Auditdienst Rijk
Ministerie van Financiën

Onderzoeksrapport
Bekendheid, toepasbaarheid en toegevoegde waarde handreiking "non-discriminatie by design

Definitief

Colofon

Titel	Bekendheid, toepasbaarheid en toegevoegde waarde handreiking 'non-discriminatie by design'
Uitgebracht aan	CIO Rijk
Datum	26 juni 2023
Kenmerk	2023-0000154060

Inlichtingen
Auditdienst Rijk
070-342 7700

Inhoud

Inleiding

Aanleiding opdracht
In januari 2021 is de handreiking 'non-discriminatie by design' vastgesteld. Volgens de kamerbrief 'Voortgang algoritmen en artificiële intelligentie'[1] van 10 juni 2021 is de handreiking een praktisch toepasbaar ontwerpkader dat ontwikkelaars helpt om al in de ontwikkelfase van een Artificiële Intelligentie (AI)-systeem[2] discriminerende patronen zoveel mogelijk te identificeren, te voorkomen en te bestrijden.
Op 28 oktober 2021 heeft de Kamer de volgende motie aangenomen: "verzoekt de regering om één jaar na de verspreiding van deze handreiking te onderzoeken in hoeverre de handreiking bekend is bij relevante overheidsdiensten en lokale overheden en te onderzoeken in hoeverre de aanbevelingen worden geïmplementeerd, en de Kamer over de resultaten te informeren." In haar reactie heeft de minister van BZK aangegeven dat de ADR zal worden gevraagd "te adviseren hoe de verspreiding en de impact van de handreiking respectievelijk uitgebreid en vergroot kan worden."

In september 2022 heeft de staatssecretaris Koninkrijksrelaties en *Digitalisering* in een reactie op het onderzoeksrapport "Algoritmes afwegen[3]" van het Rathenau Instituut aangegeven: "Omdat momenteel nog onvoldoende duidelijk is hoe deze handreiking in de praktijk gebruikt wordt en de impact daardoor lastig vast te stellen is, heb ik de Auditdienst Rijk gevraagd om een advies uit te brengen hoe de bekendheid van deze handreiking vergroot kan worden. Deze resultaten neem ik mee in het genoemde implementatiekader.[4]"

De reacties van de minister en staatssecretaris vormden de aanleiding voor BZK om de ADR onderhavige opdracht te verstrekken.

Context Implementatiekader voor algoritmen
Het ministerie van BZK werkt aan een kader voor de inzet van algoritmen met heldere juridische eisen en ethische beginselen. Met dit implementatiekader wordt voor overheidsorganisaties duidelijk wat de eisen zijn voor verantwoorde inzet van algoritmen. Er zijn al verschillende instrumenten ontwikkeld die helpen wet- en regelgeving beter toe te passen. Het implementatiekader zou bestaande verplichtingen, hulpmiddelen en handreikingen voor overheden moeten stroomlijnen, prioriteren en concretiseren, zodat overheden in alle fasen van de levenscyclus van algoritmische toepassingen praktische handvatten hebben. Daarnaast worden overheidsorganisaties verplicht een algoritmeregister[5] in te stellen.
In Europees verband zullen de juridische eisen en ethische beginselen vastgelegd gaan worden in een Europese AI-verordening.

1 Kamerstukken 2020/2021, 27529, nr 240

2 AI staat voor artificial intelligence / kunstmatige intelligentie

3 Ratenau Instituur (2022). *Algoritmes afwegen - Verkenning naar maatregelen ter bescherming van mensenrechten bij profilering in de uitvoering*. Den Haag.

4 Implementatiekader voor algoritmen

5 Algoritmes.overheid.nl is het centrale algoritmeregister voor de Nederlandse overheid. Hier worden door overheden gepubliceerde gegevens over hun algoritmes op één plek doorzoekbaar voor iedereen.

Doelstelling en onderzoeksvragen
Doelstelling van het onderzoek is inzicht geven in de bekendheid van de handreiking 'non-discriminatie by design' (verder: de handreiking) en in de mate waarin de handreiking daadwerkelijk wordt gebruikt inclusief afwegingen daarbij, teneinde adviezen te kunnen geven voor het vergroten van de impact van de handreiking.

Met het verkregen inzicht wil de opdrachtgever een strategie opzetten om de impact van de handreiking te vergroten. Verder wil de opdrachtgever de resultaten gebruiken bij het door BZK op te stellen implementatiekader voor algoritmen. Daarnaast kan het onderzoek een bijdrage leveren aan de bewustwording bij departementen over de komst van de Europese AI-Verordening en het (nationale) implementatiekader.

De centrale onderzoeksvraag luidt:
In welke mate is de handreiking 'non-discriminatie by design' bekend en wordt deze gebruikt en wat zijn mogelijkheden om de impact van de handreiking te vergroten?

Voor de beantwoording van deze centrale onderzoeksvraag zijn de volgende deelvragen geformuleerd:
1. Welk inzicht kan gegeven worden in de bekendheid en het gebruik van de handreiking?
2. Welk inzicht kan gegeven worden in de toepasbaarheid van de handreiking?
3. Wat is de toegevoegde waarde van de handreiking (hoe verhoudt deze zich tot alternatieve instrumenten)?
4. Welke mogelijkheden zijn er om de impact van de handreiking te vergroten?

De antwoorden op de deelvragen zijn tot stand gekomen door een documentstudie en interviews bij acht uitvoeringsorganisaties van de overheid en twee functionarissen van BZK.

Hoofdboodschap

Het onderzoek omvatte een documentstudie en interviews met acht organisaties. De onderzoeksresultaten zijn veelal gebaseerd op percepties van de geïnterviewden. In interviews is aangegeven dat AI nog volop in ontwikkeling is en ook nog een leerproces is. Geïnterviewden geven aan dat kaders van toezichthouders nog in ontwikkeling zijn en ook onderling verschillen. Daarnaast ervaart men een veelvoud aan publicaties en komt er met de Europese AI-verordening nog een kader bij. Verder is politiek veel aandacht voor de inzet van algoritmen en wijzen sommige geïnterviewden erop dat angst de ontwikkeling en kansen van de inzet van AI kan beperken.
Meerdere van de acht organisaties zijn nog zoekende naar de balans tussen alle kaders en hoe daar uitvoering aan te geven. Onze indruk daarbij is dat de ene organisatie hierin verder is dan de andere. Mogelijk mede afhankelijk van de context waarin ze werken en daarmee de urgentie voor het onderwerp non-discriminatie.

Hierna volgen de bevindingen over bekendheid, toepasbaarheid en de toegevoegde waarde van de handreiking. Tot slot geven wij inzichten en suggesties mee om de impact van de handreiking te vergroten.

Bekendheid
Meerdere geïnterviewden waren bekend met de handreiking en de inhoud ervan. Geen van de geïnterviewden gebruikt de handreiking expliciet bij de ontwikkeling van algoritmen. Wel hebben geïnterviewden aangegeven dat de handreiking of delen ervan zijn verwerkt in eigen richtlijnen en kaders. De handreiking wordt op deze wijze wel (indirect) gebruikt.

Kijkend naar randvoorwaardelijke aspecten om de bekendheid te vergroten, dan komen wij tot de volgende bevindingen. Wie verantwoordelijk is voor verspreiding, beheer en onderhoud van de handreiking is bij betrokkenen niet duidelijk. Wel zijn initiatieven ondernomen om de handreiking onder de aandacht te brengen, maar een concreet plan voor de implementatie en blijvende aandacht ontbreekt.

Toepasbaarheid
Hoewel de handreiking als begrijpelijk wordt ervaren, komt uit interviews naar voren dat het niet duidelijk is bij welke soorten algoritmen de handreiking toegepast moet worden. Het algemene beeld is dat het voor organisaties een uitdaging is om alle verschillende kaders en richtlijnen toe te passen. Daarnaast blijkt er behoefte te zijn aan helderheid over taak- en verantwoordelijkheidsverdeling bij de ontwikkeling van een algoritme.

Toegevoegde waarde
De meerwaarde van de handreiking ten opzichte van andere instrumenten zit volgens geïnterviewden vooral in praktische voorbeelden en de benadering vanuit drie perspectieven (technisch, organisatorisch en juridisch). Daarbij is wel aangegeven is dat de handreiking (weer) een extra instrument is.

