheidsinstellingen dat het nooit is toegestaan om etniciteit, nationaliteit en andere afkomstgerelateerde gegevens te gebruiken als criterium in geautomatiseerde risicoprofielen die worden gebruikt in de zoektocht naar potentiële normovertreders.
- Een verbod in te stellen op het gebruik van zelflerende algoritmen in de gebruiksfase van algoritmische besluitvormingssystemen in de publieke sector als het gaat om a) beslissingen met rechtsgevolgen, b) beslissingen die de rechten en vrijheden van burgers aantasten, en c) beslissingen die een grote impact hebben op de samenleving.
- Een verbod in te stellen op het gebruik van black box-algoritmen voor beslissingen in risicovolle contexten zoals criminaliteitsbestrijding, sociale bescherming en migratie.
- Transparant te zijn over het gebruik van algoritmen voor beslissingen over controles van burgers, onder meer door verplichte en complete publicatie in het algoritmeregister van alle algoritmische systemen, inclusief de algoritmen die worden ingezet in de context van opsporing, rechtshandhaving en nationale veiligheid.
- De algoritmetoezichthouder en andere relevante toezichthouders uit te rusten met adequate bevoegdheden en voldoende middelen.

¹ Een algoritmisch systeem dat wordt gebruikt bij (de ondersteuning van) verschillende stappen in besluitvormingsprocessen. Onder 'besluit' wordt in deze briefing verstaan: alle beslissingen van overheidsinstellingen over burgers, waaronder formele besluiten om burgers te controleren, maar ook beslissingen om burgers te controleren die geen rechtsgevolgen hebben of niet kenbaar zijn gemaakt aan burgers.

1. AANLEIDING: DISCRIMINERENDE ALGORITMEN

Overheden wereldwijd maken in toenemende mate gebruik van algoritmen bij de uitvoering van hun beleid. Deze briefing gaat over algoritmen die door overheden worden gebruikt om beslissingen te nemen om burgers te controleren in de context van bijvoorbeeld criminaliteitsbestrijding, migratie en sociale bescherming.² Ook in Nederland ondersteunen algoritmen zulke beslissingen, bijvoorbeeld wie door de Belastingdienst wordt geselecteerd voor onderzoek naar fraude met kinderopvangtoeslagen, welke personen door de politie wordt gestopt voor een controle, of welke visumaanvraag door ambtenaren extra wordt bestudeerd op misbruik. De praktijk laat zien dat het vrijwel onmogelijk is om algoritmische systemen te ontwikkelen die vrij zijn van vooringenomenheid of discriminatie.³ In het geval van de Belastingdienst heeft het gebruik van een discriminerend algoritme voor het zoeken naar potentiële fraudegevallen mede geleid tot het toeslagenschandaal.⁴ Het toeslagenschandaal heeft de impact van algoritmen zichtbaar gemaakt en het maatschappelijk-politieke debat hierover gevoed.

Ondanks de risico's heeft de regering nog altijd weinig grip op de inzet van algoritmen en schiet zij tekort in het beschermen van burgers tegen discriminatie en etnisch profileren door de overheid. Sinds het toeslagenschandaal heeft de regering diverse acties ondernomen met als doel om te waarborgen dat algoritmen met respect voor mensenrechten worden ontwikkeld en gebruikt (zie kader 1). Geen van deze instrumenten heeft echter een verplichtend of bindend karakter: ze zijn gebaseerd op de vrijwilligheid van overheidsinstellingen. Ook de nieuwe algoritmetoezichthouder, de Directie Coördinatie Algoritmes, stelt vast dat het ontbreekt aan "volwassen, structurele en bindende algoritmespecifieke instrumenten om alle risicovolle algoritmen adequaat te monitoren en algoritmerisico's te beheersen".⁵ De DCA constateert dat initiatieven om waarborgen te garanderen zich vaak nog in de pilotfase bevinden.

Het is positief dat de regering stappen in de goede richting heeft gezet, maar de praktijk laat zien dat de genomen maatregelen onvoldoende zijn om een nieuw toeslagenschandaal of vergelijkbare mensenrechtenschendingen te voorkomen. Sinds het toeslagenschandaal zijn er meerdere onthullingen geweest over overheidsinstellingen die discriminerende algoritmen gebruik(t)en voor beslissingen over controles van burgers (zie kader 2 voor voorbeelden). Dit onderstreept de noodzaak van effectieve en bindende maatregelen voor regulering van algoritmen in de publieke sector.

Deze briefing: focus en methodologie

In deze briefing analyseert Amnesty de door de regering genomen en toegezegde maatregelen die als doel hebben om te waarborgen dat algoritmen in lijn met de mensenrechten worden ingezet. Amnesty beoordeelt in hoeverre deze maatregelen daadwerkelijk bijdragen aan het voorkomen van discriminatie en etnisch profileren door overheden. De briefing neemt als uitgangspunt de aanbevelingen voor de regulering van algoritmegebruik uit Amnesty's rapport *Xenofobe machines* (2021) over het discriminerend algoritme dat door de Belastingdienst werd gebruikt. De analyse in deze briefing is gebaseerd op parlementaire en ministeriële documenten die in het afgelopen twee jaar, sinds de publicatie van het rapport *Xenofobe machines*, zijn gepubliceerd.

² Sociale bescherming omvat alle regelingen die tot doel hebben om de lasten van burgers te verlichten, waaronder sociale zekerheid en sociale voordelen.

³ Amnesty International, Digitally Divided: Technology, inequality and human rights (Index: POL 40/7108/2023), 2 October 2023, <https://www.amnesty.org/en/documents/pol40/7108/2023/en/>.

⁴ Amnesty International publiceerde in oktober 2021 het rapport *Xenofobe Machines* over het algoritme dat door de Belastingdienst werd gebruikt en de impact daarvan op de mensenrechten, zie <https://www.amnesty.nl/actueel/toeslagenaffaire-is-mensenrechtenschending-zegt-amnesty-international> en https://www.amnesty.nl/content/uploads/2021/10/NL_Xenophobe-Machines_Amnesty-International_EUR-35_4686_2021.pdf.

⁵ Directie Coördinatie Algoritmes, *Rapportage Algoritmerisico's Nederland*, juli 2023, p. 5, <https://www.autoriteitpersoonsgegevens.nl/actueel/primeur-eerste-rapportage-algoritmerisicos-nederland>.

KADER 1: VOORBEELDEN VAN MAATREGELEN EN INSTRUMENTEN VAN DE OVERHEID

- Oprichting van de Directe Coördinatie Algoritmes (DCA)⁶
- Lancering van het Algoritmeregister⁷
- Oprichting van de Nationaal Coördinator tegen Discriminatie en Racisme⁸
- Impact Assessment Mensenrechten en Algoritmes (IAMA)⁹
- Handreiking 'non-discriminatie by design'¹⁰
- Implementatiekader 'Verantwoorde inzet algoritmen'¹¹

KADER 2: VOORBEELDEN VAN DISCRIMINERENDE ALGORITMEN

- De Nederlandse politie experimenteerde in Roermond met algoritmen om voertuigen te selecteren voor controle, met als doel het voorkomen en opsporen van 'mobiel banditisme'. Uit onderzoek van Amnesty bleek dat dit experiment leidde tot geautomatiseerd etnisch profileren, omdat het algoritmisch model tot doel had om vast te stellen of de personen in de voertuigen van Oost-Europese afkomst zijn.¹² De politie zette het experiment in 2020 stop. In een reactie op het rapport ontkent de minister van Justitie en Veiligheid dat sprake is van etnisch profileren en wijst de aanbevelingen af.¹³
- Immigratiedienst IND ontwikkelde het Risicomodel Erkend Referenten met als doel fraude met verblijfsvergunning voor kennismigranten tegen te gaan. Bedrijven met bestuursleden van 'niet-westerse achtergrond' kregen een hogere risicoscore.¹⁴
- De politie gebruikte sinds 2015 het risicotaxatie-instrument RTI-Geweld om de kans te berekenen wie in de toekomst geweld gaat plegen. Tot 2017 kregen mensen met een Antilliaanse, Marokkaanse of Somalische achtergrond een hogere score dan iemand met een Nederlandse achtergrond.¹⁵ Na kritiek zette de politie het instrument in 2023 stop.¹⁶

⁶ De DCA is ondergebracht bij de Autoriteit Persoonsgegevens, <https://autoriteitpersoonsgegevens.nl/themas/algoritmes-ai/coordinatie-toezicht-algoritmes-ai/directie-coordinatie-algoritmes-dca>.

⁷ Het algoritmeregister van de Nederlandse overheid, <https://algoritmes.overheid.nl>.

⁸ Nationaal Coördinator tegen Discriminatie en Racisme, <https://www.bureauncdr.nl/>.

