3 Verplichtingen en richtlijnen

3.1 Introductie

Dit hoofdstuk geeft een overzicht van de belangrijkste verplichtingen en richtlijnen voor de verantwoorde inzet van algoritmen. In lijn met de ethische richtsnoeren die mede ten grondslag liggen aan de AI Verordening betekent het dat die inzet:

A. wettig is, door te voldoen aan alle toepasselijke wet- en regelgeving
B. ethisch is, door naleving van ethische beginselen en waarden te waarborgen, en
C. robuust is uit zowel technisch als sociaal oogpunt.

Deze drie componenten zijn verder uitwerkt in zeven thema's, in lijn met publieke waarden en grondrechten zoals menselijke controle, rechtvaardigheid en non-discriminatie.

Deze onderverdeling in de zeven thema's is in dit hoofdstuk aangehouden. Per thema zijn de belangrijkste normen, risico's en maatregelen opgenomen.

Het gaat hier om zowel verplichte als niet-verplichte normen en maatregelen om de inzet van algoritmen te reguleren. Verplichte normen en maatregelen zijn voorschriften die wettelijk bindend zijn en moeten worden nageleefd bij de inzet van algoritmen. Ze hebben een dwingend karakter en het niet naleven ervan kan juridische consequenties hebben. Niet-verplichte normen en maatregelen zijn richtlijnen, aanbevelingen of best practices die niet juridisch bindend zijn, maar worden aanbevolen als goede praktijken bij de inzet van algoritmen. In onderstaand overzicht van normen en maatregelen zijn verplichtingen te herkennen aan de '(*)' die is toegevoegd.

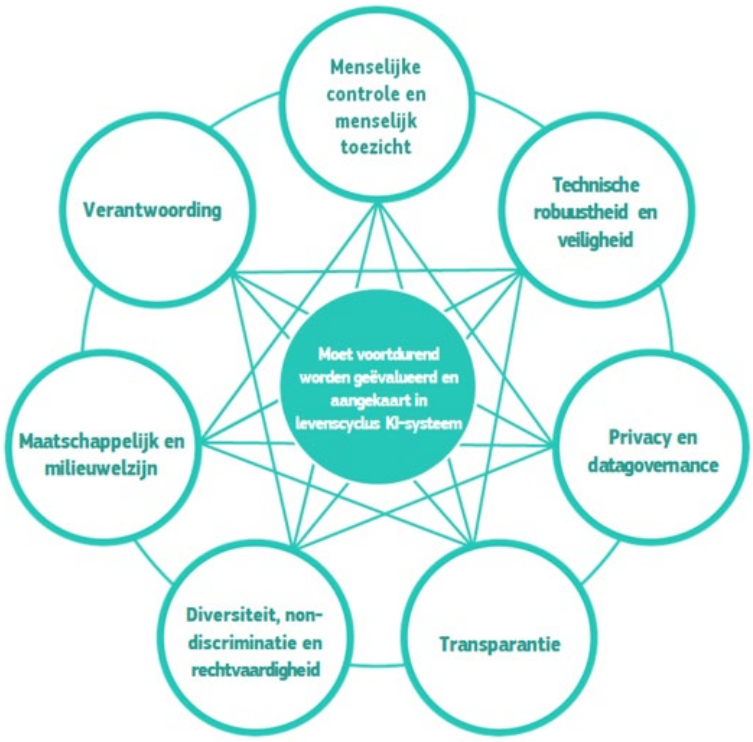
Onderstaand overzicht is zoals gezegd niet bedoeld als 'checklist', maar als tool voor de verantwoorde inzet van algoritmen. Het blijft altijd nodig om zelf tot een afweging te komen en om waar nodig advies in te winnen.

Het overzicht zal in de vorm van een continu proces verder verrijkt en aangepast worden als gevolg van juridische, technologische of organisatorische ontwikkelingen.

De zeven thema's zijn:

1. Menselijke controle en menselijk toezicht
   Omvat grondrechten, menselijke controle en menselijk toezicht.
2. Technische robuustheid en veiligheid
   Omvat weerbaarheid tegen aanvallen en beveiliging, continuïteit en algemene veiligheid, nauwkeurigheid, betrouwbaarheid en reproduceerbaarheid.
3. Privacy en data governance
   Omvat respect voor privacy, de kwaliteit en integriteit van gegevens en toegang tot gegevens.

4. Transparantie
   Omvat traceerbaarheid, verklaarbaarheid en communicatie.
5. Diversiteit, non-discriminatie en rechtvaardigheid
   Omvat het voorkomen van onrechtvaardige vertekening, toegankelijkheid en universeel ontwerp en participatie van belanghebbenden.
6. Maatschappelijk en milieuwelzijn
   Omvat duurzaamheid en milieuvriendelijkheid, sociale gevolgen, samenleving en democratie.
7. Verantwoording
   Omvat controleerbaarheid, minimalisering en verslaglegging van negatieve gevolgen, afwegingen en beroep.



Figuur 3 Onderlinge verhouding van de zeven thema's (overgenomen uit de Ethische richtsnoeren)

Centraal bij de thema's staat de continue evaluatie van de inzet en de governance daaromheen. Naast de normen en maatregelen zijn daarom ook de organisatorische aspecten van de inzet van algoritmen van belang, zoals bewustzijn, cultuur, governance en mogelijkheden voor ethische reflectie. Deze aspecten zijn verder uitgewerkt in hoofdstuk 4 van dit implementatiekader.

3.2 Menselijke controle

Deze paragraaf bevat de verplichting om bij algoritmen een gepaste mate van menselijke controle te borgen voor het specifieke algoritme en de specifieke gebruikssituatie.

3.2.1 Norm: Er is een gepaste mate van menselijke controle voor het specifieke algoritme en de specifieke gebruikssituatie.

Risico: Gebruikers kunnen geen onderbouwde, autonome beslissingen nemen ten aanzien van een algoritmisch systeem.

Maatregel: Zorg dat bij uitsluitend geautomatiseerde besluitvorming wordt voldaan aan de vereisten die volgen uit artikel 22 van de AVG. Zie hierbij norm 3.4.10

Zorg dat goede toezichtmechanismen zijn gewaarborgd. Deze kunnen worden bereikt door middel van human-in-the-loop (HITL), human-on-the-loop (HOTL) en human-in-command (HIC) benaderingen.

- HITL: het vermogen van de mens om in te grijpen in elke beslissingscyclus van het systeem dat wordt gemonitord. Dit betekent dat bij elke beslissing een mens actief betrokken is geweest.
- HOTL: menselijke interventie tijdens de ontwerp- en monitoringcycli van het op AI gebaseerde systeem. De rol van de mens hierbij is die van toezichthouder op een in principe automatisch verlopend proces, waarbij de mens de mogelijkheid heeft om in grijpen
- HIC: het vermogen van de mens om toezicht te houden op de algehele activiteit van het AI-systeem, inclusief de bredere economische, maatschappelijke, juridische en ethische gevolgen ervan, en ervoor te zorgen dat beslissingen die door het AI-systeem worden geproduceerd, door de mens kunnen worden overschreven. De ultieme controle over het proces blijft in deze opzet altijd in handen van een mens.

Afhankelijk van de toepassing die wordt overwogen, kunnen mechanismen worden ontworpen die een van de bovengenoemde niveaus van menselijk toezicht ondersteunen. De tot nu toe voorgestelde methoden zijn grotendeels domein-specifiek, omdat gebruikers-algoritme-interfaces variëren afhankelijk van de mogelijkheden en achtergrond van de op AI gebaseerde oplossing

Indien het algoritme of de gebruikssituatie zelflerend of autonoom is, dan is het van belang om specifiekere controle- en toezichtmechanismen in te stellen:

- Zorg voor detectie- en responsmechanismen om te controleren of er iets mis kan gaan
- Zorg voor een "stopknop" of procedure waarmee een activiteit indien nodig veilig kan worden afgebroken.
- Beoordeel of met deze procedure het proces volledig wordt afgebroken, het gedeeltelijk wordt afgebroken of de controle wordt overgedragen aan een mens

3.3 Technische robuustheid en veiligheid

Dit onderdeel omvat weerbaarheid tegen aanvallen en beveiliging, continuïteit en algemene veiligheid, nauwkeurigheid, betrouwbaarheid en reproduceerbaarheid. Dit is uitgewerkt in vijf normen, met daaronder een aantal concrete risico's en maatregelen. De eerste norm verwijst naar de verplichte toepassing van de Baseline Informatiebeveiliging Overheid (BIO). De daaropvolgende normen hebben betrekking op de kwaliteit van de data en het algoritme, de herhaalbaarheid van de uitkomsten en inbedding van het algoritme in de beheerorganisatie.