Vergroten impact
De impact van de handreiking kan worden vergroot door verantwoordelijkheden over het beheer van de handreiking te beleggen en blijvende aandacht hiervoor te organiseren.

De handreiking is omvangrijk en toepassing van de handreiking op alle algoritmen kan een (te) grote belasting voor organisaties zijn. Daarom is er behoefte aan een handvatten voor een risicogerichte benadering, waarmee organisaties op basis van

criteria en weging het risico op discriminatie bij gebruik van een algoritme kunnen inschatten.

Voor het verantwoord gebruik van algoritmen bestaan al veel verschillende kaders en instrumenten, waarbij niet duidelijk is hoe die kaders en instrumenten zich tot elkaar verhouden. Er is dan ook geen behoefte aan een nieuw instrument, maar wel aan een integraal kader of routekaart voor het gebruik van de bestaande instrumenten.

Daarnaast vragen geïnterviewden aandacht voor helderheid omtrent de verantwoordelijkheid voor de meer beleidsmatige keuzes die gemaakt moeten worden als het gaat om mogelijke discriminatie (welke selectiecriteria zijn al dan niet ethisch aanvaardbaar). Zij vinden dat een handreiking niet een instrument is relatie met het begrip 'fairness', waarbij het gaat over de redelijkheid en billijkheid om onderscheid te maken.

Ook is het voor het vergroten van de impact van de handreiking belangrijk dat het bewustzijn voor algoritmen en (data-)ethiek in organisaties wordt vergroot.

Tot slot geven geïnterviewden van meerdere organisaties aan hun organisatie al een heel eind gevorderd is zijn met het maken van een eigen beleidsinstrumentarium om discriminatie by design tegen te gaan. Daarbij putten zij soms uit (delen van) de handreiking. Het eventueel verplicht stellen van het gebruik van de handreiking kan dergelijke initiatieven frustreren en als demotiverend worden ervaren.

1 Bekendheid

Dit hoofdstuk geeft antwoord op de eerste deelvraag van het onderzoek:
Welk inzicht kan gegeven worden in de bekendheid en het gebruik van de handreiking?

Als eerste geven we aan in hoeverre de handreiking bekend is bij de geïnterviewden en of de handreiking ook wordt gebruikt bij de ontwikkeling van algoritmen. Daarnaast geven wij inzicht in de organisatorische aspecten die zijn getroffen om bekendheid en blijvende bekendheid van de handreiking te borgen.

1.1 Handreiking is beperkt bekend en wordt veelal indirect toegepast

In deze eerste paragraaf geven wij inzicht in de bekendheid van de handreiking onder de geïnterviewden van de acht betrokken organisaties. Vervolgens geven of de handreiking ook daadwerkelijk wordt gebruikt bij deze organisaties.

1.1.1 Deel van de geïnterviewden is bekend met de handreiking

Meerdere geïnterviewden waren bekend met de handreiking en de inhoud ervan. De geïnterviewden die de handreiking nog niet inhoudelijk kenden hebben de handreiking naar aanleiding van ons onderzoek voorafgaand aan het interview doorgelezen.
Van de geïnterviewden die vooraf niet inhoudelijk bekend waren met de handreiking gaven meerdere personen aan dat de inhoud van de handreiking wel herkenbaar is. Dit komt mede door de toepassing van de fasering conform het CRISP-DM[6] wat onder data scientists een bekend model is.

1.1.2 Gebruik van de handreiking is veelal indirect via eigen beleid

Geen van de geïnterviewden gebruikt de handreiking expliciet bij de ontwikkeling van algoritmen. Wel hebben geïnterviewden van meerdere organisaties aangegeven dat de handreiking of delen ervan zijn verwerkt in eigen richtlijnen en kaders. De handreiking wordt op deze wijze (deels) indirect gebruikt. De handreiking vormt daarbij niet de enige bron, ook zijn genoemd het IAMA[7], Ethische Richtsnoeren van de High Level Expert Group van de Europese Commissie, het toetsingskader algoritmen van de Algemene Rekenkamer en het normenkader van de ADR om de beheersing van een algoritme te onderzoeken.
Bij een aantal organisaties is een verwijzing opgenomen naar de handreiking, bijvoorbeeld in een eigen kader of op intranet.
Zoals uit het volgende hoofdstuk (paragraaf 2.4) zal blijken vinden organisaties het ook niet realistisch de handreiking volledig toe te passen voor elk algoritme.

1.2 Implementatie en ondersteuning voor blijvende aandacht is niet ingericht

In het onderzoek hebben wij naar een aantal randvoorwaardelijke aspecten gekeken voor het organiseren van draagvlak voor en blijvende bekendheid van de handreiking, Hierna geven wij inzicht in hoeverre hier uitvoering aan gegeven is.

1.2.1 Verantwoordelijkheden voor de handreiking zijn binnen BZK niet duidelijk belegd

Informatie bij de publicatie van de handreiking op de website open.overheid.nl geeft nu aan dat het ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK)

[6] De relevante vragen en uitgangspunten worden volgens de handreiking weergegeven aan de hand van fasen die zijn gebaseerd op, maar niet precies hetzelfde zijn als, het CRoss-Industrie StandaardProces voor DataMining (CRISP-DM).
[7] IAMA staat voor 'Impact Assessment voor Mensenrechten bij de inzet van Algoritmes'.

verantwoordelijk is[8]. Wij merken op dat bij aanvang van ons onderzoek ook de ministeries van Economische Zaken en Klimaat en Justitie en Veiligheid als verantwoordelijke ministeries waren opgenomen op deze website. In de beleidsbrief "AI, publieke waarden en mensenrechten" van 8 oktober 201915 heeft het ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK) toegezegd te willen onderzoeken hoe publieke waarden en mensenrechten geoperationaliseerd kunnen worden tot AI-systeemprincipes, te beginnen met non-discriminatie. Het onderzoeksrapport dat ten grondslag ligt aan de 'Handreiking non-discriminatie by design' volgt uit die toezegging.[9] De verantwoordelijkheid voor de handreiking blijkt binnen BZK niet duidelijk te zijn belegd. Ook hebben wij geen informatie ontvangen waarin verantwoordelijkheden zijn uitgewerkt. Aannemelijk is volgens de gesproken functionarissen van BZK dat de directeur van de Directie Digitale Samenleving van het Directoraat-Generaal Digitalisering en Overheidsorganisatie (DGDOO) verantwoordelijk is voor de handreiking. Aangegeven is dat deze directie de initiator en opdrachtgever was voor het opstellen van de handreiking en dat de SG van BZK eindverantwoordelijk is. Verder is opgemerkt dat CIO-Rijk pas is betrokken nadat de handreiking was gepubliceerd (juni 2021) en aandacht ontstond voor de verdere verspreiding van de handreiking.

1.2.2 Initiatieven voor bekendheid zijn ondernomen, maar een concreet plan voor de implementatie en blijvende aandacht ontbreekt

BZK geeft aan dat voor de implementatie van de handreiking geen plan is opgesteld, zoals bijvoorbeeld een communicatieplan of implementatie strategie. Aangegeven is dat gaandeweg wel een aantal initiatieven zijn ontplooid om de handreiking onder de aandacht te brengen, zoals:

- Ten tijde van de publicatie van de handreiking in juni 2021 is ook de Tweede Kamer hiervan op de hoogte gesteld.
- De handreiking is gepubliceerd op www.digitaleoverheid.nl; de website van de Rijksoverheid voor professionals die aan digitalisering werken. Daarbij is de handreiking opgenomen in de toolbox "Ethisch Verantwoorde Innovatie[10]". Ook is de handreiking terug te vinden op andere websites zoals open.overheid.nl.
- Presentaties zijn gegeven bij verschillende instituten en congressen (deels online):
    - conferentie Nederland Digitaal, 10 februari 2021
    - presentatie voor Verbond van Verzekeraars, 10 november 2021
    - presentatie voor VNG, 24 juni 2021
    - presentatie voor topambtenaren, 3 februari 2021
    - presentatie op TILT seminar, 13 juli 2021
- Het Rijks-ICT-gilde heeft sessies georganiseerd over IAMA / 'ethics by design', waarbij ook de handreiking is genoemd.