⁹ Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK), Impact Assessment Mensenrechten en Algoritmes, 31 juli 2021, <https://www.rijksoverheid.nl/documenten/rapporten/2021/02/25/impact-assessment-mensenrechten-en-algoritmes>.

¹⁰ Ministeries van BZK, Justitie & Veiligheid (J&V) en Economische Zaken en Klimaat (EZK), Handreiking non-discriminatie by design, 10 juni 2021, <https://www.rijksoverheid.nl/documenten/rapporten/2021/06/10/handreiking-non-discriminatie-by-design>

¹¹ Ministerie van BZK, Implementatiekader Verantwoorde inzet van algoritmen, 30 juni 2023, <https://www.rijksoverheid.nl/documenten/rapporten/2023/06/30/implementatiekader-verantwoorde-inzet-van-algoritmen>.

¹² Amnesty International, *We Sense Trouble: Automated Discrimination and Mass Surveillance in Predictive Policing*, september 2020, https://www.amnesty.nl/content/uploads/2020/09/Amnesty-International_We-Sense-Trouble_EUR-35_2971_2020.pdf.

¹³ Brief Regering, Kamerstukken II, 2020/2021, 26643/32761, nr. 724, https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2020Z24743&did=2020D51904. Zie ook de reactie van Amnesty op de regeringsbrief, 22 januari 2021, https://www.amnesty.nl/content/uploads/2021/01/20210122_Kamerbrief-en-notitie-We-Sense-Trouble.pdf?x62925.

¹⁴ Dit risicomodel is volgens het kabinet uiteindelijk niet ingezet. Zie NRC, "IND verzweg een dikke error met kennismigranten", 6 mei 2022, <https://www.nrc.nl/nieuws/2022/05/06/ind-verzweg-een-dikke-error-met-kennismigranten-a4123661>.

¹⁵ Zie Follow The Money, "Dubieus algoritme van de politie 'voorspelt' wie in de toekomst geweld zal plegen", 23 augustus 2023, <https://www.ftm.nl/artikelen/nederlandse-politie-gebruikt-minority-report-algoritme>.

¹⁶ Zie Follow The Money, "Politie stopt met gewraakt algoritme dat 'voorspelt' wie in de toekomst geweld gebruikt", 25 augustus 2023, <https://www.ftm.nl/artikelen/politie-stopt-met-voorspellend-algoritme-geweld>.

- Vijf Utrechtse gemeenten gebruikten tot medio 2022 de ‘fraudescorekaart’ om bijstandsuitkeringen te controleren. Mensen die bijvoorbeeld in een woonwagengemeente wonen of een bepaald beroep uitoefenen hadden een grotere kans om als ‘fraudeur’ te worden aangemerkt.¹⁷ In 2020 riep de VNG gemeenten al op de fraudescorekaart niet meer te gebruiken. De Autoriteit Persoonsgegevens (AP) doet onderzoek naar de vijf gemeenten.¹⁸
- De gemeente Rotterdam gebruikte tussen 2017 en 2021 een algoritme om bijstandsfraude op te sporen. Het algoritme kende risicoscores toe aan uitkeringsontvangers op basis van bijvoorbeeld hun taalvaardigheid, geslacht en woonwijk.¹⁹ Na een kritisch rapport van de Rotterdamse Rekenkamer werden gegevens die tot discriminatie zouden kunnen leiden verwijderd. In 2023 werden de inzet en verdere ontwikkeling van het algoritme gestaakt.²⁰
- Voor de beoordeling van aanvragen van Schengenvisa, gebruikt een door het ministerie van Buitenlandse Zaken ingezet bedrijf een profilerend algoritme. Mensen met specifieke nationaliteiten, bijvoorbeeld de Surinaamse, worden door het algoritme geselecteerd voor een intensievere beoordeling van hun aanvraag.²¹ De AP onderzoekt het algoritme.²²

Voorbeelden van algoritmegebruik waarbij de overheid heeft nagelaten om vooraf het risico op (indirecte) discriminatie te onderzoeken:

- De gemeente Nissewaard gebruikte tot medio 2021 een algoritme om bijstandsfraude op te sporen. Na een kritisch onderzoek door TNO stopte de gemeente met dit systeem.²³
- De Nederlandse politie rolde in 2017 het Criminaliteits Anticipatie Systeem (CAS) landelijk uit. CAS heeft tot doel te voorspellen op welke locaties een verhoogde kans is op criminaliteit en overlast. De Algemene Rekenkamer concludeerde in 2022 dat systematische afwijkingen (bias), en daarmee (indirecte) discriminatie, niet konden worden uitgesloten.²⁴

Sinds 2012 werkt de Dienst Uitvoering Onderwijs (DUO) met een algoritme om fraude met de studiebeurs op te sporen. In juni 2023 zette minister Dijkgraaf van Onderwijs het algoritme stop vanwege aanwijzingen voor mogelijke (indirecte) discriminatie. Er loopt een onafhankelijk onderzoek en ook de AP onderzoekt de gegevensverwerking.²⁵

¹⁷ Zie NOS, “Utrechtse gemeenten werkten met omstreden fraudescorekaart”, 25 juni 2022,

<https://nos.nl/artikel/2434066-utrechtse-gemeenten-werkten-met-omstreden-fraudescorekaart>.

¹⁸ Autoriteit Persoonsgegevens, “AP wil opheldering over fraude-algoritme gemeenten”, 16 mei 2023,

<https://autoriteitpersoonsgegevens.nl/actueel/ap-wil-opheldering-over-fraude-algoritme-gemeenten>.

¹⁹ Zie Argos, “Fraude-algoritme op de snijtafel”, 6 maart 2023,

<https://www.vpro.nl/argos/lees/onderwerpen/algoritme/2023/fraude-algoritme-op-de-snijtafel-rotterdam-verdenkt-vooral-jonge-moeders.html>.

²⁰ Zie Rijnmond, “Rotterdamse ambtenaren sleutelden aan algoritmen om bijstandsfraude op te sporen”, 24 oktober 2023, <https://www.rijnmond.nl/nieuws/1729938/rotterdamse-ambtenaren-sleutelden-aan-algoritme-om-bijstandsfraude-op-te-sporen>.

²¹ Zie NRC, “‘Pas op met deze aanvraag’, waarschuwt het algoritme dat discriminatie in de hand werkt. Het ministerie negeert kritiek”, 23 april 2023, <https://www.nrc.nl/nieuws/2023/04/23/beslisambtenarenblijven-profileren-met-risicoscores-a4162837>.

²² Zie NRC, “Autoriteit Persoonsgegevens roept minister Hoekstra op het matje om algoritme”, 1 mei 2023, <https://www.nrc.nl/nieuws/2023/05/01/minister-moet-uitleg-geven-over-algoritme-voor-visa-a4163510>.

²³ Zie Trouw, “Nissewaard stopt met omstreden methode om fraude met bijstand op te sporen”, 7 juli 2022, <https://www.trouw.nl/economie/nissewaard-stopt-met-omstreden-methode-om-fraude-met-bijstand-op-te-sporen-b985a30d/>.

²⁴ Algemene Rekenkamer, Algoritmes getoetst, 2022,

<https://www.rekenkamer.nl/publicaties/rapporten/2022/05/18/algoritmes-getoetst>.

²⁵ Zie Investico, “DUO negeerde waarschuwingen over discriminatie bij fraudeonderzoek”, 17 juli 2023,

<https://www.platform-investico.nl/artikel/duo-negeerde-waarschuwingen-over-discriminatie-bij-fraudeonderzoek/>.

2. AMNESTY'S RAPPORT *XENOFIBE MACHINES* (2021)

De grootste risico's voor mensenrechten doen zich voor wanneer overheden algoritmen gebruiken voor risicoprofilering met het doel om potentiële normovertreders te identificeren of op te sporen. Dit zijn algoritmen die worden ingezet om te selecteren welke burgers aan controle worden onderworpen, zonder dat er concrete en geïndividualiseerde signalen zijn over mogelijke regelovertrading. Algoritmen worden dus gebruikt voor geautomatiseerde risicoprofilering: op basis van data worden berekeningen en risicoanalyses gemaakt, om vervolgens te beoordelen of iemand een hoge of lage kans heeft dat die een regel of wet aan het overtreden is. Dergelijke algoritmen worden met name binnen de maatschappelijke terreinen van sociale bescherming, criminaliteitsbestrijding en migratie gebruikt. Hierdoor is de impact op mensen potentieel groot, omdat het gaat om ingrijpende beslissingen over burgers, niet zelden met rechtsgevolgen. Door de inzet van geautomatiseerde risicomodellen is het voor burgers steeds moeilijker om inzicht te krijgen in de manier waarop overheidsbeslissingen over hen tot stand zijn gekomen. Wanneer zulke systemen worden ingezet voor preventieve of proactieve controles, is er vaak geen sprake van een formeel besluit en blijft de inzet van algoritmen voor de burger vaak onzichtbaar. De manier waarop algoritmen worden gebruikt is bovendien voor de burger meestal moeilijk te doorgronden.

