Ministerie van Binnenlandse Zaken
en Koninkrijksrelaties

Toelichtingsdocument

Impact Assessment Mensenrechten en Algoritmes

IAMA Versie 2 | 2026

Prof. mr. Janneke Gerards, Iris Muis, Julia Straatman, Arthur Vankan en Max Boiten.

IAMA Versie 1 | 2021

Prof. mr. Janneke Gerards, Dr. Mirko Tobias Schäfer, Arthur Vankan en Iris Muis.

Het IAMA is ontwikkeld door Universiteit Utrecht in opdracht van het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties.

Inhoud

Inleiding

Het IAMA-toelichtingsdocument kan gebruikt worden ter verheldering en/of ter verdieping van het IAMA-proces. Allereerst worden er in dit document praktische handvatten gegeven, zoals een houvast welke functie er in welk onderdeel van het IAMA uitgenodigd kan worden en een instructie voor het gebruik van het IAMA. Daarna volgt een toelichting over de relatie van het IAMA tot de Europese AI-verordening, evenals een aantal vaak gebruikte definities. Ten derde biedt dit document per IAMA-subonderdeel een uitleg over de IAMA-vragen en de daarin gebruikte termen. Waar relevant wordt de relatie tot artikel 27 van de AI-verordening duidelijk gemaakt. Tot slot biedt dit document een overzicht van de verscheidene grondrechtenclusters, die als houvast kunnen dienen bij het beantwoorden van de vragen van deel 4.

Het huidige IAMA v2 is een actualisatie van de oorspronkelijke versie. Het is gestroomlijnd op basis van gebruikersfeedback en in lijn gebracht met de vereisten vanuit artikel 27 van de Europese AI-verordening.

Eerder ingevulde IAMA's (v1) hoeven niet herzien te worden.

Uitvoeringstips

Voorwerk

Om het IAMA-proces soepel te laten verlopen, kan het handig zijn om vooraf wat zaken paraat te hebben.
Loop daarvoor de volgende punten door:

1. Is er al documentatie over de casus beschikbaar?
In het IAMA wordt wellicht iets gevraagd wat al beschikbaar is in andere documentatie. In dat geval kan er gekozen worden om de tekst – na het met het projectteam goed gecontroleerd te hebben op juistheid en passendheid – te knippen en plakken in het IAMA-document om dubbel werk te voorkomen. Ook kan het soms volstaan om in het antwoord op een IAMA-vraag louter te verwijzen naar dit andere document. Let er hierbij op dat alleen doorverwezen wordt naar documentatie als deze ook beschikbaar is voor de doelgroep of de uiteindelijke lezers van het ingevulde IAMA. Bijvoorbeeld: wordt het ingevulde IAMA uiteindelijk gepubliceerd buiten de organisatie, maar staan er verwijzingen in naar documenten die alleen binnen de organisatie te raadplegen zijn? Vervang de verwijzingen dan met lopende tekst (al dan niet overgenomen van de brondocumentatie, uiteraard met bronvermelding).

2. Zijn de doelstelling en algemene werking van het algoritme helder en begrijpelijk gedocumenteerd?
Als dit niet zo is, zal het team al snel vastlopen in het IAMA-proces. Zorg ervoor dat dit bekend is voordat wordt gestart met het IAMA.

3. Is in de casus sprake van een hoog-risico AI-systeem in de zin van de AI-verordening?
Als de casus een hoog-risico AI-systeem betreft in de zin van de AI-verordening, is er bijzondere aandacht vereist voor verscheidene IAMA-vragen. Meer uitleg hierover is te vinden onder het kopje ‘Het IAMA en de AI-verordening’. Daarnaast staat bij verscheidene IAMA-vragen een icoontje, waarmee aangegeven wordt dat bijzondere aandacht voor artikel 27 van de AI-verordening is vereist.

4. Is er een verantwoordelijke voor het IAMA-proces aangewezen?
Wie is verantwoordelijk voor het uitvoeren van het IAMA, ziet toe op het uitvoeren van de actiepunten en mitigerende maatregelen en hoe is de actualisatie van het document ingeregeld? Als dit van tevoren helder is, kan dit vertraging voorkomen.

5. Worden de juiste mensen aan tafel gevraagd?
Op deze vraag wordt hieronder verder ingegaan.

IAMA-functietabel

Om een houvast te creëren, is hieronder schematisch aangegeven uit welke functies per IAMA-onderdeel in elk geval mensen uitgenodigd kunnen worden om deel te nemen aan het IAMA-traject. Veel voorkomende functies hebben hierin een plaats gekregen, maar de lijst is niet uitputtend. Ook kunnen de benamingen van de functies per organisatie verschillen.
Let ook op: deze lijst is niet bedoeld als gouden standaard, maar geeft louter een indicatie.

Legenda:

- Lichtgroen = essentieel
- Lichtpaars = gewenst
- Lichtgeel = optioneel

Functies:	**Deel 1**	**Deel 2**	**Deel 3**	**Deel 4**
Projectleider of product owner	•	•	•	•
Data Scientist*	•	•	•	•
Jurist (kan ook een privacy officer of Functionaris Gegevensbescherming zijn)	•	•	•	•
Indien aanwezig in de organisatie: Strategisch adviseur ethiek	•	•	•	•
Domeinexpert (medewerker met kennis over het domein waarin het algoritme toegepast gaat worden)	•	•	•	•
Indien van toepassing: externe ontwikkelaar algoritme	•	•	•	•
Opdrachtgever	•	•	•	•
Externe betrokkenen (zoals belangengroepen of burgervertegenwoordiging	•		•	•
CISO		•		

zorg dat er minimaal 2 mensen aanwezig zijn die de techniek achter het algoritme snappen en kunnen bevragen.

Essentieel is de betrokkenheid van een jurist die uitzoekwerk kan doen op het gebied van grondrechten (voornamelijk deel 4). Daarnaast is het van belang dat er altijd minimaal twee mensen aanwezig zijn die de techniek achter het algoritme snappen en kunnen bevragen. Alleen op die manier kan er een wezenlijke dialoog ontstaan en leunt het projectteam niet op de input van slechts één persoon.

In de tabel hierboven worden ook externe betrokkenen genoemd. Hierbij kan gedacht worden aan een stakeholderpanel, belangengroep of burgervertegenwoordiging. Deze zouden direct bij de toepassing van het IAMA betrokken kunnen worden, maar zouden ook op een parallel spoor input kunnen geven voor de grondrechtenbeoordeling in het IAMA.
Ook het Stakeholder Engagement Process (SEP) uit het HUDERIA zou ingezet kunnen worden.

Gebruiksinstructies en praktische uitvoering

Op verschillende punten in het IAMA-vragendocument zijn gebruiksinstructies opgenomen. In aanvulling daarop kunnen de volgende suggesties de praktische uitvoering van het IAMA te vergemakkelijken:

- Laat het IAMA niet door één persoon uitvoeren, maar voer dit uit met een team. (*Zie de functietabel voor houvast voor welke functies wanneer uitgenodigd kunnen worden*)
- Voer het IAMA uit in een vergaderzaal met scherm, waarop een persoon diens laptop kan aansluiten met daarop het IAMA-vragendocument. Wanneer er gediscussieerd wordt over de IAMA-vragen, typt iemand het uiteindelijke antwoord in het IAMA-vragendocument, zichtbaar voor alle aanwezigen. Op die manier is de gehele groep zich bewust van het genoteerde antwoord en kan het door de groep geaccordeerd worden voordat wordt doorgegaan wordt naar de volgende vraag.
- Het zal voorkomen dat er een vraag niet (volledig) te beantwoorden is. Maak hier dan een actiepunt van in de bijbehorende actiepunt velden en noteer hier een actiehouder bij.
- Voer het IAMA uit in verschillende sessies, verspreid over meerdere dagen met idealiter één of twee weken daartussen. Een voorbeeld van een sessiestructuur is: twee sessies van drie uur, of drie sessies van twee uur. Probeer het IAMA niet in één dag uit te voeren.
- Wijs een gespreksleider aan die kan sturen op timemanagement en die geen inhoudelijk belang heeft bij de casus.
- Voor subonderdeel 4.2 zal een jurist waarschijnlijk wat uitzoekwerk moeten verrichten. Het is handig om hier rekening mee te houden bij het plannen van de sessies, door bijvoorbeeld subonderdeel 4.1 te betrekken bij een eerdere sessie en de daaropvolgende sessie te starten vanaf subonderdeel 4.2.
- Sla het IAMA regelmatig tussentijds op.
- Voer het IAMA idealiter uit voordat het algoritme daadwerkelijk ingezet gaat worden.
- Voer het IAMA uit op dat punt in het proces dat er nog veranderingen doorgevoerd kunnen worden.
- Zorg voor goed versiebeheer en duidelijkheid over eigenaarschap en beheer van het document.

Voor meer informatie over de praktische uitvoering van een IAMA, kunt u het rapport IAMA in Actie raadplegen:
Straatman, J., Muis, I., Bössenecker, G., Koole, R., Draijer, R., van Deijck, N., & van Vledder, I. (2024). IAMA in actie: Lessons learned van 15 IAMA-trajecten bij Nederlandse overheidsorganisaties. Universiteit Utrecht & Rijks ICT Gilde. https://www.rijksoverheid.nl/documenten/rapporten/2024/06/20/iama-in-actie-lessons-learned-van-15-iama-trajecten-bij-nederlandse-overheidsorganisaties

Scope IAMA

Het IAMA is bedoeld voor **algoritmes die een medium tot hoge impact zouden kunnen hebben op grondrechten**. Deze kunnen technisch zeer variëren van aard: van simpele regelgebaseerde algoritmes tot complexe AI-systemen. De relatie van het IAMA tot de Europese AI-verordening wordt verderop in dit document uiteengezet.

De scope van het IAMA is breder dan louter de technische component van het algoritme. Bij het IAMA is **het gehele algoritmische proces** onderdeel van de beoordeling. Zaken als de doelstelling, de technische werking en ook de organisatorische context en beslissingen die op basis van het algoritme genomen worden, vallen binnen scope.

Mochten er bij de casus problemen geïdentificeerd worden die er in de situatie voorafgaand aan de implementatie van het algoritme ook al waren, dan rijst nog wel eens de vraag of die problemen ook meegenomen dienen te worden in het IAMA-proces. Het is van belang om bij het IAMA **álle problemen mee te nemen die bij het algoritme spelen**, ook als deze problemen er al waren en niet specifiek veroorzaakt zijn door het algoritme zelf.

Notitie meerdere algoritmes

In veel toepassingen worden meerdere algoritmes gecombineerd. Denk hier bijvoorbeeld aan een algoritme dat gebruikt wordt om fraude te detecteren op basis van verschillende indicatoren. Als die indicatoren zelf het resultaat zijn van een algoritme, betekent dat dat er meerdere algoritmes gebruikt worden om tot één uitkomst te komen. De vragen in deel 2 van het IAMA zijn opgesteld om een enkel algoritme te bevragen. Is er sprake van een combinatie van algoritmes, dan is het waardevol om alle algoritmes in het systeem te bespreken. Daarvoor worden de volgende tips meegegeven:

- **Vul deel 2 per algoritme los in.** Hiervoor is het handig om meerdere documenten te gebruiken. Dat maakt het makkelijker om achteraf de antwoorden per algoritme na te slaan. Als alternatief kan in één document gewerkt worden. Signaleer dan heel duidelijk in de beantwoording van de vragen welk antwoord over welk algoritme gaat. NB: vul de overige IAMA-delen (deel 1, deel 3 en deel 4) niet los in per algoritme. Neem hierbij de volledige scope van het systeem als geheel.
- **Bespreek de algoritmes één voor één.** Bespreek steeds alle vragen van deel 2 voor één algoritme en begin dan pas aan de volgende. Dat voorkomt verwarring en zo krijgt elke vraag voor elk algoritme de aandacht die het verdient.
- **Begin bij het laatste algoritme in de keten.** De vragen in deel 2 maken de koppeling tussen techniek van het algoritme en het beschreven doel. Het 'laatste' algoritme is het algoritme dat het dichtst bij het eindresultaat – en dus het bereiken van het doel – staat. De andere algoritmes leveren bijvoorbeeld input aan dit 'laatste' algoritme of werken in een keten naar de output toe. Als het laatste algoritme dan al goed beschreven is, volstaat bij de voorafgaande algoritmes om te beschrijven hoe zij bijdragen aan het doel.

Notitie General Purpose AI (GPAI)

General Purpose AI (GPAI) is AI die voor meerdere doelen gebruikt kan worden. Doorgaans gaat het hierbij om generatieve AI-modellen zoals chatbots. Deze modellen kunnen voor veel toepassingen worden ingezet, wat het ook moeilijker kan maken om een goed impact assessment uit te voeren. Voor een goed impact assessment is het noodzakelijk om het doel duidelijk te definiëren, zodat bepaald kan worden wanneer het doel behaald is en hoe effectief het voorgestelde algoritme daaraan bijdraagt. Een toepassing als een chatbot kan voor veel verschillende doeleinden worden gebruikt, waardoor het soms moeilijk is om het precieze doel te formuleren op dusdanige wijze dat getoetst kan worden of het algoritme aan dit doel bijdraagt.

We raden daarom bij dergelijke toepassingen aan om voor het invullen van het IAMA concrete *use cases* te formuleren met bijbehorende doelstellingen. Dit maakt het IAMA eenvoudiger in te vullen. Houd bij rekening met de mogelijkheid om het model voor andere toepassingen te gebruiken als die mogelijkheid bestaat *(function creep*), en bespreek bij de mitigerende maatregelen wat er gedaan kan worden om het gebruik van het model te beperken tot deze 'goede' *use cases*.

General Purpose AI zal daarnaast in bijna alle gevallen een model van een externe leverancier zijn. In dat geval is het nodig om documentatie op te vragen bij de aanbieder/ontwikkelaar om de IAMA-vragen goed te kunnen beantwoorden. Indien mogelijk kan de aanbieder/ontwikkelaar ook zelf aansluiten bij het IAMA-proces, in ieder geval bij de vragen over de techniek in deel 2.

Notitie generatieve AI

Generatieve AI is bezig met een sterke opmars. Modellen die tekst, afbeeldingen of video genereren zijn voor de meeste mensen al onderdeel van het dagelijks leven. Het IAMA is ook geschikt voor het beoordelen van toepassingen waarin deze modellen worden ingezet, maar in sommige gevallen maakt het de beantwoording van de vragen ingewikkelder. De uitdagingen die dat brengt, worden daarom verder toegelicht.

Bij generatieve AI wordt meestal gebruik gemaakt van een voorgetraind model. Die modellen worden getraind door leveranciers, bijvoorbeeld OpenAI/Microsoft of Mistral AI. In beide gevallen is het model kant-en-klaar: het is al getraind en wordt als zodanig in een toepassing ingezet. Dit geeft geen controle over het ontwerp van het model en vaak is het ook niet duidelijk welke trainingsdata precies zijn gebruikt. Het is ook niet te herleiden hoe het model tot een uitkomst komt. Het model kan daardoor bias introduceren in de toepassing en die bias is moeilijk vooraf in te schatten.

Dit kan bijvoorbeeld gebeuren doordat het model bestaande patronen voortzet. Europese of Amerikaanse modellen die afbeeldingen genereren, beelden bijvoorbeeld vaker witte mensen af. Verpleegkundigen zullen er doorgaans als vrouw uitkomen en bouwvakkers als man. Op diezelfde manier zullen taalmodellen zoals ChatGPT bestaande vooroordelen herhalen. Het versterkt zo bestaande maatschappelijke patronen, omdat die nu eenmaal vaker voorkomen in de trainingsdata.

Dit betekent niet dat generatieve AI niet gebruikt kan worden. Het vraagt wel om een ander soort reflectie, omdat het bestaan van bias in het model een gegeven is. De uiteindelijke reflectie moet dan plaatsvinden op de output: Welke vormen van bias treden op? Wat kan worden gedaan om dat te ondervangen?

Bij de beantwoording van de vragen van het IAMA zal deze verschuiving van aandacht ook terugkomen. Sommige vragen uit deel 2 zijn makkelijker te beantwoorden en andere moeilijker. In de aanwijzingen en toelichtingen van deel 2 staat steeds apart aangegeven hoe de beantwoording mogelijk anders wordt bij de inzet van generatieve AI.

Het IAMA en de AI-verordening

De Europese AI-verordening is gefaseerd ingegaan vanaf 2 februari 2025. Dit betekent dat verschillende transitieperiodes zijn voorzien en niet alle verplichtingen meteen zijn ingegaan. Toch is het wenselijk om alvast na te denken over de relevantie en toepassing van deze verordening. Bepaalde onderdelen van de AI-verordening kunnen immers vergen dat een toepassing substantieel moet worden aangepast. Als de transitieperiode voor dat onderdeel verstrijkt, moet de toepassing meteen daarna aan de verplichtingen voldoen. Het loont dan om het werk al eerder gedaan te hebben.

Eén van de verplichtingen uit de AI-verordening is om bij hoog-risico AI-systemen een FRIA (Fundamental Rights Impact Assessment) uit te voeren, op grond van artikel 27. Hoewel er op het moment van schrijven geen formeel template beschikbaar is voor de FRIA, bevat artikel 27 een aantal eisen waaraan een dergelijk FRIA moet voldoen. Deze eisen zijn geïntegreerd in het IAMA en gemarkeerd met een **'artikel 27-icoontje'**. 

Om de impact van de AI-verordening goed te kunnen inschatten is het van belang om een aantal vragen te beantwoorden:

- Is bij een toepassing sprake van *artificial intelligence (AI)*? Het antwoord op deze vraag bepaalt of de AI-verordening van toepassing is op het algoritme.
- In welke risicoklasse van de AI-verordening valt het algoritme? Het antwoord op deze vraag bepaalt aan welke eisen uit de AI-verordening het algoritme moet voldoen.
- Welke rol heeft de organisatie ten opzichte van het algoritme? Het antwoord op deze vraag bepaalt welke verantwoordelijkheden gelden bij productie en inzet van het algoritme.

Op deze vragen wordt hieronder verder ingegaan.