De initiatieven waren niet alleen gericht op de overheid, omdat systemen ook van externe ontwikkelaars kunnen worden afgenomen. Aangegeven is dat het daarom ook van belang was dat de markt op de hoogte is van de handreiking.
Een ander belangrijk aspect is zorgen voor blijvende aandacht. Bij afwezigheid van een plan is ook dit onderdeel nog niet vormgegeven.

1.2.3 De faciliterende rol is nog niet uitgewerkt in een concreet plan

Wij hebben geen informatie aangetroffen over met wie een organisatie contact kan opnemen bij vragen over de handreiking. Ook in de handreiking zelf is dit niet vermeld. Hierdoor ontbreekt een mogelijkheid voor organisaties om vragen te stellen over de handreiking, melding te maken van een niet functionerende link of actuele ontwikkelingen die mogelijk meegenomen kunnen worden in de handreiking. Vanuit BZK is, in aanvulling op de initiatieven uit paragraaf 1.2.2, nog geen opleiding of ander initiatief georganiseerd om de bekendheid te vergroten.

---

8 Stand van zaken op 23 maart. Op het moment van start van het onderzoek stond op de betreffende website dat drie ministeries verantwoordelijk waren. Naast BZK ook het ministerie van Economische Zaken en Klimaat en het ministerie van Justitie & Veiligheid.
9 Onderzoeksrapport 'Non-discriminatie by design',
10 Publieke waarden centraal Toolbox Ethisch Verantwoorde Innovatie - Digitale Overheid.

1.2.4 Delen van successen en kennis is niet vormgegeven

De rol van BZK bij het delen van kennis tussen organisaties of successen is niet vormgegeven. Successen of ervaringen met de toepassing van de handreiking worden nog niet gedeeld.

2 Toepasbaarheid

Dit hoofdstuk geeft antwoord op de tweede deelvraag van het onderzoek:
Welk inzicht kan gegeven worden in de toepasbaarheid van de handreiking?
Wij gaan achtereenvolgens in op de (ervaren) begrijpelijkheid, diepgang en volledigheid van de handreiking en in hoeverre de toepassing van de handreiking bij algoritmen in de praktijk ook realistisch is.

2.1 De handreiking is begrijpelijk, maar verhouding met andere instrumenten/kaders is niet duidelijk

Geïnterviewden geven aan dat de handreiking begrijpelijk is. Over het algemeen wordt het onderscheid naar een juridisch, technisch en organisatorisch perspectief als prettig ervaren. De verdeling draagt eraan bij dat meerdere perspectieven bij het maken van afwegingen en uitwerkingen bewuster worden meegenomen bij de ontwikkeling van AI-systemen.
De gehanteerde fasering is herkenbaar door de aansluiting op de levenscyclus van een algoritme (volgens CRISP-DM[11]).
Afhankelijk van de doelgroep bevat het document volgens geïnterviewden echter te veel vakjargon, wat ten koste kan gaan van de begrijpelijkheid. Hierbij past de kanttekening dat geïnterviewden een bredere doelgroep voor ogen hebben, dan de doelgroep zoals beoogd in de handreiking: "het document is bedoeld voor projectleiders die sturing geven aan systeembouwers, data-analisten en AI-experts" (zie ook paragraaf 2.2).

Uit de gesprekken komt verder naar voren dat het niet duidelijk is bij welke algoritmen de handreiking toegepast moet worden. Het algemene beeld is dat het voor organisaties een uitdaging is om alle verschillende kaders en richtlijnen toe te passen. Niet helder is welke publicatie wanneer toegepast kan of moet worden. Daarbij hebben wij wel gemerkt dat meerdere organisaties (onderdelen van) de handreiking herkennen of daadwerkelijk hebben verwerkt in eigen beleid.

2.2 Onduidelijkheid over doelgroep leidt tot verschillend beeld over gewenste diepgang/omvang van de Handreiking

Over de diepgang van de handreiking bestaat onder geïnterviewden een wisselend beeld, maar dit hangt samen met de doelgroep die geïnterviewden voor ogen hebben. Over de doelgroep wordt niet consistent gecommuniceerd. In de handreiking staat dat de handreiking is bedoeld voor projectleiders die sturing geven aan systeembouwers, data-analisten en AI-experts. In de aanbiedingsbrief[12] van de Handreiking aan de Tweede Kamer staat echter dat de handreiking ontwikkelaars helpt om al in de ontwikkelfase van een AI-systeem discriminerende patronen in gegevens zoveel mogelijk te identificeren, te voorkomen en te bestrijden.
De opmerking in meerdere interviews dat de handreiking te weinig concreet is voor ontwikkelaars, moet ook in dat perspectief worden geplaatst. Ontwikkelaars verwachten concrete normen, dus meer absolute termen over wat wel of niet is toegestaan. Dat lijkt ook niet het primaire doel van de handreiking. In de handreiking staat: "De handreiking moet dan ook met name

[11] De relevante vragen en uitgangspunten worden volgens de handreiking weergegeven aan de hand van fasen die zijn gebaseerd op, maar niet precies hetzelfde zijn als, het CRoss-Industrie StandaardProces voor DataMining (CRISP-DM).
12 BZK, 10 juni 2021, kenmerk 2021-0000291680

worden gezien als document dat dit vraaggesprek kan faciliteren en dat er voor zorg kan dragen dat alle relevante vragen in de juiste fasen van het project worden gesteld".
Met inachtneming van het bovenstaande is daarnaast in interviews opgemerkt dat het detailniveau voor de datawetenschappers te laag is, terwijl voor de minder technisch onderlegde personen het detailniveau juist te hoog is met te veel vakjargon. Als voorbeeld is gegeven dat de handreiking aangeeft dat je moet checken of het algoritme discrimineert, maar ontbreekt hierin de uitwerking, hoe je daar concreet uitvoering aan moet geven. De hyperlinks naar meer informatie in de handreiking zijn nuttig, maar niet alle geïnterviewden zijn zich hier bewust van. Ook werken niet alle links in het document. Voor de technische diepgang met het oog op 'fairness' en 'bias' is het hoofdstuk over the Bias Analysis in The Fairness Handbook[13] volgens een geïnterviewde concreter. Daarbij is opgemerkt dat in de handreiking een hele algemene definitie van discriminatie wordt gehanteerd, die te weinig concreet is om te bepalen of een bepaalde selectie wel of niet is toegestaan. In de handreiking worden wel voorbeelden van directe of indirecte discriminatie genoemd, maar er is behoefte aan een concretere definitie.

Verder is aangegeven dat meer diepgang wel zorgt voor een omvangrijker document, terwijl de huidige handreiking meermaals als te omvangrijk is ervaren (zie ook paragraaf 2.4). Ook is aangegeven dat het document lastig doorzoekbaar is. De mate van gewenste diepgang (en volledigheid) zorgt daarmee dus voor een spanningsveld met het oog op de omvang van de handreiking en is afhankelijk van de beoogde doelgroep.

2.3 De handreiking wordt inhoudelijk als volledig ervaren, informatie over verdeling van taken en verantwoordelijkheden wordt gemist

De handreiking ervaart men inhoudelijk over het algemeen als een volledig document, in die zin dat geen onderwerpen worden gemist. Gegeven kanttekeningen hebben veelal betrekking op de diepgang van de handreiking (zoals weergegeven in paragraaf 2.2). Zo is aangegeven dat de handreiking erg gericht is op ontwikkelaars en AI-experts, terwijl volgens geïnterviewden veel vragen meer algemene vragen zijn die ontwikkelaars en AI-experts niet moeten beantwoorden. Aangegeven is dat de eigenaar van een algoritme uiteindelijk een besluit moet nemen over toepassing van het algoritme en niet de ontwikkelaar. Daarbij gaat het ook om vragen over welke selecties toelaatbaar zijn en welke niet. Het betreft dan beleidsmatige/politieke keuzes.
In het verlengde hiervan geeft een deel van de geïnterviewden aan dat een verdeling van verantwoordelijkheden wordt gemist. Men zou dan graag zien dat in de handreiking wordt aangegeven wie wat moet doen in het proces: wat is wiens verantwoordelijkheid, wie betrek je wanneer bij welke fase (projectleider, data-scientist, jurist, business owner etc) en wie zou moeten controleren om vast te stellen of een juiste invulling is gegeven aan 'non-discriminatie by design'.