In oktober 2021 publiceerde Amnesty het rapport *Xenofobe machines*. Hoewel veel verschillende factoren hebben geleid tot het toeslagenschandaal, richt het rapport zich op de rol van het algoritme, het zogenoemde risicoclassificatiemodel, dat de Belastingdienst vanaf 2013 gebruikte om fraude met kinderopvangtoeslagen op te sporen. Het risicoclassificatiemodel was een geautomatiseerd besluitvormingssysteem²⁶ dat onder andere bestond uit een zelflerend algoritme.²⁷ Op basis van verschillende indicatoren bepaalde het systeem welke ouders en verzorgers een verhoogd risico op fraude zouden hebben. Eén van de indicatoren die werd gebruikt was iemands nationaliteit. Aanvragers met een niet-Nederlandse nationaliteit kregen een hogere risicoscore; ze hadden daarmee dus een grotere kans op controle van hun dossier en uiteindelijk om onterecht als fraudeur te worden bestempeld. Daarnaast kregen huishoudens met een laag inkomen een hogere risicoscore dan die met een hoog inkomen. De oorzaak hiervan is dat het zelflerende model zelfstandig berekende welke indicatoren een verhoogde kans op fraude gaven. Het algoritme leerde zichzelf dat er een grotere kans was dat huishoudens met een laag inkomen fraudeerden.

Amnesty heeft de inzet van het risicoclassificatiemodel door de Belastingdienst getoetst aan het internationale mensenrechtenkader en concludeerde in het rapport *Xenofobe machines* dat het algoritme van de Belastingdienst heeft geleid tot etnisch profileren. Daarnaast stelde Amnesty vast dat er sprake was van discriminatie op basis van sociale klasse en van intersectionele discriminatie, vanwege de selectie van aanvragers met lagere inkomens door het zelflerende systeem. Het gebruik van het black box-systeem²⁸ en het zelflerend algoritme belemmerde bovendien de transparantie en uitvoering van de verantwoordingsplicht en was daarom onverenigbaar met de beginselen van behoorlijk bestuur, legaliteit en de rechtsstaat. De Belastingdienst heeft ook nagelaten voorafgaand en tijdens het gebruik van het risicomodel de risico's voor mensenrechten te toetsen. Tot slot constateerde Amnesty dat het toezicht ernstig faalde.

Om de gesignaleerde tekortkomingen weg te nemen en het gebruik van algoritmen voor beslissingen over burgers in de publieke sector in lijn te brengen met mensenrechten, moet de regering een kader tot stand brengen dat a) mensenrechtenschendingen voorkomt bij het gebruik van algoritmische

²⁶ Zie noot 1.

²⁷ Zelflerende algoritmen geven algoritmische systemen het vermogen om zelfstandig te leren van ervaringen over de loop van tijd, en om veranderingen aan te brengen in de manier waarop de algoritmen werken zonder dat deze veranderingen expliciet door mensen worden geprogrammeerd

²⁸ Een algoritmisch systeem waarvan de gegevensinvoer en de berekende uitkomst kunnen worden bekeken, maar waarbij de interne werking van het systeem onbekend is voor de gebruiker.

besluitvormingssystemen; b) monitoring- en toezichtmechanismen instelt als waarborg; c) degenen die verantwoordelijk zijn voor schendingen ter verantwoording roept; en d) voorziet in effectieve rechtsmiddelen voor individuen en groepen wier rechten zijn geschonden. Daartoe geeft het rapport specifieke aanbevelingen, die zijn gericht aan de regering (zie kader 3).

Als reactie op het rapport en de daarin gedane aanbevelingen, kondigde het kabinet in 2022 een aantal maatregelen aan om herhaling van het toeslagenschandaal en het leed dat mensen is aangedaan te voorkomen.²⁹ Tegelijkertijd erkende het kabinet dat er veel werk te verzetten is op het terrein van de digitale overheid.³⁰ In het coalitieakkoord van kabinet Rutte-IV werd vastgelegd dat het kabinet wettelijk wil regelen dat algoritmen worden gecontroleerd op transparantie, discriminatie en willekeur.³¹ In de correspondentie richting de Tweede Kamer blijft het kabinet het belang van beleid en regelgeving rondom algoritmen benadrukken. Volgens het kabinet is een proactieve houding noodzakelijk om problemen met algoritmen te voorkomen.³² Het kabinet heeft sindsdien een aantal concrete instrumenten in het leven geroepen die als hulpmiddelen moeten dienen voor overheidsorganisaties die algoritmen willen inzetten.³³

Amnesty constateert echter dat ruim twee jaar na publicatie van het rapport *Xenofobe machines* de tot dusver genomen maatregelen ernstig tekortschieten en mensen nog altijd onvoldoende beschermd zijn tegen de risico's van geautomatiseerde risicoprofilering door de overheid. In deze briefing wordt nader toegelicht waarom de maatregelen onvoldoende zijn en wat er nodig is voor effectieve waarborgen. De aanbevelingen uit *Xenofobe machines* zijn als volgt geclusterd: a) algoritmegebruik in lijn met mensenrechten, b) transparantie, en c) toezicht.

KADER 3: AANBEVELINGEN IN *XENOFOBE MACHINES* VOOR REGULERING ALGORITMEGEBRUIK³⁴

1. Bescherm de mensenrechten bij het gebruik van algoritmische besluitvormingssystemen.
2. Gebruik geen black box-systemen voor risicovolle situaties in de publieke sector.
3. Gebruik geen zelflerende algoritmen voor impactvolle besluitvorming.
4. Stel een openbaar register in voor algoritmische besluitvormingssystemen, dat gedetailleerde informatie bevat over het gebruik van algoritmen in de publieke sector.
5. Biedt volledige transparantie over (semi-)geautomatiseerde besluitvormingssystemen.
6. Beoordeel het effect op de mensenrechten en betrek hierbij relevante rechthebbenden. Deze mensenrechtentoets moet bindend en verplicht zijn.
7. Zet alomvattende en onafhankelijke toezichtsmechanismen op met voldoende mandaat, bevoegdheden en capaciteit.
8. Verbied het gebruik van nationaliteit, etniciteit en andere afkomstgerelateerde gegevens als indicator in risicoprofilering gericht op het opsporen van potentiële regel- of wetsovertreders.
9. Waarborg een zinvolle verantwoordingsplicht, doeltreffende rechtsmiddelen en verhaalsmogelijkheden voor slachtoffers van mensenrechtenschendingen.

²⁹ Amnesty ontving een eerste reactie op 26 oktober 2021. Op 4 juli 2022 kwam de regering met een uitgebreidere reactie waarin per aanbeveling werd toegelicht welke acties en interventies er zijn ondernomen, zie Kamerstukken II, 2021/22, 33826/26643, nr. 45, https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2022D28856&did=2022D28856.

³⁰ Zie brief regering 4 juli 2022, p. 1.

³¹ Coalitieakkoord 'Omzien naar elkaar, vooruitkijken naar de toekomst', 10 januari 2022, bijlage bij Kamerstuk 35 788, nr. 77, p. 33, <https://open.overheid.nl/documenten/ronl-f3cb0d9c-878b-4608-9f6a-8a2f6e24a410/pdf>.

³² Brief van de ministers van BZK en voor de Rechtsbescherming en van de Staatssecretarissen van BZK en EZK, Kamerstukken II, 2020/21, 26643, nr. 765, p. 2, <https://zoek.officielebekendmakingen.nl/kst-26643-765.pdf>.

³³ Deze instrumenten zijn samengebracht in de online *Toolbox Ethisch verantwoord innoveren voor overheden*. Zie <https://www.digitaleoverheid.nl/dossiers/toolbox-ethisch-verantwoorde-innovatie/>.

³⁴ Amnesty International, *Xenofobe Machines*, 2021, p.40-41. Amnesty deed nog enkele andere aanbevelingen; deze briefing beperkt zich tot de aanbevelingen die zijn gericht op de regulering van algoritmen.

3. ALGORITMEGEBRUIK IN LIJN MET MENSENRECHTEN

Om te voorkomen dat het gebruik van algoritmen leidt tot discriminatie en etnisch profileren door de overheid, is het noodzakelijk om de inzet ervan te reguleren. Een essentieel onderdeel daarvan is dat er een voortdurend proces van risicobeoordeling plaatsvindt, waarbij de impact op mensenrechten wordt getoetst. Daarnaast moet een aantal risicovolle praktijken worden verboden, omdat deze in strijd zijn met de mensenrechten en beginselen van behoorlijk bestuur.

