Ministerie van Infrastructuur
en Waterstaat

AI Impact Assessment

Het hulpmiddel voor een betrouwbaar AI-project

In samenwerking door collega's van het ministerie Infrastructuur en Waterstaat (IenW) bij Concerndirectie Informatiebeleid (CDIB), Inspectie Leefomgeving en Transport (ILT IDlab en afdeling analyse) en Rijkswaterstaat (RWS Datalab)

Inhoudsopgave

Toelichting

ARTIFICIAL INTELLIGENCE (AI)[1] kan worden ingezet om handelingen sneller of veiliger uit te voeren, zoals het inspecteren van asfaltkwaliteit of overtredingen op zee handhaven. AI biedt kansen, maar het brengt ook gevaren met zich mee. De cio-raad van IenW heeft in november 2020 akkoord gegeven om aan de slag te gaan met een concept AI Impact Assessment (AIIA), zodat er meer aandacht komt voor verantwoorde AI. Het IDlab ILT, het RWS Datalab en Concerndirectie Informatiebeleid van IenW (CDIB) hebben dit opgepakt en gezamenlijk deze nieuwe versie van het AIIA ontwikkeld. Het AIIA is vastgesteld door de Bestuursraad van IenW op 4 juli 2022.

Het AI Impact Assessment (AIIA) wordt gebruikt voor het discussiëren over AI-systemen. Hierbij wordt gekeken naar obstakels in de data, het systeem, de algoritmiek en wordt rekening gehouden met geldende wet- en regelgeving. Het AIIA dient als instrument voor het gesprek en het vastleggen van het denkproces zodat onder andere de verantwoording, kwaliteit en **REPRODUCEERBAARHEID** worden vergroot. Het verwachte resultaat van het AIIA is een helder ingevuld document waarin duidelijk zichtbaar is welke afwegingen gemaakt zijn bij het inzetten van AI in een project.

Primair is de **OPDRACHTGEVER VERANTWOORDELIJK** voor het (laten) uitvoeren van het AIIA. Er **moet** een AIIA worden opgesteld voor elk **AI-SYSTEEM**. Het invullen van het AIIA gebeurt nadrukkelijk proportioneel, passend bij de impact en het risicoprofiel van de toepassing. De verantwoordelijkheid voor wat **proportioneel** is, ligt bij de projectleiders en de opdrachtgever.

AI kan ook worden gebruikt in onderzoek. Daarbij geldt dat het van belang is om onder andere te kijken naar zaken als false positives en false negatives en de verantwoording over en uitlegbaarheid van de resultaten. Ook kan AI worden gebruikt om hypotheses te genereren, die dan met AI of andere technieken verder worden uitgewerkt. Kortom ook voor onderzoekers geldt: denk goed na over AI en gebruik daar deze AIIA voor. Uiteraard vallen niet van toepassing zijnde vragen af, bijvoorbeeld als het systeem niet in beheer wordt genomen.

De proportioneel ingevulde AIIA moet af zijn voor het **IN GEBRUIK NEMEN** van een AI-systeem. Het is belangrijk dat het AIIA daarna regelmatig wordt bijgesteld, bijvoorbeeld als het doel van het AI-systeem wordt gewijzigd of er veranderingen aan het AI-systeem plaatsvinden. Daarnaast kunnen zich in de loop van de tijd nieuwe risico's voordoen. Controleer dit als **PROJECTLEIDER**/opdrachtgever periodiek.

Om een AI-project goed en verantwoord uit te voeren, is er meer nodig dan een AI Impact Assessment. Denk aan het organiseren van een **MOREEL BERAAD**, het zorg dragen voor een zorgvuldige inbeheername en het informeren van de omgeving. Zie hiervoor de meest recente handreiking 'AI voor opdrachtgevers' (op te vragen bij de Concerndirectie Informatiebeleid).

Past het AIIA niet bij een AI-project of systeem waaraan je denkt? Of heb je andere opmerkingen of vragen over het AIIA? Neem contact op met de CIO-office bij Concerndirectie Informatiebeleid.