Definities: algoritme, modellen en AI

Bij het invullen van het IAMA lopen soms verschillende termen door elkaar. Er kan worden gesproken van algoritmes, modellen of AI. De algemene term die in het IAMA wordt gebruikt, is 'algoritme'. Hoewel beide andere termen specifieker zijn, is bij zowel AI als modellen ook sprake van een algoritme. Het Algoritmeregister van de Nederlandse overheid hanteert de algoritme-definitie van de Algemene Rekenkamer, die een algoritme definieert als "sets van regels en instructies die een computer geautomatiseerd volgt bij het maken van berekeningen om een probleem op te lossen of een vraag te beantwoorden" (Algemene Rekenkamer, 2021). Hieronder valt alles waarbij een computer een vorm van logica volgt, of die logica nu heel eenvoudig is of heel complex.

Simpele algoritmes kunnen beslisbomen of beslisregels zijn. Bij dit soort algoritmes volgt uit een bepaalde input een vooraf vastgestelde output. Modellen zijn algoritmes waarbij de patronen geleerd worden uit data, bijvoorbeeld statistische modellen of *machine learning*. Hierbij is er ook een set van regels en instructies, maar leert de computer zelf hoe die te gebruiken. De term 'model' wordt alleen gebruikt in deel 2 van het IAMA, namelijk in de vragen die uitdrukkelijk betrekking hebben op modellen. In alle andere situaties kunnen de termen inwisselbaar worden gebruikt.

Onder artificiële intelligentie (AI) worden soms zeer complexe modellen verstaan, zoals neurale netwerken. In de praktijk wisselt het echter wat mensen onder AI scharen, en verdere informatie vanuit de EU moet hier helderheid over bieden. In het IAMA wordt de term 'AI' alleen gebruikt waar de AI-verordening mogelijk op het algoritme van toepassing is.

In de AI-verordening wordt een AI-systeem gedefinieerd als

> *Een op een machine gebaseerd systeem dat is ontworpen om met verschillende niveaus van autonomie te werken en dat na het inzetten ervan aanpassingsvermogen kan vertonen, en dat, voor expliciete of impliciete doelstellingen, uit de ontvangen input afleidt hoe output te genereren zoals voorspellingen, inhoud, aanbevelingen of beslissingen die van invloed kunnen zijn op fysieke of virtuele omgevingen.* (art. 3 lid 1)

Centraal in deze definitie staat dat het model uit trainingsdata heeft geleerd om een voorspelling te doen, beslissingen te nemen of aanbevelingen of inhoud te genereren. Hieronder valt een verscheidenheid aan statistische modellen en *machine learning* modellen. De definitie uit de AI-verordening is niet volledig uitgekristalliseerd, en het is nog onduidelijk wat er precies wel en niet onder zal kunnen vallen. Verdere informatie vanuit de EU moet hier uitsluitsel over geven.

Om te bepalen of het algoritme volgens de AI-verordening een AI-systeem is, wordt verwezen naar de beslishulp AI-verordening van het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties.

Risicoklasse

De regels die onder de AI-verordening voor een AI-systeem gelden, zijn gebaseerd op de risicoklasse waarbinnen de toepassing kan worden ondergebracht:

- **Verboden AI-systemen:** Deze toepassingen mogen niet worden aangeboden of in gebruik genomen (artikel 5 AI-verordening). Een verdere toelichting hierop volgt bij de aanwijzingen en toelichting op subonderdeel 1.3 van het IAMA.
- **Hoog-risico AI-systemen:** Deze toepassingen moeten aan strenge eisen onder de AI-verordening voldoen omdat er een hoog risico is dat grondrechten erdoor geraakt worden. Artikel 6 en bijlage III van de AI-verordening bevatten een overzicht van toepassingen die hieronder kunnen vallen.
- **Lagere risicoklassen:** Het is mogelijk dat de toepassing niet in de hoog-risico klasse valt en ook niet als verboden AI-systeem geldt. In een dergelijk geval is de FRIA-verplichting uit artikel 27 niet van toepassing. Het kan echter nog steeds raadzaam zijn om het IAMA te doorlopen als er een redelijke kans is dat grondrechten worden geraakt.

Rollen AI-verordening

De verplichtingen die bij AI-systemen horen zijn afhankelijk van de rol van degene die verantwoordelijk is voor het AI-systeem. In deze paragraaf worden twee relevante rollen uitgelicht:

De **aanbieder** (***provider***) is de natuurlijke of rechtspersoon die een AI-systeem ontwikkelt en op de markt brengt. Deze moet, in het geval van hoog-risico AI-systemen, onder andere goede technische documentatie aanleveren. In de documentatie moet onder andere zijn toegelicht voor welke toepassingen het AI-systeem bedoeld is.

De **gebruiksverantwoordelijke** (***deployer***) is de natuurlijke of rechtspersoon die een AI-systeem inzet. De gebruiksverantwoordelijke heeft minder verplichtingen dan de aanbieder, mits de inzet van het systeem in lijn is met de technische documentatie. Bij afwijking hiervan moet de gebruiksverantwoordelijke ook alle verantwoordelijkheden van de aanbieder dragen.

Hulp bij de AI-verordening

Om te bepalen of er sprake is van een AI-systeem dat onder de AI-verordening valt, wordt doorverwezen naar de beslishulp AI-verordening van het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties en naar de EU AI Act Compliance Checker.

Aanwijzingen en toelichting bij de IAMA-vragen

Als er tijdens het doorlopen van het IAMA onduidelijkheid is of meer toelichting gewenst is over de vragen, kan onderstaande toelichting geraadpleegd worden.

Deel 1: Waarom?

Aanwijzingen en toelichting 1.1

Bij de eerste vraag van dit deel gaat het erom na te denken over wat nu eigenlijk de **aanleiding** is om een algoritme in te willen zetten: wat is het probleem waarvoor de beoogde inzet van het algoritme een oplossing zou moeten vormen? Het gaat hierbij dus om **probleemdefinitie en -afbakening**. Daarbij is het essentieel om het probleem zo concreet en precies mogelijk te krijgen. Soms kan het probleem of de aanleiding een interne aangelegenheid zijn: interne processen verlopen niet efficiënt of kunnen efficiënter worden gemaakt door de inzet van een algoritme. In andere gevallen kan een algoritme worden ingezet om een maatschappelijk probleem of een probleem bij een bepaalde bevolkingsgroep op te lossen.

Voortbouwend daarop is het zaak om bij 1.1.2 **het doel van de inzet van een algoritme zo concreet en specifiek mogelijk te definiëren**. Dus niet (alleen): 'bescherming van de nationale veiligheid', maar (ook) 'het geautomatiseerd in kaart kunnen brengen en analyseren van indicatoren voor terrorismerisico's op bepaalde locaties'. Ook doelen van efficiëntie of kostenbesparing kunnen worden geïdentificeerd, bijvoorbeeld: 'geautomatiseerde tekstanalyse om de werkdruk voor medewerkers significant terug te dringen en in het volgende jaar 4 fte minder aan personeel nodig te hebben'. Waar mogelijk is het goed om de doelstellingen te kwantificeren.

Het zo expliciet mogelijk maken van de doelstelling is belangrijk, omdat in een later stadium de prestaties van het algoritme getoetst moeten kunnen worden aan de doelstelling. Een algoritme ontwikkelen vanuit goede intenties alleen is niet voldoende om het ontbreken van ongewenste effecten te waarborgen. Daarnaast moet het beoogde doel legitiem zijn.

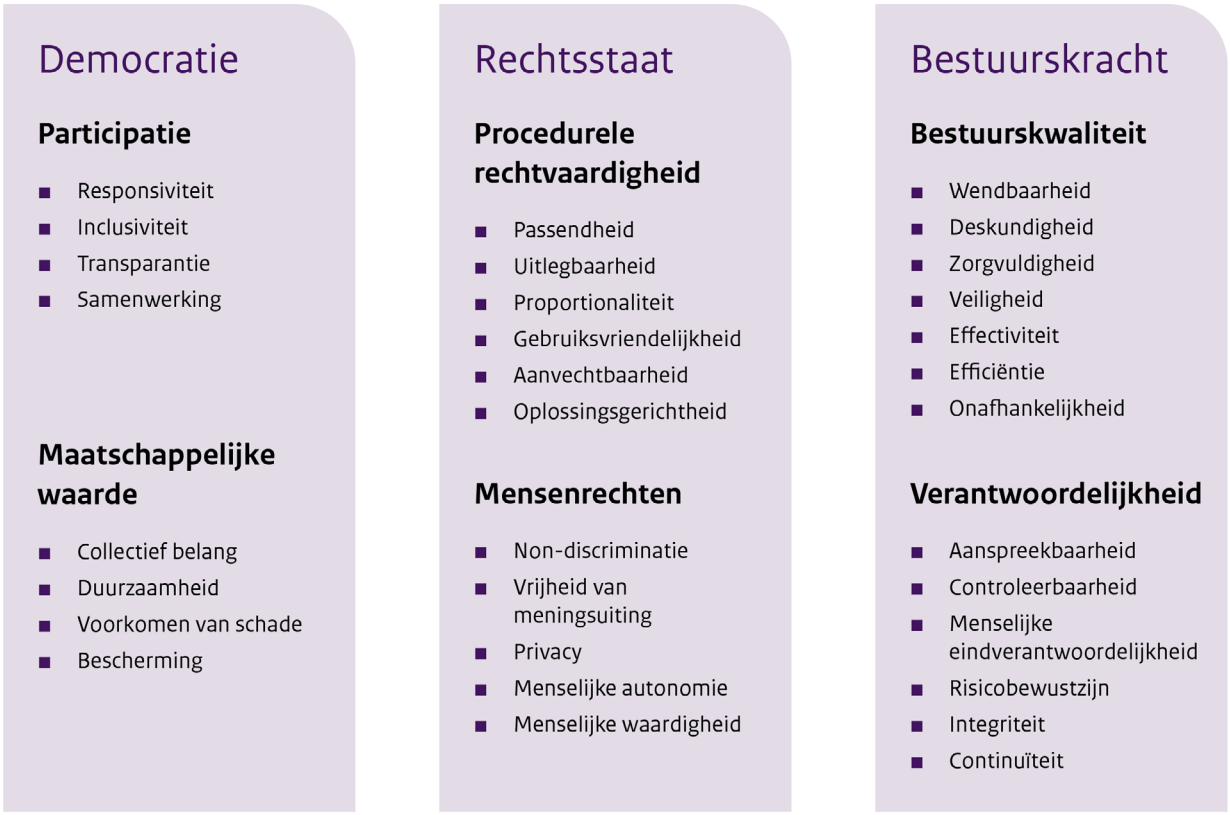
Daarnaast is het belangrijk om een zekere rangorde te maken als er meer doelen zijn (dat is bijna altijd zo): wat zijn de belangrijkste doelen en waarom? Welke doelen zijn 'subdoelen', waarvoor het niet zo erg is als ze niet voor 100% kunnen worden gerealiseerd?

Als ook een DPIA wordt uitgevoerd – bijvoorbeeld omdat verwacht wordt dat bij de ontwikkeling of inzet van het algoritme persoonsgegevens worden gebruikt – is het daarbij nodig om de doelen van gegevensverwerking te definiëren. Deze doelen kunnen overlappen met de doelen van het inzetten van een algoritme (al is dat niet noodzakelijk). Het kan dus handig zijn om de bij het IAMA gedefinieerde doelen te bewaren en ze ook in te zetten bij een DPIA, of andersom.

> **NOTITIE AI-VERORDENING**
> Indien gebruik gaat worden gemaakt van een hoog-risico AI-systeem, is het voor artikel 27 van de Europese AI-verordening van belang om het beoogde doel van het AI-systeem uiteen te zetten. Deze vraag is daarom gemarkeerd met een icoontje. AI

Aanwijzingen en toelichting 1.2

Figuur: CODIO-waardenkader



De vragen bij subonderdeel 1.2 zijn bedoeld om expliciet stil te staan bij de **publieke waarden** waar het algoritme impact op heeft. Concreet is het de bedoeling dat de waarden die de ontwikkeling en het gebruik van het betreffende algoritme ingeven, expliciet worden gemaakt. Het expliciet maken van de waarden die besloten liggen in het algoritme, kan helpen om de evaluatie van de effecten van het algoritme, in een latere fase van het proces, te vergemakkelijken. Deze vraag kan dan ook input bieden voor de beantwoording van de vragen over de impact van het algoritme in subonderdeel 3.5.

Algoritmes kunnen dienen om bepaalde publieke waarden te vertalen naar concrete besluitvorming. Daarbij kunnen algoritmes bepaalde waarden versterken, maar ze kunnen ook publieke waarden – zoals grondrechten – aantasten. Juist daarom is het belangrijk om in kaart te brengen welke publieke waarden aan de orde kunnen zijn bij de inzet van het algoritme. Het CODIO-waardenkader en het daarin gegeven overzicht van publieke waarden kan als houvast dienen bij de beantwoording van deze vragen.

Houd bij de beantwoording van deze vraag tevens rekening met de aanleiding voor de inzet van het algoritme (zie vraag 1.1.1) en de doelen ervan (zie vraag 1.1.2), evenals met de kernwaarden van de organisatie of de afdeling die het algoritme gaat gebruiken.

Aanwijzingen en toelichting 1.3

Bij vraag 1.3.1 moet eerst worden bepaald of het betreffende algoritme een AI-systeem is dat verboden is in de zin van artikel 5 van de AI-verordening. Wanneer dit het geval is, hoeft bij deze vraag niet lang te worden stilgestaan. **Wanneer het AI-systeem in de klasse van verboden AI-systemen valt, mag het niet worden ingezet**. Er kan dan op dit punt worden gestopt met het invullen van het IAMA.

Om te weten of het betreffende AI-systeem in de zin van de AI-verordening een verboden systeem is, moet artikel 5 van deze verordening worden bekeken. In dit artikel wordt benoemd welke AI-systemen in de verboden categorie vallen. Te denken valt hierbij aan de volgende systemen:

- AI-systemen die schade kunnen veroorzaken door mensen opzettelijk te misleiden of te manipuleren;
- AI-systemen die schade kunnen veroorzaken door misbruik te maken van kwetsbaarheden van mensen en/of groepen vanwege hun leeftijd, handicap of specifieke sociale of economische situatie;
- AI-systemen die gebruik maken van *social scoring* (d.w.z. het evalueren of classificeren van mensen op basis van gedrag en/of persoonskenmerken) dat leidt tot ongerechtvaardigde of onevenredige nadelige effecten voor mensen.

In artikel 5 van de AI-verordening staan meer systemen die tot de categorie verboden AI-systemen behoren, inclusief uitzonderingsgronden. Bestudeer daarom dit artikel nauwkeurig voor de beantwoording van deze vraag.

Bij vraag 1.3.2 wordt nagegaan of het algoritme al dan niet gebruikt wordt om een wettelijke taak uit te voeren en wat de bijbehorende wettelijke grondslag is. De inzet van een algoritme vindt plaats als onderdeel van een bepaald proces. Bij de beantwoording van deze vraag gaat het erom uit te zoeken of er een wettelijke grondslag bestaat die – concreet en in heldere bewoordingen – de taak of het proces formuleert dat de overheidsorganisatie dient uit te voeren. Als de verwachting is dat een algoritme tot gevolg heeft dat wordt ingegrepen in het leven of de vrijheid van mensen, en zeker als de verwachting is dat er grondrechten worden geraakt, moet er een wettelijke grondslag bestaan voor de taak/het proces waarin het algoritme wordt ingezet. **Ontbreekt in een dergelijk geval een concrete en helder geformuleerde wettelijke grondslag, dan mag het algoritme niet worden ingezet**.

Let op: *wanneer het algoritme slechts wordt gebruikt voor interne processen en geen directe impact heeft op mensen, is geen wettelijke grondslag vereist.*

In een aantal gevallen – in het bijzonder als het algoritme een inbreuk kan maken op een van de grondrechten zoals die in de Grondwet zijn vastgelegd – moet sprake zijn van een grondslag in een **formele wet**. (Vetzo, Gerards & Nehmelman 2018, p. 53 e.v.; Koops e.a. 2017) In andere gevallen kan een **lagere regeling** (bijvoorbeeld een AMvB) als grondslag volstaan.

Aanwijzingen en toelichting 1.4

De vragen bij subonderdeel 1.4 gaan over de **verantwoordelijkheden** ten aanzien van het algoritme. Een duidelijke verdeling en borging van (verschillende niveaus van) verantwoordelijkheden is cruciaal voor het effectief en verantwoord verloop van het proces rondom de inzet van een algoritme. Zeker wanneer ongewenste effecten (kunnen) optreden bij de implementatie van een algoritme zijn korte lijnen en een duidelijke taakverdeling belangrijk voor het doen van aanpassingen of voor de inzet van mitigerende maatregelen. Hierover moet op voorhand al worden nagedacht; voorzien moet worden in goede besluitvormingsstructuren en organisatorische inbedding van het algoritme. Eventueel kan het RASCI-model worden gebruikt als hulpmiddel om verantwoordelijkheden expliciet te maken. Dit model maakt onderscheid in verschillende lagen van verantwoordelijkheid: R*esponsible, Accountable, Supporting, Consulted, en Informed*. Wat behulpzaam is aan dit model, is dat het benadrukt dat degenen die verantwoordelijk zijn, niet altijd dezelfde zijn als degenen die rekenschap dienen af te leggen.

Als het niet mogelijk is om de verantwoordelijkheden van betrokkenen voldoende te borgen, mag het algoritme niet worden ingezet.

Deze duidelijke verdeling van verantwoordelijkheden dient ook in de toekomst geborgd te blijven. Het is bijvoorbeeld niet ondenkbaar dat bepaalde collega's van baan wisselen of om andere redenen hun functie niet kunnen voortzetten. Ook kan het voorkomen dat (onderdelen binnen) de organisatie gereorganiseerd worden. Centraal in de discussie over deze thema's staat dan tevens de vraag hoe er binnen de organisatie voor wordt gezorgd dat verantwoordelijkheden en relevante contextuele kennis goed belegd zijn en blijven, ook in de toekomst.

Wanneer, om welke reden dan ook, blijkt dat ontwikkeling of implementatie van het algoritme niet meer gewenst is, moet het mogelijk zijn om een **exitstrategie** te volgen. Dit is belangrijk, omdat het soms voor kan komen dat een organisatie al 'te diep' in het project zit en het lijkt alsof er geen weg terug meer mogelijk is. Dit kan schadelijke consequenties tot gevolg hebben, zoals de verspilling van publieke middelen of de inzet van een inaccuraat algoritme. Met een exitstrategie wordt dit voorkomen. De exitstrategie hoeft niet van tevoren al volledig uitgewerkt te worden; hoofdlijnen kunnen volstaan.