3.1 EEN VERPLICHTE EN BINDEnde MENSENRECHTENTOETS

De regering moet een verplichte en bindende mensenrechtentoets invoeren bij het gebruik van algoritmische besluitvormingssystemen³⁵ in de publieke sector, ook wanneer deze systemen worden gebruikt door wetshandavingsinstanties. De mensenrechtentoets moet worden uitgevoerd tijdens de ontwerp-, ontwikkelings-, gebruiks- en evaluatiefase van algoritmische besluitvormingssystemen. Het effect op alle relevante mensenrechten, met inbegrip van sociale en economische rechten, moet vakkundig worden beoordeeld. Bij het proces moeten relevante belanghebbenden worden betrokken, waaronder onafhankelijke mensenrechtendeskundigen, toezichthouders, mensen uit potentieel getroffen, gemarginaliseerde en/of achtergestelde gemeenschappen, en technische deskundigen.

In opdracht van het Ministerie van BZK is in 2021 het Impact Assessment Mensenrechten en Algoritmes (IAMA) ontwikkeld.³⁶ Dit kan een geschikt en effectief middel zijn voor overheden om een risicobeoordeling van de inzet van algoritmen uit te voeren. In maart 2022 heeft de Tweede Kamer een motie aangenomen die het gebruik van het IAMA verplicht stelt voorafgaand aan het gebruik van algoritmen die worden ingezet om evaluaties van of beslissingen over mensen te maken.³⁷ Daarnaast wordt in de motie verzocht om het ingevulde assessment openbaar te maken. De motie is echter nog niet door het kabinet uitgevoerd en vooralsnog ontbreekt de specifieke invulling van de verplichtstelling van het IAMA. In juni 2023 heeft de Kamer een motie aangenomen waarin wordt verzocht vanaf dan alle nieuwe hoogrisicoalgoritmen die binnen de overheid worden gebruikt verplicht te onderwerpen aan het IAMA.³⁸

Hoewel het IAMA momenteel door een aantal overheidsorganisaties wordt getest, lijkt het kabinet de verplichtstelling te willen uitstellen tot de inwerkingtreding van de voorgestelde Europese AI-Verordening.³⁹ De AI-verordening zal waarschijnlijk (voor hoog-risico AI-systemen) een zogenoemde conformiteitsbeoordeling bevatten.⁴⁰ Het is onwenselijk om de AI-Verordening af te wachten, omdat deze op verschillende plekken een uitzondering zou kunnen toelaten voor relevante terreinen zoals wetshandhaving en grensbewaking.⁴¹ Als de verplichtstelling van het IAMA ook onderworpen wordt aan deze uitzondering, dan ontstaat een zeer zorgwekkende situatie, waarin de meest risicovolle algoritmen met de meest vergaande gevolgen niet hoeven te worden beoordeeld op hun effect op de mensenrechten en niet openbaar hoeven te worden gemaakt. Het is daarom van groot belang dat het kabinet de komst van de AI-Verordening niet afwacht, maar tegemoetkomt aan de eis van de Tweede

³⁵ Zie noot 1.

³⁶ Zie Impact Assessment Mensenrechten en Algoritmes.

³⁷ Motie Bouchallikh en Dekker-Abdulaziz, 29 maart 2022, Kamerstukken II 2021-2022, 26 643, nr. 835, <https://www.tweedekamer.nl/kamerstukken/moties/detail?id=2022Z06024&did=2022D12329>.

³⁸ Motie Dekker-Abdulaziz, 13 juni 2023, Kamerstukken II 2022/23, 36360 VI, nr. 14.

³⁹ Zie brief van de minister van BZK, Kamerstukken II, 26643, nr. 924, p. 9, https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2022D40581&did=2022D40581.

⁴⁰ Zie art 19 van de Verordening op artificiële intelligentie (AI Act), versie van de Europese Commissie, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>.

⁴¹ Zie hierover het statement van o.a. Amnesty International, https://www.amnesty.eu/wp-content/uploads/2022/12/Open-letter_EU-AI-Act_migration_December-2022.pdf.

Kamer om de IAMA te verplichten voor de inzet van alle algoritmen, zoals verzocht in de Tweede Kamermotie.

Een andere tekortkoming in de implementatie van de mensenrechtentoets is dat het momenteel onvoldoende duidelijk is wie deze toets moet uitvoeren. Het kabinet moet daarom vastleggen dat de mensenrechtentoets zowel door de ontwikkelaar als gebruiker van het systeem wordt uitgevoerd. De ontwikkelaar heeft immers niet altijd goed inzicht in hoe het systeem in de praktijk gebruikt zal worden en welke onvoorziene effecten zich kunnen voordoen. Het is daarom belangrijk dat, naast de ontwikkelaar, ook de gebruiker het IAMA uitvoert.

3.2 VERBOD OP AFKOMST ALS CRITERIUM IN RISICOPROFIELEN

Het discriminatieverbod is vastgelegd in artikel 1 van de Grondwet en in internationale mensenrechtenverdragen. Het gebruik van etniciteit, nationaliteit en andere afkomstgerelateerde gegevens als criterium in risicoprofielen die worden gebruikt voor het identificeren of opsporen van potentiële normovertreders komt neer op etnisch profileren.⁴² Etnisch profileren is in strijd met het discriminatieverbod zoals vastgelegd in de Grondwet en in internationale mensenrechtenverdragen. Afkomst mag geen enkele rol spelen in beslissingen om burgers te selecteren voor controle en mag dus geen criterium zijn in selectiebeslissingen en (geautomatiseerde) risicoprofielen. Dit is bevestigd door het Gerechtshof Den Haag in de rechtszaak van Amnesty International en anderen tegen de Staat over etnisch profileren door de Koninklijke Marechaussee (zie kader 4).⁴³

Toch laat het kabinet ruimte voor het gebruik van afkomstgerelateerde gegevens als criterium in (geautomatiseerde) risicoprofielen. De redenering van het kabinet is dat er in een concreet geval zwaarwegende redenen kunnen bestaan waardoor het gebruik van etniciteit of nationaliteit als criterium in een risicoprofiel noodzakelijk zou kunnen zijn.⁴⁴ Het discriminatieverbod zou volgens het kabinet ruimte laten voor een verschil in behandeling als dit slechts ‘in enige mate’ is gebaseerd op afkomst, wanneer er naast afkomst meerdere en/of meer zwaarwegende indicatoren worden gebruikt.⁴⁵ Dit is een onjuiste uitleg van het discriminatieverbod en niet in lijn met de uitspraak van het Gerechtshof over de Koninklijke Marechaussee (KMar). Het kabinet verwijst hierbij naar het toetsingskader ‘Discriminatie door risicoprofielen’ dat het College van de Rechten van de Mens in 2021 heeft gepubliceerd.⁴⁶ Dit toetsingskader laat onterecht ruimte voor etnisch profileren en is bovendien niet aangepast na de uitspraak van het hof in de KMar-rechtszaak. Het toetsingskader is vanaf het najaar van 2022 onder de aandacht gebracht bij verschillende overheidsinstellingen en de politie.⁴⁷

In 2022 is op verzoek van de Tweede Kamer de Staatscommissie tegen Discriminatie en Racisme. Deze commissie doet vier jaar onderzoek en zal een advies uitbrengen over onder meer discriminatie en racisme door de overheid. De commissie zal in elk geval ook de wenselijkheid onderzoeken van

⁴² Zie Amnesty's standpunt over discriminerende risicoprofielen, <https://www.amnesty.nl/content/uploads/2023/01/infosheet-discriminerende-risicoprofielen-jan23.pdf?x76291>. Meer gedetailleerde uitleg: T. Abdul-Aliyeva & G. van Eijk, *Nederlands Juristenblad*, 2023, “Discriminerende risicoprofielen”, https://www.amnesty.nl/content/uploads/2023/01/Amnesty-NL-2023-NJB04_Discriminerende-risicoprofielen.pdf?x54998.

⁴³ Gerechtshof Den Haag, 14 februari 2023, ECLI:NL:GHDHA:2023:173.

⁴⁴ Brief van de minister van BZK, Kamerstukken II 2021/22. 30 950, nr. 281, <https://zoek.officielebekendmakingen.nl/kst-30950-281.html> en brief van de minister van BZK en de staatssecretaris van J&V, Kamerstukken II 2022/23. 30 950, nr. 349, https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2023Z08553&did=2023D20409.

⁴⁵ Kamerstukken II 2021/22, 30 950, nr. 281, <https://zoek.officielebekendmakingen.nl/kst-30950-281.html>.

⁴⁶ Zie College voor de Rechten van de Mens, *Discriminatie door risicoprofielen – Een mensenrechtelijk toetsingskader*, 2021, <https://publicaties.mensenrechten.nl/publicatie/61a734e65d726f72c45f9dce>.