2.4 Toepassen van de handreiking bij elk algoritme acht men niet realistisch

Als onderdeel van de toepasbaarheid hebben wij ook gevraagd of toepassing van de handreiking bij elk algoritme realistisch is. Daarbij merken wij op dat sommige bevindingen mogelijk onvermijdbaar zijn gezien het uitgangspunt van de handreiking, zoals ook in de handreiking is opgenomen: "Afhankelijk van de context van het AI-project kan de ene vraag of opmerking relevanter zijn dan de andere. Aangezien dit een generieke handreiking is met als doel toepasbaar te zijn in verschillende contexten, is dit iets waar gebruikers zelf rekening mee dienen te houden".

2.4.1 Toepassen van de handreiking voor alle algoritmen is te belastend voor organisaties

[13] The Fairness Handbook, gemeente Amsterdam, mei 2022.

Geïnterviewden geven aan dat het niet realistisch is om de handreiking bij de ontwikkeling van een algoritme in alle gevallen even grondig te doorlopen. De handreiking wordt daarvoor als te omvangrijk ervaren (69 pagina's). Daarbij is ook opgemerkt dat de definitie van een algoritme een breed begrip is. Vanuit een brede definitie van een algoritme zouden kaders op alle algoritmen toepast moeten worden wat leidt tot een situatie die erg belastend en onwerkbaar is voor organisaties. Een risicogerichte benadering voor toepassing van de handreiking wordt gemist.
Ook is aangegeven dat de handreiking geschreven lijkt te zijn voor grote projecten, terwijl organisaties in de praktijk ook te maken hebben met kleine projecten met een hele specifieke toepassing. In die gevallen wordt het niet werkbaar en zinvol geacht om de handreiking volledig toe te passen.

2.4.2 Zoals ook opgemerkt in de handreiking sluit de fasering niet altijd aan op de praktijk

De ontwikkeling van een algoritme doorloopt niet altijd de fasering zoals die wordt gehanteerd in de handreiking. Soms ontwikkelen datalabs een tool (innovatie) en wordt daarna nagedacht of de tool in de praktijk van toegevoegde waarde kan zijn. Vanaf dat moment kan het algoritme in de context worden geplaatst voor de ethische reflectie. Een andere voorkomende situatie is dat incrementele aanpassingen plaatsvinden aan een bestaand algoritme. Het is niet werkbaar de handreiking voor elke aanpassing opnieuw te doorlopen.
In de handreiking is hierover opgenomen: "In deze handreiking zijn zes stappen onderscheiden en lineair gepresenteerd, terwijl het bouwen van een AI-systeem in werkelijkheid een iteratief proces is. Bovendien kan het zijn dat er al data zijn verzameld, een AI-systeem wordt gekocht van een externe partij of de probleemdefinitie al vast staat. In zulke gevallen kan het voorkomen dat niet alle stappen hoeven te worden doorlopen of in een andere volgorde."

3 Toegevoegde waarde

Dit hoofdstuk geeft antwoord op de derde deelvraag van het onderzoek:
Wat is de toegevoegde waarde van de handreiking (hoe verhoudt deze zich tot alternatieve instrumenten)?

In de handreiking is het doel en de doelgroep als volgt verwoord:

- "De handreiking heeft ten doel te helpen bij het inrichten van het data-management, het bouwen van een algoritme en het ordenen van processen om tot besluiten te komen."[14]
- "Deze handreiking is geschreven voor de projectleider van een AI-systeem en heeft ten doel de projectleider de juiste fasen te laten onderscheiden, de juiste mensen op juiste momenten bij elkaar te zetten en hen de juiste vragen te laten stellen. Als de technische experts en de data-analisten samen met de juristen en de functionaris gegevensbescherming aan tafel zitten, aangevuld met relevante stakeholders, domeinexperts en data stewards, zijn de vragen die in dit document zijn vervat leidend voor de discussie. De handreiking moet dan ook met name worden gezien als document dat dit vraaggesprek kan faciliteren en dat er voor zorg kan dragen dat alle relevante vragen in de juiste fasen van het project worden gesteld."[15]
- "Het document is bedoeld voor projectleiders die sturing geven aan systeembouwers, data-analisten en AI-experts. Stel je wilt een AI-systeem zo non-discriminatoir mogelijk maken, waar moet je dan aan denken en welke discussies moet je binnen je team voeren?"[16]

3.1 De relevantie van de handreiking wordt verschillend ervaren

Het is ons opgevallen dat het doel zoals geformuleerd in de handreiking niet eenduidig is. Dat maakt het lastiger te duiden of de handreiking bijdraagt aan het beoogde doel. De mate van relevantie ervaren geïnterviewden verschillend. Dit lijkt deels afhankelijk van de context van de organisatie (welke informatie wordt verwerkt) en van de mate waarin algoritmen worden toegepast. Verder lijkt een persoonlijke overtuiging van invloed te zijn op de veronderstelde relevantie. De één vindt de handreiking een veelbelovend document en de ander vindt deze te omvangrijk of is van mening dat een ethische discussie niet in een vragenlijst is te vatten.

Onze indruk is dat het volwassenheidsniveau van de betrokken organisaties verschilt als het gaat om 'fairness' en voorkomen van discriminatie bij toepassing van algoritmen. Mogelijk heeft dit ook te maken met de context van een organisatie en het soort algoritme. Daarbij hebben wij gemerkt dat bepaalde organisaties zelf op zoek zijn gegaan naar handvatten om invulling te geven aan deze aspecten bij de toepassing van algoritmen. Sommige betrokkenen kwamen tijdens deze zoektocht uit bij de handreiking en vonden deze een veelbelovend document waar ze ook uit geput hebben voor een eigen beleid. De aandacht voor non-discriminatie en 'fairness' lijkt intrinsiek gedreven te worden door individuen die meer bewust zijn van de risico's, gecombineerd met de politieke aandacht voor het onderwerp.

Meerdere geïnterviewden geven aan dat de handreiking kan helpen om het bewustzijn van mensen van discriminatie te vergroten. De handreiking bevat veel vragen die kunnen helpen bij een goede dialoog over discriminatie. Ook de in de handreiking genoemde voorbeelden worden door meerdere personen als waardevol

14 Blz. 11 van de handreiking

15 Blz. 15 van de handreiking

16 Blz. 2 van de handreiking

ervaren.
Als kritische noot is meegegeven dat een ethische discussie niet in een vragenlijst is te vatten en dus ook niet in een handreiking. Het risico is dat de algemene vragen niet raken aan de morele vragen die relevant zijn in de specifieke casus.

De relevantie hangt ook nauw samen met de toepasbaarheid, zoals in het vorige hoofdstuk is besproken. Indien de handreiking belemmeringen kent voor de toepasbaarheid, dan zal de handreiking niet (optimaal) bijdragen aan het beoogde doel. Onder andere zijn genoemd de technische diepgang voor data scientists, de omvang van het document, de hoeveelheid aan publicaties en de vermeende onduidelijkheid wanneer de handreiking moet worden toegepast.

3.2 Meerwaarde handreiking versus alternatieve instrumenten zit vooral in praktische voorbeelden

De meerwaarde van de handreiking ten opzichte van alternatieve instrumenten zit volgens geïnterviewden vooral in de praktische voorbeelden van de handreiking en in de onderverdeling naar de drie perspectieven (juridisch, technisch en organisatorisch). Aangegeven is dat de vele voorbeelden in de handreiking bijdragen aan bewustwording over het onderwerp discriminatie bij inzet van algoritmen. De handreiking kan zorgen voor waardevolle discussies in teams die bezig zijn met ontwikkeling van algoritmen. De handreiking bevat veel vragen die kunnen helpen bij een goede dialoog over discriminatie, niet alleen voor ontwikkelaars en data-scientists, maar ook voor bijvoorbeeld bestuurders. Geïnterviewden geven aan dat met name de eerste 20 pagina's zich hier goed voor lenen. Ook is opgemerkt dat de handreiking geschikt is om te gebruiken als naslagwerk bij de ontwikkeling van algoritmen.