Deel 2: Wat?

Aanwijzingen en toelichting 2.1

Afhankelijk van het antwoord op de vraag of er al een beeld is van het in te zetten algoritme, zijn er voor de discussie verschillende opties:

- Als er al een ruw of meer omlijnd idee is van het in te zetten algoritme, kunnen de vragen hierna voor dat algoritme worden bediscussieerd.
- Als er verschillende te overwegen typen van algoritmes zijn, kunnen de vragen aan de orde komen voor de verschillende typen.
- Als er nog geen idee is over welk type algoritme gebruikt gaat worden, kan op dit punt gestopt worden met het IAMA. Pas als er een ruw idee is over het type algoritme, of als er verschillende opties zijn, kan het IAMA verder worden ingezet.

ZELFLEREND OF NIET-ZELFLEREND?

Voor het type algoritme maken we onderscheid tussen **zelflerende** en n**iet-zelflerende algoritmes**. Daaronder verstaan we voor het IAMA het volgende:

A. Bij een **niet-zelflerend algoritme** specificeert de mens de regels die de computer moet volgen. Hierbij kan gedacht worden aan beslisregels. Van een niet-zelflerend algoritme is ook sprake als de regels gebaseerd worden op data of (statistische) analyses, zolang een mens maar bepaalt wat deze regels zijn.

B. Bij een **zelflerend algoritme** leert de machine zelf over de patronen in de data. Voorbeelden hiervan zijn statistische modellen en *machine learning*. Het is ook mogelijk om beslisregels automatisch te trainen, in welk geval deze hieronder vallen. Onder de zelflerende algoritmes valt ook alle generatieve AI. Deze algoritmes worden ook wel **modellen** genoemd.

Een aantal voorbeelden illustreert dit verder.

Een voorbeeld van een **niet-zelflerend algoritme** is een algoritme dat controleert of een parkeerboete correct is betaald. Hierbij zou een mens bijvoorbeeld kunnen specificeren dat ‘als het betalingskenmerk bij de transactie overeenkomt met een openstaande boete én het bedrag dat gestort is op de rekening gelijk is aan het boetebedrag, de parkeerboete correct is betaald’. Op basis van dit algoritme kan automatisch een betalingsbevestiging gestuurd worden aan de persoon waar de boete aan gekoppeld was. Een ander voorbeeld is een algoritme voor fraudesignalering, waarbij een expert specificeert dat ‘als er in het afgelopen jaar meer dan drie fouten zitten in de aangeleverde administratieve informatie, het systeem een signaal moet afgeven dat er nader onderzoek gedaan moet worden naar dit dossier’.

Een voorbeeld van een **zelflerend algoritme** is een algoritme dat op basis van gelabelde voorbeeldfoto’s leert of een foto een hond of een kat toont. Een ander voorbeeld is een algoritme dat op basis van verkoopcijfers en klantgegevens leert hoe bepaalde klanteigenschappen samenhangen met de verkoop aan deze klanten. In het voorbeeld over fraudesignalering hierboven kan ook een zelflerend algoritme worden ingezet, bijvoorbeeld doordat het systeem, op basis van voorbeelden van fraude en niet-fraude, zo wordt ingericht dat het kan leren welke eigenschappen en gedragskenmerken samen lijken te hangen met fraude.

Het **onderscheid tussen deze twee hoofdtypen is relevant** omdat ze verschillende vragen oproepen over de inzet van een algoritme. Als een mens specificeert wat de machine moet doen, is het van belang om te reflecteren op de mensen die deze regels specificeren, op het (leer)proces dat zij doorlopen hebben om tot deze kennis te komen, en op de kwaliteit en legitimiteit van de keuzes die ze hierbij maken. Hier gaat de reflectie dus over de achtergrond van de totstandkoming en de mate waarin er consensus is over de correctheid van de regels. Voor de zelflerende algoritmes moet de reflectie vooral gaan over de 'machine', hoe deze kan leren en op basis waarvan, en of dat proces geschikt is voor de context waarin het algoritme gebruikt zou moeten worden.

NOTITIE GENERATIEVE AI

Generatieve AI is altijd zelflerend: Het gaat om een model dat is getraind om bijvoorbeeld het volgende woord te voorspellen of middels tekst een afbeelding of video te genereren. In de toelichting van subonderdeel 2.2 wordt verder ingegaan op de uitdagingen van generatieve AI. Omdat het model zelflerend is, maar de gebruiker ervan doorgaans niet weet op welke data het getraind is, is het niet eenvoudig om biases van het model in kaart te brengen.

Aanwijzingen en toelichting 2.2

Voor subonderdeel 2.2 worden drie concepten toegelicht. Bias en dataminimalisatie zijn relevant voor 2.2A én 2.2B.
De toelichting over trainings- en testdata is alleen relevant voor 2.2B.

Bias

Bias (of '**vooringenomenheid**') is een breed begrip: **bias in data kan verschillende verschijningsvormen hebben**. Het is belangrijk om bij de discussie naar deze verschillende verschijningsvormen te kijken en te onderzoeken hoe daarmee moet worden omgegaan. In de onderstaande box worden enkele voorbeelden en verschijningsvormen van bias gegeven.

Voorbeelden en verschijningsvormen van bias

Bias kan op verschillende momenten en verschillende manieren in een algoritme sluipen:

1. Gekozen *metric* of proxy. Algoritmes worden gebruikt om een keuze eenvoudiger te maken. Hierbij wordt gebruik gemaakt van benaderingen (proxy's). Proxy's die heel logisch lijken, kunnen wel bias introduceren. Dat kan direct zijn. Zoekt iemand bijvoorbeeld een huurder voor een gemeubileerde woning, dan geeft zoeken naar expats een bias richting mensen zonder Nederlands paspoort. Een bias kan ook indirect worden geïntroduceerd. Als een organisatie de inschatting van de kans op arbeidsongeschiktheid baseert op werkonderbrekingen, dan zullen vrouwen een hogere score krijgen omdat zij bijvoorbeeld langduriger verlof hebben als zij kinderen krijgen.

2. Bias vanuit de data of observaties. Bij het gebruik van data of observaties wordt de blik beperkt tot datgene wat is gemeten in de data of is geobserveerd. Zijn de data bijvoorbeeld verzameld middels een online survey? Dan beperkt dat waarschijnlijk de personen erin tot hen die digitaal vaardig genoeg zijn om hieraan deel te nemen. De groep die onvoldoende digitaal vaardig is, wordt mogelijk geraakt door het algoritme, omdat de methode van dataverzameling niet op hen is afgestemd. Dit geldt niet alleen voor data, maar ook voor ervaringen. Context kan ervaringen vormen: een leraar in de Randstad loopt mogelijk tegen andere uitdagingen aan dan een leraar buiten de Randstad. Als de ervaring van de ene groep in beslisregels wordt gegoten, dan levert dat dus een bias op naar hun ervaringen.

Bij gebruikmaking van trainingsdata zijn ook nog andere factoren van belang om bias te voorkomen. Trainingsdata kunnen bijvoorbeeld een goede kwaliteit hebben, maar alsnog niet geschikt zijn voor het specifieke algoritme en de doelstelling die met inzet van het algoritme wordt beoogd. Om te kunnen beoordelen of de trainingsdata geschikt zijn, moet worden gekeken naar de doelstelling van het betreffende algoritme en naar de aard van de trainingsdata. De vraag is daarbij of de trainingsdata inderdaad in staat zijn om het algoritme zodanig te trainen dat het de doelstelling kan vervullen.

Een voorbeeld dat laat zien waarom het belangrijk is om trainingsdata en doelstelling van het algoritme goed op elkaar aan te laten sluiten, is het voorbeeld van een algoritme dat had geleerd om husky's en wolven van elkaar te onderscheiden. Daarbij was het algoritme getraind met evenwichtige data die voor de helft uit afbeeldingen van husky's en voor de helft uit afbeeldingen van wolven bestonden. Het neurale netwerk had echter geleerd om het verschil te herkennen door te signaleren dat afbeeldingen van wolven vaker een groene omgeving hebben (bijvoorbeeld doordat ze worden gefotografeerd in het gras) en de afbeeldingen van husky's vaker in een witte omgeving (sneeuw). Er was dus een bias, maar dan bias die was gericht op de achtergrond van de afbeeldingen. Daardoor leerde het algoritme weliswaar verschillen in de afbeeldingen herkennen, maar ging het niet om verschillen die relevant waren vanuit het perspectief van het doel (het onderscheiden van husky's en wolven). Met name bij *machine learning* is het van belang om de dataset heel precies vorm te geven om voldoende specifiek te kunnen definiëren wat de machine leert.

Van belang is hiernaast of de dataset compleet en accuraat is. Als enerzijds een algoritmische regel wordt bedacht om mensen een uitkering te geven, maar niet alle groepen op wie het algoritme van toepassing is ook in de trainingsdata voorkomen, kan de toepassing van het algoritme alsnog tot bias leiden. Hoe goed de regel ook werkt op de trainingsdata, als de data een bias hebben, werkt dat vervolgens door bij inzet in de praktijk.

NOTITIE AI-VERORDENING

Als een AI-systeem wordt ingekocht, moet de ontwikkelaar worden verzocht om de **gebruiksinstructies** te overleggen. In deze instructies moet onder andere goed worden beschreven voor welke context het model is getraind; dit is een verantwoordelijkheid van de ontwikkelaar. De gebruiker van het systeem heeft de verantwoordelijkheid om te zorgen dat het model alleen voor die toepassingen worden gebruikt waarvan in de instructies staat dat het daarvoor is bedoeld. Wordt het AI-systeem gebruikt in een andere situatie, dan bestaat de kans dat er bias geïntroduceerd wordt, omdat het dan minder passend kan zijn bij de trainingsdata. Afwijken mag niet zomaar: wie een model buiten de context gaat gebruiken, neemt de verantwoordelijkheid over van de ontwikkelaar om aan te tonen dat het goed genoeg werkt, met alle verplichtingen van dien.

In de gebruiksinstructies heeft de aanbieder ook de verantwoordelijkheid om te reflecteren op bias in de trainingsdata en modellen. De informatie uit de gebruiksinstructie kan helpen bij het beantwoorden van deze vragen.

NOTITIE GENERATIEVE AI

Bij generatieve AI is vaak sprake van voorgetrainde modellen, waarbij het vaker niet dan wel inzichtelijk is op welke data het model is getraind. Daarom is het voor de invuller van een IAMA vrijwel onmogelijk om vooraf aan te wijzen welke bias hierin zit. Waar wel naar gekeken kan worden, is bias in de output. Besteed tevens extra aandacht aan bias in de output bij de beantwoording van de vragen onder subonderdeel 2.4.

Dataminimalisatie

Met dataminimalisatie doelen we op het beperken van de data die gebruikt worden tot alleen die data die (a) geoorloofd zijn en (b) een daadwerkelijke bijdrage leveren. Het gebruik van meer data kan een systeem beter laten presteren, maar hoeft een systeem niet daadwerkelijk beter te maken. Bij een algoritme dat zich op minder data baseert, is de uitkomst makkelijker uitlegbaar. Extra data kunnen dan wellicht een kleine winst opleveren in prestaties, maar ten koste gaan van die uitlegbaarheid.

Bij gebruik van persoonsgegevens is dataminimalisatie een vereiste vanuit de AVG: het is niet toegestaan om meer persoonsgegevens te gebruiken dan nodig is om het systeem goed te laten werken. Het is bij persoonsgegevens daarom van belang om toe te lichten waarom elk van de gebruikte datapunten van zodanige toegevoegde waarde is voor de uitkomst dat het gebruik ervan gerechtvaardigd is.

Trainings- en testdata

Tegenwoordig is bijna iedereen bekend met het principe '**garbage in = garbage out**'. Deze uitspraak reflecteert het gegeven dat de data die gebruikt worden als input voor het algoritme en de kwaliteit daarvan, bepalend zijn voor de uitkomsten uit het algoritme. Daarom is het belangrijk dat de data compleet en accuraat zijn. De datakwaliteit kan gecontroleerd worden op verschillende manieren, bijvoorbeeld met behulp van steekproeven. De vraag die hierbij centraal staat, is: beschrijft de data het fenomeen dat onderzocht dient te worden? Met andere woorden, de data die worden verzameld, dienen de juiste proxy te zijn voor hetgeen geïdentificeerd dient te worden.

Als het gaat om een algoritme dat gebruik maakt van **trainingsdata** is het nodig om te onderzoeken waar de trainingsdata van afkomstig zijn en of deze van goede kwaliteit zijn. Ook is het van belang om na te gaan of de set met trainingsdata al eerder is bekritiseerd als foutief of zelfs is teruggetrokken.

Naast trainingsdata kan er ook sprake zijn van **testdata**. Dit zijn de data waarop de prestaties van het algoritme worden gemeten. Deze data mogen niet onderdeel zijn van de trainingsdata, omdat het algoritme in de praktijk ook gaat werken op 'nieuwe' data: data die het nog nooit eerder heeft gezien. De testdata worden daarom apart gehouden van de trainingsdata, zodat zo goed mogelijk te bepalen is hoe goed het model gaat zijn met nieuwe data.

In de praktijk komen testdata en trainingsdata vaak uit dezelfde bron. Is dit in het besproken systeem niet zo? Bespreek dan beide databronnen afzonderlijk.

Aanwijzingen en toelichting 2.3

Bij een hoog-risico AI-systeem moeten er vanuit de AI-verordening gebruiksinstructies aanwezig zijn, die toelichten voor welke context en wijze van inzet het AI-systeem gemaakt is. Het is van belang dat om, ook als gebruiksverantwoordelijke, altijd deze gebruiksinstructies te raadplegen. Bij twijfel moeten ze worden opgevraagd bij de aanbieder van de toepassing. Als er geen sprake is van een hoog-risico AI-systeem, dan zijn er weliswaar geen formele eisen verbonden aan de gebruiksinstructies, maar is het alsnog wenselijk om van de aanbieder een uitleg te ontvangen over de context en inzet waar het AI-systeem voor ontwikkeld is.

Aanwijzingen en toelichting 2.4

Om de kwaliteit of accuraatheid van een algoritme te bepalen wordt vaak gekeken naar *metrics* die meten hoe goed het presteert in een testsituatie. Afhankelijk van de toepassing en de ernst van (gevolgen van) fouten kunnen er verschillende eisen worden gesteld aan de kwaliteit of accuraatheid van het algoritme. Dit is een subjectieve kwestie waarvoor een zorgvuldige afweging moet worden gemaakt. Afhankelijk van het soort voorspelling van een model kunnen andere *metrics* gebruikt worden. Een model dat categorieën voorspelt zal bijvoorbeeld *precision* en *recall* gebruiken. Een model dat een waarde voorspelt zal een *metric* nodig hebben dat de gemiddelde afstand tot de juiste uitkomst gebruikt. Betrek de ontwikkelaars van het model voor de bepaling en inschatting van deze *metrics*.

Vraag 2.4.3 vraagt om een uitwerking van mogelijke **ongewenste uitkomsten**. Dit gaat een stap verder dan de *metrics* op kwaliteit. Het vraagt om te reflecteren op de gevolgen van fouten voor betrokkenen. Verderop in het IAMA wordt dieper ingegaan op wat dit betekent en of er sprake is van een impact op grondrechten. Deze vraag is opgenomen om met de aanwezigen bij deel 2 (inclusief de ontwikkelaars dus) een eerste reflectie te doen op de waarde van de gekozen *metrics* en kwaliteitscriteria.

In de vragen hiervoor is gekeken naar de uitkomsten van de *metrics*. Ook als deze voor een specifiek soort model (erg) goed zijn, moet worden gekeken naar de consequenties van fouten om de wenselijkheid in de praktijk te schatten. Om te helpen bij het beantwoorden van vraag 2.4.3 worden hieronder twee methoden toegelicht om te kijken naar welke ongewenste uitkomsten er bestaan: *false positives/false negatives* en onder-/ overschattingen.

False positives* en *false negatives

Voor zowel **zelflerende** als **niet-zelflerende algoritmes** kan gekeken worden naar *false positives* (valse positieven) en *false negatives* (valse negatieven).

Als voorbeeld kan een algoritme worden genomen dat controleert of een verkeersboete correct is betaald. Hierbij zou een mens bijvoorbeeld kunnen specificeren dat 'de boete correct is betaald als het betalingskenmerk bij de transactie overeenkomt met een openstaande boete én het bedrag dat gestort is op de rekening gelijk is aan het boetebedrag'.
Het algoritme kan één van twee uitkomsten geven: [1] de boete is wél correct betaald of [2] de boete is niet correct betaald.
Om een uitspraak te doen over de accuraatheid van dit algoritme, zou gekeken moeten worden naar vier situaties die zich kunnen voordoen:

		Uitkomst van het algoritme	
		Boete is correct betaald	Boete is niet correct betaald
Werkelijkheid	Boete is correct betaald	***True positive***	***False positive***
	Boete is niet correct betaald	***False negative***	***True negative***

1. Gevallen waarin de boete in werkelijkheid correct betaald is én het algoritme aangeeft dat het correct betaald is (***true positive***)
2. Gevallen waarin de boete in werkelijkheid niet correct betaald is, maar het algoritme aangeeft dat het wel correct betaald is (***false positive***)
3. Gevallen waarin de boete in werkelijkheid niet correct is betaald én het algoritme aangeeft dat het niet correct betaald is (***true negative***)
4. Gevallen waarin de boete in werkelijkheid wel correct betaald is, maar het algoritme aangeeft dat het niet correct betaald is (***false negative***).

De bruikbaarheid van deze methode is het grootst als de uitkomstmaat binair is, dus als er twee opties zijn. Hiervan is bijvoorbeeld sprake bij een model dat voorspelt of iemand wel of geen recht heeft op een uitkering. Het model geeft hier twee mogelijke uitkomsten, waardoor het eenvoudig is om te beoordelen of deze goed of fout zijn. In sommige andere gevallen is het ook bruikbaar, zoals bij een chatbot. In dat geval kan namelijk beoordeeld worden of het antwoord goed of fout is, of bijvoorbeeld behulpzaam of onbehulpzaam is. Zo ontstaat er alsnog een binaire uitkomstmaat en kan deze worden beoordeeld door te kijken naar *false positives* en *false negatives*.