3.3.1 Norm: De informatiebeveiliging is op orde. (*)

Risico: Wanneer de informatiebeveiliging m.b.t. de omgeving waarin het algoritme ontwikkeld en gebruikt wordt onvoldoende op orde is, brengt dit grote risico's met zich mee voor de beschikbaarheid, de integriteit en de vertrouwelijkheid van gegevens.

Maatregel: Zorg ervoor dat zowel bij de ontwikkeling van het algoritme als bij het gebruik ervan wordt voldaan aan de eisen uit de BIO. (*)

Toelichting: De gebruikte technische infrastructuur, de inrichting en het beheer ervan moeten voldoen aan de daarvoor geldende eisen op het gebied van informatiebeveiliging. Voor de overheid zijn deze vastgelegd in de BIO. De BIO is in zijn geheel van toepassing.

Bron: Baseline Informatieveiligheid Overheid.

3.3.2 Norm: De data die worden gebruikt is eenduidig en representatief voor de populatie waarop het algoritme wordt toegepast.

1. Risico: Door training van het model met een niet-evenwichtige dataset presteert het model in de praktijk minder goed dan bij de tests.

Maatregel: Leg vast welke keuzes zijn gemaakt bij het samenstellen van de datasets waarmee het algoritme getraind is en bij de waarborging van de kwaliteit ervan. Omschrijf de populatie waarop het algoritme wordt toegepast dusdanig duidelijk dat deze statistisch gevalideerd kan worden. Leg statistieken vast van inputdatasets die zijn gebruikt bij de onderbouwing van data- en modelkeuzes. Bij de keuze voor training- en testdata in de ontwikkelfase moet zowel under- als overfitting worden voorkomen.

Toelichting: De dataset waarmee het algoritme wordt getraind moet passend zijn bij het beoogde gebruik ervan. Trainings-, test- en validatiedata moeten zijn afgestemd op de populatie waarop het algoritme wordt toegepast, inclusief daarin voorkomende subgroepen.

Bron: EC/AI HLEG April 2019 - hoofdstuk II. 1.2, 2.1§100.

2. Risico: De input- en outputdata voldoen qua kwaliteit, volledigheid en betrouwbaarheid niet aan de functionele eisen, waardoor verkeerde beslissingen kunnen worden genomen.

Maatregel: Stel documentatie op waarin de eisen aan de input- en outputdata zijn opgenomen en implementeer structurele controles om dit te toetsen. Ga voor alle inputdata na of deze tijdig, volledig en definitief beschikbaar is, een herleidbare afkomst heeft, consistent is gecodeerd en duidelijke metadata bij de variabelen bevat. Kijk daarnaast naar het risico van datavervuiling in het proces.

Toelichting: Als input- en outputdata van onvoldoende kwaliteit zijn of een andere betekenis hebben dan verwacht, dan zijn de uitkomsten van het model onbetrouwbaar, wat kan leiden tot verkeerde beslissingen.

Bron: EC/AI HLEG April 2019 - hoofdstuk II. 1.3.

3.3.3 Norm: Het algoritme dat wordt gebruikt is geschikt voor het doel waarvoor het wordt ingezet.

1. Risico: Het algoritme dat wordt gebruikt is niet geschikt voor het doel waarvoor het wordt ingezet.

Maatregel: Documenteer de grenzen van de toepasbaarheid van het algoritme en de voorwaarden waaronder het kan worden gebruikt.

Toelichting: Het moet duidelijk zijn onder welke voorwaarden een algoritme wel of niet gebruikt kan worden. Denk hierbij bijvoorbeeld aan eisen voor de verdeling van data of afwijkingen van de productieset ten opzichte van de testset. Deze voorwaarden moeten dusdanig helder zijn gedocumenteerd dat deze informatie ook toegankelijk is voor medewerkers die niet bij de ontwikkeling van het algoritme betrokken waren.

Bron: EC/AI HLEG April 2019 - hoofdstuk II. 1.1, 1.2.

2. Risico: De hyperparameters zijn niet goed gekozen, waardoor niet goed gestuurd kan worden op het leerproces en het model suboptimaal presteert.

Maatregel: Leg de keuze voor de hyperparameters vast en onderbouw deze, bijvoorbeeld in Git en/of het technisch of functioneel ontwerp. Het uitvoeren van een peer review (vierogenprincipe) kan hiervan onderdeel zijn.

Toelichting: Een hyperparameter is een parameter waarmee kan worden gestuurd op een trainings- en leerproces.

Bron: EC/AI HLEG April 2019 - hoofdstuk II. 1.2.

3.3.4 Norm: De uitkomsten van het algoritme zijn eenduidig en betrouwbaar.

1 Risico: Als trainings-, validatie- en testdata door elkaar lopen ("data leakage"), kan dit leiden tot overfitting, waardoor het model beter lijkt te presteren dan in werkelijkheid het geval is.

Maatregel: Scheid de datasets voor training, validatie en testen en leg vast welke datasets voor welk doel gebruikt zijn. Let in het bijzonder op wanneer het model geüpdatet wordt aan de hand van (deels) eerder gebruikte data.

Toelichting: Vermenging van trainings-, validatie- en testdata kan leiden tot overfitting van het model, waardoor het goed werkt voor de data waarmee het is getraind, maar niet geschikt is voor nieuwe observaties. Het gevolg van overfitting is dat prestaties van het model worden overschat. Dit komt doordat het lijkt alsof wordt getest op basis van data die het model nog niet eerder heeft gezien, terwijl in werkelijkheid wordt getest met data waarmee het model ook is getraind.

Bron: EC/AI HLEG April 2019 - hoofdstuk II. 1.2, 2.1§100.

2. Risico: Het doel van het algoritme, de functionele eisen van het model en de performance metrics sluiten onvoldoende op elkaar aan, waardoor niet kan worden beoordeeld of het voldoende presteert en of de kwaliteit ervan op orde is.

Maatregel: Documenteer alle prestatiecriteria, inclusief de relatie met de doelstellingen en de functionele eisen van het algoritme. Bewaar testresultaten waaruit blijkt dat het algoritme aan de criteria voldoet.

Toelichting: De functionele eisen moeten zijn uitgewerkt in meetbare eisen, onder andere met betrekking tot de prestaties, robuustheid, bias en uitlegbaarheid van het model. Deze eisen moeten passen bij de modelspecifieke context.

Bron: EC/AI HLEG April 2019 - hoofdstuk II. 1.1 t/m 1.6.

3. Risico: Door veranderingen in de data presteert het algoritme niet meer zoals verwacht.

Maatregel: Leg voor elk algoritme de kwaliteits- en prestatiedoelen vast. Stel een procesbeschrijving op voor het monitoren hiervan. Evalueer bij veranderingen in de data of het algoritme nog aan de vastgestelde doelen voldoet en neem indien nodig maatregelen.

Toelichting: Veranderingen in de data kunnen ertoe leiden dat de prestaties van het algoritme achteruitgaan. Wanneer veranderingen in de data niet direct duidelijk zijn, zal periodiek een evaluatie moeten plaatsvinden.

Bron: EC/AI HLEG April 2019 - hoofdstuk II. 1.2, 1.3.

4. Risico: Beslissingen worden genomen op basis van outputdata die verkeerd zijn begrepen.

Maatregel: Zorg voor een eenduidige beschrijving van alle variabelen. Neem indien nodig extra toelichting op in de metadata.

Toelichting: De interpretatie van de uitkomsten van het model moet eenduidig zijn en onafhankelijk van de persoon die de beoordeling uitvoert.