Diverse alternatieven zijn genoemd voor de handreiking. Het gaat dan om algemene instrumenten die meerdere organisaties kunnen gebruiken en instrumenten die specifiek op de eigen organisatie en eigen werkwijze zijn toegesneden. Ook is opgemerkt dat overlap zit in de verschillende instrumenten.
Het IAMA, de DPIA, het toetsingskader algoritmen van de Algemene Rekenkamer, de Ethische Richtsnoeren voor Betrouwbare AI van de Europese Commissie, het Handbook fairness, de Begeleidingsethiek methode en het ADR normenkader voor algoritmen zijn genoemd als algemene instrumenten of kaders (zie voor een beknopte omschrijving bijlage 1 van dit rapport).
Meerdere organisaties werken met een specifiek op de eigen organisatie toegesneden instrument, waarbij in sommige gevallen is aangegeven dat daarbij gebruik is gemaakt van de handreiking. Soms is geput uit meerdere van de bovengenoemde instrumenten.

De meerwaarde van het IAMA, het Handbook fairness en de Begeleidingsethiek methode zijn specifieker naar voren gekomen tijdens het onderzoek en beschrijven wij daarom hierna.

Als meerwaarde van het IAMA ten opzichte van de handreiking is aangegeven dat het een bredere scope heeft dan de handreiking. Waar de handreiking zich beperkt tot discriminatie richt het IAMA zich op meerdere aspecten van de mensenrechten. Bovendien ziet men het IAMA meer als een instrument om algoritmen te toetsen. Als nadeel van het IAMA is opgemerkt dat het theoretischer en abstracter is, waar de handreiking juist veel praktische voorbeelden bevat. Het IAMA is een soort kapstok, waarvan voor de verdere uitwerking wordt verwezen naar andere instrumenten, zoals de handreiking. Ook is opgemerkt dat het IAMA wordt gebruikt zodra een eerste concept van een algoritme gereed is. De handreiking is al vanaf het begin bij de ontwikkeling van een algoritme te gebruiken.

Over het Handbook fairness is aangegeven dat het een technischere uitwerking bevat dan de Handreiking. Het Handbook biedt wellicht meer houvast voor iemand die analyses maakt om inzicht te geven in de prestaties van het algoritme met het

oog op 'fairness' en 'bias'. Dit komt doordat het Handbook meer details geeft over de technische stappen die nodig zijn om te komen tot bepaalde analyses.

De Begeleidingsethiek methode gaat uit van ethische gesprekken over een verantwoorde toepassing van technologie. Het bestaat uit een workshop waarin verschillende betrokkenen de dialoog met elkaar voren over de toepassing van een concrete technologie in een specifieke context om gezamenlijk tot een aantal concrete handelingsopties te komen over hoe dat ethischer verantwoord kan[17]. Het is dus breder dan discriminatie; het gaat over ethiek in breder perspectief.

3.3 Handreiking is (weer) een extra document, terwijl meer behoefte is aan een integraal kader / instrument

Zoals in de vorige paragraaf is aangegeven, wordt veel waarde toegekend aan de praktische voorbeelden die in de handreiking staan. De handreiking wordt ook gebruikt als naslagwerk waarbij ook gebruik wordt gemaakt van de verwijzingen (links) in de handreiking. De handreiking kan ervoor zorgen dat er waardevolle discussies starten binnen een team en draagt daarmee bij aan bewustwording. Praktijkvoorbeelden zijn daarbij waardevol om meer gevoel te krijgen bij risico's en afwegingen met het oog op 'bias' en discriminatie. Opgemerkt is dat met name de eerste 20 bladzijdes van de handreiking een goede toelichting bevat voor degene die nog niet erg bekend zijn met het onderwerp.
De handreiking richt zich op het hele ontwikkeltraject van een algoritme. Naast gebruik als naslagwerk kan de handreiking tijdens de ontwikkeling van een algoritme ook geraadpleegd worden als checklist om vast te stellen of aan alle relevante aspecten is gedacht.

Niettemin wordt de toegevoegde waarde van de handreiking over het algemeen als beperkt ervaren omdat het een extra document is naast alle andere publicaties. In sommige gevallen worden verschillende instrumenten ook wel naast elkaar gebruikt, afhankelijk van de specifieke situatie en het type algoritme. Meermaals is opgemerkt dat er meer behoefte is aan en integraal kader of routekaart, waarbij dan duidelijk wordt in welke situaties (en op welk moment) een bepaald instrument kan of moet worden gebruikt.

17 Flyer begeleidingsethiek, ECP,2019 www.begeleidingsethiek.nl.

4 Adviezen voor het vervolg

Dit hoofdstuk geeft antwoord op de laatste deelvraag:
Welke mogelijkheden zijn er om de impact van de handreiking te vergroten?

De adviezen hierna passen in de context van de ontwikkeling van een implementatiekader voor algoritmen. Een belangrijke kanttekening is dat de bevindingen en dus ook de aanbevelingen zijn gebaseerd op de perceptie van één of twee medewerkers bij 8 overheidsorganisaties (zie ook de onderzoeksverantwoording in hoofdstuk 5).

4.1 Plaats de handreiking in een kader in relatie tot andere instrumenten

Organisaties lijken meer behoefte te hebben aan een integraal kader of routekaart dan dat ze een aanpassing wensen van de handreiking non-discriminatie by design. Door de hoeveelheid aan verschillende publicaties is het niet altijd helder waar organisaties aan moeten voldoen. Daarnaast geeft men aan dat tussen verschillende publicaties ook overlap zit. Wij adviseren om te komen tot een integraal kader of routekaart, dat kan helpen om inzicht te geven welke publicatie kan of moet worden gebruikt. Waarbij tevens aansluiting gevonden kan worden met eigen organisatiegebonden kaders. Het integraal kader of de routekaart kan onderdeel zijn van het eerdergenoemde implementatiekader voor algoritmen.

4.2 Overweeg een risicogerichte benadering voor de toepassing van de handreiking

Uit hoofdstuk 2 en 3 is naar voren gekomen dat de handreiking omvangrijk wordt gevonden en dat de toepassing van de handreiking voor elk algoritme een te zware belasting voor de organisatie zou zijn. Ook is het niet voor iedereen duidelijk wanneer de handreiking het beste toegepast kan worden. Wij adviseren daarom een risicogerichte benadering te overwegen en organisaties handvatten te bieden op basis van welke criteria en weging het risico van een algoritme kan worden ingeschat. Dit kan ook onderdeel zijn van het in de vorige paragraaf genoemd integraal kader of routekaart.

4.3 Werk aan het vergroten van bewustzijn voor algoritmen en (data-)ethiek in de organisatie

Uit interviews komt naar voren dat bewustwording van AI-risico's een onderwerp is dat breder binnen de organisatie aandacht behoeft voor een goede ethische reflectie. Wij adviseren dan ook te werken aan bewustwording van AI-risico's breed in de organisatie. De handreiking bevat nuttige voorbeelden en een algemene introductie zoals opgemerkt in hoofdstuk 3. Alleen is de handreiking niet specifiek bedoeld om bewustwording te vergroten onder een bredere doelgroep binnen een organisatie.
Ook is in interviews opgemerkt dat met enkel de handreiking een organisatie niet zal komen tot succesvolle implementatie van een algoritme. Voor het succesvol ontwikkelen en implementeren van betrouwbare algoritmen moeten ook aanvullende randvoorwaarden aanwezig zijn, zoals beleid, governance en bewustzijn voor algoritmen en ethiek in de organisatie.

4.4 Zorg voor duidelijkheid in taken en verantwoordelijkheden van verschillende betrokkenen

De handreiking is geschreven voor de projectleider van een AI-systeem en heeft ten doel de projectleider de juiste fasen te laten onderscheiden, de juiste mensen op de juiste momenten bij elkaar te zetten en hen de juiste vragen te laten stellen. Wij adviseren helderheid te geven over taken en verantwoordelijkheden bij de ontwikkeling van algoritmen. Verantwoord gebruik van algoritmen is niet alleen een

taak en verantwoordelijkheid van projectleiders en ontwikkelaars. Er is behoefte aan meer helderheid over beleidsmatige keuzes, over wat wel of niet acceptabel is bij het maken van onderscheid. Dit zijn geen keuzes die ontwikkelaars en AI-projectleiders moeten maken. Benadrukt is dat de handreiking niet toereikend is om dergelijke onderwerpen te verankeren.