Het algoritme kan als 100% accuraat getypeerd worden als het altijd *true positives* en *true negatives* teruggeeft. Met andere woorden: als de boete correct betaald is, komt het algoritme ook altijd tot die conclusie, en als de boete niet correct betaald is, komt het algoritme ook altijd tot die conclusie. In de praktijk kan het echter zo zijn dat de geprogrammeerde regels niet altijd tot de gewenste resultaten leiden. Stel bijvoorbeeld dat het algoritme enkel naar het betalingskenmerk kijkt in het specifiek daarvoor bestemde betalingskenmerkveld, en niet naar het opmerkingenveld. Alle transacties van personen die in de praktijk het betalingskenmerk typen in het opmerkingenveld zouden dan onterecht als 'niet correct betaald' worden aangemerkt (*false negative*). Het is dan nodig om te onderzoeken en te bediscussiëren of [1] aanpassingen aan het algoritme noodzakelijk zijn en/of [2] dit niveau van accuraatheid acceptabel zou zijn in de context waarin het algoritme ingezet zou worden.

De eenvoudigste maatstaf voor accuraatheid is om te kijken naar het aantal *true positives* en *true negatives* ten opzichte van alle beoordeelde gevallen. Deze maatstaf voor accuraatheid is echter niet altijd de meest wenselijke. Stel bijvoorbeeld dat een algoritme moet inschatten of een persoon een dodelijke ziekte heeft. In dat geval is het voor mensen essentieel dat het algoritme niet aangeeft dat er niks aan de hand is, terwijl de persoon dodelijk ziek is. Een *false negative* weegt duidelijk heel zwaar in deze situatie. Een ogenschijnlijk hoge accuraatheid van 99% zegt dan ook weinig, als dit in de praktijk betekent dat er vele mensenlevens verloren gaan. Voor een dergelijk algoritme kan het dus nodig zijn om de kwaliteit niet te bepalen op het totale aandeel fouten, maar om het zwaarder te beoordelen op specifieke fouten, de *false negatives* in dit geval.

Tot slot is belangrijk om op te merken dat gebruikers van een algoritme niet altijd weten welke gevallen als een *false positive* te bestempelen zijn. Een algoritme kan bijvoorbeeld een leerling op een school aanmerken als risicovol voor schooluitval, terwijl deze leerling slechts tijdelijk slecht presteerde door ziekte. Een medewerker van de school ziet vervolgens de markering, maar kan niet direct vaststellen of het daadwerkelijk om een probleem gaat en dus al dan niet een *false positive* is. Ook weten gebruikers niet altijd in welke gevallen er *false negatives* gemist worden door het systeem. Als een algoritme bijvoorbeeld een voorsortering maakt op dossiers die speciale aandacht vereisen van een ambtenaar, is het niet te weten voor de ambtenaar welke dossiers deze niet te zien krijgt, die eigenlijk wel handmatig beoordeeld hadden moeten worden (een *false negative*).

Een reflectie op de gevolgen van *false positives* en *false negatives* dient dus goed te gebeuren voordat het systeem in gebruik genomen wordt, omdat hier tijdens het gebruik vermoedelijk niet of beperkt zicht op bestaat.

Onder- en overschattingen

Niet in alle toepassingen zijn *false positives* en *false negatives* werkbaar. Er zijn namelijk ook modellen die een numerieke uitkomst teruggeven, zoals een model dat het aantal aangevraagde uitkeringen voor komend jaar voorspelt of een model dat de kosten voor het onderhoud van groenvoorzieningen voorspelt. In deze situaties is het niet zo zinvol om te bekijken of het model exact het goede antwoord biedt, maar is het relevanter om te kijken hoe dicht het model bij de werkelijkheid komt. Een manier om naar de fouten van een dergelijk model te kijken en de impact van deze fouten in te schatten, is door de aandacht te leggen op onder- en overschattingen en de gevolgen hiervan. Hoe erg is het als het model een overschatting of onderschatting teruggeeft?

In het voorbeeld van het voorspellen van de kosten voor onderhoud van groenvoorzieningen zou een onderschatting kunnen betekenen dat er mogelijk te weinig middelen gereserveerd zijn om de groenvoorzieningen goed te onderhouden, met een slecht onderhouden buitenruimte als gevolg. Een overschatting zou kunnen betekenen dat de middelen nu 'onterecht' gereserveerd zijn en deze niet op een andere manier ingezet kunnen worden binnen de gemeente. Ook kan het zo zijn dat men binnen een gemeente het 'overschot' aan gereserveerde middelen hoe dan ook zal gaan inzetten, met beperkte doelmatige inzet van publieke middelen als gevolg. In het voorbeeld van de groenvoorziening zijn de fouten van het model wellicht te overzien, maar modellen met numerieke uitkomsten kunnen ook in meer precaire contexten ingezet worden. Een voorbeeld is een algoritme dat voorspelt hoeveel beroep er gedaan gaat worden op de Wmo en deze voorspelling de basis vormt voor de beschikbare ondersteuning aan inwoners. In dit geval kan een onderschatting grote gevolgen hebben: mensen die de hulp hard nodig hebben, krijgen die hulp niet. Een overschatting zou in dit geval waarschijnlijk minder kwalijke directe gevolgen hebben, maar als er te veel budget gereserveerd is, kan dat toch leiden tot een gebrek aan efficiëntie.

Het is aan het projectteam om in iedere casus te beoordelen hoe erg het is als het model ernaast zit, zowel bij een overschatting als een onderschatting.

NOTITIE GENERATIEVE AI

Bij subonderdeel 2.3 werd al aangegeven dat het voor generatieve AI niet mogelijk is om de bias van modellen in kaart te brengen aan de hand van een blik op de trainingsdata. Daarom is het voor deze modellen van belang om meer aandacht te geven aan de weging van de uitkomsten. Dit betekent dat de vragen van subonderdeel 2.4 extra gewicht toekomen, omdat hier de uitkomsten gewogen worden.

Aanwijzingen en toelichting 2.5

In sommige gevallen kiest een overheidsorganisatie ervoor om een algoritme te laten ontwikkelen door een marktpartij. In dat geval moeten er duidelijke afspraken worden gemaakt over eigenaarschap van het algoritme en de controle daarover. Hierbij moet ook worden gedacht aan de impact van mogelijke updates van het algoritme die (moeten) worden uitgevoerd door de externe partij.

Daarnaast is het eigenaarschap belangrijk voor de uitlegbaarheid van het algoritme. Ook wanneer het algoritme is ontwikkeld door een derde partij, moet de organisatie die het algoritme uiteindelijk inzet, uitleg kunnen geven over de werking ervan. Op dit punt kan ook worden teruggekeken naar subonderdeel 1.4, waarin discussie is gevoerd over de verantwoordelijkheden rondom de ontwikkeling en de inzet van het algoritme.

Om de vraag wat 'voldoende beveiligd' betekent te kunnen beantwoorden, moet worden gekeken naar de binnen de organisatie gebruikte standaarden of richtlijnen. Hieraan kunnen technische en organisatorische maatregelen worden getoetst.

NOTITIE AI-VERORDENING

Het is voor de beantwoording van vraag 2.5.1 noodzakelijk om vast te leggen wie welke rol heeft (gebruiksverantwoordelijke en aanbieder) en welke verantwoordelijkheden daaruit volgen.

Deel 3: Hoe?

Aanwijzingen en toelichting 3.1

Een algoritme in isolatie zorgt niet voor ongewenste effecten. Die worden altijd veroorzaakt door implementatie, inzet of toepassing van het algoritme en door de beslissingen en maatregelen die worden gekoppeld aan de output van het algoritme – oftewel de gebruikscontext. Dit subonderdeel ziet op die gebruikscontext en vraagt uit hoe deze context eruitziet. Pas als dat helder is, kan later in dit deel geanalyseerd worden wat de impact is op de betrokken (rechts)personen.

Bij vraag 3.1.1 is het vooral belangrijk om te bediscussiëren welke (typen) beslissingen gebaseerd kunnen worden op de output van een algoritme. Die beslissingen kunnen divers van aard zijn: het kan gaan om besluiten van bestuursrechtelijke aard, maar bijvoorbeeld ook om het nemen van beleidsmaatregelen of om de inzet van publieke middelen. Daarnaast is het van belang om het proces te relateren aan de doelstelling die eerder is gedefinieerd. Een beknopte procesbeschrijving volstaat bij de beantwoording van deze vraag. Ook een visuele weergave zoals een procesplaat kan ter beantwoording dienen. Eventueel kan bij de vraag worden verwezen naar een projectplan of andere documentatie als daarin de procesbeschrijving al is vastgelegd.

Vraag 3.1.5 is bedoeld om expliciet stil te staan bij het effect van het algoritme op de processen eromheen. Soms heeft de toevoeging van een algoritme aan een proces weinig of vooral positieve effecten op dat proces, maar ontstaan er onbedoelde neveneffecten voor andere algoritmes of (organisatorische) processen. Dit fenomeen wordt ook wel het **waterbedeffect** genoemd.

Een voorbeeld hiervan is de introductie van scanauto's. Hoewel de introductie hiervan positieve gevolgen kan hebben voor het parkeerbeleid van een gemeente, kan dit tegelijkertijd een onbedoeld neveneffect hebben op de processen rondom de gehandicaptenparkeerkaart. Dit bleek toen in de praktijk scanauto's de parkeerkaart achter de voorruit van een auto niet als zodanig konden herkennen. Om dit op te lossen, moest een aparte database worden ingericht waarin mensen met een gehandicaptenparkeerkaart hun kenteken moeten registreren. Deze database is niet landelijk georganiseerd: elke gemeente heeft een eigen database en een eigen beleid eromheen. Voor mensen met een beperking die de auto gebruiken om naar verschillende gemeenten te reizen, evenals voor mensen met een hulpbehoevend kind met meerdere verzorgers, is dit een zware belasting. Hierdoor is in feite dus een administratieve last verlegd van de gemeente naar een groep mensen die al te maken heeft met veel bureaucratie.

Het is van belang om te beseffen dat een algoritme/algoritmisch proces niet in isolatie opereert, maar ook effect kan hebben op de bredere organisatorische structuur. De inzet van een algoritme verandert niet alleen de workflow van het proces waarin het wordt ingezet, maar wellicht ook dat van andere processen eromheen, waardoor bijvoorbeeld beleid moet worden aangepast.

NOTITIE AI-VERORDENING

Artikel 27 van de AI-verordening stelt verplichtingen vast voor hoog-risico AI-systemen. Een van die verplichtingen is om een beschrijving te geven van de processen van het algoritme in lijn met het doel ervan. Het is eveneens verplicht om de tijdsperiode en frequentie waarmee het hoog-risico AI-systeem ingezet zal worden te beschrijven. Deze vragen in het IAMA zijn voorzien van een icoontje, waardoor de overlap expliciet gemaakt wordt. Wanneer de AI-verordening van toepassing is, is het bij vraag 3.1.1 ook nodig om duidelijk te maken dat het proces verzekert dat het AI-systeem niet voor andere doelen wordt gebruikt dan waarvoor het primair is bedoeld.

Aanwijzingen en toelichting 3.2

Subonderdeel 3.2 is erop gericht te bepalen welke rol de verschillende medewerkers hebben in het proces waar het algoritme een onderdeel van is. Ook is het doel te bekijken hoe deze mensen in staat worden gesteld hun kennis en kunde toe te passen in dit proces. Zij moeten beslissingen over de werking, toepassing en implementatie van het algoritme op een verantwoorde en zorgvuldige manier kunnen nemen.

Hoewel het in sommige gevallen mogelijk en wettelijk toegestaan is om besluitvorming volledig te automatiseren, wordt in het algemeen vereist dat mensen voldoende betrokken zijn, al is het maar in de vorm van voldoende controle en toezicht. Eigenlijk is het noodzakelijk dat mensen in alle fasen van de levenscyclus van een algoritme (dus van ontwerp tot implementatie en toezicht) betrokken zijn en kunnen interveniëren. Dit heet '***human in the loop***'. In het geval van geautomatiseerde besluitvorming in de zin van de AVG is in bepaalde gevallen '**betekenisvolle menselijke tussenkomst**' vereist. Louter een mens een symbolische plek in het proces geven, is niet voldoende. Mensen moeten ook in staat worden gesteld om betekenisvol te kunnen interveniëren. Voor meer informatie kan worden gekeken naar de Handvatten betekenisvolle menselijke tussenkomst van de Autoriteit Persoonsgegevens.

Het is bij de beantwoording van deze vraag ook belangrijk om te reflecteren op de impact die het algoritme heeft op de huidige kennis en kunde van de medewerkers. Het automatiseren van specifieke processen of handelingen kan wellicht een efficiëntieslag opleveren, maar kan er ook voor zorgen dat medewerkers bepaalde vaardigheden verliezen. Dit kan vervolgens impact hebben op de expertise van de medewerker, maar ook op de ervaring van het werk en het werkplezier. Een voorbeeld hiervan is de introductie van de zelfscankassa in supermarkten: waar de kassamedewerker eerst een praatje kon maken met klanten, wordt de medewerker bij de zelfscankassa gestuurd door een computersysteem dat hen zegt welke klant gecontroleerd moet worden. Wanneer er geen controle plaatsvindt, verveelt de medewerker zich vooral. Een ander voorbeeld is het gebruik van een algoritme om juridische teksten snel te kunnen vinden en door te spitten. Hoewel dit een grote efficiëntieslag kan opleveren voor juristen, kan dit ook zorgen voor een vermindering aan vaardigheden. Immers, als de jurist zelf had gezocht naar het relevante stuk, had deze gewerkt aan kritische onderzoeksvaardigheden en wellicht tijdens het lezen nieuwe inhoudelijke kennis opgedaan. Onderzoek wijst uit dat met name het gebruik van *Large Language Models* (zoals ChatGPT) als vervanging van traditionelere onderzoeksvaardigheden een negatieve impact heeft op hersenactiviteit en eigenaarschap van het geleverde werk (Kosmyna et al 2025). Kortom: door introductie en afhankelijkheid van bepaalde vormen van technologie kunnen de werkervaring, het werkplezier, maar ook de vaardigheden van de werknemer worden aangetast.

NOTITIE AI-VERORDENING

Gelet op artikel 27 – in samenhang met artikel 14 – van de AI-verordening is het van belang voor gebruiksverantwoordelijken om maatregelen te treffen voor menselijk toezicht. Daarvoor is het relevant om te beschrijven welke rol de mens speelt in het algoritmische proces en wat de ruimte is voor deze medewerker om af te wijken van het systeem.

Daarnaast is voor AI-systemen '**AI-geletterdheid**' een belangrijk begrip. Dit begrip is omschreven in artikel 4 van de AI-verordening. Om goed in staat te kunnen zijn om te interveniëren, moeten medewerkers steeds geïnformeerde, onderbouwde, autonome beslissingen kunnen nemen ten aanzien van de algoritmische output. Ook moeten ze de juiste kennis en kunde hebben om dit te kunnen doen. Het niveau van AI-geletterdheid kan verschillen op basis van het soort en de mate van contact van een werknemer met een AI-systeem. Mocht een organisatie geen AI-systeem gebruiken of ontwikkelen (en daarmee buiten de scope van de AI-verordening vallen), dan is het nog steeds aan te raden AI-geletterdheid te stimuleren om de verantwoorde inzet van algoritmes te faciliteren.

Aanwijzingen en toelichting 3.3

De vragen bij 3.3 gaan over de communicatie over het algoritme, zowel binnen de eigen organisatie als naar buiten toe. Anders gezegd gaat het om de mate van openheid die de organisatie wil bieden over de inzet, werking en effecten van het algoritme. '**Openheid**' moet daarbij worden beschouwd als een spectrum: organisaties kunnen ervoor kiezen om volledig open of volledig gesloten te zijn over het algoritme, maar ook daartussen zijn allerlei vormen denkbaar. In welke mate en op welke manier openheid wordt gegeven, kan per algoritme verschillen. Het is vooral van belang dat de mate en wijze van communicatie gerechtvaardigd kan worden in het licht van het algoritme en de doelstellingen ervan. Dit kan in sommige gevallen leiden tot vraaggestuurde, passieve informatieverstrekking en in andere gevallen tot meer actieve informatieverstrekking. Dit laatste kan op allerlei manieren, zoals door het organiseren van informatieavonden, het creëren van een 'dashboard' met informatie over de werking van een algoritme, het maken van informatieve filmpjes, het publiceren van een algoritme in een algoritmeregister of het geven van informatie in (individuele) overheidsbesluiten. Ook de manier waarop op de communicatie gereageerd kan worden, kan sterk uiteenlopen, van het bieden van geen enkele reactiemogelijkheid tot het actief ophalen van reacties van bepaalde doelgroepen. Sommige organisaties kiezen er bijvoorbeeld voor om een speciale mailbox in te richten waar mensen terecht kunnen met vragen over algoritmes.

De mate en de vorm van communicatie zijn uiteraard ook afhankelijk van het beoogde publiek. Dat kan bijvoorbeeld een professioneel publiek zijn (rechters, advocaten, belangenbehartigingsorganisaties, toezichthouders), maar ook een algemener publiek (burgers die geconfronteerd gaan worden met geautomatiseerde vormen van besluitvorming, andere geïnteresseerden). Om een voorbeeld te noemen: toezichthouders willen graag inzage kunnen hebben in de technische werking van een algoritme, terwijl geraakte burgers meer behoefte hebben aan informatie over waar ze terecht kunnen met vragen en aan een begrijpelijke uitleg over hoe het algoritme leidt tot besluiten.

Begrippen die vaak in het kader van communicatie opkomen, zijn **transparantie** en **uitlegbaarheid**. Hoewel de begrippen transparantie en uitlegbaarheid veel met elkaar te maken hebben, hebben ze een iets ander doel. **Transparantie** ziet vooral op technische informatie over het algoritme: hoe werkt het precies, welke data worden gebruikt, wat zijn de details van de organisatorische processen waarin het opereert? Volledige transparantie kan erg detailgericht en extreem technisch zijn, waardoor vaak alleen experts in staat zijn de informatie te begrijpen. **Uitlegbaarheid** ziet op een ander aspect, namelijk begrijpelijkheid. De vraag is dan dus of mensen kunnen begrijpen wat het algoritme doet en hoe het werkt. Daarbij is ook van belang voor welke doelgroepen (professionals, toezichthouders, rechters, burgers) deze uitlegbaarheid er moet zijn en welke voorkennis zij hebben. Om een algoritme uitlegbaar te maken, moet er wellicht ingeleverd worden op het geven van technische details over de complexiteit van een systeem of moet de communicatie rondom het systeem worden vereenvoudigd. Pure gerichtheid op uitlegbaarheid kan daardoor zorgen voor een verminderde transparantie.