⁴⁷ Brief van de minister van BZK, Kamerstukken II, 2022/23, 30950, nr. 319, https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2022D43313&did=2022D43313.

een verbod op het gebruik van etniciteit in fraudebestrijding door overheidsinstanties. Ook onderzoekt de commissie of het mogelijk is dat onderscheid op grond van etniciteit of nationaliteit in risicoprofielen alleen mag worden toegepast ter bescherming dan wel ondersteuning van mensen, en dus in ieder geval niet ter bestrijding van fraude.⁴⁸

Uit de verschillende onthullingen over discriminerende risicomodellen in de afgelopen jaren blijkt dat het voor overheden en uitvoeringsorganisaties niet duidelijk is dat het gebruik van afkomstgerelateerde gegevens in risicoprofielen altijd neerkomt op etnisch profileren en dus in strijd is met het discriminatieverbod. Doordat het kabinet de deur openlaat voor etnisch profileren, zijn burgers van kleur en met een migratieachtergrond niet beschermd tegen discriminatie door de overheid. Amnesty roept het kabinet op om het onderzoek van de Staatscommissie niet af te wachten, maar nu een duidelijke norm te stellen en vast te leggen dat etniciteit, nationaliteit en andere afkomstgerelateerde gegevens nooit gebruikt mogen worden als criteria in (semi-geautomatiseerde) risicoprofielen die worden gebruikt voor het zoeken naar potentiële normovertreders.⁴⁹

KADER 4: RECHTSAAK AMNESTY C.S. OVER ETNISCH PROFILEREN BIJ GRENSCONTROLES

Samen met een coalitie van twee ‘niet-witte’ burgers, Controle Alt Delete en RADAR en ondersteund door advocaten van PILP en Houthoff voerde Amnesty sinds 2020 een rechtszaak tegen de Nederlandse staat over etnisch profileren door de Koninklijke Marechaussee (KMar). De rechtszaak draaide om de vraag of de KMar bij grenscontroles mensen mede vanwege hun (vermeende) etniciteit mag controleren. Bij deze grenscontroles is er geen sprake van een concrete en individuele verdenking tegen de mensen die gecontroleerd worden. De KMar werkt hierbij met risicoprofielen zoals de “Nigeriaanse geldsmokkelaar”: snellopend, goed gekleed met een “niet-Nederlands uiterlijk.” Volgens de coalitie leidt dit gebruik van (vermeende) etniciteit in selectiebeslissingen en risicoprofielen tot discriminatie. Volgens het kabinet en de Koninklijke Marechaussee mag het wel.

Op 14 februari 2023 heeft het gerechtshof Amnesty c.s. in het gelijk gesteld. Het hof oordeelde dat selectie mede op grond van (vermeende) etniciteit een vorm van rassendiscriminatie is. Volgens het hof mag etniciteit op geen enkele manier, ook niet een klein beetje, een rol spelen in beslissingen om mensen controleren. Het maakt daarbij niet uit dat naast etniciteit ook andere factoren een rol spelen in de selectiebeslissing. Door afkomst als criterium in een selectiebeslissing op te nemen, worden immers twee verder vergelijkbare burgers verschillend behandeld vanwege hun afkomst. Dit is directe discriminatie. Het hof bepaalde dat de KMar haar werkwijze per direct moet aanpassen. De Nederlandse Staat is niet in cassatie gegaan tegen de uitspraak en daarmee is de uitspraak definitief. Deze uitspraak is een belangrijke stap in het uitbannen van discriminatie en etnisch profileren door overheden als gevolg van (geautomatiseerde) risicoprofilering.⁵⁰

⁴⁸ De Staatscommissie tegen Discriminatie en Racisme is een adviescollege dat voor vier jaar is ingesteld op verzoek van de Tweede Kamer. De Staatscommissie doet wetenschappelijk onderzoek en maakt onderdeel uit van een pakket aan initiatieven om discriminatie en racisme in Nederland aan te pakken. Staatscommissie tegen Discriminatie en Racisme, *Werkprogramma op hoofdlijnen*, zie

<https://www.staatscommissietegendiscriminatieenracisme.nl/publicaties/publicaties/2023/02/02/werkprogramma>; zie ook Motie Mutluer c.s., Kamerstukken II, 2021/22, 30 950, nr. 293 en Motie Belhaj c.s., Kamerstukken II, 2020/21, 30 950, nr. 237.

⁴⁹ Zie verder: <https://www.amnesty.nl/wat-we-doen/tech-en-mensenrechten/discriminerende-risicoprofielen>.

⁵⁰ Meer over de rechtszaak: <https://www.amnesty.nl/wat-we-doen/landen/mensenrechten-nederland/rechtszaak-koninklijke-marechaussee>.

KADER 5: GEEN ZICHT OP INDIRECTE DISCRIMINATIE DOOR ALGORITMEN

Wanneer een ongelijke behandeling niet direct is gebaseerd op een verdachte of verboden grond, zoals etniciteit of nationaliteit, maar die wel het effect heeft dat leden van een groep die worden gekenmerkt door een verdachte grond in het bijzonder benadeeld worden, is sprake van indirect onderscheid. Wanneer hiervoor geen objectieve en redelijke rechtvaardiging bestaat, leidt het onderscheid tot verboden indirecte discriminatie. Er is sprake van een objectieve en redelijke rechtvaardiging wanneer het onderscheid een gerechtvaardigd doel dient en er sprake is van proportionaliteit tussen het gehanteerde middel (de ongelijke behandeling) en het nagestreefde doel.⁵¹ In het vereiste van proportionaliteit ligt besloten dat de maatregel geschikt en noodzakelijk is om het nagestreefde doel te bereiken.

Momenteel is er geen overzicht van algoritmen die indirect onderscheid maken op basis van afkomst. De discussie in de politiek spitst zich toe op de vraag of het gebruikte systeem expliciet besluiten neemt op basis van discriminerende gronden zoals etniciteit of nationaliteit.⁵² Het is daarom ook niet eenvoudig om vast te stellen hoe wijdverbreid indirecte discriminatie binnen het overheidshandelen is. Het is van groot belang dat ook deze vorm van discriminatie door de overheid serieus wordt onderzocht en aangepakt. Een van de manieren waarop indirecte discriminatie voorkomen kan worden, is bijvoorbeeld het verplicht stellen van de mensenrechtentoets voorafgaand aan het ontwerp en gebruik van een algoritme. Zo kunnen ook onbedoelde discriminerende effecten van een algoritme aan het licht komen.

3.3 VERBOD OP ZELFLERENDE ALGORITMEN IN DE UITVOERENDE FASE

De overheid maakt gebruik van algoritmen in de vorm van eenvoudige beslisbomen met criteria, maar ook ingewikkelde toepassingen met een zelflerend vermogen (machine learning). Het risicoclassificatiemodel dat de Belastingdienst gebruikte en het Rotterdamse bijstandsfraudealgoritme (zie kader 2) zijn voorbeelden van een zelflerend algoritme. Het gebruik van zelflerende algoritmen in de publieke sector brengt uitdagingen en risico's met zich mee. Zelflerende algoritmen kunnen zichzelf aanpassen aan de context waarbinnen ze worden ingezet. Hoe het algoritme zich ontwikkelt is vooraf niet bekend of te voorspellen. Dit kan ertoe leiden dat de geautomatiseerde besluitvorming in eerste instantie niet discriminerend is, maar na verloop van tijd dat wel kan worden, omdat het algoritme zichzelf aanleert om in te zoomen op bijvoorbeeld mensen met een migratieachtergrond. Discriminerende effecten kunnen bij het gebruik van zelflerende algoritmen daarom moeilijk te ontdekken zijn.

Daarnaast geldt dat overheidshandelen bij het gebruik van zelflerende algoritmen niet goed controleerbaar, uitlegbaar en voorspelbaar is.⁵³ Een ambtenaar kan immers niet goed uitleggen hoe

⁵¹ Gerechtshof Den Haag, 14 februari 2023, ECLI:NL:GHDHA:2023:173, r.o. 8.4.

⁵² Zo ziet 'Motie 21' toe op het inventariseren en opruimen van het gebruik van etniciteit, nationaliteit en geboorteplaats in risicomodellen, dat leidt tot direct onderscheid op grond van afkomst, maar wordt mogelijk indirect onderscheid op grond van afkomst buiten beschouwing gelaten (motie Marijnissen c.s. Kamerstukken II, 2020/21, 35510, nr. 21 en motie Klaver c.s. Kamerstukken II, 2020/21, 35510, nr. 16).