4.5 Beleg verantwoordelijkheid voor de handreiking en borg de (blijvende) aandacht ervoor

Uit hoofdstuk 1 is naar voren gekomen dat bepaalde randvoorwaarden bij BZK niet zijn ingericht om (blijvende) bekendheid te kunnen organiseren. Zo heeft BZK geen implementatieplan of communicatieplan opgesteld (ook al werd er gaandeweg wel meer ruchtbaarheid aan de handreiking gegeven) en is onduidelijkheid over belegging van verantwoordelijkheden. Wij adviseren om duidelijk te beleggen wie verantwoordelijk is voor het beheer van de handreiking (denk hierbij aan het actualiseren van de inhoud en het informeren van de doelgroep over relevante wijzigingen) en voor de ondersteuning bij de toepassing van de handreiking. Ook is het van belang om initiatieven te stroomlijnen die zijn gericht op het vergroten van de (blijvende) bekendheid en de toepassing van de handreiking. Een planmatige aanpak kan daarbij helpen.

In interviews zijn mogelijkheden genoemd om desgewenst de bekendheid van de handreiking te verbeteren. BZK zou hierbij een faciliterende rol op zich kunnen nemen. De volgende suggesties zijn gegeven en delen wij ter overweging:

- Aandacht besteden aan de handreiking bij congressen, zoals bijvoorbeeld de Dataconferentie van JenV en het NVBB congres (voor gemeenteambtenaren).
- Ontwikkelen van een e-learning. Bijvoorbeeld online trainingen aanbieden op de website van de Nederlandse AI-coalitie. Eventueel verschillende trainingen voor verschillende doelgroepen.
- Meer kennisdelen door het uitwisselen van ervaringen met de toepassing van de handreiking.

4.6 Verplichte toepassing van de handreiking kan bestaande initiatieven tenietdoen

Hoewel het verplicht stellen van de handreiking de bekendheid en de toepassing ervan naar verwachting zal vergroten, zien geïnterviewden dit niet als een gewenste optie. Dit heeft deels te maken met de belasting die dit mee zou brengen voor de organisaties. Daarnaast kan het verplichten van de handreiking ook een negatief effect hebben op organisaties die verder zijn in de ontwikkeling van algoritmen en eigen beleid en werkwijze hebben samengesteld. Het eventueel verplicht stellen van het gebruik van de handreiking kan dergelijke initiatieven frustreren en als onredelijk worden ervaren.

5 Verantwoording onderzoek

5.1 Doelstelling en onderzoeksvragen

Doelstelling van het onderzoek is inzicht geven in de bekendheid van de handreiking 'non-discriminatie by design' (verder: de handreiking) en in de mate waarin de handreiking daadwerkelijk wordt gebruikt inclusief afwegingen daarbij, teneinde adviezen te kunnen geven voor het vergroten van de impact van de handreiking. Met het verkregen inzicht wil de opdrachtgever een strategie opzetten om de impact van de handreiking te vergroten. Tevens wil de opdrachtgever de resultaten gebruiken bij het door BZK op te stellen implementatiekader voor algoritmen. Daarnaast kan het onderzoek een bijdrage leveren aan de bewustwording bij departementen over de komst van de Europese AI-Verordening en het (nationale) implementatiekader.

De centrale onderzoeksvraag luidt:
In welke mate is de handreiking non-discriminatie by design bekend en wordt deze gebruikt en wat zijn mogelijkheden om de impact van de handreiking te vergroten?

Hiertoe zijn de volgende deelvragen beantwoord:
1. Welk inzicht kan gegeven worden in de bekendheid en het gebruik van de handreiking?
2. Welk inzicht kan gegeven worden in de toepasbaarheid van de handreiking?
3. Wat is de toegevoegde waarde van de handreiking (hoe verhoudt deze zich tot alternatieve instrumenten)?
4. Welke mogelijkheden zijn er om de impact van de handreiking te vergroten?

5.2 Object van onderzoek, afbakening en definities

Object van onderzoek
Object van onderzoek is het gebruik van de handreiking 'non-discriminatie by design' door organisaties van de rijksoverheid. Daarbij focussen wij ons op de factoren: bekendheid, toepasbaarheid en toegevoegde waarde van de handreiking.

Afbakening
Met het onderzoek wordt geen rijksbreed dekkend inzicht gegeven. In het onderzoek richten wij ons op een beperkt aantal organisaties (acht), die in overleg met de opdrachtgever zijn geselecteerd op basis van de volgende drie criteria:
- uitvoeringsorganisaties of toezichthoudende organisaties die met hun werkzaamheden een directe werking hebben op burgers,
- intensief werken met persoonsgegevens,
- werken met algoritmen.

De inzichten in de rapportage zijn veelal gebaseerd op percepties van geïnterviewde functionarissen die betrokken zijn bij de ontwikkeling van AI-systemen, zoals onder andere projectleiders die sturing geven aan systeembouwers, data-analisten en AI-experts.

Dit onderzoek geeft *geen* inzicht in de vraag of de AI-systemen van de geselecteerde organisaties non-discriminatoir zijn. Indien de handreiking niet bekend is en/of niet toegepast wordt, hoeft dat dus niet te betekenen dat discriminatieregels worden overtreden.

Het onderzoek beperkt zich, voor wat betreft alternatieve instrumenten, tot door geïnterviewden genoemde instrumenten als alternatief voor de handreiking. We zijn niet zelf op zoek gegaan naar mogelijke alternatieven.

Definities
Onder 'non-discriminatie by design' verstaan wij in dit onderzoek het identificeren, voorkomen en bestrijden van discriminatie bij het ontwikkelen van een algoritme.

Onder de toegevoegde waarde van de handreiking verstaan wij de ervaren relevantie, de meerwaarde ten opzichte van alternatieve instrumenten en nut en noodzaak van de handreiking.

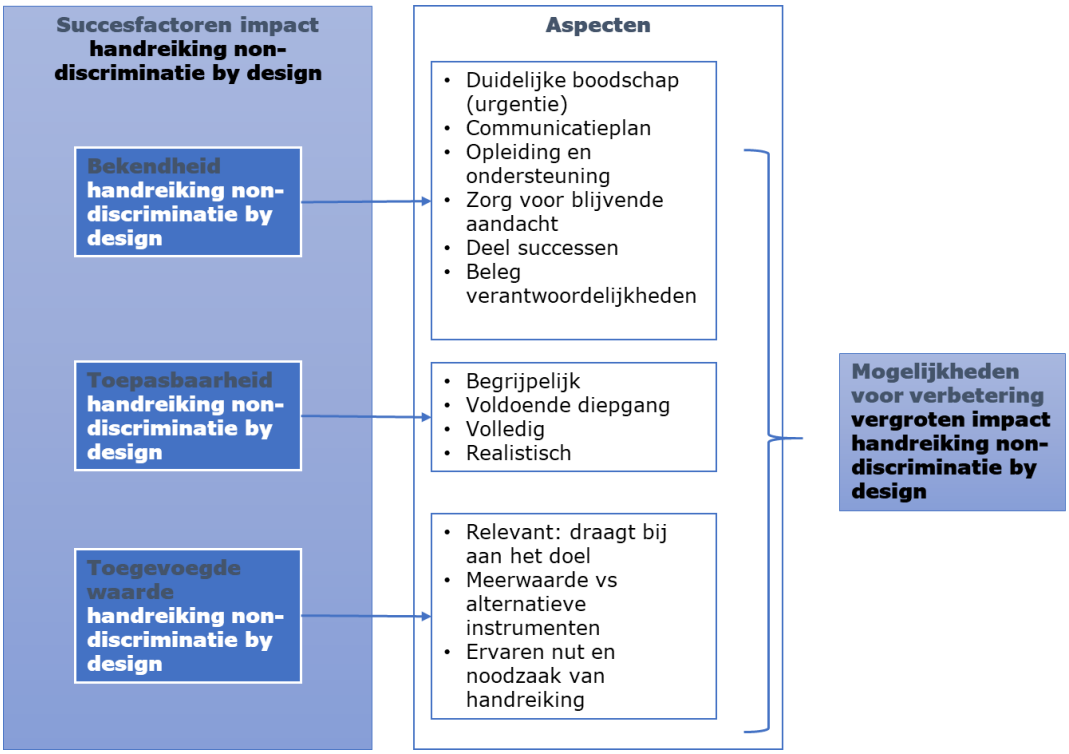
Onder de impact van de handreiking verstaan wij het effect van de handreiking op het ontwikkelen van zo non-discriminatoir mogelijke algoritmen.