Het is daarom mogelijk om in te zetten op twee strategieën: een uitleg- of communicatiestrategie die is gericht op niet-technisch onderlegde doelgroepen (in het kader van uitlegbaarheid), en een strategie die vooral is gericht op experts in het vakgebied (in het kader van transparantie).

In een communicatiestrategie waarin veel accent wordt gelegd op uitlegbaarheid moet goed rekening worden gehouden met wat de doelgroepen nodig hebben. Daarbij moet worden stilgestaan dat veel doelgroepen vooral behoefte hebben aan informatie die in grote lijnen duidelijk maakt hoe een systeem werkt, zonder al te veel (technische) details. Tegelijkertijd kan het bieden van té weinig technische informatie wantrouwen wekken. Het is hier dus goed zoeken naar de beste middenweg, waarbij ook differentiatie naar doelgroepen nodig is.

Aanwijzingen en toelichting 3.4

Bij deze vragen gaat het primair om het garanderen van '***accountability***' of het afleggen van rekenschap over de werking en effecten van het algoritme. Dit betreft de mogelijkheid om vragen te stellen of aan de bel te trekken, de correctheid van handelen te kunnen bediscussiëren en eventuele bijsturing te overwegen.

Als een algoritme eenmaal is ontwikkeld en wordt ingezet, is het belangrijk om de werking van het algoritme goed te blijven controleren. Door contextuele factoren of veranderingen in data kunnen algoritmes anders uitwerken dan verwacht. Ook kan hun werking door de tijd heen veranderen. Vaak worden allerlei zaken bij de start van de ontwikkeling of inzet van het algoritme goed geregeld, maar verslapt tijdens latere fases in het proces de aandacht hiervoor. Daarom is het noodzakelijk dat al voordat het algoritme wordt ingezet, wordt nagedacht over stappen om de werking van het algoritme in de gaten te blijven houden en het zo nodig bij te stellen. Daarbij kan zowel worden gedacht aan interne processen van evaluatie, auditing en borging in kwaliteits- en risicomanagement (dus processen die worden ingericht binnen de organisatie die het algoritme toegepast) als aan externe processen (bijvoorbeeld toezicht door een externe toezichthouder).

In het kader van kwaliteits- en risicomanagement zijn een aantal standaarden ontwikkeld of nog in ontwikkeling, waarbij aangesloten kan worden. Voor kwaliteitsmanagement zijn dat ISO/IEC 42001 en prEN 18286 en voor risicomanagement zijn dat ISO/IEC 23894 en prEN 18228.

Ook van belang is continue monitoring van algoritmes die in productie zijn, zodat fouten tijdig worden gedetecteerd en hierop kan worden geacteerd (een voorbeeld van een maatregel die hierop kan worden genomen, is het gebruik van het Functioneringsgesprek voor AI van Universiteit Utrecht).

Het is essentieel dat mensen handelingsmogelijkheden hebben als een algoritme niet naar behoren functioneert. De formele mogelijkheden van bezwaar en beroep vallen hieronder, die openstaan voor belanghebbenden en die nader zijn uitgewerkt in de Algemene wet bestuursrecht. Daarnaast is elke overheidsorganisatie verplicht om een klachtenprocedure te hebben, waar een bredere groep dan alleen belanghebbenden terecht kan.

Voor medewerkers zijn dit vaak geen bruikbare of passende routes, terwijl de kans bestaat dat juist zij incidenten rond het algoritme opmerken, omdat zij hier regelmatig mee werken. Daarom is het belangrijk dat er een sfeer op de werkvloer heerst waarin zorgen op een veilige manier geuit kunnen worden richting collega's en/of leidinggevenden, en dat bij terechte zorgen actie wordt ondernomen.

Het kan daarnaast zo zijn dat een algoritme is ontwikkeld voor een bepaald doel, maar gedurende de tijd langzaamaan ook wordt ingezet voor andere doelen waar het algoritme oorspronkelijk niet voor bedoeld was. Dit wordt ook wel '***function creep***' genoemd. *Function creep* draagt risico's met zich mee, omdat eerdere controles en waarborgen louter uitgevoerd waren over het originele doel en niet over dit 'verschoven' doel. Denk hier bijvoorbeeld aan risico's voor privacy, autonomie of rechtvaardigheid. Een hypothetisch voorbeeld hiervan is een kantoor waarin een beveiligingssysteem wordt geïnstalleerd dat vergt dat werknemers in- en uitchecken om toegang te kunnen krijgen tot het gebouw. Als dit systeem vervolgens ook wordt gebruikt om in de gaten te houden of individuele werknemers niet te laat op het werk komen, is er sprake van *function creep*. Dit is namelijk een gebruik van het systeem dat voorbijgaat aan het oorspronkelijke doel van het creëren van een beveiligde toegang.

> **NOTITIE AI-VERORDENING**
> Als het algoritme een hoog-risico AI-systeem betreft, moet een aantal vragen van subonderdeel 3.4 worden beantwoord in het licht van artikel 27 van de AI-verordening. Op grond van artikel 27 moet dan worden beschreven hoe het menselijk toezicht wordt uitgevoerd en welke maatregelen genomen worden wanneer bijbehorende risico's zich voordoen. Dit is inclusief de regelingen voor interne governance en klachtenregelingen.

Aanwijzingen en toelichting 3.5

De vragen uit subonderdeel 3.5 staan in het teken van de impact van het algoritme op betrokkenen en de samenleving. De term 'betrokkenen' moet hierbij breed worden gelezen: het gaat om al diegenen die impact ervaren van het algoritme. In het geval van een overheidsorganisatie zal het vaak gaan om (specifieke groepen) burgers, terwijl een zorgorganisatie bijvoorbeeld eerder zal spreken over cliënten. In andere gevallen kan het gaan om ondernemingen of om verdachten. Het is van belang om scherp te hebben welke mensen, groepen en/of rechtspersonen geraakt zullen worden door de inzet van het algoritme, om vervolgens na te kunnen gaan wat die impact is en hoe negatieve gevolgen gemitigeerd kunnen worden. Daarvoor kan ook worden gekeken naar de al gegeven antwoorden bij subonderdeel 1.2 (publieke waarden). De geraakte publieke waarden kunnen als hulpmiddel dienen om de positieve en negatieve effecten van het algoritme in te schatten.

Bij de beantwoording van de vragen is extra aandacht nodig voor **historisch kwetsbare groepen** en voor de manier waarop **intersectionaliteit** kan leiden tot versterkte negatieve effecten. Een voorbeeld hiervan zijn gezichtsherkenningssystemen die beter werken bij mensen met een lichte huidskleur dan op mensen met een donkere huidskleur. Als een dergelijk systeem ook nog eens beter werkt bij mannen dan bij vrouwen, ervaren donkere vrouwen extra negatieve effecten. Met intersectionaliteit wordt de wijze bedoeld waarop verschillende eigenschappen (in het voorbeeld huidskleur en gender) overlappen en invloed hebben op de werking en uitkomsten van een systeem.

De impact van een algoritme kan ook door andere factoren dan het algoritme zelf worden bepaald. Wanneer een algoritme bijvoorbeeld alleen invloed heeft op de efficiëntie van de interne werkprocessen van een rechtbank, is de impact anders dan wanneer een algoritme wordt ingezet voor een vergaand geautomatiseerd besluitvormingsproces waardoor miljoenen burgers worden geraakt. Het gaat dus niet alleen om de gevolgen van het algoritme zelf, maar ook om de gevolgen van het proces of systeem waarin het algoritme opereert.

Bij vraag 3.5.2 is het van belang om, wanneer sprake is van 'bijzondere schade', deze schade en de mensen, groepen en/of rechtspersonen die deze schade lijden expliciet te beschrijven. Bijzondere schade is schade die daadwerkelijk wordt geleden, zoals financiële schade of ontslag.

Bij vraag 3.5.6 wordt expliciet ruimte geboden om het gesprek breder te trekken dan louter het voorliggende, specifieke algoritme. Soms heeft een enkel algoritme nauwelijks tot geen negatieve effecten, terwijl een grootschalige toepassing ervan toch een aanzienlijke impact op de samenleving kan hebben. Een voorbeeld hiervan is de inzet van chatbots door overheidsorganisaties. Het inzetten van een enkele chatbot heeft wellicht weinig effect op de maatschappij in het algemeen, maar door de menselijke medewerker overal uit te halen en te vervangen door chatbots, kan er een stukje menselijkheid en toegankelijkheid voor de burger verloren raken en kan zelfs de relatie tussen overheid en burger worden geraakt. Het is bij deze vraag van belang om zo open mogelijk dit bredere gesprek over maatschappelijke impact te voeren, waarbij de bij de vraag gegeven voorbeelden kunnen dienen als houvast voor de discussie.

> **NOTITIE AI-VERORDENING**
>
> Als een hoog-risico AI-systeem ingezet zal worden, is een aantal van de vragen uit subonderdeel 3.5 verbonden met de verplichtingen uit artikel 27 van de AI-verordening. Op grond van artikel 27 moet worden beschreven welke categorieën van natuurlijke personen en groepen naar verwachting gevolgen zullen ondervinden van het gebruik van het systeem in een specifieke context. Daarnaast moet worden benoemd wat de specifieke risico's op schade zijn voor deze personen en groepen.

Deel 4: Mensenrechten

Aanwijzingen en toelichting 4.1

Achtergrond van de vragen

Zoals vermeld bij de vragen voor dit subonderdeel, is het de bedoeling dat bij vraag 4.1.1 het team open brainstormt over de grondrechten die aan de orde kunnen zijn bij de inzet van het algoritme. Sommige algoritmes kunnen bedoeld zijn of het effect hebben om bepaalde grondrechten te beschermen (positieve impact). Te denken is bijvoorbeeld aan een algoritme dat zorgvuldig differentieert tussen de behoeftes van gezinnen aan maatschappelijke ondersteuning. Een dergelijk algoritme kan positieve effecten hebben op hun recht op respect voor het gezinsleven, hun autonomie en de mogelijkheden voor de voorziening in hun levensonderhoud. Algoritmes kunnen echter ook nadelige gevolgen hebben voor grondrechten (negatieve impact of een 'inbreuk'). Hetzelfde differentiërende algoritme bij maatschappelijke ondersteuning kan bijvoorbeeld leiden tot problematische profilering of ongelijke behandeling. Als het ertoe leidt dat maatschappelijke ondersteuning juist niet meer wordt geboden, kan bovendien sprake zijn van negatieve impact op het eigendomsrecht of op de persoonlijke autonomie. Natuurlijk kan een dergelijk algoritme ook leiden tot verwerking van persoonsgegevens, zodat ook het recht op bescherming van persoonsgegevens aan de orde kan zijn.

Doel van de vragen; werkwijze

Het doel van dit subonderdeel is om voor het concreet voorgestelde algoritme te achterhalen of er een verwachte positieve of negatieve impact is op grondrechten. Lastig daarbij is dat er veel verschillende grondrechten zijn en dat het niet altijd gemakkelijk is om te bepalen of bepaalde belangen of effecten echt met een grondrecht te maken hebben.

Hieronder staan drie grondrechten benoemd waarop een algoritme vaak inbreuk maakt: het recht op gelijke behandeling, het recht op persoonsgegevensbescherming en procedurele grondrechten en behoorlijk bestuur. Het team zal in ieder geval moeten nagaan of deze drie grondrechten bij het algoritme inderdaad aan de orde zijn.

Daarnaast zijn er nog veel andere grondrechten waar een algoritme positieve of negatieve impact op kan hebben. Deze grondrechten (en aspecten daarvan) zijn opgenomen in een lijst aan het eind van het IAMA, met een bepaalde logische indeling. Het team kan deze lijst doornemen en onderling bespreken welke van de genoemde grondrechten aan de orde zouden kunnen zijn.

Potentieel leidt dit tot een vrij lange groslijst van grondrechten die aan de orde zouden kunnen zijn, maar het gaat erom te achterhalen op welke grondrechten écht inbreuk dreigt te worden gemaakt, of welke grondrechten écht beter worden van het algoritme. Als vuistregel kan gelden dat naast de genoemde drie grondrechten nog een of twee andere grondrechten kunnen worden geselecteerd waarop het algoritme vanwege het beleidsterrein en het onderwerp echt impact kan hebben. Het gaat er bij deze vraag (4.1.2) om in onderling overleg te bepalen welke dat zijn.

Grondrechten die vrijwel altijd aan de orde zijn

In de praktijk blijkt dat algoritmes bijna altijd impact hebben op een paar grondrechten:

- **Het recht op gelijke behandeling en non-discriminatie**. Het is eigen aan de categorisering en profilering waarvoor veel algoritmes worden ingezet dat ze onderscheid maken. Ook is een bekend risico dat van bias in datasets, bijvoorbeeld doordat ze maatschappelijk bestaande stereotypen of historische patronen van structurele discriminatie reflecteren.
- **Het recht op persoonsgegevensbescherming.** Veel algoritmes worden gevoed met persoonsgegevens, soms zelfs gevoelige persoonsgegevens. Daardoor worden die persoonsgegevens verwerkt. De AVG en aanverwante regelgeving stellen bijzondere eisen aan deze verwerking.
- **Procedurele rechten en het recht op behoorlijk bestuur.** Voor mensen kan het lastig zijn om te detecteren dat het algoritme een nadelige uitkomst voor hen heeft en waardoor dat nadeel precies wordt veroorzaakt. Dat maakt het moeilijker om naar de rechter te stappen om een besluit aan te vechten en om te bewijzen dat sprake is van een door het algoritme veroorzaakt ongerechtvaardigd onderscheid. Verschillen in kennis van de werking van het algoritme tussen de procespartijen kunnen bovendien het recht op *equality of arms* raken. Bij geautomatiseerde besluitvorming met behulp van algoritmes kunnen daarnaast het zorgvuldigheids- en motiveringsbeginsel of de rechten van de verdediging op het spel te staan, die onderdeel vormen van het recht op behoorlijk bestuur.

De vragen in de delen 1-3 van het IAMA hebben vaak ook al betrekking op deze grondrechten. Zo zijn kenbaarheid, uitlegbaarheid, transparantie, zorgvuldige totstandkoming en communicatie van belang al aan de orde geweest. Hetzelfde geldt voor bias in datasets of in het algoritme zelf. Het is dus goed om even terug te kijken naar de antwoorden op de vragen in deel 1-3. Als al heel goed is gezorgd voor het wegnemen van bias of stevige maatregelen worden getroffen voor de uitlegbaarheid en communicatie, is de kans veel kleiner dat het algoritme echt nog inbreuk maakt op een van de genoemde drie rechten. In dat geval kan dat hier worden vermeld.

Andere grondrechten

Behalve op de drie genoemde grondrechten kan een algoritme soms ook impact hebben op andere grondrechten. Welke dit zijn is afhankelijk van het beleidsterrein en de aard en doelstellingen van het algoritme. Een paar voorbeelden kunnen dit illustreren:

- Een algoritme om berichten op sociale media te doorzoeken om te bepalen of wordt opgeroepen tot wanordelijkheden zal al snel raken aan de vrijheid van meningsuiting, het recht op informatie en de demonstratievrijheid.
- Een algoritme om preventief gezondheidsrisico's te detecteren zal impact hebben (positief én negatief) op privacyrechten en het recht op gezondheid.
- Een algoritme om te bepalen wie extra ondersteuning nodig heeft bij het vinden van een baan zal impact hebben (positief én negatief) op het recht om professionele/zakelijke relaties aan te gaan en op sociale en economische grondrechten.
- Een algoritme dat helpt om de hoogte van een boete of naheffing te bepalen kan impact hebben op het eigendomsrecht.
- Een algoritme dat helpt om dynamische energieprijzen te bepalen kan (positieve of negatieve) impact hebben op het recht op een duurzame leefomgeving, in het bijzonder het recht op toegang tot energie.

Aanwijzingen en toelichting 4.2

Voor alle bij vraag 4.1.2 geïdentificeerde grondrechten moet worden nagegaan of *specifieke wetgeving* van toepassing is waarin dat grondrecht nader wordt uitgewerkt. Als dat zo is, moet die wetgeving worden toegepast. Dit is echt juristenwerk: juristen kunnen uitpluizen welke specifieke wetgeving van toepassing is en hoe de toepassing van die regels op het algoritme uitpakt. Ook kunnen zij nagaan of er concrete rechtspraak is waarin aanwijzingen te vinden zijn over de toepassing van de wetgeving op het algoritme.

Let op: als een van de hierna genoemde stukken specifieke wetgeving *niet* van toepassing zijn, bijvoorbeeld omdat een kwestie buiten het toepassingsbereik van een wet valt, kan het onderliggende grondrecht natuurlijk nog wel in algemene zin van toepassing zijn. Als bijvoorbeeld een ongelijke behandeling niet onder de Algemene wet gelijke behandeling valt, kan worden teruggevallen op artikel 1 Grondwet, artikel 20 EU-Grondrechtenhandvest en/of artikel 14 of 1 Twaalfde Protocol EVRM. In dat geval moeten de vragen uit subonderdelen 4.3-4.5 worden beantwoord voor dit algemene grondrecht.