Bron: EC/AI HLEG April 2019 - hoofdstuk II. 1.4.

3.3.5 Norm: De continuïteit van het algoritme is gewaarborgd.

1 Risico: De organisatie is voor de data en/of het gebruik van het algoritme afhankelijk van derden en kan daardoor de reproduceerbaarheid, het prestatieniveau en de continuïteit ervan niet garanderen.

Maatregel: Alle gebruikte data moeten traceerbaar of reproduceerbaar zijn. In geval van uitbesteding van het beheer aan derden moeten hierover heldere afspraken gemaakt worden gemaakt.

Toelichting: De data en het model zijn bij voorkeur in eigen beheer. Wanneer dit niet mogelijk is, moeten afspraken zijn gemaakt om de functionele eisen die hieraan gesteld zijn te waarborgen.

Bron: EC/AI HLEG April 2019 hoofdstuk II 1.7.

2 Risico: Er vindt na ingebruikname van het algoritme onvoldoende monitoring plaats op de werking ervan, waardoor fouten of ongewenste effecten in de toepassing van het algoritme niet of niet tijdig worden opgemerkt.

Maatregel: Richt een proces in rondom monitoring van het algoritme.

Toelichting: Na ingebruikname van een algoritme moet periodiek worden beoordeeld of het nog doet wat het zou moeten doen. Denk hierbij aan monitoring op beschikbaarheid, prestaties/kwaliteit en compliance met (actuele) wet- en regelgeving.

Bron: COBIT APO11 / BAI04 / DSS04

3. Risico: Er is onvoldoende capaciteit beschikbaar in de beheerorganisatie, waardoor benodigde aanpassingen op het algoritme niet of niet tijdig worden doorgevoerd.

Maatregel: Zorg voor heldere afspraken op het gebied van onderhoud en beheer op het algoritme, o.a. met betrekking tot de technische

componenten, de gebruikte data, het model en de daarin gebruikte parameters.

Toelichting: Het risico bestaat dat bij het in productie nemen van het algoritme onvoldoende aandacht wordt besteed aan de overdracht aan de beheerorganisatie. Gevolg hiervan kan zijn dat in de beheerorganisatie onvoldoende capaciteit en/of kennis van het algoritme beschikbaar is om eventuele aanpassingen tijdig door te voeren.

Bron: COBIT APO09 / APO14 / BAI06.

3.4 Privacy en datagovernance

Als persoonsgegevens worden verwerkt en is vastgesteld dat de AVG van toepassing is[14], dan moet de algoritmische toepassing in overeenstemming zijn met de vereisten uit de AVG. Hierbij geldt dat een verwerking van persoonsgegevens altijd aan de eisen van proportionaliteit en subsidiariteit moet voldoen. Voor verwerkingsverantwoordelijken die in het kader van een taak in het publieke belang persoonsgegevens verwerken, geldt bovendien dat in overeenstemming met de algemene beginselen van behoorlijk bestuur moet worden gehandeld. Deze paragraaf geeft een overzicht van twaalf normen die hun grondslag vinden in de AVG. Naast de beoordeling van de normen en risico's die aan de verwerking verbonden zijn, kan een DPIA verplicht zijn om de beoordeling en de getroffen maatregelen te documenteren. Er is een handreiking gemaakt die uitlegt wat de AVG betekent voor partijen die persoonsgegevens verwerken. En aan welke regels deze gegevensverwerking moet voldoen.[15]

3.4.1 Norm: Rollen en verantwoordlijkheden m.b.t. de verwerking van persoonsgegevens zijn gespecificeerd. (*)

Risico: Rollen en verantwoordelijkheden m.b.t. de verwerking van persoonsgegeven in het algoritme zijn niet helder belegd. (*)

Maatregel:
Ga van alle betrokken partijen na wat hun rol is en documenteer deze. Zorg ervoor dat met partijen duidelijke afspraken zijn gemaakt.

Toelichting:
De verplichtingen uit de AVG zijn van toepassing op de verwerkingsverantwoordelijke.

Wanneer een partij samen met een of meerdere andere partijen de doelen en essentiële middelen bepaalt voor de verwerking, dan is er sprake van gezamenlijke verwerkingsverantwoordelijkheid. Bij gezamenlijke verantwoordelijkheid moeten de partijen onderling duidelijke afspraken maken over wie invulling geef aan de diverse rechten en plichten uit de AVG.

[14] Zoals aangegeven in paragraaf 1.2 zijn er bepaalde sectoren waar een afwijkend wettelijk kader geldt. Op deze organisaties is dit hoofdstuk niet van toepassing, maar moet worden voldaan aan de respectievelijke geldende wettelijke kaders zoals de Wpg en de Wjsg.
[15] Handleiding Algemene Verordening Gegevensbescherming (AVG) | Rapport | Rijksoverheid.nl

Verwerkingsverantwoordelijken schakelen regelmatig personen of organisaties in die voor hen persoonsgegevens verwerken. De verwerker verwerkt persoonsgegevens uitsluitend ten behoeve van de verwerkingsverantwoordelijke. De taken van de verwerker jegens de verwerkingsverantwoordelijke moeten in een verwerkersovereenkomst worden gespecificeerd.

Bron: Artikel 24,26,27, 28,29 AVG.

3.4.2 *Norm: Een gegevensbeschermingseffectbeoordeling / Data Protection Impact Assessment (GEB / DPIA) is verplicht, indien een verwerking van persoonsgegevens waarschijnlijk een hoog risico inhoudt voor de rechten en vrijheden van natuurlijke personen. (*)*

Risico: Bij de verwerking van persoonsgegevens zijn de risico's voor de rechten en vrijheden van betrokkenen niet bekend en niet gemitigeerd.

Maatregel:
Beoordeel of de gegevensverwerking waarschijnlijk een hoog risico voor de rechten en vrijheden oplevert voor de mensen van wie gegevens worden verwerkt. Met behulp van bijvoorbeeld de Pre-scan DPIA[16] of de lijst verplichte DPIA van de AP[17] kan worden vastgesteld of het uitvoeren van een DPIA noodzakelijk is.

Indien blijkt dat de gegevensverwerking waarschijnlijk een hoog privacyrisico oplevert, dan moet een DPIA worden uitgevoerd.(*) De Rijksoverheid heeft het Model DPIA Rijksdienst ontwikkeld.[18]

Toelichting:
Een DPIA is een beoordeling van de effecten van een voorgenomen verwerkingsactiviteit op de bescherming van persoonsgegevens en de rechten en vrijheden van betrokkenen. Op basis hiervan worden maatregelen getroffen om deze effecten voor betrokkenen te voorkomen of te verkleinen. Hoewel een DPIA een verplichting is op basis van de AVG en voornamelijk ziet op het beoordelen van privacyrisico's, dient een DPIA dus breder opgevat te worden.

Bron: Artikel 35 AVG.

3.4.3 *Norm: De verwerking van persoonsgegevens is rechtmatig, behoorlijk en transparant. (*)*

Risico: De verwerking van persoonsgegevens gebeurt niet rechtmatig, behoorlijk of transparant.

Maatregel:
De betrokkenen zijn geïnformeerd over de verwerking van persoonsgegevens. Bijvoorbeeld middels een privacyverklaring op de

[16] Producten/diensten - cip-overheid
[17] Besluit lijst verplichte DPIA | Autoriteit Persoonsgegevens
[18] Producten/diensten - cip-overheid

website en indien relevant opgenomen in individuele communicatie naar betrokkene(n). (*)

Organisaties die gegevens verwerken zoals bedoeld in artikel 22 AVG moeten naast de algemene vereisten nuttige informatie over de onderliggende logica verstrekken. Nuttige informatie over de onderliggende logica betreft bijvoorbeeld:

- Dat de verstrekte informatie volledig genoeg moet zijn voor de betrokkene om de redenen van het besluit te kunnen begrijpen.
- Gedetailleerde informatie over de belangrijkste kenmerken die bij de besluitvorming in beschouwing worden genomen, de bron van deze informatie en het belang ervan.