5.3 Onderzoeksaanpak

Als referentiekader dient de handreiking 'non-discriminatie by design'.

Naast de handreiking hebben we gewerkt met een conceptueel model (zie onderstaande figuur). Het model geeft de samenhang weer tussen de succesfactoren die bepalend zijn voor de impact van de handreiking, de te onderzoeken aspecten per succesfactor en de verbetermogelijkheden om de impact van de handreiking 'non-discriminatie by design' te vergroten.

Figuur: conceptueel model onderzoek



Werkzaamheden
Voor het onderzoek is een documentstudie en een interview met twee medewerkers van BZK uitgevoerd gericht op de handreiking en activiteiten die waren ondernomen om de handreiking te verspreiden onder de doelgroep (potentiële gebruikers).

Daarna zijn interviews gehouden met twee medewerkers (in twee interviews met met één medewerker) van de volgende acht uitvoeringsorganisaties van de Rijksoverheid: Belastingdienst, Dienst Toeslagen, Centraal Justitieel Incassobureau, Nederlands Forensisch Instituut, Politie, Rijksdienst voor Identiteitsgegevens, Dienst Uitvoering Onderwijs en UWV.

Hiermee zijn de overeengekomen werkzaamheden, zoals vastgelegd in de opdrachtbevestiging, uitgevoerd.

De bevindingen zijn middels hoor en wederhoor afgestemd met de (contactpersoon van) de opdrachtgever.

5.4 Gehanteerde Standaard

Deze opdracht is uitgevoerd in overeenstemming met de Internationale Standaarden voor de Beroepsuitoefening van Internal Auditing. Dit onderzoek verschaft geen zekerheid in de vorm van een oordeel of conclusie, omdat het een onderzoeksopdracht betreft en geen controle-, beoordelings- of andere assurance-opdracht. Als hier wel sprake van was geweest, dan zouden we wellicht andere zaken hebben geconstateerd en gerapporteerd.

De opdracht is uitgevoerd conform de algemene uitgangspunten voor de uitoefening van de interne auditfunctie bij de rijksdienst. Daarbij hoort ook een stelsel van kwaliteitsborging. Een onderdeel daarvan is dat er een onafhankelijke kwaliteitstoetsing heeft plaatsgevonden op deze onderzoeksopdracht.

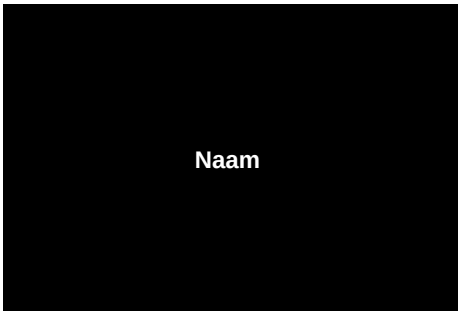
5.5 Verspreiding rapport

De opdrachtgever, CIO Rijk, is eigenaar van dit rapport. Dit rapport is primair bestemd voor de opdrachtgever met wie wij deze opdracht zijn overeengekomen. Hoewel het rapport de context van het onderzoek zo goed mogelijk probeert te beschrijven, is het mogelijk dat iemand die de context niet (volledig) kent, de uitkomsten anders interpreteert dan bedoeld.

De ADR is de interne auditdienst van het Rijk. Dit rapport is primair bestemd voor de opdrachtgever met wie wij deze opdracht zijn overeengekomen. Voor openbaarmaking door het opdrachtgevende ministerie van door de ADR aan dit ministerie uitgebrachte rapporten gelden de voorschriften uit de Wet open overheid. De minister van Financiën stuurt elk halfjaar een overzicht van door de ADR uitgebrachte rapporten naar de Tweede Kamer.

6 Ondertekening

Den Haag, 26 juni 2023



Naam

Projectleider
Auditdienst Rijk

Bijlage 1: de meest genoemde alternatieve instrumenten voor de Handreiking

IAMA
IAMA[18] staat voor 'Impact Assessment voor Mensenrechten bij de inzet van Algoritmes'. IAMA is in 2021 ontwikkeld op verzoek van het ministerie van BZK en bevat net als de handreiking een groot aantal vragen. Over die vragen moet dan discussie plaats vinden en vervolgens een antwoord worden geformuleerd in alle gevallen waarin een overheidsorgaan overweegt een algoritme in te gaan zetten. Ook wanneer een algoritme al wordt ingezet kan het IAMA dienen als instrument voor reflectie. De discussie over de verschillende vragen moet, evenals bij de handreiking, plaatsvinden in een breed samengesteld team waarin mensen met verschillende specialisaties en achtergronden zitting hebben. Per vraag is in het IAMA aangegeven wie in ieder geval bij de discussie moet zijn betrokken. Alle in het schema opgenomen functies of rollen binnen een multidisciplinair team komen in dit instrument aan bod. Veel voorkomende functies hebben hierin een plaats gekregen, maar de lijst is niet uitputtend. Ook de benamingen van de functies kunnen per organisatie verschillen.

Het is de bedoeling dat per vraag de antwoorden en de belangrijkste overwegingen en gemaakte keuzes worden vastgelegd. Het ingevulde IAMA kan dienen als naslagwerk en ter verantwoording van het besluitvormingsproces rondom de ontwikkeling en de implementatie van een algoritme. IAMA is dus een meer integraal kader dat kan worden gebruikt bij de (voorbereiding van de) juiste vragen. IAMA legt verbanden tussen relevante regels, instrumenten en toetskaders op het gebied van algoritmen. In IAMA wordt ook verwezen naar de 'handreiking non-discriminatie by design'.

DPIA of GEB
DPIA staat voor Data Protection Impact Assessment. Het behandelt het doel, de noodzaak en proportionaliteit van data, oftewel het begrenzen van het datagebruik voor de AI-applicatie.
Een DPIA wordt ook wel een gegevensbeschermingseffectbeoordeling (GEB) genoemd. Het is een instrument om vooraf de privacyrisico's van een gegevensverwerking in kaart te brengen. En om daarna maatregelen te kunnen nemen om de risico's te verkleinen. In de AVG (Algemene Verordening Gegevensbescherming), Wpg (Wet politiegegevens) en Wjsg (Wet justitiële en strafvorderlijke gegevens) is op hoofdlijnen aangegeven wanneer een DPIA verplicht is. Dat is het geval als een gegevensverwerking waarschijnlijk een hoog privacyrisico oplevert voor de mensen van wie de organisatie gegevens verwerkt. Een DPIA bestaat in ieder geval uit een systematische beschrijving van de beoogde gegevensverwerking, een beoordeling van de privacyrisico's en de te treffen maatregelen om de risico's aan te pakken.

Toetsingskader algoritmen van de Algemene Rekenkamer
Het toetsingskader algoritmen van de Algemene Rekenkamer (AR) is volgens de website van de AR een praktisch en integraal instrument dat overheidsorganisaties kunnen gebruiken om te toetsen of algoritmen aan bepaalde kwaliteitscriteria voldoen én of de risico's voldoende in beeld zijn en/of worden beperkt.

18 Impact Assessment Mensenrechten en Algoritmes | Rapport | Rijksoverheid.nl

Het toetsingskader gaat uit van 5 perspectieven. Ethiek (ethische richtlijnen) is als perspectief verbonden met de andere 4 perspectieven: sturing en verantwoording (eenduidig doel), model en data (in lijn met doelstellingen), privacy (onder meer wettelijke verplichting verwerkingsregister) en ITGC (toegankelijke loginformatie). De perspectieven geven een vertaling van normenkaders en richtlijnen; het toetsingskader maakt gebruik van beschikbare informatie, kaders en raamwerken.