[1] Dit document hanteert de omschrijving van AI door de algemene rekenkamer, zie de Begrippenlijst.

Leeswijzer

Voor het invullen van het AIIA is het volgende van belang:
- Het AIIA dient als instrument voor gesprek en controle.
- Wijzigt het doel van het **AI-SYSTEEM**, dan moet het AIIA bijgesteld worden.
- **DIKGEDRUKTE** woorden zijn aanklikbare begrippen, gedefinieerd in bijlage Begrippenlijst.
- Het AIIA moet af zijn voor het **IN GEBRUIK NEMEN** van een AI-systeem.
- Het AIIA is verplicht, maar de mate waarin het wordt ingevuld ligt bij de expertise van de **PROJECTLEIDER**.
- Enkel 'ja' of 'nee' volstaat niet als antwoord op de vragen.
- Vragen zijn gecodeerd en genummerd, de eerste letter verwijst naar het hoofdstuk (bv. T voor Technische robuustheid), de letter 'o' wordt hieraan toegevoegd als het een groene ● hulpvraag betreft.

Vragenlijst

Het volledige AIIA omvat zo'n 100 vragen. Het AIIA is verplicht bij het maken of inkopen van AI-systemen, maar het invullen gebeurt nadrukkelijk proportioneel, door de opdrachtgever en projectleider zelf te bepalen. Dit noemen we: verplicht, maar soepel. Dit betekent vooral dat je met gezond verstand moet nadenken over hoeveel impact jouw **AI-SYSTEEM** heeft. Bij alle vormen van **ARTIFICIAL INTELLIGENCE** zijn de blauwe ● overkoepelende vragen verplicht, deze helpen bij het faciliteren van de discussie over de wenselijkheid van het AI-systeem. De groene ● hulpvragen helpen om hier een concrete invulling aan te geven. Niet alle groene vragen zijn bij iedere casus relevant, deze zijn dan ook niet verplicht (soepel). De **OPDRACHTGEVER** en **PROJECTLEIDER** kunnen op basis van een eigen risico-inschatting besluiten om deze vragen wel in te vullen. Houd er rekening mee dat de Auditdienst Rijk en de Algemene Rekenkamer het systeem kunnen controleren op correctheid en veiligheid. Ook betekent een volledig ingevuld AIIA niet per definitie dat de AI veilig is. Wanneer de AI Act in werking treedt (COM/2021/206 final), moeten ook de vragen met een rode ster ✶ verplicht worden ingevuld voor **AI MET HOOG RISICO**. Het is wenselijk om dat nu ook al te doen. Voor iedere vraag geldt dat het antwoord moet worden toegelicht. Enkel een 'ja' of 'nee' volstaat dus nooit als antwoord.

In bijlage 'Wie doet wat' tref je een overzicht dat kan helpen bij het bepalen wie welke vraag moet invullen. Sommige vragen kunnen bijvoorbeeld beter door een data scientist worden ingevuld, en andere vragen door een jurist.

Inleidende vragen

De inleidende vragen gaan over de algemene aspecten van het **AI-SYSTEEM** dat je gaat ontwikkelen. Deze vragen gaan over het doel van het systeem, en de rol die het systeem binnen de organisatie gaat hebben. Denk aan vragen over de gebruikte technieken, of wie er verantwoordelijk gaat zijn.

Doel van het systeem

Deze vragen zijn basis vragen om weer te geven wat je voor welk doel aan het doen bent. Het antwoord op deze vragen heeft relevantie voor de rest van het AIIA.

i 1. Geef een korte beschrijving van het beoogde **AI-SYSTEEM** (titel, algemene omschrijving, probleemstelling, en het domein)

i 2. Waarom is er voor de huidige techniek gekozen? (hierbij is het van belang dat alle afwegingen van robuustheid tot mensenrechten, impact op gebruiker en eindgebruiker, verantwoordingsplicht etc. zijn meegenomen in het antwoord)

i 3. Wat is het doel en beoogde resultaat van het AI-systeem?

i 4. Welk doel wordt er aan het AI-systeem gekoppeld volgens het rapport Aandacht voor Algoritmes van de Algemene Rekenkamer[2]? **DOEL 1**, **DOEL 2** of **DOEL 3**?