Leg dit vast in de DPIA.

Toelichting:
De inzet van een algoritme is voor de betrokkene vaak onzichtbaar. De mate waarin mensen dergelijke processen begrijpen, verschilt per persoon. Voor sommigen kan het moeilijk zijn de complexe technieken te begrijpen.

Bron: Artikel 5, lid 1, onder a AVG, Artikel 12-14 AVG.

3.4.4 *Norm: Persoonsgegevens mogen alleen voor welbepaalde, uitdrukkelijk omschreven en gerechtvaardigde doeleinden worden verzameld en mogen vervolgens niet verder op een met die doeleinden onverenigbare wijze worden verwerkt (doelbinding)**

Risico: Er worden persoonsgegevens verwerkt zonder dat hiervoor een doel is bepaald. De verwerking van persoonsgegevens in het algoritme valt niet onder het doel waarvoor zij verzameld zijn of een hiermee verenigbaar doel.

Maatregel:
Stel welbepaalde, uitdrukkelijk omschreven en gerechtvaardigde doelen vast voor de verwerking van persoonsgegevens. Check of de verwerking past binnen de wettelijke grondslag die geldt voor het uitvoeren van de taak.

Persoonsgegevens die reeds verzameld zijn mogen wel verder worden verwerkt voor andere doelen, maar alleen als die doelen verenigbaar zijn met het oorspronkelijke verzameldoel. Om te bepalen of een nieuw doel verenigbaar is, moet een verenigbaarheidstoets worden uitgevoerd.

Leg dit vast in de DPIA.

Toelichting:

Uitgangspunt is dat persoonsgegevens alleen mogen worden verwerkt als de verwerking noodzakelijk is met het oog op het bereiken van specifieke doelen. Wanneer het gerechtvaardigd is om persoonsgegevens te verwerken, moet de verwerking ervan vervolgens netjes en verantwoord gebeuren. Ten slotte moet duidelijk zijn voor welke doelen persoonsgegevens worden verwerkt en hoe dat gebeurt.

Bron: Artikel 5, lid 1, onder b AVG en Overweging 50 AVG.

3.4.5 *Norm: Elke verwerking van persoonsgegevens moet gerechtvaardigd zijn.(*)*

Risico: De verwerking is niet gerechtvaardigd omdat het doel van de verwerking niet kan worden gebaseerd op één van de zes rechtsgrondslagen die in de AVG worden gegeven.

Maatregel:
Bepaal of de gegevensverwerking gerechtvaardigd is. De verwerking is gerechtvaardigd wanneer het doel van de verwerking kan worden gebaseerd op één van de zes rechtsgrondslagen die in de AVG worden gegeven. Kan dat niet, dan is het niet toegestaan persoonsgegevens te verwerken. De lijst van rechtsgrondslagen is limitatief, er kunnen geen andere gronden worden aangevoerd.*

Toelichting:
In de AVG staan de volgende 6 grondslagen voor het verwerken van persoonsgegevens:

1. Toestemming van de persoon om wie het gaat.
2. Het is noodzakelijk om persoonsgegevens te verwerken om een overeenkomst uit te voeren.
3. Het is noodzakelijk om persoonsgegevens te verwerken omdat dit wettelijk verplicht is.
4. Het is noodzakelijk om gegevens te verwerken om vitale belangen te beschermen.
5. Het is noodzakelijk om gegevens te verwerken om een taak van algemeen belang of openbaar gezag uit te oefenen.
6. Het is noodzakelijk om gegevens te verwerken om gerechtvaardigde belang te behartigen

Wanneer er sprake is van een afhankelijkheidsrelatie, bijvoorbeeld in de relatie overheid-burger, zal toestemming niet snel vrij zijn gegeven. Dit betekent dat dergelijke situaties geen geldige toestemming kan worden gevraagd.

Vanwege het feit dat de wetgever de rechtsgrond bepaalt voor de verwerking van persoonsgegevens door overheidsinstanties, is de rechtsgrond 'gerechtvaardigd belang' niet van toepassing op verwerkingen door overheidsinstanties in het kader van de uitvoering van hun taken.

Voor de verwerking van bijzondere categorieën van persoonsgegevens en persoonsgegevens van strafrechtelijke aard gelden aanvullende voorwaarden, deze bepalingen worden beschreven norm 3.4.10.

Leg dit vast in de DPIA.

Bron: Artikel 6 AVG.

3.4.6 *Norm: De verwerkte persoonsgegevens zijn toereikend, ter zake dienend en beperkt tot het noodzakelijke (dataminimalisatie). (*)*

Risico: De verwerkte persoonsgegevens zijn niet proportioneel en relevant in relatie tot het doel.

Maatregel: Beoordeel welke persoonsgegevens voor het algoritme nodig zijn en hoe lang deze moeten worden bewaard. Gebruik waar mogelijk geanonimiseerde of gepseudonimiseerde gegevens. (*)

Leg dit vast in de DPIA

Bron: artikel 5 lid 1 onder c AVG.

3.4.7 Norm: De gegevens zijn juist en zo nodig geactualiseerd. (*)

Risico: De kwaliteit en integriteit van data zijn niet voldoende geborgd.

Maatregel: (verplicht)
Verwerkingsverantwoordelijken moeten in alle stappen van het algoritmeproces juistheid in het oog houden, in het bijzonder bij:

- Het verzamelen van gegevens;
- Het analyseren van gegevens;
- Het opstellen van model; of
- Het toepassen van een model om een besluit met betrekking tot de persoon te nemen

Leg dit vast in de DPIA.

Toelichting: Als de gegevens die in het algoritme worden gebruikt onjuist zijn, zullen ook besluiten of profielen die daaruit voortkomen onjuist zijn. Besluiten zijn mogelijk genomen op basis van verouderde gegevens of de onjuiste interpretatie van externe gegevens. Zelfs wanneer ruwe gegevens correct worden opgeslagen, is de gegevensreeks mogelijk niet geheel representatief of kunnen de analysegegevens ongemerkt vooroordelen bevatten.

Bron: artikel 5 lid 1 sub d AVG.

*3.4.8 Norm: Gegevens worden niet langer worden bewaard dan nodig (opslagbeperking).**

Risico: Persoonsgegevens worden langer bewaard dan nodig.

Maatregel: De bewaartermijnen moeten aansluiten op de selectielijsten (Archiefwet). Zorg voor een vernietigingsprocedure, een document waaruit blijkt dat dit ingeregeld en/of uitgevoerd is.*

Leg dit vast in DPIA.

Toelichting: Persoonsgegevens mogen niet langer bewaard worden dan noodzakelijk voor het doel van de verwerking. Wanneer de gegevens niet langer noodzakelijk zijn, dan moeten zij worden vernietigd of gewist.

Bron: Artikel 5, lid 1 AVG; Archiefwet.

3.4.9 Norm: Het algoritme verwerkt alleen bijzondere persoonsgegevens, strafrechtelijke gegevens of nationale identificatienummers (o.a. BSN) als deze op basis van een wettelijke uitzondering verwerkt mogen worden.

Risico: Verwerking van bijzondere persoonsgegevens (o.a. gegevens m.b.t. ras of afkomst, religie, gezondheid of seksuele geaardheid), strafrechtelijke gegevens of nationale identificatienummers (o.a. BSN) is alleen toegestaan als hierop een wettelijke uitzondering van toepassing is.

Maatregel:
Op het verwerken van bijzondere categorieën van persoonsgegevens rust een verwerkingsverbod. Hierop is echter wel een beperkt aantal uitzonderingen geformuleerd. Een deel van de uitzonderingen is geregeld in de AVG. De UAVG bevat daarnaast specifieke uitzonderingen per categorie. (*)

De verwerking van strafrechtelijke gegevens, is alleen toegestaan als dat gebeurt onder toezicht van de overheid, of als het specifiek bij wet is geregeld.

Nationale identificatienummers mogen alleen worden gebruikt voor in de wet voorgeschreven doelen.