⁵³ Het beginsel van *equality of arms* vereist dat (deels) geautomatiseerde besluitvorming inzichtelijk en controleerbaar moet zijn om de burger in staat te stellen weloverwogen rechtsmiddelen te kunnen aanwenden om zichzelf te verdedigen tegen besluiten. Het zorgvuldigheidsbeginsel en het motiveringsbeginsel en daaraan gerelateerde eisen van transparantie en uitlegbaarheid verlangen dat een besluit dat genomen is op basis van een geautomatiseerd systeem met toepassing van algoritmen transparant, verifieerbaar en toetsbaar moet zijn, zodat voor de burger duidelijk is welke gegevens zijn gebruikt en wat de relatie is tussen die gegevens en het besluit, om te voorkomen dat het besluit een 'black box' wordt. Zie Raad van State, *Digitalisering. Wetgeving en bestuursrechtspraak*, mei 2021, https://www.raadvanstate.nl/publish/library/13/digitalisering_wetgeving_en_bestuursrechtspraak.pdf. Daarnaast vereist de Algemene Verordening Gegevensverwerking ook dat (overheids)organisaties die geautomatiseerd besluiten nemen in bepaalde gevallen voldoen aan concrete eisen, zoals uitlegbaarheid en transparantie.

het algoritme tot een bepaalde output is gekomen, omdat de interne werking van het algoritme voor de ambtenaar niet inzichtelijk is. Wanneer die output invloed heeft op een uiteindelijk besluit, dan is dat besluit dus niet volledig uit te leggen aan burgers. Dergelijk algoritmegebruik is onverenigbaar met de beginselen van behoorlijk bestuur. Ook het College voor de Rechten van de Mens legt in zijn recente *position paper* over transparantie rondom algoritmen uit dat een overheid het motiveringsbeginsel uit de Algemene wet bestuursrecht schendt, wanneer die gebruik maakt van algoritmen zonder hiervan melding te maken en uit te leggen.⁵⁴ Daarom moet de overheid in de gebruiksfase geen zelflerende algoritmen inzetten als het gaat om besluitvorming met rechtsgevolgen of besluitvorming die anderszins van invloed is op de rechten en vrijheden van burgers of de samenleving als geheel.

Momenteel zijn er geen concrete beperkingen met betrekking tot het gebruik van zelflerende algoritmen. Met verwijzing naar algemene juridische kaders stelt het kabinet dat de nodige waarborgen voor een verantwoorde inzet van algoritmen aanwezig zijn.⁵⁵ Het kabinet laat echter na concreet aan te geven welke juridische kaders er zijn om de negatieve effecten van zelflerende algoritmen weg te nemen. Elk bestuursbesluit moet deugdelijk worden gemotiveerd op grond van algemene bestuursrechtelijke normen. Dit is niet goed mogelijk wanneer ondoorzichtige zelflerende systemen worden gebruikt voor besluitvorming. Het kabinet moet daarom een duidelijke norm stellen door de inzet van zelflerende algoritmen bij impactvolle besluitvorming in de gebruiksfase te verbieden.

KADER 6: HET ONDERSCHIED TUSSEN DE ONTWIKKELFASE EN DE GEBRUIKSFASE VAN ALGORITMEN

Er kan globaal gezien onderscheid worden gemaakt tussen twee fasen van de levenscyclus van algoritmen: de ontwikkelfase en gebruiksfase. In de ontwikkelfase wordt het algoritme ontworpen, getraind en getest. Het algoritme krijgt grote hoeveelheden data gevoed, op basis waarvan het wordt getraind om bijvoorbeeld verbanden te herkennen tussen de verschillende gegevens om risico-inschattingen te doen. In deze fase wordt het algoritme nog niet 'live' of 'realtime' ingezet voor het doen van risico-inschattingen. Dit gebeurt wel in de gebruiksfase, waarin het algoritme in gebruik wordt genomen om besluiten te nemen. De besluiten die in de gebruiksfase door het algoritme worden genomen kunnen grote impact hebben op mensen.

3.4 VERBOD OP BLACK BOX-ALGORITMEN

Wanneer algoritmen worden ingezet is er een groot risico dat de modellen niet begrijpelijk of transparant zijn. Het kabinet erkent dat algoritmen die de overheid gebruikt niet altijd transparant zijn, waardoor burgers de digitale wereld ervaren als een 'black box'.⁵⁶ Er is sprake van een black box wanneer de interne werking van het model voor de burger of de ambtenaar niet bekend is, waardoor je als het ware in een zwarte doos kijkt en daarin niks kunt zien. Er kunnen verschillende oorzaken aan een black box ten grondslag liggen. Ten eerste kan een model technisch een black box zijn, omdat het te complex is voor mensen om te begrijpen. Daarnaast kan een model een black box zijn omdat het een beleidskeuze is om niet transparant te zijn en de werking van het model niet openbaar te maken. Daarnaast brengen black box-systemen extra risico's met zich mee als uitvoerende ambtenaren niet weten of begrijpen dat ze met een algoritme werken die hun besluiten stuurt. Black

⁵⁴ College voor de Rechten van de Mens, *In alle openheid: transparant algoritmegebruik door de overheid*, p. 6, <https://publicaties.mensenrechten.nl/file/c96620d1-07d7-3b84-7bcc-da90a845fca2.pdf>.

⁵⁵ Brief van de minister van BZK, Kamerstukken II, 30950, nr. 319, p. 5, https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2022D43313&did=2022D43313.

⁵⁶ Rijksoverheid, *Werkagenda Waardengedreven Digitalisering*, 2022, p. 37, <https://open.overheid.nl/documenten/ronl-0de79e5c4c0c9b203c0a1c263efca7eca410958b/pdf>.

box-algoritmen zijn niet verenigbaar met het principe dat overheidshandelen controleerbaar, uitlegbaar en voorspelbaar moet zijn. Beide vormen van black box-systemen mogen door de overheid dan ook niet worden gebruikt om impactvolle beslissingen over mensen te maken, bijvoorbeeld of zij al dan niet nader onderzocht moeten worden op fraude met sociale regelingen. Burgers hebben immers het recht op uitlegbare en transparante besluitvorming.

4. TRANSPARANTIE: EEN VERPLICHT EN BINDEND ALGORITMEREGER

Overheden moeten transparant zijn over het gebruik van algoritmen voor beslissingen over burgers, door zinnvolle informatie te verstrekken over de onderliggende logica, het belang en de verwachte gevolgen van beslissingen, zelfs indien de besluitvorming niet volledig geautomatiseerd is, en ongeacht de mate van menselijke betrokkenheid bij het besluitvormingsproces.

Het kabinet onderkent dat transparantie over algoritmen en hun toepassing een belangrijke voorwaarde is voor het tegengaan van discriminatie en willekeur.⁵⁷ Tegelijkertijd erkent het kabinet dat de transparantie over de inzet van algoritmen door de overheid nog vaak tekortschiet.⁵⁸ Daarom heeft de Tweede Kamer moties aangenomen voor het oprichten van een algoritmeregister voor overheden⁵⁹, om het gebruik van dit register verplicht te stellen⁶⁰ en specifiek om vanaf juni 2023 alle nieuwe hoogrisicoalgoritmen die binnen de overheid worden gebruikt verplicht op te nemen.⁶¹ Burgers moeten immers weten welke algoritmen de overheid gebruikt en de overheid desgewenst kritisch bevragen wanneer zij zich niet aan de regels houdt. Een algoritmeregister is een instrument dat hieraan een bijdrage kan leveren en beter inzichtelijk en uitlegbaar kan maken hoe algoritmen worden toegepast en welke impact ze hebben op de levens van mensen. Ter uitvoering van deze moties heeft het kabinet op 21 december 2022 het centrale algoritmeregister gelanceerd waarin ruim honderd algoritmen die de overheid gebruikt zijn opgenomen.⁶² Begin november 2023 bevat het register ruim tweehonderd algoritmen.

Amnesty constateert echter drie tekortkomingen van dit register: het algoritmeregister is niet verplicht, de informatie is niet compleet en ontoereikend, en er bestaan uitzonderingen voor opsporing en rechtshandhaving.

4.1 HET ALGORITMEREGER MOET VERPLICHT WORDEN

Alhoewel dit wel de eis van de Tweede Kamer was bij het aannemen van de motie over verplichting van het algoritmeregister, is het registreren van algoritmen in dit register niet verplicht. De verplichtstelling kan volgens het kabinet op zijn vroegst pas in 2024 ingaan, omdat in de periode daarvoor wordt onderzocht welke beleidsmatige en juridische keuzes gemaakt moeten worden om de reikwijdte van het register te bepalen en een definitie van algoritmen die daarin moeten worden

⁵⁷ Brief van de Staatssecretaris van BZK, Kamerstukken II 2022/23, 26643, nr. 950, p. 1, https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2022D55799&did=2022D55799.

⁵⁸ Brief van de Staatssecretaris van BZK, Kamerstukken II 2022/23, 26643, nr. 950, p. 1.

⁵⁹ Motie Klaver c.s., 19 januari 2021, Kamerstukken II 2020/21, 35510, nr. 16.

⁶⁰ Motie van Dassen c.s., 28 oktober 2021, Kamerstukken II 2021/22, 35925 VII, nr. 26.