Rol binnen de organisatie

Deze vragen leveren vaak discussie op. Naast vragen over de bouw en details van een AI-systeem, moet er goed nagedacht worden over de impact van het AI-systeem in zijn geheel. Dit zijn fundamentele vragen. Zorg daarom dat je ze goed uitdenkt. Probeer dit goed te nuanceren. Stel dat het AI-systeem een positieve impact heeft op duizenden burgers, maar een negatieve impact op tien burgers, dan is het systeem niet meteen onbruikbaar, maar moeten er wel goede maatwerkoplossingen komen voor de tien burgers waarop negatieve impact is.

Bij deze vragen stel je ook de rolverdeling binnen het ontwikkelen en gebruiken van je systeem vast. Deze rollen staan gedefinieerd in de begrippenlijst. Houd deze definities aan.

i 5. Waar in de organisatie is beoogd het AI-systeem te gebruiken en welke beoogde impact is er voor de organisatie?

i 6. Beschrijf de rolverdeling binnen het opzetten van het AI-systeem (zoals de **ONTWIKKELAAR**, **OPDRACHTGEVER**, **PROJECTLEIDER**, **BEHEERORGANISATIES** en **EINDVERANTWOORDELIJKE**).

i 7. Wie is de **GEBRUIKER** van het AI-systeem, wie zijn de **EINDGEBRUIKERS** die met het systeem werken en welke **BETROKKENEN** ondervinden impact van het AI-systeem?

[2] Algemene Rekenkamer (2021), Aandacht voor Algoritmes.

Fundamentele rechten & fairness

Zoals veel technologieën kunnen AI-systemen grondrechten zowel bevorderen als benadelen. Gelet op het grote belang van de bescherming van grondrechten, en de bijzondere risico's die kunnen bestaan voor de aantasting van die grondrechten door inzet van AI-systemen, is het van belang om aan dit onderwerp afzonderlijk aandacht te besteden. Dit hoofdstuk hangt nauw samen met het hoofdstuk data governance, waarin privacy wordt behandeld. Het recht op privacy is een grondrecht, maar door het karakter van het onderwerp privacy is het onderverdeeld in een eigen hoofdstuk.

Grondrechten

f 1. Wat is de mogelijke impact op de grondrechten van burgers door het gebruik van het **AI-SYSTEEM**?

f 2. Is het **PROPORTIONEEL** en **SUBSIDIAIR** om dit systeem in te zetten om de gestelde doelen te realiseren? Oftewel: is de impact in verhouding met de beoogde doelen en zijn er geen andere minder ingrijpende manieren om deze doelen te behalen?

f 3. Wat is de wettelijke grondslag van de inzet van het AI-systeem en van de beoogde besluiten die genomen worden op basis van het AI-systeem?

Mensen die belang hebben bij de werking van het **AI-SYSTEEM** moeten goed behandeld worden. Dat betekent dat de (fundamentele) rechten van alle **BETROKKENEN** gewaarborgd moeten worden. Voor een goede invulling van dit onderwerp verwijzen wij je naar de Impact Assessment Mensenrechten en Algoritmes.[3] De vragen die in deze AIIA gesteld worden zijn dan ook niet voldoende voor het afbakenen van dit onderwerp.

Bij het beantwoorden van de vragen zijn de grondrechten van de mens van toepassing, deze staan vastgelegd in de Grondwet en het Europese Verdrag voor de Rechten van de Mens.

fo 1. Welke grondrechtelijke bepalingen zijn mogelijk van toepassing?

fo 2. Op welk van deze grondrechtelijke bepalingen kan mogelijk een inbreuk worden gemaakt bij verkeerde uitvoering van het **AI-SYSTEEM**? Welke acties worden genomen om dit te voorkomen?