Leg dit vast in de DPIA

Toelichting:
Bijzondere categorieën van persoonsgegevens zijn gegevens die gezien hun aard extra gevoelig zijn. Het gaat specifiek om: gegevens waaruit ras of etnische afkomst blijkt, politieke opvattingen, religieuze of levensbeschouwelijke overtuigingen, het lidmaatschap van een vakbond, genetische gegevens, biometrische gegevens met het oog op de unieke identificatie van een persoon, gegevens over gezondheid en gegevens met betrekking tot iemands seksueel gedrag of seksuele gerichtheid.

Bron: Artikel 9, 10 en 87 AVG, Hoofdstuk 3 UAVG.

3.4.10 Norm: Betrokkenen kunnen een beroep doen op hun privacyrechten. (*)

Risico: Bij het verwerken van persoonsgegevens in een algoritme is voor gebruikers niet duidelijk dat hier sprake van is waardoor ze geen beroep kunnen doen op hun privacyrechten.

Maatregel: Geef aan hoe invulling wordt gegeven aan de rechten van betrokkenen.(*) Stel een procedure vast waarmee o.a. aan verzoekers informatie kan worden verstrekt over geautomatiseerde verwerking van persoonsgegevens en de onderliggende logica en het belang en de verwachte gevolgen van de verwerking.

Indien de rechten van de betrokkene worden beperkt, bepaal dan op grond van welke wettelijke uitzonderingen dat is toegestaan en leg dit vast.

Toelichting: Om een eerlijke verwerking van persoonsgegevens te waarborgen geeft de AVG diverse rechten aan de betrokkene. De

betrokkene kan deze rechten uitoefenen tegen de verwerkingsverantwoordelijke.

Bron: Artikel 15-22 AVG.

3.4.11 *Norm: Betrokkenen hebben het recht om niet onderworpen te worden aan een enkel op geautomatiseerde verwerking, waaronder proflering, gebaseerd besluit, wanneer dit rechtsgevolgen heeft voor hen of het hen anderszins in aanzienlijke mate tref. (*)*

Risico: Een betrokkene ondervindt aanmerkelijke gevolgen door een geautomatiseerd besluit, zonder dat deze een beroep op menselijke tussenkomst kan doen.

Maatregel:
Ga na of sprake is van een verboden vorm van geautomatiseerde verwerking van persoonsgegevens.

Deze maatregelen moeten tenminste het volgende omvatten:
- Het recht op menselijke tussenkomst;
- Het recht voor de betrokkene om zijn standpunt kenbaar te maken; en
- Het recht om het besluit aan te vechten.

Leg dit vast in de DPIA.

Toelichting:
Bij uitsluitend geautomatiseerde individuele besluitvorming is er géén sprake van (noemenswaardige) menselijke tussenkomst.

Bron: artikel 22 AVG en artikel 40 UAVG.

3.4.12 *Norm: Privacy en gegevensbescherming is meegenomen als eis bij het ontwerp van nieuwe systemen waarmee persoonsgegevens worden verwerkt. ('privacy door ontwerp en standaardinstellingen/ Privacy by Design en Privacy by Default)*

Risico: Ontwerp en opzet van het algoritme zijn onvoldoende gericht op de bescherming van privacy.

Maatregel:
Neem technische en organisatorische maatregelen om invulling te geven aan het uitgangspunt van privacy door ontwerp en door standaardinstellingen. Bij het bepalen van de maatregelen moet rekening worden gehouden met de volgende elementen:
- De stand van de techniek;
- De uitvoeringskosten;
- De aard, omvang, context en het doel van de verwerking;
- De risico's voor de betrokkene.

Leg dit vast in de DPIA.

Toelichting:
Privacy door ontwerp en door standaardinstellingen houdt in dat privacy en gegevensbescherming worden meegenomen als eisen bij de ontwikkeling het ontwerp van nieuwe systemen waarmee persoonsgegevens worden verwerkt, waarbij een zo klein mogelijke inbreuk op de persoonlijke levenssfeer wordt gemaakt, bijvoorbeeld door het toepassen van pseudonimisering en het inbouwen van andere technische waarborgen.

Bron: Artikel 25 AVG.

3.5 Transparantie

Deze paragraaf bevat vier onderdelen. Deze onderdelen houden nauw verband met het motiveringsbeginsel en gaan over de transparantie van elementen die relevant zijn voor de inzet van algoritmen. Algoritmen dienen traceerbaar, verklaarbaar en transparant te zijn, zodat achterhaald kan worden op basis van welke afwegingen een besluit is genomen.

3.5.1 Norm: Besluitvorming dient transparant te zijn en moet zorgvuldig tot stand komen.

Risico: Gebrek aan transparantie, waardoor niet kan worden achterhaald of aan een besluit een zorgvuldig proces vooraf is gegaan en of het voldoet aan andere vereisten.

Maatregel:
Algoritmen en modellen dienen transparant, betrouwbaar en controleerbaar te zijn, zodat achterhaald kan worden op basis van welke afwegingen een besluit is genomen. (*)

Richt een procedure in waardoor kan worden voldaan aan het recht op toegang tot publieke informatie.

Toelichting:
Bij de voorbereiding van een besluit dient het bestuursorgaan alle relevante informatie te verzamelen en de belangen van betrokkenen zorgvuldig af te wegen. Het besluit mag geen onredelijke of discriminerende gevolgen voor betrokkenen hebben.

Bron: Artikel 3:2 en 3:46 Awb, artikel 1.1 en 2.5 Woo.

*3.5.2 Norm: Een besluit berust op een deugdelijke motivering.**

Risico: Het is niet duidelijk dat het besluit (gedeeltelijk) op een algoritme is gebaseerd.

Maatregel: Richt een procedure in waarmee bij een besluit dat is gebaseerd op een algoritme uit eigen beweging gemaakte keuzes en gebruikte gegevens en aannames aan de betrokkene inzichtelijk worden gemaakt of worden gedocumenteerd. *

Een bestuursorgaan moet inzichtelijk maken:

1. dat een besluit tot stand is gekomen met behulp van een algoritme;
2. van welke feiten het is uitgegaan en welke gegevens van de burger gebruikt c.q. verwerkt zijn;
3. welke relevante belangen tegen elkaar zijn afgewogen en hoe die afweging is verlopen (bijvoorbeeld het gewicht dat wordt toegekend aan elk afgewogen kenmerk; welke analytische technieken gebruikt zijn; welke specifieke voorspellende data; wat de belangrijkste policy-keuzes waren; een uitleg van het voorspellende algoritme);
4. hoe het algoritme werkt (niet de techniek, maar hoe de uitkomsten van het algoritme tot stand komen).

Ofwel, welke keuzes en aannames zijn gemaakt en welke invoergegevens zijn gebruikt.

Toelichting: Transparantie vormt de kern van maatregelen om grip te krijgen op algoritmische besluitvorming. Bij het bepalen van deze transparantie kan worden aangesloten bij het motiverings- en het zorgvuldigheidsbeginsel. Een besluit dat is gebaseerd op een algoritme dient te berusten op een deugdelijke motivering. De motivering wordt vermeld bij de bekendmaking van het besluit.

Bron: Artikel 3:46 - 3:50 Awb, Jurisprudentie over AERIUS (ABRvS 18 juli 2018, ECLI:NL:RVS:2018:2454), Hof van Justitie C-274/18.

3.5.3 Norm: De besluitvorming door het algoritme is traceerbaar.

Risico: Het is niet goed na te gaan op welke manier het algoritme tot het besluit is gekomen.

Maatregel:
Neem maatregelen om de traceerbaarheid in de besluitvorming te waarborgen. Daarbij kan het gaan om de documentatie van:

- Het ontwerp en de ontwikkeling van door het algoritme gebruikte methoden.
- Het testen en valideren van door het algoritme gebruikte methoden
- Resultaten van het algoritme
- Onderzoek in hoeverre de door het algoritme genomen beslissingen en dus het resultaat kunnen worden begrepen.
- Zorg ervoor dat er voor alle eindgebruikers die een verklaring wensen, een voor hen begrijpelijke verklaring kan worden gemaakt
- Onderzoek in hoeverre de beslissingen van het systeem van invloed zijn op het besluitvormingsproces binnen de organisatie.
- Onderzoek en toon aan welk model het eenvoudigst zou zijn voor de betreffende toepassing.
- Controleer of trainings- en testgegevens kunnen worden geanalyseerd. Kunnen deze in de loop van de tijd worden veranderd en bijgewerkt?
- Onderzoek of na de training en ontwikkeling van het model er mogelijkheden zijn om de interpreteerbaarheid te beoordelen en of er toegang is tot de interne werkstroom van het model.