Wetgeving die altijd moet worden onderzocht

Bij algoritmes zijn een paar stukken wetgeving over grondrechten bijna altijd van belang. Daarover gaan de vragen 4.2.1, 4.2.2 en 4.2.3. Het gaat dan om de:

1. **AI-verordening.** Zeker als het gaat om hoog-risico AI-systemen, maar ook bij sommige andere AI-systemen, biedt de AI-verordening bescherming aan bijvoorbeeld het non-discriminatierecht, het recht op behoorlijk bestuur en bepaalde procedurele rechten. Het is dan ook altijd zaak om te onderzoeken of aan de eisen van de AI-verordening is voldaan. Sommige eisen zijn al eerder in het IAMA aan de orde gekomen. Het gaat er nu om te checken of ook aan andere, nog niet specifiek al benoemde eisen wordt voldaan.
2. **Algemene Verordening Gegevensbescherming (AVG) en aanverwante wetgeving over persoonsgegevensbescherming (bijvoorbeeld Wet politiegegevens (Wpg) of Wet justitiële en strafvorderlijke gegevens (Wjsg))**. Omdat bij algoritmes bijna altijd persoonsgegevens worden verwerkt, moet worden onderzocht of wordt voldaan aan de Europese en nationale wetgeving daarover. Als vaststaat dat persoonsgegevens worden verwerkt, kan een pre-scan DPIA worden uitgevoerd of kan op een andere manier worden onderzocht of een DPIA moet worden uitgevoerd. Dat kan leiden tot twee bevindingen:
   a. Als blijkt dat het nodig is om een DPIA uit te voeren, is het nodig om op dit punt even uit het IAMA te stappen en een DPIA te maken. Daarbij kan gebruik worden gemaakt van de eerder al gegeven antwoorden in het IAMA. Als de DPIA klaar is, kunnen de belangrijkste antwoorden in het IAMA worden geïntegreerd. Hierbij kan de Handreiking IAMA/DPIA worden gebruikt, die een uitgebreid overzicht geeft van de manier waarop de beide instrumenten in elkaar grijpen.
   b. Als een DPIA volgens de pre-scan DPIA niet nodig is, is het goed om te checken of er toch bepaalde eisen voortvloeien uit de wetgeving over persoonsgegevensbescherming waar het algoritme aan moet voldoen. De uitkomsten van dat onderzoek kunnen dan ook weer in het IAMA worden verwerkt.
3. **Gelijkebehandelingswetgeving (Algemene wet gelijke behandeling, Wet gelijke behandeling op grond van handicap en chronische ziekte, etc.)**. Let op: deze wetgeving is niet altijd van toepassing op eenzijdig overheidshandelen, zoals het opleggen van boetes of vergunningverlening. Wel kan de gelijkebehandelingswetgeving van toepassing zijn bij dienstverlening door de overheid en in arbeidsrechtelijke kwesties. Het verbod op discriminatie op grond van ras/etniciteit is bovendien van toepassing op besluiten of regelgeving op het terrein van sociale voorzieningen en sociale zekerheid. De gelijkebehandelingswetgeving heeft veel nadere invulling gekregen in rechtspraak van het Hof van Justitie van de EU (HvJ EU) en het Europees Hof voor de Rechten van de Mens (EHRM), en daarnaast in de oordelen van het College voor de Rechten van de Mens. Het College heeft in dit verband ook een besliskader ontwikkeld voor risicoprofielen.

Let op: van belang is nogmaals om te benadrukken dat als een algoritme buiten het toepassingsbereik van deze specifieke regelgeving valt, moet worden teruggevallen op de algemene grondrechtenbepalingen uit de Grondwet, het EU-Grondrechtenhandvest, het EVRM en andere internationale verdragen. In dat geval moeten de vragen uit subonderdelen 4.3-4.5 worden beantwoord.

Sectorspecifieke wetgeving en rechtspraak

Ook voor veel andere grondrechten is er bijzondere wetgeving of rechtspraak voorhanden. Die kan soms al de nodige duidelijkheid bieden, omdat de wetgever bepaalde grondrechtelijke keuzes al heeft gemaakt en het resultaat daarvan in de wetgeving heeft neergelegd. Dat is zeker het geval in situaties waarin impact te zien is op het recht op behoorlijk bestuur of procedurele rechten. Daarvoor bieden nationale wetgeving (Wetboek van Strafvordering, Algemene wet bestuursrecht) en de rechtspraak veel aanknopingspunten, al dan niet gelezen in het licht van de rechtspraak over artikel 6 EVRM en artikelen 47-50 van het EU-Grondrechtenhandvest. Voor bijzondere toepassingen – bijvoorbeeld in relatie tot het analyseren van onderschepte data bij bescherming van de nationale veiligheid of bij verwerking van passagiersgegevens voor de bestrijding van terrorisme – is bovendien uitgebreide EU-regelgeving voorhanden. Om de procedurele zorgvuldigheid te beschermen heeft het HvJ EU in zijn rechtspraak heel precieze eisen gesteld aan de inzet van een algoritme.

Als dit soort bijzondere wetgeving of duidelijke rechtspraak beschikbaar is, kan de vraag of het algoritme verenigbaar is met het grondrecht meestal al aan de hand daarvan worden beantwoord. In dat geval is het niet nodig om nog verder te gaan met de open vragen van subonderdelen 4.3-4.5. Het is dan voldoende om de wet toe te passen, gelezen in het licht van de relevante rechtspraak.

In andere gevallen is er soms wel sectorale of specifieke wetgeving voorhanden, maar zegt die niets over de concrete situatie die bij het algoritme aan de orde is. In dat geval is het wel nodig om verder te gaan met de vragen uit subonderdelen 4.3-4.5.

Voorbeelden

Voor de hiervoor genoemde voorbeelden kan beantwoording van deze vraag leiden tot bevindingen als de volgende:

- Algoritme om berichten op sociale media te doorzoeken om te zien of wordt opgeroepen tot wanordelijkheden: aanknopingspunten hiervoor zijn te vinden in het Wetboek van Strafrecht (haatzaaien, opruiing) en de Gemeentewet (noodbevoegdheden voor de burgemeester). Daarnaast is er over de vrijheid van meningsuiting, het recht op informatie en de demonstratievrijheid veel rechtspraak van het EHRM beschikbaar. Die wetgeving en rechtspraak kan op het algoritme worden toegepast. Als er dan nog punten open blijven, kunnen daarvoor subonderdelen 4.3-.4.5 worden behandeld.
- Algoritme om preventief gezondheidsrisico's te detecteren: hierover is niet heel veel wetgeving beschikbaar, al bevat de Wet publieke gezondheid enige aanknopingspunten en keuzes. Voor zover die wet onvoldoende antwoord biedt op de vraag of het algoritme verenigbaar is met het recht op gezondheid en de persoonlijke autonomie, is het nodig om subonderdelen 4.3-4.5 te behandelen.
- Algoritme om te bepalen wie extra ondersteuning nodig heeft bij het vinden van een baan: hiervoor kan arbeidsrechtelijke wetgeving en rechtspraak daarover een handvat bieden, maar voor zover die geen duidelijke keuzes voorschrijft, kan het goed zijn om subonderdelen 4.3-4.5 in te vullen.
- Algoritme dat helpt om de hoogte van een boete of naheffing te bepalen: de toepasselijke bepalingen uit het Wetboek van Strafvordering en de Algemene wet bestuursrecht zijn hierbij leidend. Die bepalingen moeten uiteraard worden gelezen in het licht van (de rechtspraak over) artikel 6 EVRM en/of artikel 47-50 EU-Grondrechtenhandvest (voor zover van toepassing). Waarschijnlijk volstaat een toepassing van deze wetgeving en rechtspraak om te kunnen vaststellen of de procedurele rechten bij dit algoritme voldoende worden beschermd.
- Algoritme dat helpt om dynamische energieprijzen te bepalen: de Energiewet bevat hierover relevante bepalingen, maar gaat niet heel specifiek op grondrechten in. Hier kan verdere toetsing (subonderdelen 4.3-4.5) dus nuttig zijn.

Aanwijzingen en toelichting 4.3

Waarom vaststelling van de ernst van de inbreuk?

Grondrechten zijn meestal heel ruim geformuleerd. Een grondrecht als het recht op privacy omvat heel veel verschillende aspecten. Het kan gaan over seksuele of genderidentiteit, maar ook over het kunnen aangaan van professionele relaties. Het kan gaan over cameratoezicht in de openbare ruimte, maar ook over bescherming tegen huiszoekingen. En het kan gaan over beschikking over de eigen gegevens, maar ook over reputatierechten. Niet al die aspecten van het recht op privacy zijn even belangrijk of zwaarwegend.

Van een ernstige inbreuk op het recht op privacy is bijvoorbeeld sprake als, om te bepalen of iemand mag optreden als bloeddonor, gegevens over iemands seksuele gedrag worden gebruikt om te bekijken of er een verhoogd risico is dat het bloed van die persoon besmet is met een overdraagbare ziekte, zoals HIV/aids. Dat wil niet zeggen dat een dergelijk algoritme niet mag worden gebruikt, maar wel dat er zwaarwegende maatschappelijke belangen tegenover de aantasting van het recht op privacy moeten staan. Ook betekent dit dat zorgvuldig onderzoek wordt gedaan naar de doeltreffendheid van precies *dit* algoritme en naar de beschikbaarheid van minder vergaande alternatieven.

Van een niet zo ernstige inbreuk is sprake als (volledig geanonimiseerde en geaggregeerde) gegevens worden gebruikt om via een app de meest optimale en minst drukke fietsroute door het centrum van een stad aan te geven, om op die manier het fietsverkeer beter te spreiden en de verkeersveiligheid te beschermen. Bij een dergelijke app kan sprake zijn van 'nudging' en gedragsbeïnvloeding, en daarmee heeft de app een zekere impact op de persoonlijke autonomie als onderdeel van het recht op privacy. Heel vergaand is die impact tegelijkertijd niet, omdat het vrij gemakkelijk zal zijn om de aanwijzingen van de app te negeren. Het is dan minder erg als de app niet zo heel effectief is, juist omdat de aanwijzingen gemakkelijk zijn te negeren. Bovendien gaat het hier alleen om het kiezen van een fietsroute, wat weliswaar een autonomie-aspect is, maar niet het allerbelangrijkste onderdeel van het recht op privacy. Omdat de inbreuk niet zo ernstig is, kan een wat minder grote effectiviteit gemakkelijker op de koop toe worden genomen en zal het maatschappelijke belang al snel zwaarder wegen dan het individuele recht op privacy.

Factoren om de ernst van de inbreuk te bepalen

De 'ernst van de inbreuk' wordt meestal bepaald aan de hand van twee factoren, die in de praktijk nauw met elkaar samenhangen:

1. Hoe belangrijk is het aspect van het grondrecht dat aan de orde is in vergelijking met andere aspecten van dit grondrecht?
2. Hoe vergaand is de impact van het algoritme op dit aspect van het grondrecht?

Deze vragen zijn *in abstracto* natuurlijk moeilijk te beantwoorden. Daarom volgt hierna wat extra informatie om goed met deze vragen te kunnen werken.

Let op: als in het algoritme sprake is van ongelijke behandeling, moet ook worden gekeken naar grond van onderscheid. Als bijvoorbeeld sprake is van risicoprofilering op grond van ras, wordt dat gezien als een zeer ernstige vorm van discriminatie, waarvoor (vrijwel) nooit een rechtvaardiging kan worden gegeven. Hetzelfde kan aan de orde zijn als het gaat om (direct of indirect) onderscheid op grond van de in artikel 1 Grondwet of andere gelijkebehandelingsbepalingen genoemde gronden, zoals geslacht, seksuele oriëntatie, (vermeende) handicap of chronische ziekte of godsdienst. Als een ongelijke behandeling op een van deze beschermde of 'verdachte' gronden is gebaseerd, is het alleen aanvaardbaar als daar zeer zwaarwegende redenen voor zijn te geven. De inbreuk op het gelijkheidsbeginsel is dan dus vrijwel altijd te zien als ernstig, ongeacht de vraag welke individuele of maatschappelijke belangen verder aan de orde zijn.

Bepalen van het belang van het grondrecht

Hiervoor werd opgemerkt dat bepaalde grondrechtenaspecten 'belangrijker' zijn dan andere. Dat is natuurlijk moeilijk objectief vast te stellen: iedereen zal eigen beelden hebben bij welke (aspecten van) grondrechten er écht toe doen en welke minder belangrijk zijn. Het gaat er dan ook om dat het team hierover met elkaar in discussie gaat en de teamleden samen – dus door **intersubjectiviteit** – een conclusie bereiken over de vraag of het gaat om een *belangrijk* aspect van het grondrecht, een *niet zo belangrijk* aspect, of iets ertussenin.

Een hulpmiddel hierbij kan het maken van een schaal zijn. Elk grondrecht heeft een zogenaamde 'kern': de wezenlijke reden waarom dat grondrecht wordt erkend als *fundamenteel* recht, dus een recht dat zo belangrijk is dat het extra zware bescherming krijgt via de Grondwet of internationale verdragen. Een nader overzicht van die wezenlijke redenen is gegeven in de lijst met grondrechtenclusters aan het eind van dit document. Kort gezegd gaat het hierbij bij het recht op privacy om de bescherming van de persoonlijke autonomie en de menselijke waardigheid, bij gelijkheidsrechten om gelijkwaardigheid, bij de vrijheid van meningsuiting en de verenigingsvrijheid om democratie, autonomie en pluralisme, en bij de procedurele rechten om eerlijkheid en openheid.

Sommige aspecten van grondrechten liggen heel dicht aan tegen de kern. Bijvoorbeeld:

- **Recht op privacy** – kern is: **autonomie en waardigheid** – dicht bij de kern liggen bijvoorbeeld identiteit (dan gaat het om je wezen, om wie je bent en om wie je wil zijn; denk bijvoorbeeld aan seksuele of genderidentiteit) of de bescherming van de integriteit van lichaam en geest.
- **Gelijkheidsrechten** – kern is: **gelijkwaardigheid** – dicht bij de kern ligt: het verbod van discriminatie op grond van historische of diepgewortelde vooroordelen tegenover bepaalde groepen.
- **Vrijheid van meningsuiting** – kern is: **democratie en pluralisme** – dicht bij de kern liggen: politieke uitingsvrijheid en bijdragen aan discussies over onderwerpen van algemeen belang, persvrijheid.
- **Procedurele rechten** – kern is: **openheid en eerlijkheid** – dicht bij de kern liggen: toegang tot de rechter, onafhankelijkheid en onpartijdigheid van de rechter.

Voor alle grondrechten is het ook wel mogelijk om aspecten van grondrechten te bedenken die veel verder van de kern van die grondrechten vandaan liggen. Bijvoorbeeld:

- **Recht op privacy** – kern is: autonomie en waardigheid – ver van de kern liggen: vrijheid om je eigen fiets-/autoroute naar het werk te bepalen, keuze om je hond los te laten lopen in het park.
- **Gelijkheidsrechten** – kern is: gelijkwaardigheid – ver van de kern ligt: ongelijke behandeling op grond van 'niet-verdachte' persoonskenmerken die logisch gerelateerd zijn aan het onderwerp van een regeling; bijvoorbeeld inkomen bij de inkomensbelasting of bij toeslagen, of opleidingsniveau bij toegang tot vervolgopleidingen.
- **Vrijheid van meningsuiting** – kern is: democratie en pluralisme – ver van de kern liggen: commerciële uitingen zoals reclame, scheldwoorden of opruiing.
- **Procedurele rechten** – kern is: openheid en eerlijkheid – ver van de kern liggen: praktische eisen om een beroepschrift op een bepaalde manier aan te leveren, verplichting om in hoger beroep grieven aan te voeren, verplichting om in cassatie te procederen met een cassatieadvocaat.

Deze exercitie levert voor het betreffende grondrecht een soort schaal op, waarbij aan de ene kant de aspecten van het grondrecht staan die heel dicht bij de kern van het grondrecht staan, en aan de andere kant de aspecten die er veel verder vandaan liggen.

Bij privacy kan de schaal er bijvoorbeeld als volgt uitzien:

Als op die manier een schaal is gemaakt, kan het team bekijken of het aspect van het grondrecht waarop impact wordt gemaakt, kan worden gezien als 'belangrijk', 'niet zo belangrijk', of iets ertussenin. Het kunnen doneren van bloed in het voorbeeld van de bloeddonor zou bijvoorbeeld kunnen worden gezien als een grondrecht van 'medium' belang. Het staat niet zo dicht bij de kern van autonomie en menselijke waardigheid als aspecten van identiteit of integriteit, maar duidelijk dichterbij dan het kunnen bepalen van de eigen route door het centrum van de stad. Tegelijkertijd is in dit voorbeeld duidelijk dat mensen kunnen worden uitgesloten als bloeddonor vanwege hun seksuele gedrag of seksuele identiteit, en dat daarover gegevens moeten worden aangeleverd. Die seksuele identiteit vormt natuurlijk wel een kernaspect van het privacyrecht en het moeten geven van informatie over een intiem iets als het seksuele gedrag kan als heel ingrijpend worden ervaren. Daardoor kan in dit specifieke geval toch worden gezegd dat het algoritme een belangrijk aspect van het grondrecht raakt.

Bepalen van de grootte van de impact

Op dezelfde manier kan een rangschikking worden gemaakt van de mate waarin inbreuk op dit grondrecht wordt gemaakt. Een criterium kan daarbij zijn of de uitoefening van het aspect van het grondrecht helemaal onmogelijk wordt gemaakt (heel vergaande inbreuk) of dat er nog heel veel mogelijkheden overblijven om het grondrecht uit te oefenen (nauwelijks een inbreuk). Bij het algoritme om de optimale fietsroute door het centrum te bepalen, is de inbreuk op het recht op privacy bijvoorbeeld heel beperkt, omdat de app-gebruiker vrij is om de *nudge* van de app te negeren en af te wijken van de aangeraden route. Bij het algoritme dat bepaalt of iemand bloeddonor kan worden, is het effect veel groter: als het algoritme een risico detecteert, kan iemand zijn wens om bloed te geven helemaal niet meer waarmaken. Ook hiervoor kan een schaal worden gemaakt, waarbij de grootte van de verwachte negatieve impact kan worden bepaald als 'groot', 'klein' of 'medium'.

Bepalen van de ernst van de inbreuk

Het is uiteindelijk natuurlijk wel zaak om het belang van het grondrechtenaspect en de grootte van de impact in onderlinge samenhang te bekijken.

Bij het algoritme voor het bepalen van risico's bij het doneren van bloed geldt dat – door de inrichting en de gebruikte criteria – een belangrijk aspect van het grondrecht aan de orde is. Bovendien is de impact groot. Dat betekent dat een dergelijk algoritme een ernstige inbreuk oplevert op de privacy.

Bij het algoritme voor de app om de optimale fietsroute door het centrum te bepalen is de gemaakte inbreuk op de privacy juist licht. Niet alleen gaat het hier om een minder belangrijk aspect van het grondrecht, maar ook is de *nudge* die de app geeft gemakkelijk te vermijden en is de impact dus niet zo groot.