⁶¹ Motie Dekker-Abdulaziz, 13 juni 2023, Kamerstukken II 2022/23, 36360 VI, nr. 14.

⁶² Zie <https://algoritmes.overheid.nl>

opgenomen op te stellen.⁶³ In de *Werkagenda Waardengedreven Digitalisering* van de Rijksoverheid staat dat de verplichtstelling pas “voor einde 2025” zal zijn gerealiseerd.⁶⁴

Het kabinet verwijst hierbij opnieuw naar de voorgestelde Europese AI-verordening en de keuzes die daarin worden gemaakt ten aanzien van een verplicht algoritmeregister. De AI-verordening is echter nog niet definitief vastgesteld en het is dan ook niet duidelijk wanneer de verordening in werking zal treden. Daarnaast is het niet duidelijk wat de reikwijdte van de registratieplicht in de AI-Verordening zal zijn. Het is daarom van cruciaal belang dat het kabinet, in afwachting van de ontwikkelingen op Europees niveau, overheidsinstellingen verplicht om in ieder geval algoritmen met grote impact voor mensenlevens, bijvoorbeeld in de opsporing van fraude in de sociale zekerheid of criminaliteit door politie, in het register op te nemen.⁶⁵

Er is geen enkele goede reden om de implementatie van nationale waarborgen uit te stellen, des te meer omdat het de bedoeling is dat informatie over algoritmen uit de decentrale registers van overheidsorganisaties wordt opgehaald en in het centrale register wordt opgenomen.⁶⁶ Het probleem is echter dat veruit de meeste overheidsorganisaties momenteel geen decentrale algoritmeregisters hebben, waardoor er dus ook geen informatie over de gebruikte algoritmen beschikbaar is. Het voornemen van het kabinet om de totstandkoming van decentrale registers te “stimuleren” zal niets opleveren zolang het niet verplicht is.

4.2 HET ALGORITMEREGEREESTER MOET COMPLEET ZIJN

De informatie die in het algoritmeregister is opgenomen, is niet compleet en ontoereikend voor het doel, namelijk om de werking en impact van algoritmen transparant te maken. Zo zijn niet alle algoritmen geregistreerd en ontbreekt nog veel informatie over de algoritmen die wel zijn geregistreerd. Het kabinet benadrukt weliswaar dat dit register een eerste versie is en dat verdere doorontwikkeling nodig is⁶⁷, maar dit ontslaat de regering niet van de verplichting om van overheidsinstellingen te eisen dat zij het register volledig inrichten. Op basis van het algoritmeregister is het nog steeds niet mogelijk om van buitenaf te controleren of een algoritme dat door de overheid wordt gebruikt discriminerend is of kan zijn.

In het rapport *Xenofobe machines* heeft Amnesty een lijst opgenomen met de informatie die minimaal in het register moet worden opgenomen.⁶⁸ Hier wordt momenteel slechts ten dele aan voldaan. Voor effectieve mensenrechtenwaarborgen is het nodig dat bovenop de huidige informatie, de volgende elementen in het algoritmeregister worden opgenomen:

- Aannames, uitgangspunten en ontwerpbeginselen waarop het algoritme is gebaseerd;
- Nauwkeurigheid van het systeem;
- Datasets die gebruikt worden om het model te trainen en het doel van de oorspronkelijke verzameling van deze data;
- Wie het algoritme evalueert en hoe vaak en wat het resultaat is van de periodieke evaluatie;
- Juridische en technische waarborgen die ingebouwd zijn om risico's te beperken.

⁶³ Brief van de Staatssecretaris van BZK, Kamerstukken II 2022/23, 26643, nr. 950, p. 3.

⁶⁴ Zie Rijksoverheid, *Werkagenda Waardengedreven Digitalisering*, p. 38, <https://open.overheid.nl/documenten/ronl-0de79e5c4c0c9b203c0a1c263efca7eca410958b/pdf>.

⁶⁵ Het kabinet stelt dat het in de aanloop naar verplichtstelling van het algoritmeregister overheidsorganisaties zal stimuleren om waar mogelijk hun (hoog-risico)algoritmen te publiceren, zie brief van de Staatssecretaris van BZK, Kamerstukken II 2022/23, 26643, nr. 950, p. 3.

⁶⁶ brief van de Staatssecretaris van BZK, Kamerstukken II 2022/23, 26643, nr. 950, p. 3.

⁶⁷ brief van de Staatssecretaris van BZK, Kamerstukken II 2022/23, 26643, nr. 950, p. 1-2.

⁶⁸ Amnesty International, *Xenofobe Machines*, 2021, p. 31-33.

4.3 REGISTRATIE OOK VOOR RECHTSHANDHAVING EN NATIONALE VEILIGHEID

Het kabinet is van plan een uitzondering op de registratieplicht te introduceren voor opsporing, rechtshandhaving, defensie en inlichtingenverzameling.⁶⁹ Dit is een zorgwekkend voornemen, omdat juist het gebruik van algoritmen in de context van opsporing, rechtshandhaving en nationale veiligheid vergaande gevolgen kunnen hebben voor mensenlevens. Het gebruik van algoritmen in deze contexten vereist dus niet minder waarborgen, maar juist meer, zoals transparantie en toezicht. Wanneer risicovolle algoritmen worden gebruikt in de context van opsporing, rechtshandhaving en nationale veiligheid, kan dit vergaande gevolgen hebben voor mensenrechten, zoals het recht op vrijheid, het recht op een eerlijk proces, de onschuldpresumptie en het recht op gelijke behandeling. Over het gebruik van algoritmen voor opsporing en de gevolgen hiervan voor verschillende mensenrechten heeft Amnesty in september 2020 het rapport *We Sense Trouble* gepubliceerd.⁷⁰ De potentiële omvang en ernst van de impact van geautomatiseerde risicoprofieling in de context van opsporing, rechtshandhaving en nationale veiligheid vereisen een hoge norm van bescherming. Het kabinet moet daarom afzien van de voorgenomen uitzonderingsposities.

5. TOEZICHT: EEN ALGORITMEWAAKHOND MET MANDAAT EN MIDDELEN

De regering dient alomvattende en onafhankelijke mechanismen op te zetten voor mensenrechtentoezicht op het gebruik van algoritmen voor beslissingen over burgers, waarbij toezicht wordt gehouden op de naleving van zowel burger- en politieke rechten als sociale en economische rechten, zodat de verantwoordingsmechanismen worden versterkt en de mensenrechten beter beschermd. Het toezichtsorgaan dient het mandaat, de bevoegdheden en de capaciteit krijgen om toe te zien op de bescherming van de mensenrechten. Zij dient richtsnoeren uit te kunnen vaardigen, en ontwerpers, ontwikkelaars en gebruikers ter verantwoording te kunnen roepen in bindende besluiten.

Amnesty pleit al geruime tijd voor een aparte algoritmetoezichthouder met de juiste bevoegdheden en voldoende capaciteit om erop toe te zien dat mensenrechten, zowel burger- en politieke rechten als sociaaleconomische rechten, bij de inzet van algoritmen worden nageleefd.⁷¹ Het kabinet kondigde in het coalitieakkoord Rutte-IV aan dat een algoritmetoezichthouder zou worden ingericht.⁷² Begin 2023 loste het kabinet deze belofte in en werd de oprichting van de algoritmetoezichthouder officieel gemaakt.⁷³ De oprichting van de algoritmetoezichthouder is onderdeel van de bredere opgave om

⁶⁹ Brief van de Staatssecretaris van BZK, Kamerstukken II 2022/23, 26643, nr. 950, p. 4. Zie ook *Werkagenda Waardengedreven Digitalisering*, p. 38. In het kader van de EU AI-Verordening wordt ook gesproken over uitzonderingen voor nationale veiligheid, rechtshandhaving en migratie. Zie Amnesty International, EU AI Act Trilogues: Status of Fundamental Rights Recommendations, 16 november 2023, <https://www.amnesty.eu/news/eu-ai-act-trilogues-status-of-fundamental-rights-recommendations/>.

⁷⁰ Amnesty International, *We Sense Trouble*, 2020, https://www.amnesty.nl/content/uploads/2020/09/Amnesty-International_We-Sense-Trouble_EUR-35_2971_2020.pdf

⁷¹ Zie Amnesty, *Xenofobe Machines*, p. 33-35. Zie ook NRC, "Nederland heeft een algoritme waakhond nodig", 14 oktober 2020, <https://www.nrc.nl/nieuws/2020/10/14/nederland-heeft-een-algoritme-waakhond-nodig-a4016038>.

⁷² Coalitieakkoord 'Omzien naar elkaar, vooruitkijken naar de toekomst', 10 januari 2022, bijlage bij Kamerstuk 35 788, nr. 77, p. 33, <https://open.overheid.nl/documenten/rnl-f3cb0d9c-878b-4608-9f6a-8a2f6e24a410/pdf>.