Toelichting: Indien de totstandkoming van een besluit dat met behulp van een algoritme is genomen niet traceerbaar is, kan het algoritme niet worden gecontroleerd.

Bron: EC/AI HLEG April 2019 - Hoofdstuk II.1.4.

3.5.4 *Norm: De inzet en werking van het algoritme is gepubliceerd in een register en inzichtelijk voor belanghebbenden.*

Risico: Ontbreken transparantie voor burgers/bedrijven/stakeholders (belanghebbenden)

Maatregel:
Opname van het algoritme in het centrale algoritmeregister of op sites als github.com, inclusief beschrijving van werking, gebruikte data en/of beschrijving daarvan. Belanghebbenden worden zo op een begrijpelijke manier geïnformeerd over onderliggende logica van het algoritme, alsmede het belang en de verwachte gevolgen van die verwerking voor de betrokkene.

Toelichting:
Doel van publicatie van informatie over het algoritme is het bieden van transparantie naar betrokkenen. Het zorgt ervoor dat het voor de vooraf bepaalde personen/doelgroepen duidelijk is dat zij met een algoritme te maken hebben, welke consequenties dat heeft en welke beperkingen het algoritme kent. De gewenste mate van transparantie (technische transparantie vs. uitlegbaarheid) is weloverwogen; het hangt af van 1) de impact van het algoritme op de beslissing, uitkomst en burger, (2) de mate van autonomie bij de besluitvorming en (3) het type en de complexiteit van het algoritme. De informatie dient voldoende begrijpelijk te zijn voor de doelgroep(en).

Bron: EC/AI HLEG April 2019 - Hoofdstuk II.1.4.

3.6 Diversiteit, non -discriminatie en rechtvaardigheid

Deze paragraaf bevat vier onderdelen binnen het verbod op discriminatie. De risico's die worden geadresseerd zijn discriminatie op basis van beschermde persoonskenmerken, andere data dan beschermde persoonskenmerken die leiden tot discriminatie, het gebruiken van proxy variabelen die leiden tot discriminatie en bias in het algoritme. Een van de maatregelen is bij het vaststellen van verdacht onderscheid als ultieme tegenmaatregel het (tijdelijk) stopzetten van het algoritme. De normen, risico's en maatregelen voor diversiteit, non-discriminatie en rechtvaardigheid komen voort uit de Grondwet, EVRM, Algemene wet gelijke behandeling.

3.6.1 *Norm: Verbod op ongelijke behandeling in gelijke omstandigheden. Discriminatie wegens godsdienst, levensovertuiging, politieke gezindheid, ras, geslacht of op welke grond dan ook, is niet toegestaan. (*)*

1 Risico:
Toepassing van het model leidt tot discriminatie op basis van beschermde persoonskenmerken.

Maatregel:
Stel vast of bij het doel, design of de uitkomst van het algoritme sprake is van (direct of indirect) onderscheid. Maak bij de uitkomst ook gebruik van controlevariabelen (bijvoorbeeld nationaliteit/ras). Bepaal of er een wettelijke uitzondering of een objectieve rechtvaardiging is. Maak de consequentie expliciet en leg deze op het juiste niveau vast. Neem indien nodig tegenmaatregelen. Een ultieme tegenmaatregel kan zijn het (tijdelijk) stopzetten van het algoritme.

Toelichting:

Er is sprake van een objectieve rechtvaardiging als sprake is van een legitiem doel en het middel dat wordt gebruikt om het doel te bereiken is passend, noodzakelijk en evenredig. Om te bepalen of een algoritme verboden onderscheid maakt, moet worden bekeken of door het algoritme sprake is van een ongelijke behandeling van een persoon of groep personen in verhouding tot anderen in een vergelijkbare situatie.

Bron: Grondwet Art. 1 EVRM Art. 1 en 14 Algemene wet gelijke behandeling, Protocol 12.

Risico:
Andere data dan beschermde persoonskenmerken leiden tot discriminatie in de uitkomsten.

Maatregelen:

- Zorg voor gelijke mate van vertegenwoordiging relevante groepen. Selecteer een afgebakende toepassing om het systeem te testen; zorg dat deze afgebakende toepassing representatief is voor het gehele domein waarop het AI-systeem later wordt ingezet. Maak de consequenties expliciet en leg deze op het juiste niveau vast. Neem indien nodig tegenmaatregelen. Een ultieme tegenmaatregel kan zijn het (tijdelijk) stopzetten van het algoritme.
- Documenteer de mate van afhankelijkheid van historische data. Weeg af of de geconstateerde afhankelijkheid wenselijk is en of deze op discriminatie duidt. Maak de consequenties expliciet en leg deze op het juiste niveau vast. Neem indien nodig tegenmaatregelen. Een ultieme tegenmaatregel kan zijn het (tijdelijk) stopzetten van het algoritme.

Toelichting:

- Onvoldoende representativiteit in de trainingsdata kan leiden tot ongelijke uitkomsten.

- Onvoldoende representativiteit in de trainingsdata kan leiden tot ongelijke uitkomsten. Kan bij voorspellende algoritmen ook leiden tot ongewenste feedbackloops. (Gebruik van data van bijvoorbeeld vroegere surveillance en criminaliteitscijfers in nieuwe algoritmen algoritmen tot link aan wijken, personen met een

immigratieachtergrond. Indien het geval, is dit niet representatief en neutraal)

Bronnen:
Grondwet Art. 1.
EVRM Art. 1 en 14 Algemene wet gelijke behandeling, Protocol 12.
Artikel 9 AVG.
Artikel 14 EVRM jo. 21 HvEU jo. 1 GW.

3 Risico: Het gebruik van proxy variabelen leidt tot indirecte discriminatie.

Maatregel:

Identificeer bij een onrechtmatigheid in de uitkomst van het algoritme ogenschijnlijk neutrale data (proxies). Maak de consequenties expliciet en leg deze op het juiste niveau vast. Neem indien nodig tegenmaatregelen. Een ultieme tegenmaatregel kan zijn het (tijdelijk) stopzetten van het algoritme.

Toelichting:

Ogenschijnlijk neutrale data die op het eerste gezicht bijvoorbeeld geen enkele link hebben met afkomst of nationaliteit kunnen leiden tot indirect onderscheid op grond van ras of nationaliteit. Voorbeelden zijn postcode, hoogte van inkomen, hoogte van inkomensafhankelijke toeslagen, kinderopvang door een gastouderbureau met een 'homogeen' klantenbestand, een familielid in het buitenland, kenteken en laaggeletterdheid.

Bron: Artikel 1 lid 1 sub c Awgb.

4 Risico: Bias in het algoritme leidt tot discriminatie.

Maatregelen:

1. Beoordeel of de geconstateerde bias wenselijk is en of deze op discriminatie duidt. Maak de consequenties expliciet en leg deze op het juiste niveau vast. Neem indien nodig tegenmaatregelen. Een ultieme tegenmaatregel kan zijn het (tijdelijk) stopzetten van het algoritme. Van belang bij de gemeten bias die duidt op discriminatie is het kijken vanuit wetgeving én wenselijkheid.

2. Documenteer in de functionele eisen de definitie van acceptabele bias. In de documentatie waarin deze eisen tot meetbare prestatiecriteria zijn uitgewerkt is te vinden welke fairness metrics hierbij horen. Maak de consequenties expliciet. Eerst moet duidelijk worden wat eerlijk is voor het proces. Dan kunnen fairness metrics worden opgesteld om eerlijkheid te meten.

3. Documenteer in de functionele eisen de methoden van voorkomen, detecteren en corrigeren van bias. Maak de consequenties expliciet. Documentatie van de aanpak van bias bevordert de controleerbaarheid.