Aanwijzingen en toelichting 4.4

De vraag naar doeltreffendheid (of geschiktheid, of effectiviteit) is een empirische vraag, die uitgaat van '*evidence based*'-beleid en -regelgeving. Beantwoording ervan vergt een vooruitziende blik: er moet een realistische prognose worden gemaakt van wat de inzet van een algoritme concreet zal opleveren en welke verschillende consequenties die inzet zal hebben. Zal een algoritme dat gezondheidsrisico's bij het doneren van bloed voorspelt echt kunnen leiden tot een verbetering van de kwaliteit en veiligheid van bloed? Leidt een app die '*nudget*' tot het nemen van een andere fietsroute echt tot minder drukte op bepaalde routes en daarmee tot grotere verkeersveiligheid? Hoe waarschijnlijk is het dat het algoritme echt tot een kostenbesparing of efficiëntieverbetering zal leiden, zeker als ook rekening wordt gehouden met de kosten van het ontwikkelen ervan en het auditen achteraf?

Een algoritme zal (net als ieder ander beleidsinstrument) nooit 100% doeltreffend zijn. Of een maatregel effectief is, hangt vaak af van heel veel andere (context- en omgevings)factoren. Daardoor kunnen bij het vormgeven van een algoritme alleen prognoses worden uitgesproken over de effectieve werking van het algoritme. De vraag is dan vooral ook welke onzekerheidsmarge acceptabel is als het gaat om het kunnen realiseren van de gestelde doelen met het voorgestelde algoritme. Daarom wordt hier gevraagd of het vrijwel zeker is dat de doelen zullen worden bereikt (of daaraan in ieder geval een belangrijke bijdrage zal worden geleverd) of dat hierover nog enige of zelfs veel onzekerheid bestaat.

Zelfs als er grote onzekerheid is over de effecten van het algoritme, kan dit vanuit grondrechtenperspectief soms acceptabel zijn. Of dat zo is, hangt ook af van de ernst van de inbreuk. Uit subonderdeel 4.3 is gebleken dat de vraag naar de verwachte doeltreffendheid of geschiktheid van het algoritme om de doelen van de inzet ervan te bereiken niet altijd even nauwkeurig hoeft te worden beantwoord. Als de inbreuk licht is, kan best genoegen worden genomen met een grote onzekerheidsmarge over de verwachte effecten. Maar andersom, als de inbreuk ernstig is (zoals bij het bloeddonatie-algoritme, of bij een algoritme dat onderscheid maakt op een beschermde grond), moet grondig onderzoek worden gedaan naar de verwachte effectiviteit.

Technisch gezien gaat het bij doeltreffendheid om de vraag of een algoritme betrouwbaar en accuraat is. Als er veel *false positives* of *false negatives* worden verwacht, kan dat zeker bij een ernstige inbreuk zo problematisch zijn dat moet worden besloten dat het algoritme niet kan worden ingezet of ontwikkeld. Hetzelfde geldt als blijkt dat een algoritme dat een ernstige inbreuk maakt op een grondrecht ingezet gaat worden in een context die maakt dat nog heel onzeker is of de doelen wel gehaald gaan worden.

Aanwijzingen en toelichting 4.5

Om beleidsdoelen te bereiken kunnen allerlei verschillende instrumenten worden ingezet, waaronder algoritmes. Het doel van verbeteren van veiligheid van gedoneerd bloed kan bijvoorbeeld ook worden bereikt met schriftelijk of mondeling in te vullen vragenlijsten (dat gebeurde vroeger), maar ook met de inzet van technisch geavanceerde testmogelijkheden om besmettingen vroegtijdig te detecteren. Een vragenlijst maakt waarschijnlijk ongeveer net zoveel inbreuk op de privacy als een algoritme dat gebruikmaakt van gevoelige gegevens. De keuze tussen die instrumenten is vanuit grondrechtenperspectief dan dus neutraal. Als een technisch geavanceerde testmogelijkheid beschikbaar is en die voldoende betrouwbaar blijkt te zijn, dan kan dat vanuit grondrechtenperspectief wel een goed alternatief zijn. In dat geval is het beter om daarvoor te kiezen, ook al zijn de kosten voor een dergelijke testmogelijkheid misschien iets hoger.

Als ondanks de beschikbaarheid van realistische, minder belastende alternatieven toch gebruik wordt gemaakt van een algoritme dat een ernstige inbreuk maakt op een grondrecht, dan kan niet worden gezegd dat het algoritme een 'noodzakelijk' middel is om de gestelde doelen te bereiken.

Daarnaast kunnen er soms compenserende/ mitigerende maatregelen mogelijk blijken om de nadelen van het algoritme te verzachten. Als er meer opties zijn, moet die keuze worden gemaakt die (a) zo goed mogelijk tegemoetkomt aan technische en financiële randvoorwaarden, die (b) de doelen zo goed mogelijk bereikt, en die (c) grondrechten die op het spel staan zo min mogelijk belast.

De vragen bij 4.5 zijn erop gericht om de verschillende beschikbare alternatieven zo goed mogelijk in beeld te krijgen en om te onderzoeken of bepaalde maatregelen ook echt realistisch zijn als alternatief. De vraag kan natuurlijk rijzen wat de *trade-off* moet zijn als een alternatief iets minder effectief lijkt te zijn of als er aan dit alternatief hoge kosten zijn verbonden, maar wel het grondrecht minder lijkt aan te tasten. Het zou bijvoorbeeld kunnen zijn dat er wel bepaalde tests voor de kwaliteit van gedoneerd bloed beschikbaar zijn, maar dat die risico's van besmetting met specifieke overdraagbare aandoeningen toch niet helemaal kunnen beperken. Het kan zijn dat vragenlijsten of een algoritme dan de enige manier zijn om het risico voor de volksgezondheid in te perken. Vraag 4.5.4 richt zich op die situatie. Het gaat er dan om gezamenlijk na te denken over de vraag hoe groot de kansen en risico's van het algoritme zijn en of die – ook vanuit het perspectief van grondrechtenbescherming – acceptabel kunnen worden gevonden. Naarmate de inbreuk op het grondrecht ernstiger is, zal het wenselijker zijn om te kiezen voor een alternatief, maar dat hangt ook af van de context, de omvang van de restrisico's en het belang van wat er op het spel staat.

Restrisico's zijn risico's die overblijven wanneer alle mitigerende maatregelen en eventuele andere actiepunten om nadelige gevolgen te beperken zijn doorgevoerd. De specifieke restrisico's hebben te maken met de aard en context van de ontwikkeling en/of inzet van het algoritme in kwestie. Als bijvoorbeeld gebruik wordt gemaakt van een *Large Language Model* dat door een Amerikaanse partij is getraind op grote hoeveelheden data waarvan de herkomst onduidelijk is, zal altijd een restrisico blijven bestaan op bias in het model vanwege mogelijke bias in de trainingsdata. De verantwoordelijke binnen de organisatie zal dit restrisico moeten bekijken en besluiten om dit al dan niet te accepteren.

Grondrechtenclusters

Inleiding

In dit deel wordt ten behoeve van deel 4 van het IAMA een nader overzicht gegeven van grondrechten. Daarbij zijn de verschillende grondrechten ingedeeld in vier clusters:

1. **Aan de persoon gerelateerde grondrechten**
2. **Vrijheidsrechten**
3. **Gelijkheidsrechten**
4. **Procedurele grondrechten**

Voor ieder cluster wordt hierna steeds eerst een korte samenvatting gegeven van de relevante kernwaarden; die kern kan van belang zijn bij het beantwoorden van de vraag hoe groot de impact is van een algoritme op een bepaald grondrecht. Daarbij wordt ook een aantal voorbeelden gegeven van grondrechten die nauw met deze kernwaarden samenhangen.

Deze samenvatting wordt gevolgd door een schematische weergave van de diverse grondrechten die binnen een bepaald cluster thuishoren. Daarbij wordt in de linker kolom een aantal grondrechten weergegeven die tot dit cluster kunnen worden gerekend. In de rechter kolom worden op basis van de rechtspraak en de literatuur verschillende aspecten van de grondrechten benoemd. Deze overzichten zijn niet volledig en er kan overlap bestaan tussen de verschillende grondrechten en de daartoe gerekende aspecten. Er is bovendien altijd discussie mogelijk over de vraag of niet nog extra aspecten kunnen tot een grondrecht moeten worden gerekend, of nieuwe grondrechten moeten worden erkend, en of bepaalde grondrechten (of deelgrondrechten) niet beter zouden kunnen worden samengevoegd.

Deze lijst is dan ook niet bedoeld als een definitieve, uitputtende, volledige lijst van grondrechten die niet voor discussie vatbaar is. Eerder gaat het om een startpunt voor een discussie over de vraag op welke (aspecten van) grondrechten een algoritme impact kan hebben, hetzij omdat het algoritme bijdraagt aan de bescherming van het grondrecht (positieve impact), hetzij omdat het een inbreuk of beperking oplevert op het grondrecht (negatieve impact).

Van belang is tot slot dat de genoemde grondrechten zelden ‘absoluut’ zijn. In de meeste gevallen is het mogelijk om redelijke beperkingen te stellen of de uitoefening van grondrechten te reguleren. Om te beoordelen of dat het geval is, moeten de andere vragen uit deel 4 van het IAMA worden beantwoord.

Een toelichting op veruit de meeste in dit overzicht genoemde grondrechten, met nadere verwijzing naar rechtspraak en literatuur, is te vinden in Gerards e.a. 2020 en Gerards e.a. 2023. De grondrechten zelf zijn ontleend aan de volgende primaire bronnen:

- De VN-mensenrechtenverdragen – zie voor een overzichtje van de belangrijkste https://www.ohchr.org/en/core-international-human-rights-instruments-and-their-monitoring-bodies
- De Raad van Europa-verdragen, waarvan het Europees Verdrag voor de Rechten van de Mens en het Europees Sociaal Handvest het meest relevant zijn – zie voor een overzicht https://www.coe.int/en/web/conventions/full-list
- Het EU-Handvest voor de Grondrechten – zie https://eur-lex.europa.eu/NL/legal-content/summary/charter-of-fundamental-rights-of-the-european-union.html

1. Aan de persoon gerelateerde grondrechten

Inleiding

De brede categorie van de aan persoon gerelateerde grondrechten is nauw verbonden met verschillende kernwaarden, in het bijzonder met de menselijke waardigheid, de persoonlijke autonomie, de fysieke en geestelijke integriteit en de eigen en de sociale identiteit. Heel kort en grof samengevat houden die kernwaarden in dat moet worden gerespecteerd dat mensen zijn wie ze (willen) zijn, omdat ze allemaal als mens hun eigen waarde hebben. Ook moeten mensen (uiteraard binnen redelijke grenzen) de mogelijkheid hebben om zichzelf te ontplooien en hun eigen keuzes te maken.

Het gaat bij deze waarden allereerst om 'interne' aspecten van het mens-zijn en van de eigen persoonlijkheid, die moeten worden beschermd tegen inmenging en bemoeienis door anderen. Typische voorbeelden van rechten die nauw met deze kernwaarden samenhangen zijn het recht op respect voor de persoonlijke leefomgeving (zoals de eigen woning), het recht op bescherming van persoonsgegevens, rechten die te maken hebben met lichamelijke en geestelijke integriteit, de bescherming van eer en goede naam en de gewetensvrijheid (Vetzo, Gerards & Nehmelman 2018, p. 53 e.v.; Koops e.a. 2017).

Minstens zo belangrijk zijn meer externe en op sociale identiteit gericht aspecten van de persoonlijke levenssfeer en de individuele autonomie, die gaan over de interactie met anderen. Mensen kunnen vaak pas zichzelf zijn en zichzelf ontwikkelen als ze relaties kunnen aangaan en onderhouden. Concrete uitwerkingen van deze grondwaarde zijn het recht op respect voor het gezinsleven, het recht om te huwen, en in meer algemene zin het recht om relaties met de buitenwereld aan te gaan, zowel in het privéleven als in de werksfeer.

Tot het brede cluster van de aan de persoon gerelateerde grondrechten kan verder een aantal rechten worden gerekend die kunnen worden gezien als voorwaardenscheppend voor de uitoefening van de genoemde kernwaarden. Het ultieme voorbeeld is het recht op leven: als dit recht niet effectief wordt beschermd, is het op geen enkele manier mogelijk om een van de andere genoemde rechten te kunnen uitoefenen en om waarden als die van menselijke waardigheid of autonomie te kunnen verwezenlijken.

Belangrijke voorwaardenscheppende grondrechten zijn ook 'sociale' of economische grondrechten zoals het recht op water en voedsel, het recht op goede arbeidsomstandigheden, het recht op een bestaansminimum, het recht op een toegankelijke gezondheidszorg en het recht op een gezonde leefomgeving. Als dergelijke grondrechten onvoldoende worden gerespecteerd en beschermd, is het immers niet mogelijk (of in ieder geval minder gemakkelijk) om een menswaardig leven te kunnen leiden waarbinnen autonome keuzes mogelijk zijn. (Nickel 2007)

Aan de persoon gerelateerde grondrechten en aspecten van deze grondrechten

1. Recht op persoonlijke identiteit / persoonlijkheidsrechten / persoonlijke autonomie	■ Recht op zelfontplooiing ■ Vrijheid om het eigen gedrag en het eigen denken te bepalen ■ Vrijheid om het eigen uiterlijk vorm te geven ■ Vrije beroepskeuze, onderwijskeuze, opleidingskeuze etc. ■ Respect voor de eigen identiteit (genderidentiteit / seksuele identiteit etc.) ■ Reproductieve rechten ■ Recht op kennis van de eigen afstamming ■ Naamrechten ■ Contractvrijheid
2. Recht op sociale identiteit / relationele privacyrechten / relationele autonomie	■ Recht op respect voor familierelaties / gezinsleven ■ Recht om te huwen ■ Recht op gezinsvorming ■ Recht om seksuele relaties aan te gaan ■ Recht om professionele/zakelijke relaties aan te gaan ■ Recht op toegang tot de arbeid/professie ■ Recht op toegang tot een land / verblijfsrechten ■ Recht op onderwijs
3. Recht op lichamelijke en geestelijke integriteit	■ Gewetensvrijheid / vrijheid van gedachte ■ Recht op leven ■ Verbod van foltering / onmenselijke of vernederende behandeling en bestraffing ■ Verbod van refoulement ■ Verbod op fouilleren: lijfsvisitatie ■ Vereiste van geïnformeerde toestemming voor medische behandeling en onderzoek ■ Recht op toegang tot de gezondheidszorg ■ Respect voor handelingsbekwaamheid ■ Recht op vrijwillige levensbeëindiging ■ Recht op abortus ■ Verbod op (moderne) slavernij/dienstbaarheid/gedwongen arbeid/mensenhandel/ uitbuiting
4. Recht op persoons-gegevensbescherming / informationele privacy-rechten	■ Bescherming tegen ongeautoriseerde/onzorgvuldige persoonsgegevensverwerking ■ Recht op toegang tot persoonsgegevens ■ Recht op correctie van persoonsgegevens ■ Recht op vergetelheid
5. Communicatierechten	■ Briefgeheim ■ Bescherming tegen afluisteren/aftappen/interceptie ■ Verbod op ongeautoriseerde doorgifte van communicatiegegevens ■ Recht op vertrouwelijkheid van communicatie met advocaat, arts etc.

6. Ruimtelijke privacyrechten	■ Bewegingsvrijheid ■ Habeas-corpusrechten (verbod van vrijheidsbeneming, huisarrest etc.) ■ Recht op vrije woonplaatskeuze ■ Vrijverkeersrechten (EU-recht) ■ Recht om het land te verlaten ■ Verbod op volgen van personen (bijv. met een GPS-tracker)
7. Verbod op cameratoezicht	■ Eigendomsgebonden privacyrechten ■ Huisrecht (bescherming tegen invallen/doorzoekingen) ■ Bescherming tegen doorzoeking kleding/tassen/laptop/computer etc. ■ Recht op vrije beschikking over eigendom ■ Bescherming tegen onteigening ■ Intellectueel-eigendomsrechten
8. Reputatierechten	■ Verbod van strafbare belediging / smaad / laster ■ Bescherming van eer en goede naam
9. Recht op een gezonde/ duurzame leefomgeving	■ Recht op duurzame ontwikkeling ■ Recht op milieubescherming ■ Bescherming tegen uitstoot van schadelijke stoffen ■ Recht op toegang tot schoon drinkwater ■ Recht op toegang tot sanitatie (riolering) ■ Recht op toegang tot energie
10. Sociale en economische grondrechten	■ Recht op een minimum bestaansniveau ■ Recht op sociale zekerheid en bijstand ■ Recht op (toegang tot) arbeid ■ Vakbondsgerelateerde rechten (bijv. recht op lidmaatschap van een vakbond, rechten verbonden met collectieve onderhandelingen, stakingsrecht) ■ Recht op onderwijsvoorzieningen (in ieder geval basis- en voortgezet onderwijs)

2. Vrijheidsrechten

Inleiding

Vrijheidsrechten hebben een sterke relatie met de kernwaarden van waardigheid, autonomie, identiteit en integriteit. Zij betreffen vooral de mogelijkheid om de eigen keuzes, identiteit en opvattingen onbelemmerd naar buiten toe uit te dragen en zich naar de buitenwereld toe te gedragen zoals men wil. (Vetzo, Gerards & Nehmelman 2018) Zo zorgt de vrijheid van meningsuiting ervoor dat mensen hun eigen opvattingen of gevoelens kunnen delen met anderen, terwijl de godsdienstvrijheid ruimte laat om een bepaalde geloofsovertuiging (of juist het ontbreken daarvan) actief te belijden en uit te dragen. De verenigingsvrijheid laat toe dat mensen samenkomen met gelijkgestemden, terwijl de demonstratievrijheid betrekking heeft op het op bepaalde manieren kenbaar maken van bepaalde opvattingen aan de buitenwereld, bijvoorbeeld via een mars of een fluitconcert. De bewegingsvrijheid en het 'habeas corpus'-recht maken het mogelijk om zich ongehinderd te verplaatsen. Het kiesrecht maakt het mogelijk dat mensen vrij kunnen kiezen wie hun zal vertegenwoordigen bij het maken van wetgeving en beleid. Enzovoort.