⁷³ Zie Rijksoverheid, "Algoritmetoezichthouder gelanceerd", 17 januari 2023, <https://www.digitaleoverheid.nl/nieuws/algoritmetoezichthouder-gelanceerd/>.

discriminatie en willekeur tegen te gaan en transparantie te bevorderen.⁷⁴ Het toezicht op algoritmen is belegd bij een nieuw organisatieonderdeel van de Autoriteit Persoonsgegevens, genaamde de Directie Coördinatie Algoritmes (DCA).

Amnesty constateert na bestudering van de activiteiten, processen en taken die de algoritmetoezichthouder krijgt, dat deze niet voldoende mandaat en middelen heeft om mensenrechtenschendingen als gevolg van de inzet van algoritmen effectief te kunnen aanpakken. De DCA zal in zijn huidige vorm enkel de samenwerking tussen de bestaande sectorale toezichthouders stimuleren en coördineren.⁷⁵ Daarnaast gaat de DCA sectoroverstijgende risico's signaleren en gezamenlijke normuitleg in de toezichtspraktijk ten aanzien van algoritmen bevorderen.⁷⁶ Met dit pakket aan taken en activiteiten is de algoritmetoezichthouder slechts een papieren tijger. Daarnaast is het onwenselijk de verantwoordelijkheid om toezicht te houden op juist algoritmegebruik bij de individuele toezichthouders te laten. Zo zijn toezichthouders zoals de Autoriteit Persoonsgegevens, de Nationale Ombudsman en het College voor de Rechten van de Mens niet in staat gebleken om adequaat te reageren op het toeslagenschandaal, bijvoorbeeld omdat ze geen grip hadden het algoritmegebruik in hun sector, niet berust zijn met de noodzakelijke bevoegdheden en mandaat om effectief op te treden, of specifieke expertise over alle relevante mensenrechten missen.⁷⁷

Om effectief te kunnen optreden tegen mensenrechtenschendingen, is meer nodig. De algoritmetoezichthouder moet in staat zijn om zelfstandig onderzoek te doen naar algoritmegebruik, te handhaven door sancties op te leggen of verkeerd algoritmegebruik aan banden te leggen.⁷⁸ De algoritmetoezichthouder moet naast adequaat mandaat en voldoende middelen ook specifieke deskundigheid hebben om de mogelijke gevaren van het gebruik van algoritmen voor beslissingen over burgers te begrijpen en onderzoeken. Dit vereist onder meer expertise op het gebied van alle relevante mensenrechtennormen, waaronder gegevensbeschermingsnormen, datawetenschap, het ontwerp van algoritmische systemen en (semi-)geautomatiseerde besluitvorming.

KADER 7: DE ALGORITMETOEZICHTHOUDER OVER HET GEVAAR VAN DISCRIMINATIE

De DCA schrijft in haar eerste Rapportage Algoritmerisico's van juli 2023 over de risico's van algoritmen die door overheden worden ingezet voor risicoprofilering: "Voorbeelden hiervan zijn de Toeslagenaffaire en frauderisicosystemen die door gemeenten worden ingezet. Vaak gaat het om systemen en toepassingen gebaseerd op historische data of impliciete veronderstellingen waarmee risicofactoren uit die data worden afgeleid. Hiermee wordt niet daadwerkelijk fraude of ander te signaleren gedrag vastgesteld, maar wordt enkel het risico op dat gedrag afgeleid uit data over vastgestelde gedragingen in het verleden. Niet eerder geconstateerde gedragingen, spuriuze verbanden (schijnverbanden) of onterechte aannames leiden vaak tot een vertekend beeld van de werkelijkheid. Het gevaar van discriminatie, gebrek aan transparantie of onbedoelde willekeur is hier sterk aanwezig."⁷⁹

⁷⁴ Zie brief minister van BZK, Kamerstukken II, 2021/22, 35925 V, nr. 84, p. 20, <https://open.overheid.nl/repository/ronl-bbc2eeff7c47541f993ebd2f9b61508f07750dd1/1/pdf/kamerbrief-hoofdpijnen-beleid-voor-digitalisering.pdf> en brief ministerie van BZK, Kamerstukken II, 26643, nr. 953, p. 1, <https://open.overheid.nl/documenten/ronl-f06bda396eb995839765314c6ee4a872c49cad10/pdf>.

⁷⁵ Brief ministerie van BZK, Kamerstukken II, 26643, nr. 953, p. 5.

⁷⁶ Brief ministerie van BZK, Kamerstukken II, 26643, nr. 953, p. 5.

⁷⁷ Zo werden waarschuwingen van de Nationale Ombudsman genegeerd, mede omdat zijn rapport niet bindend is. Ook de adviezen en oordelen van het College voor de Rechten van de Mens zijn niet bindend. De Autoriteit Persoonsgegevens (AP) onderzocht de verwerking van nationaliteit door de Belastingdienst en verklaarde herhaaldelijk te zijn misleid door de Belastingdienst. In haar rapport hanteert de AP bovendien een onjuiste uitleg van het discriminatieverbod. Zie *Xenofobe Machines*, p.34-35.

⁷⁸ Zie nieuwsbericht Amnesty, "Algoritme-'waakhond' dreigt een papieren tijger te worden", 16 januari 2023, <https://www.amnesty.nl/actueel/algoritme-waakhond-dreigt-een-papieren-tijger-te-worden>.

⁷⁹ Directie Coördinatie Algoritmes, *Rapportage Algoritmerisico's Nederland*, juli 2023, p. 8, <https://www.autoriteitpersoonsgegevens.nl/actueel/primeur-eerste-rapportage-algoritmerisicos-nederland>.

6. AANBEVELINGEN

Amnesty roept de regering op om de volgende maatregelen nemen:

- Onmiddellijk een bindende en verplichte mensenrechtentoets in te voeren voor algoritmische besluitvormingssystemen⁸⁰ in de publieke sector.
- Vast te leggen en af te dwingen bij overheidsinstellingen dat het nooit is toegestaan om etniciteit, nationaliteit en andere afkomstgerelateerde gegevens te gebruiken als criterium in geautomatiseerde risicoprofielen die worden gebruikt in de zoektocht naar potentiële normovertreders.
- Een verbod in te stellen op het gebruik van zelflerende algoritmen in de gebruiksfase van algoritmische besluitvormingssystemen in de publieke sector als het gaat om a) beslissingen met rechtsgevolgen, b) beslissingen die de rechten en vrijheden van burgers aantasten, en c) beslissingen die een grote impact hebben op de samenleving.
- Een verbod in te stellen op het gebruik van black box-algoritmen voor beslissingen in risicovolle contexten zoals criminaliteitsbestrijding, sociale bescherming en migratie.
- Transparant te zijn over het gebruik van algoritmen voor beslissingen over controles van burgers, onder meer door verplichte en complete publicatie in het algoritmeregister van alle algoritmische systemen, inclusief de algoritmen die worden ingezet in de context van opsporing, rechtshandhaving en nationale veiligheid.
- De algoritmetoezichthouder en andere relevante toezichthouders uit te rusten met adequate bevoegdheden en voldoende middelen.

⁸⁰ Een algoritmisch systeem dat wordt gebruikt bij (de ondersteuning van) verschillende stappen in besluitvormingsprocessen. Onder 'besluit' wordt in deze briefing verstaan: alle beslissingen van overheidsinstellingen over burgers, waaronder formele besluiten om burgers te controleren, maar ook beslissingen om burgers te controleren die geen rechtsgevolgen hebben of niet kenbaar zijn gemaakt aan burgers.

Amnesty International is een wereldwijde beweging van 10 miljoen mensen die een beroep doen op de menselijkheid van iedereen en campagne voert voor verandering, zodat we allemaal van onze mensenrechten kunnen genieten. Onze visie is een wereld waarin machthebbers zich aan hun beloftes houden, het internationale recht respecteren en verantwoording afleggen over hun beleid. We zijn onafhankelijk van enige regering, politieke ideologie, economisch belang of religie en worden voor het overgrote deel gefinancierd door lidmaatschappen en donaties. We geloven dat solidariteit met en compassie voor mensen in de hele wereld onze samenlevingen kan verbeteren.

Contact



servicecenter@amnesty.nl



**Amnesty International
Nederland
Keizersgracht 177
1016 DR Amsterdam**



www.amnesty.nl

© Amnesty International 2023

Tenzij anders aangegeven is de inhoud van dit document gelicentieerd op grond van een Creative Commons-licentie (Naamsvermelding-NietCommercieel-GeenAfgeleideWerken 4.0 Internationaal).

<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode.nl>

Raadpleeg voor meer informatie de pagina over toestemmingen op onze website: amnesty.org

Op materiaal waarvan iemand anders dan Amnesty International het auteursrecht bezit, is de Creative Commons-licentie niet van toepassing.