Ethische richtlijnen voor betrouwbare KI van de EC
De Europese Unie (EU) heeft in 2019 haar richtlijnen voor een ethisch verantwoorde toepassing van kunstmatige intelligentie (KI) gepubliceerd. In het document staan zeven grondbeginselen centraal waaraan "betrouwbare" kunstmatige intelligentie voldoet. De publicatie van de richtlijnen is onderdeel van de in april 2018 door de EU vastgestelde strategie voor kunstmatige intelligentie, die de mens vooropstelt. De niet-bindende richtlijnen zijn samengesteld door de High-Level Expert Group on Artificial Intelligence (AI HLEG) en dienen als aanvulling op bestaande wet- en regelgeving.
Aan de hand van de zeven grondbeginselen kunnen bedrijven en overheden beoordelen of zij kunstmatige intelligentie ethisch toepassen:

1. Menselijk handelen en toezicht – Kunstmatige intelligentie maakt rechtvaardige maatschappijen mogelijk door het menselijke handelen en fundamentele rechten te ondersteunen en de menselijke autonomie niet te verminderen, begrenzen of manipuleren.
2. Robuustheid en veiligheid – Algoritmen zijn veilig, betrouwbaar en robuust genoeg om fouten of inconsistenties op te lossen, gedurende de hele levenscyclus van KI-systemen.
3. Privacy- en databeleid – Burgers hebben de volledige controle over hun eigen data en data die betrekking op hen heeft, wordt niet gebruikt om kwaad te doen of te discrimineren.
4. Transparantie – De elementen waaruit een KI-systeem is opgebouwd, zijn inzichtelijk en data, algoritmen en uitkomsten zijn herleidbaar en uitlegbaar.
5. Diversiteit, non-discriminatie en rechtvaardigheid – KI-systemen zijn toegankelijk voor ieder en beschouwen het hele scala aan menselijke vermogens, vaardigheden en behoeften.
6. Maatschappelijk en ecologisch welzijn – KI-systemen versterken positieve sociale verandering en vergroten duurzaamheid en ecologische verantwoordelijkheid.
7. Verantwoording – Procedures zijn ingebouwd die zorgdragen voor de verantwoordelijkheid en aansprakelijkheid van KI-systemen en hun beslissingen.

Handbook Fairness
Het in opdracht van de gemeente Amsterdam opgestelde Engelstalige Fairness Handboek biedt een set instrumenten waarmee de rechtvaardigheid (Fairness) van een model kan wordenbeoordeeld. Het handboek geeft een basisuitleg over algoritmische rechtvaardigheid en mogelijke bias (vooroordelen). Het is bedoeld voor iedereen die in het werk te maken heeft met data of algoritmen. Het legt uit hoe vooroordelen en andere problemen in de ontwikkelcyclus van het model verschillende vormen van schade kunnen veroorzaken. Die kunnen daardoor invloed hebben op individuen of benadeelde groepen in de maatschappij. Zowel de in het handboek opgenomen biasanalyse als de Fairness Pipeline (rechtvaardigheidspijpleiding) bieden een stappenplan. Daarmee kan een model worden geëvalueerd op vooroordelen waarbij eventuele problemen kunnen worden geminimaliseerd.

Begeleidingsethiek methode
ECP[19] heeft in samenwerking met prof. dr. ir Peter-Paul Verbeek de aanpak Begeleidingsethiek ontwikkeld. Het is een methode die concrete handvatten biedt om technologie op een ethisch verantwoorde manier toe te passen. De aanpak Begeleidingsethiek ziet ethiek niet als beoordelaar, maar als ethische begeleider van de introductie van technologie in de samenleving.
De aanpak bestaat uit een workshop waarin verschillende betrokkenen de dialoog met elkaar voeren over de toepassing van een concrete technologie in een specifieke context. Gezamenlijk komen ze tot concrete handelingsopties. Die handelingsopties zorgen ervoor dat het niet alleen bij praten blijft, maar dat de betreffende technologie op een waarden-volle manier ingebed wordt in de dagelijkse gang van zaken binnen de organisatie of binnen een netwerk.

Het ADR normenkader voor algoritmen
De ADR heeft een normenkader ontworpen om de beheersing van een algoritme door een organisatie te onderzoeken. Het kader is ontwikkeld met behulp van nationale en internationale richtlijnen en rapporten, onder andere de (concept) richtlijnen van het ministerie van JenV, de AI impact assessment van het ECP, de ethics guidelines for trustworthy AI van de EC en het Privacy Control Framework van NOREA. Daarnaast is het kader afgestemd met (concept) toetsingskaders van andere partijen.

19 ECP: Electronic Commerce Platform Nederland is een platform voor een informatiesamenleving van bedrijfsleven, overheid en maatschappelijke organisaties en heeft tot doel het gebruik van ICT in de Nederlandse samenleving te versterken

Bijlage 2: Managementreactie

Koninkrijksrelaties

> Retouradres 2511 DP Den Haag

Auditdienst Rijk
[redacted]
Postbus 20201
2500 EE Den Haag

DGOO
Digitale Samenleving

Turfmarkt 147
Den Haag
Nederland
www.rijksoverheid.nl
www.facebook.com/minbzk
www.twitter.com/minbzk
www.linkedin.com/company/ministerie-van-bzk

Contactpersoon
[redacted]

Datum 5 juni 2023
Betreft Reactie op rapport Onderzoek naar de bekendheid en toepasbaarheid handreiking 'nondiscriminatie by design'

Kenmerk

Uw kenmerk

Bijlage(n)
1

Geachte mevrou[redacted]

Om overheidsorganisaties te ondersteunen bij een verantwoorde inzet van algoritmes zijn er door het vorige kabinet hulpmiddelen en instrumenten ontwikkeld waaronder een handreiking 'non-discriminatie by design'.

De handreiking is in 2021 ontwikkeld en gepubliceerd op rijksoverheid.nl. Deze handreiking is bedoeld om normen die voortvloeien uit wet- en regelgeving op het gebied van non-discriminatie inzichtelijk te maken die relevant zijn bij ontwerp en implementatiekeuzes van algoritmes en/of AI. Het moet ontwikkelaars helpen om de juiste vragen te stellen en om discriminatie zoveel mogelijk te voorkomen.

De Tweede Kamer heeft op 28 oktober 2021 gevraagd om 1 jaar na de publicatie van de handreiking een onderzoek te doen in hoeverre de handreiking bekend is bij overheidsdiensten. Het ministerie van Binnenlandse Zaken en Koninkrijksrelaties heeft de ADR gevraagd om dit onderzoek te doen.

Ik dank de ADR hiervoor. Er is aan de hand van diepte-interviews nu meer bekend over de ervaringen met het gebruik van de handreiking. Op 28 februari j.l. deed ik een uitvraag onder de CIO's van het Rijk om ook een kwantitatief inzicht te krijgen in de bekendheid met de handreking. Deze inzichten en de ervaringen van de ADR worden tezamen benut voor het ontwikkelen van een implementatiekader 'inzet van algoritmes'. Dit sluit goed aan bij de een van de aanbevelingen die de ADR heeft gedaan.

Daarnaast vormen de onderzoeksresultaten input voor de strategische publiekscampagne om kennis en ervaringen met het voorkomen van bias en/of discriminatie in algoritmische systemen te voorkomen. De ADR laat bijvoorbeeld het belang zien om met meer concrete voorbeelden te werken.

In de verzamelbrief 'algoritmes' die naar verwachting eind juni naar de Tweede Kamer wordt verstuurd, zal de Staatssecretaris van Binnenlandse Zaken en Koninkrijksrelaties, *Digitalisering en Koninkrijksrelaties* aangeven hoe de resultaten van het onderzoek verwerkt worden in toekomstig beleid.

DGOO
Digitale Samenleving

Datum
5 juni 2023

Kenmerk

[redacted]

Directie CIO Rijk | DG Digitalisering en Overheidsorganisatie

Auditdienst Rijk
Postbus 20201
2500 EE Den Haag
(070) 342 77 00