De genoemde kernwaarden – waardigheid, autonomie, integriteit en identiteit – zijn niet de enige die ten grondslag liggen aan de vrijheidsrechten. De vrije uitwisseling van standpunten en ideeën, de toegang tot informatie, een ruime verenigingsvrijheid, de mogelijkheid om ongehinderd een stem te kunnen uitbrengen op een bepaalde kandidaat tijdens de verkiezingen, de mogelijkheid om petities te kunnen indienen: dat alles is ook essentieel om een democratische rechtsstaat goed te laten functioneren en om ruimte te laten voor debat en de confrontatie van verschillende opvattingen en ideeën. Vrijheidsrechten houden dus ook verband met kernwaarden als rechtsstatelijkheid, pluralisme en democratie (Vetzo, Gerards & Nehmelman 2018).

Vrijheidsrechten en aspecten van deze grondrechten

11. Uitingsvrijheid	■ Persvrijheid/journalistieke vrijheid ■ Artistieke vrijheid ■ Vrijheid van wetenschap/academische vrijheid ■ Vrijheid tot keuze van uitingsmiddel (mondeling/schriftelijk etc.) ■ Klokkenluiden ■ Journalistieke bronbescherming/verschoningsrecht
12. Vrijheid om informatie te ontvangen	■ Passieve informatievergaring (recht op toegang tot aangeboden, bestaande informatie) ■ Actieve informatievergaring (recht op toegang tot overheidsinformatie / openbaarheid van bestuur) ■ Plicht tot voorzien in pluriforme informatie ■ Recht op vrije toegang tot het internet
13. Godsdienstvrijheid	■ Vrijheid om een religie te hebben of niet te hebben ■ Religieuze uitingsvrijheid (symbolen, rituelen) ■ Vrijheid om samen te komen met andere gelovigen ■ Vrijheden van kerkgenootschappen/religieuze gemeenschappen ■ Verplichting tot religieuze neutraliteit van de staat (scheiding religie/staat) ■ Respect voor religieuze/filosofische overtuigingen in het onderwijs
14. Demonstratievrijheid	■ Vrijheid van vergadering ■ Vrijheid om samenkomsten, protestmarsen etc. te organiseren en te laten plaatsvinden ■ Vrijheid van keuze van onderwerp, tijd, plaats en middelen van demonstraties ■ Recht op bescherming tegen een vijandig publiek (‘hostile audience’)
15. Verenigingsvrijheid	■ Vrijheid om al dan niet lid te zijn van een vereniging ■ Interne verenigingsvrijheid (eigen keuze leden, activiteiten, financiering) ■ Vrijheid van politieke partijen
16. Politieke rechten/ vrijheden	■ Recht op periodieke organisatie van vrije en geheime verkiezingen ■ Actief en passief kiesrecht ■ Petitierecht

3. Gelijkheidsrechten

Inleiding

Het recht op gelijke behandeling dient allereerst een aantal belangrijke rechtsstatelijke waarden, namelijk die van rechtsgelijkheid, rechtszekerheid, en bescherming tegen willekeur in het overheidshandelen. Iedereen wordt geacht voor de wet gelijk te zijn en als een wettelijke regeling op een bepaalde groep van toepassing is, moet die wet ook in de uitvoering daadwerkelijk op al die gevallen worden toegepast. Dat voorkomt dat uitvoeringsambtenaren naar eigen inzicht van de wet afwijken en bevordert bovendien de voorspelbaarheid van de toepassing van die wet.

Een tweede kernwaarde die ten grondslag ligt aan de gelijkheidsrechten is de waarde van gelijkwaardigheid (Gerards 2019a). De behoeften, kwaliteiten, wensen of talenten van ieder mens zijn anders, maar die verschillen mogen niet maken dat sommige mensen of groepen van mensen als 'minderwaardig' worden behandeld. Iedereen heeft het recht om op gelijke voet aan de samenleving, de economie en de arbeidsmarkt deel te nemen (de waarde van inclusiviteit) en het is niet aanvaardbaar om groepen of personen zonder goede reden uit te sluiten van de toegang tot belangrijke maatschappelijke goederen (de waarde van toegankelijkheid). Op dit punt bestaat er een nauwe relatie met de hiervoor genoemde kernwaarden van menselijke waardigheid, autonomie, identiteit en integriteit. Het ontbreken van gelijkwaardige en inclusieve toegang tot belangrijke voorzieningen maakt het immers moeilijk om aan die kernwaarden tegemoet te komen.

De noties van gelijkheid voor de wet, gelijkwaardigheid, inclusiviteit en toegankelijkheid betekenen niet dat mensen altijd volledig gelijk moeten worden behandeld (Vgl. Altman 2015). Er kunnen goede redenen zijn om rekening te houden met de verschillen tussen mensen of de situaties waarin ze zich bevinden. Ongelijke behandeling kan zelfs 'eerlijker' zijn dan gelijke behandeling als daarmee beter tegemoet wordt gekomen aan de individuele wensen, behoeften of capaciteiten. Als de verschillen tussen mensen niet terdege in besluitvorming worden betrokken, kan het gevolg bovendien zijn dat die mensen in feite alsnog worden buitengesloten.

Het recht op 'gelijke behandeling' en non-discriminatie impliceert dus vooral een 'faire' behandeling die is gebaseerd op objectieve, zoveel mogelijk rationele gronden, waarbij rekening wordt gehouden met de waarden van gelijkwaardigheid, inclusiviteit en toegankelijkheid. Is eenmaal de keuze gemaakt om bepaalde gevallen of personen op een bepaalde, gelijkwaardige manier te behandelen, dan geldt vervolgens het vereiste van gelijkheid voor de wet.

Tot slot is van belang om te benoemen dat codificaties van het gelijkheidsbeginsel er meestal van uitgaan dat een nadelige behandeling die is ingegeven door bepaalde, expliciet benoemde persoonskenmerken ('beschermde persoonskenmerken') in beginsel niet toelaatbaar is. Reden daarvoor is dat besluiten die op die persoonskenmerken zijn gebaseerd meestal niet zijn ingegeven door objectieve, neutrale overwegingen, maar samenhangen met bias of met onjuiste of al te brede stereotypen of vooroordelen. Ook kunnen dergelijke besluiten een uitvloeisel zijn van diepgewortelde patronen van systematische achterstelling en discriminatie van bepaalde groepen.

Een besluit dat (direct of indirect) op een beschermd persoonskenmerken is gebaseerd is daarmee 'verdacht', wat betekent dat het alleen toelaatbaar is als overtuigend aannemelijk kan worden gemaakt dat er goede en objectieve redenen voor het besluit bestaan. Veelgenoemde beschermde persoonskenmerken zijn ras, nationaliteit, etniciteit, geslacht/gender, seksuele geaardheid of gerichtheid, godsdienst, levensovertuiging, politieke gezindheid, geboorte (wettig/onwettig/geadopteerd…), burgerlijke staat, handicap, chronische ziekte en leeftijd. Deze lijst is zeker niet uitputtend; verschillende codificaties bevatten verschillende overzichten. Bovendien kan de lijst in de loop van de tijd worden aangevuld als het inzicht groeit dat ook andere groepen beschermd moeten worden tegen achterstelling en benadeling.

Gelijkheidsrechten en aspecten van deze grondrechten

17. Recht op gelijkheid voor de wet	■ Gelijke toepassing van algemene wetgeving op iedereen die onder haar bereik valt ■ Verbod van willekeur ■ Consistentievereiste ■ Rechtszekerheidsbeginsel
18. Verbod van direct onderscheid op bepaalde gronden	■ Beslissingen of regels mogen niet in doorslaggevende mate zijn ingegeven door of gebaseerd op beschermde persoonskenmerken
19. Verbod van indirect onderscheid op bepaalde gronden	■ Beslissingen of regels mogen niet disproportioneel benadelend uitpakken voor personen die behoren tot groepen met beschermde persoonskenmerken
20. Verbod van discriminatoir gemotiveerd handelen	■ Verbod van racistisch/xenofoob etc. gemotiveerd handelen (bijv. geweldpleging) ■ Verbod van het geven van opdracht tot discriminatie
21. Recht op materieel onderscheid / maatwerk	■ Plicht om rekening te houden met verschillen tussen mensen en groepen
22. Recht op redelijke accommodatie / positieve actie	■ Recht op voorzieningen voor mensen met een handicap die hen in staat stellen gelijkwaardig in de samenleving te participeren ■ Recht op compenserende maatregelen voor in het verleden ontstane structurele maatschappelijke ongelijkheid
23. Verbod van profilering	■ Verbod van het creëren van categorieën of profielen op basis van beschermde persoonskenmerken, die vervolgens de basis vormen voor besluitvorming of beleid
24. Verbod van segregatie	■ Verbod van ruimtelijke of andere vormen van scheiding van groepen die daarbij wel een vergelijkbare behandeling krijgen

4. Procedurele grondrechten

Inleiding

Procedurele grondrechten, zoals het recht op een effectief rechtsmiddel en het recht op een eerlijk proces voor een onafhankelijke en onpartijdige rechter, zijn wezenlijk om geschillen op een effectieve en objectieve manier te kunnen oplossen en om rechtsherstel te kunnen bieden bij aantasting van individuele belangen. Daarnaast zijn deze rechten essentieel bij de controle op de uitoefening van wetgevende en uitvoerende bevoegdheden door de overheid. Zeker als grondrechten zijn aangetast, is het belangrijk dat een objectieve derde dit kan vaststellen en dat deze eventueel een besluit kan vernietigen of een wettelijke regeling buiten toepassing kan verklaren.

Dit betekent dat procedurele grondrechten deels instrumenteel van aard zijn, in die zin dat ze helpen om de hiervoor besproken, meer materiële grondrechten te realiseren. Tegelijkertijd kan worden aangenomen dat procedurele grondrechten ook intrinsiek bepaalde kernwaarden representeren, in het bijzonder waarden als eerlijkheid, openheid, evenwichtigheid, objectiviteit en procedurele rechtvaardigheid.

Om deze kernwaarden te kunnen verwezenlijken is het uiteraard nodig dat er toegang is tot een effectief rechtsmiddel; dat kan een rechter zijn, maar in voorkomende gevallen ook een ander onafhankelijk instituut, zoals een toezichthouder of een mensenrechteninstituut. Als toegang tot een onafhankelijke en onpartijdige rechter wordt geboden, moet de procedure voor deze rechter aan een heel aantal eisen voldoen, variërend van berechting binnen redelijke termijn tot ‘equality of arms’. Een aantal van deze eisen en waarborgen geldt daarbij specifiek voor strafrechtelijke procedures, zoals de onschuldpresumptie of het verbod van terugwerkende kracht van strafwetgeving. Elementen daarvan kunnen in voorkomende gevallen echter ook een rol spelen in andere procedures, zeker waar het bijvoorbeeld gaat om (reparatoire) handhavingsmechanismen.

Van oudsher wordt aangenomen dat de procedurele rechten met name betekenis hebben voor de zogenaamde ‘contentieuze fase’, dus de fase waarin een besluit is genomen en iemand daarover wil klagen, of de fase waarin er een geschil is ontstaan tussen twee civiele partijen. In toenemende mate is echter aanvaard dat procedurele rechten en waarborgen ook al in de ‘precontentieuze’ of besluitvormingsfase moeten worden geboden. In het bijzonder is er een tendens om bepaalde algemene beginselen van behoorlijk bestuur te beschouwen als grondrechten; in het EU-Grondrechtenhandvest is een recht op behoorlijk bestuur zelfs expliciet gecodificeerd (in artikel 41). Dit betekent dat ook bijvoorbeeld beginselen van transparantie, motivering en zorgvuldigheid van besluitvorming kunnen worden beschouwd als procedurele grondrechten. Dit is voor algoritmes uiteraard van bijzonder belang, nu kwesties rondom uitlegbaarheid en kenbaarheid van algoritmische besluitvormingsprocessen daarbij regelmatig spelen. In het IAMA hebben wij dergelijke beginselen een aparte plaats toegekend in deel 1. Dit betekent dat bij het doorlopen van de grondrechtentoets in deel 4 vooral de ‘klassieke’ procedurele rechten voor de contentieuze fase van belang zijn, al is het altijd waardevol om daarnaast na te gaan of bij inzet van een algoritme het recht op behoorlijk bestuur voldoende kan worden verzekerd.

Procedurele grondrechten en aspecten van deze grondrechten

25. Recht op behoorlijk bestuur (precontentieuze fase)	■ Recht op transparantie en informatievoorziening ■ Participatie- en verdedigingsrechten (bijv. hoorrechten) ■ Recht op zorgvuldige besluitvorming ■ Motiveringsplichten ■ Verbod van willekeur ■ Verbod van abus/détournement de pouvoir
26. Recht op een effectief rechtsmiddel en toegang tot de rechter (contentieuze fase)	■ Bevoegdheid tot bieden van effectief rechtsherstel ■ Recht op toegang tot een overheidsrechter ■ Recht op inhoudelijke behandeling ■ 'Full jurisdiction' (gehele zaak moet kunnen worden beoordeeld) ■ Ius de non evocandi (recht om niet te worden afgehouden van de rechter die iemand in een bepaald geschil toekomt) ■ Verbod van hoge drempels (griffierecht, bijstand door advocaat, immuniteiten, termijnen) ■ Recht op gefinancierde rechtsbijstand ■ Recht op effectieve tenuitvoerlegging rechterlijke uitspraak
27. Recht op een onafhankelijke en onpartijdige rechter	■ Persoonlijke onafhankelijkheid (bijv. ambtsduur) ■ Zakelijke/functionele afhankelijkheid (bescherming tegen druk van buitenaf) ■ Institutionele onafhankelijkheid ■ Subjectieve onpartijdigheid (geen betrokkenheid bij een van de partijen) ■ Objectieve onpartijdigheid (geen legitieme twijfel mogelijk over onbevooroordeelde beoordeling)
28. Recht op beoordeling binnen redelijke termijn	■ Recht op mogelijkheden tot versnelling van de procedure ■ Recht op compensatie bij te lange duur van procedure
29. Recht op een eerlijk proces: algemeen	■ Recht op een procedure op tegenspraak ■ Recht op equality of arms (voorbereidingstijd, toegang tot dossiers/documenten) ■ Recht op hoor en wederhoor ■ Recht op evenwichtige en faire bewijsregels ■ Recht op gelijke mogelijkheden tot het horen getuigen of deskundigen ■ Recht op openbaarheid van de zitting en het vonnis ■ Verschoningsrechten ■ Motiveringsplicht voor de rechter ■ Recht op consistente rechtspraak ■ Recht op een eerlijk proces: strafrechtelijke waarborgen ■ Onschuldpresumptie ■ Zwijgrecht/recht om niet mee te werken ■ Uitlokverbod ■ Recht op ne bis in idem ■ Recht op bijstand door een advocaat ■ Recht op bijstand door een tolk
30. Strafrechtelijk legaliteitsvereiste (geen straf zonder wet)	■ Verbod op terugwerkende kracht van strafwetgeving ■ Lex-certa-vereiste ■ Lex-mitior-vereiste

Literatuur

Algemene Rekenkamer. (2021). Aandacht voor algoritmes.
https://www.rekenkamer.nl/publicaties/rapporten/2021/01/26/aandacht-voor-algoritmes

Autoriteit Persoonsgegevens. (2025). Handvatten betekenisvolle menselijke tussenkomst.
https://www.autoriteitpersoonsgegevens.nl/documenten/handvatten-betekenisvolle-menselijke-tussenkomst

Council of Europe, Committee on Artificial Intelligence. (2024). Methodology for the risk and impact assessment of artificial intelligence systems from the point of view of human rights, democracy and the rule of law (HUDERIA methodology). (CAI(2024)16rev2). Council of Europe.

Gerards, J. H. (2023). General principles of the European Convention on Human Rights (2nd ed.). Cambridge University Press.

Gerards, J. H. (Ed.). (2023). Fundamental rights: The European and international dimension. Cambridge University Press.
https://doi.org/10.1017/9781009255721

Gerards, J. H., & Xenidis, R. (2021). Algorithmic discrimination in Europe: Challenges and opportunities for equality and non-discrimination law. European Equality Law Network/European Commission.
https://op.europa.eu/en/publication-detail/-/publication/082f1dbc-821d-11eb-9ac9-01aa75ed71a1

Kamphorst, B., Muis, I., Straatman, J., & Schäfer, M.T. (2025). AI auditing through performance appraisals: A practice-informed approach. AI & Society. https://doi.org/10.1007/s00146-025-02674-3

Koops, B. J., Newell, B. C., Timan, T., Škorvánek, I., Chokrevski, T., & Galič, M. (2017). A typology of privacy. University of Pennsylvania International Law Review, 38(2), 483–575. https://scholarship.law.upenn.edu/jil/vol38/iss2/4/

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X. H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing tasks. arXiv.
https://arxiv.org/abs/2506.08872

Meijer, A., Ruijer, E., Postma, R., De Graaf, L., Goossens, J., & Muis, I. (2025). Handleiding: Het CODIO-instrument
https://www.rijksoverheid.nl/documenten/rapporten/2025/01/31/handleiding-het-codio-instrument

Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. (2025). Handreiking algoritmeregister: Aan de slag met het algoritmeregister.
https://aienalgoritmes.pleio.nl/wiki/view/19bb6e9e-7a97-43d5-bef3-b1d66e59f4ff/handreiking-algoritmeregister

Muis, I., Kamphorst, B., & Straatman, J. (2025). Responsible AI innovation in the public sector: Lessons from and recommendations for facilitating fundamental rights and algorithm impact assessments. Journal of Responsible Technology.
https://doi.org/10.1016/j.jrt.2025.100118

Nickel, J. W. (2007). Making sense of human rights (2nd ed.). Blackwell.

Straatman, J., Muis, I., Bössenecker, G., Koole, R., Draijer, R., Van Deijck, N., & Van Vledder, I. (2024). IAMA in actie: Lessons learned van 15 IAMA-trajecten bij Nederlandse overheidsorganisaties. Universiteit Utrecht & Rijks ICT Gilde.
https://www.rijksoverheid.nl/documenten/rapporten/2024/06/20/iama-in-actie-lessons-learned-van-15-iama-trajecten-bij-nederlandse-overheidsorganisaties

Vetzo, M., Gerards, J. H., & Nehmelman, R. (2018). Algoritmes en grondrechten. Boom Juridische uitgevers.