4. Documenteer in de functionele eisen de doelstelling voor en definitie van de verschillende groepen en gewenste prestatie van het model voor deze groepen. Maak de consequenties expliciet en bespreek dit in het ethisch gesprek. De keuze voor een gewenste prestatie per groep is een ethische afweging.

5. Documenteer de mate van bias in de (trainings)data, dataverzameling en het model. Maak de consequenties expliciet en leg deze op het juiste niveau vast. Neem indien nodig tegenmaatregelen. Een ultieme tegenmaatregel kan zijn het (tijdelijk) stopzetten van het algoritme. Bias in de trainingsdata, regels in het model en de beslissing van de eindgebruiker kunnen leiden tot bias. Bias in de inputdata kan doorwerken tot bias in de uitkomst.

6. Beoordeel tijdens de ontwikkeling van het model of een verschil bestaat tussen prestatie van het model voor verschillende subgroepen. Maak de consequenties expliciet en leg deze op het juiste niveau vast. Neem indien nodig tegenmaatregelen. Een ultieme tegenmaatregel kan zijn het (tijdelijk) stopzetten van het algoritme. Een model kan gemiddeld goed presteren, maar kan voor bepaalde subgroepen die minder in de testset aanwezig waren verkeerde uitkomsten geven.

7. Beoordeel of de uitkomstbias van de productiedata voor de verschillende subgroepen voldoet aan de productiecriteria. Doe dit ook met een testset gedurende het ontwikkelproces. Maak de consequenties expliciet en leg deze op het juiste niveau vast. Neem indien nodig tegenmaatregelen. Een ultieme tegenmaatregel kan zijn het (tijdelijk) stopzetten van het algoritme. Tijdens de ontwikkeling van het model kunnen methoden om bias te corrigeren helpen om aan de prestatiecriteria te voldoen. Testresultaten op de uitkomstbias tijdens het ontwikkelen zijn ook wenselijk. Merk op dat voor het meten van uitkomstbias alleen de inputdata en uitkomst van het model nodig is. Uitkomstbias kan zonder gelabelde dataset gemeten worden.

Toelichting:

In de hierboven beschreven maatregelen is te zien dat veel vormen van bias bestaan die elk op hun eigen manier tegengegaan kunnen worden. Vandaar dat de toelichting hierop direct achter de maatregel is geplaatst.

Bron: EC/AI HLEG April 2019 - hoofdstuk II. 1.1, 1.5.

3.7 Milieu en maatschappelijke waarden

Deze paragraaf bevat één onderdeel en dat is duurzaamheid. Hierin wordt de aanbeveling gedaan om de impact van het milieu zo laag mogelijk te houden in de modelkeuze. Het bevat expliciet géén maatregel. De normen, risico's en maatregelen voor milieu en maatschappelijke waarden komen uit de Sustainable Development Goals (SDG).

3.7.1 Norm: Bewerkstellig duurzaamheid; ook in de ontwikkeling van algoritmen.

Risico: De impact van het model op het milieu is disproportioneel hoog.

Aanbeveling: Inventariseer de impact op het milieu en neem deze mee bij de modelkeuze en ontwikkeling. Maak de consequenties expliciet en leg deze op het juiste niveau vast. Neem indien nodig tegenmaatregelen. Een ultieme tegenmaatregel kan zijn het (tijdelijk) stopzetten van het algoritme.

Toelichting: Het trainen van een neuraal netwerk kost bijvoorbeeld meer energie dan een beslisboom. Er is meegenomen of de hoeveelheid gebruikte energie proportioneel is voor het doel van het algoritme. Daarnaast is nagedacht hoe de milieu-impact bij de ontwikkeling zo laag mogelijk kan worden gehouden. Parameters worden bijvoorbeeld tijdens het trainen niet via *trial and error* bepaald als deze ook afgeleid kunnen worden.

Bron: SDG 11/ EC/AI HLEG April 2019 hoofdstuk II 1.6.

3.8 Verantwoording

Deze paragraaf bevat twee onderdelen. Het eerste is dat bij de inzet van algoritmen wordt nagedacht over zorgvuldigheid. Dit behelst doel, impact, risicomanagement, rollen en verantwoordelijkheden, grondrechten en evaluatie. Een van de maatregelen is het doen van een mensenrechtentoets. Het tweede deel gaat over motivering. Dit bevat documentatie van de wijze van inzet van algoritmen, maar ook over het maken van afspraken bij het uitbesteden van onderdelen van activiteiten aan externe partijen. De normen, risico's en maatregelen voor verantwoording komen voort uit de Awb en de EC/AI.

3.8.1 Norm: Zorgvuldig handelen bij de ontwikkeling en de inzet algoritmen (zorgvuldigheidsbeginsel).

1 Risico: Risico: Zonder eenduidigheid over het doel is geen sturing op en verantwoording over het algoritme mogelijk en is er een groter risico op fouten en/of verschillen in interpretatie.

Maatregel: Het doel en eventuele subdoelen van het algoritme moeten zo specifiek mogelijk zijn geformuleerd, en waar mogelijk gekwantificeerd. Maak hiervoor gebruik van een multidisciplinaire aanpak en betrek hierbij meerdere afdelingen en stakeholders. Maak de consequenties expliciet en leg deze op het juiste niveau vast. Neem indien nodig tegenmaatregelen. Een ultieme tegenmaatregel kan zijn het niet inzetten van het algoritme.

Toelichting: Het doel van het algoritme moet helder zijn gedefinieerd, ook in relatie tot het maatschappelijke resultaat (outcome), en wordt gedeeld door de eigenaar, ontwikkelaar en gebruiker van het algoritme. Een bewuste afweging of het algoritme het juiste middel is om het probleem op doelmatige en doeltreffende wijze op te lossen is gemaakt en vastgelegd. De publieke waarden die met de inzet van algoritmen worden ingegeven, zoals gelijkwaardigheid en veiligheid, maar ook grondrechten die worden geraakt, zijn geïnventariseerd en gewogen. Ook de doeltreffendheid, subsidiariteit en proportionaliteit van het in te zetten algoritme zijn afgewogen.

Bron: Art. 3.2 Awb (bij besluiten)

EC/AI HLEG April 2019 - Hoofdstuk II. 1.1 en 1.7.

2 Risico: De impact van het algoritme is niet inzichtelijk, waardoor niet helder is welke maatregelen moeten worden getroffen om ongewenste effecten (zoals bias en discriminatie) te voorkomen.

Maatregel: Onderzoek (aan de hand van een mensenrechtentoets) welke groepen kwetsbaar zijn voor een fout van het algoritme en welke impact dat heeft op deze groepen. Het betreft de 'menselijke maat' bij het nemen van beslissingen op basis van de output van het algoritme. Maak de consequenties expliciet en leg deze op het juiste niveau vast. Neem indien nodig tegenmaatregelen. Een ultieme tegenmaatregel kan zijn het niet inzetten van het algoritme.

Toelichting: Het gaat om zowel de impact op personen of doelgroepen, als de bredere impact op de samenleving (vanuit sociaal, democratisch en milieu/ecologisch perspectief).

Bron: Art. 3.2 Awb (bij besluiten)
EC/AI HLEG April 2019 - Hoofdstuk I. 1.1 & Hoofdstuk II. 1.6.

3 Risico: Zonder actueel beeld van risico's kan er geen goede afweging worden gemaakt of de voordelen van de toepassing van het algoritme opwegen tegen de nadelen. Risico's worden niet (tijdig) vastgesteld en adequaat geadresseerd en behandeld.

Maatregel: Zorg ervoor dat op vastgelegde (periodieke) momenten een afweging plaats vindt van de risico's over het gebruik van het algoritme. Alle stappen van risicomanagement zijn uitgevoerd en op het juiste niveau in de organisatie behandeld, waaronder het identificeren, analyseren, evalueren (t.a.v. risk appetite), behandelen (risicoreactie, o.a. maatregelen), monitoren & beoordelen en communiceren & rapporteren van risico's. Maak de consequenties expliciet en leg deze op het juiste niveau vast. Neem indien nodig tegenmaatregelen. Een ultieme tegenmaatregel kan zijn het niet inzetten of (tijdelijk) stopzetten van het algoritme.

Toelichting: Het gaat hier om het feit dat er over risico's wordt nagedacht. De organisatie moet beschikken over een ingericht en gedocumenteerd proces voor risicobeheersing.

Bron: Art. 3.2 Awb (bij besluiten)
COBIT EDM03 / APO12.
EC/AI HLEG April 2019 - Hoofdstuk II.1.7.

4 Risico: Rollen en verantwoordelijkheden zijn niet duidelijk belegd of onvoldoende geborgd.

Maatregel: De rollen en verantwoordelijkheden bij de ontwikkeling en inzet van het algoritme zijn belegd en gedocumenteerd.

Toelichting: Duidelijkheid en borging van rollen en verantwoordelijkheden zorgen voor een effectief en verantwoord verloop van het proces rondom de inzet van een algoritme. Alle relevante belanghebbenden zijn bij de ontwikkeling en inzet van het algoritme betrokken. Het personeel (intern en extern) dat werkt met het algoritme moet over voldoende deskundigheid en competenties beschikken. Dit moet in ieder geval blijken uit de risicoanalyse, waarbij de organisatie maatregelen neemt om eventuele gaten te overbruggen. De organisatie heeft een goed beeld van de beschikbare resources (kwalitatief en kwantitatief) en stuurt daarop.

Bron: EC/AI HLEG April 2019 - Hoofdstuk II.1.5, COBIT APO01.02 / EDM01.

5 Risico: Het algoritme kan negatieve gevolgen hebben voor grondrechten.

Maatregel: Onderzoek of het algoritme een ongerechtvaardigde inbreuk maakt op grondrechten (aan de hand van een mensenrechtentoets). Maak de consequenties expliciet en leg deze op het juiste niveau vast. Neem indien nodig tegenmaatregelen. Een ultieme tegenmaatregel kan zijn het niet inzetten van het algoritme of het (tijdig) stopzetten.

Toelichting: In situaties wanneer dergelijke risico's bestaan, kan een effectbeoordeling worden uitgevoerd wat grondrechten betreft. Dit kan voorafgaand aan de ontwikkeling van het algoritme worden uitgevoerd en moet een evaluatie omvatten van de vraag of de risico's kunnen worden beperkt of gerechtvaardigd. De afwegingen tussen de verschillende beginselen en rechten moeten zijn vastgesteld en gedocumenteerd.

Bron: EC/AI HLEG April 2019 - hoofdstuk II. 1.1.

6 Risico: Zonder evaluatie van de kwaliteit van het algoritme is er geen goede sturing, beheersing en verantwoording mogelijk over de inzet van het algoritme.

Maatregel: Er vindt periodiek een evaluatie plaats van de werking en de kwaliteit van het algoritme. Er kunnen namelijk wijzigingen in de werking van het model of de inputdata zijn. De analyse van klachten en incidenten is hier ook onderdeel van. Maak de consequenties expliciet en leg deze op het juiste niveau vast. Neem indien nodig tegenmaatregelen. Een ultieme tegenmaatregel kan zijn het niet inzetten van het algoritme of het (tijdig) stopzetten.

Toelichting: Kwaliteit is o.a. doeltreffendheid, doelmatigheid, betrouwbaarheid en accuraatheid (geschiktheid) en non-discriminatie. Evaluatie kan bijvoorbeeld worden gedaan met een peerreview. Een proces voor een periodieke evaluatie van de kwaliteit van het algoritme is gedocumenteerd en in werking. De resultaten worden met belanghebbenden gedeeld. De organisatie analyseert of (interne en externe) klachten en incidenten het gevolg kunnen zijn van het gebruik van het algoritme. De verantwoordelijke in de business legt verantwoording af over de ontwikkeling, inzet en werking van het algoritme.

Bron: EC/AI HLEG April 2019 - Hoofdstuk II.1.2.

3.8.2 Norm: Motiveer de deugdelijke inzet van algoritmen

Risico: Afhankelijkheid van externe deskundigen die na het ontwikkelen van het algoritme met de betreffende kennis en ervaring weggaan, waardoor continuïteit en beheersing daarna niet meer gewaarborgd is.

Maatregelen:

1. Bij uitbesteding van onderdelen of activiteiten met betrekking tot het algoritme zijn afspraken met betrokken externe partijen gemaakt en vastgelegd m.b.t. eigenaarschap, beheer, prestatiecriteria en reproduceerbaarheid van het algoritme

2. De documentatie over het model (ontwerp, werking en voorwaarden) is beschikbaar en begrijpelijk voor ontwikkelaars, gebruikers en de eigenaar van het model.

Toelichting:

1. Overeenkomst met betreffende partij met afspraken over o.a. eigenaarschap en beheer (bijv. SLA). Maatregelen om een te grote afhankelijkheid te voorkomen moeten zijn beschreven/getroffen, zoals het opstellen van een exit-strategie. Bij het maken van afspraken met leveranciers moet aandacht zijn voor het risico op *vendor lock-in*. Hiervan is sprake als de opdrachtgever afhankelijk wordt van een externe partij (ontwikkelaar), omdat deze niet in staat is om van partij te veranderen zonder substantiële omschakelingskosten of ongemak.

2. De documentatie moet dusdanig begrijpelijk zijn dat het voor (nieuwe) medewerkers duidelijk is hoe een model gebruikt moet worden (collegiale uitlegbaarheid). Ook wanneer de ontwikkelaars van het model niet meer binnen de organisatie werken. Instructies over het gebruik van het model. Bij medewerkers navragen of de instructies bekend zijn. Het algoritme moet uitlegbaar zijn en er moet een afweging plaatsvinden tussen de uitlegbaarheid van het model en de prestatie van het model.

Bron:

1.
Art. 3.46 Awb (bij besluiten).
EC/AI HLEG April 2019 - Hoofdstuk II.1.7.

2.
EC/AI HLEG April 2019 - hoofdstuk II. 1.1 en 1.4.

3.8.3 *Norm: Het algoritme is in goede geordende staat en voldoet aan de Archiefwet 1995, dit betekent dat het algoritme duurzaam toegankelijk is (vindbaar, beschikbaar, leesbaar, interpreteerbaar, betrouwbaar en toekomstbestendig. Voor iedereen die daar recht op heeft en voor zo lang als noodzakelijk. (*)*

Risico:
Als informatie over documentatie, trainingsdata, outputdata en logica van het algoritme niet beschikbaar is of niet wordt bijgewerkt, is het niet herleidbaar op welke wijze het algoritme functioneert en kan het algoritme op een gegeven moment anders functioneren dan gedocumenteerd en is het niet helder welke versie van de informatie geldig was op welk moment dat het algoritme is ingezet. .

Maatregelen:

1. Het is vastgelegd inclusief beheerprocessen, wat, hoe, hoelang, waar en in welke vorm de documentatie, het model (ontwerp, werking en voorwaarden), trainingsdata, output(data) en logica van het algoritme bewaard moet worden en beschikbaar moet worden gesteld.

2. Het is duidelijk welke versie van de documentatie geldig door het gebruik van versiebeheer voor documentatie.

3. Beheer van documentatie in een gecontroleerde omgeving.

Toelichting:

Algoritmen dienen duurzaam toegankelijk te zijn voor het uitvoeren van overheidstaken, het afleggen van verantwoording hierover, voor burgers en bedrijven in het kader van rechtszekerheid, voor het uitvoeren van onderzoek en mogelijk als erfgoed.

Bij de aankoop, bouw, aanpassing, migratie of uitfasering van een algoritme pas je de eisen van duurzame toegankelijkheid van overheidsinformatie (DUTO-eisen) toe. Door ze op te nemen als *requirements* op het –moment dat er inrichtingskeuzes voor nieuwe informatiesystemen gemaakt worden. Op deze manier wordt archiveren *by design* toegepast.

Bron:
Archiefwet 1995, artikel 3, Eisen voor de duurzame toegankelijkheid van overheidsinformatie (DUTO-eisen) | Nationaal Archief