Bias

f 4. Hoe wordt rekening gehouden met mogelijk onwenselijke **BIAS IN DE INPUT**, **BIAS IN HET MODEL** en **BIAS IN DE OUTPUT** van het **AI-SYSTEEM**?[4] ✶

BIAS betekent het doen van aannames over dingen, mensen of groepen. Dit heeft twee kanten. Enerzijds is het noodzakelijk om conclusies over data op een nieuwe situatie te projecteren. We maken in generalisaties namelijk altijd aannames. Tegelijkertijd is het van belang dat er geen onrechtmatige vertekening ontstaat met vormen van onterechte en onwenselijke bias die in strijd kunnen zijn met de rechten van de mens.

[3] Ministerie van Binnenlandse Zaken en Koninkrijksrelatie (2021), Impact Assessment Mensenrechten en Algoritmes.
[4] Artikel 14 lid 4.

Bias kan zitten alle facetten van het systeem: **BIAS IN DE INPUT**, **BIAS IN HET MODEL** en **BIAS IN DE OUTPUT**. Er zijn verschillende typen bias die relevant zijn tijdens het ontwikkelen en inzetten van AI, bijvoorbeeld **DATA BIAS** en **DESIGN BIAS**. Deze soorten bias worden vaak veroorzaakt door socio-economische aan-names en kunnen als gevolg versterkte socio-economische aannames hebben. Deze soorten bias kunnen ervoor zorgen dat AI-systemen niet voor alle **BETROKKENEN** goed werken als er niet voor wordt gecorrigeerd.

Het kernelement van dit thema is bewustzijn en integriteit. Het is onmogelijk om volledig zonder bias te werken. Vaak bestaat bias al decennia lang en zal deze (onterecht) niet als zodanig worden herkend. Dus in plaats van ons richten op bias-loze AI, moeten we ernaar streven om ons zoveel mogelijk bewust te zijn van mogelijke discriminatie. Het is verder van belang kritische vragen te stellen over de herkomst en inhoud van data en de werking van AI-systemen.

Bias hangt nauw samen met **DIVERSITEIT**, **GELIJKHEID** en **EERLIJKHEID** tussen mensen, maar het is van belang om bewust te zijn dat aannames ook over niet-menselijke aspecten kunnen gaan, zoals de natuur of leefomgeving. Daarnaast kan het voor de wijze waarop je bias wilt mitigeren relevant zijn om onderscheid te maken tussen **NEGATIEVE IMPACT**, **GEEN POSITIEVE IMPACT** en een **POSITIEVE IMPACT** die de bias kan hebben.

Denk bij positieve impact aan het volgende. In de statistiek is bias een systematische fout of afwijking. Dat is niet altijd verkeerd. Deze systematische fout wordt namelijk vaak bewust toegepast in modellen, bijvoorbeeld in de vorm van regularisatie. Op deze manier kan ervoor worden gezorgd dat de variantie wordt verkleind, ook al gaat dit ten koste van een (kleine) systematische afwijking. Bias kan ook positief worden ingezet. Een AI-systeem kan bijvoorbeeld bewust worden ontwikkeld om niet of minder te discrimineren, juist door het introduceren van bias.

Bias in de input(data)

fo 3. Is de input(data) data representatief voor het onderwerp waarover een beslissing moet worden genomen?

fo 4. Worden indien nodig subpopulaties beschermd bij het trekken van steekproeven?

fo 5. Is de keuze voor de inputvariabelen onderbouwd en afgestemd met de **BETROKKENEN**?

Bias in het model

fo 6. Op welke manier wordt er rekening gehouden met het feit dat er geen onterechte of onrechtvaardige **BIAS** in een AI-systeem wordt gecreëerd of versterkt?

fo 7. Is het AI-systeem te gebruiken door de beoogde **EINDGEBRUIKERS** (dus ongeacht diens kenmerken zoals leeftijd, geslacht of capaciteit)?

Bias in de output(data)

fo 8. Zijn er stop-, toezicht- of controle- mechanisme ingesteld om te voorkomen dat groepen in de maatschappij disproportioneel getroffen kunnen worden door de negatieve implicaties van het AI-systeem? Specifiek voor ILT: maak hier onderscheid tussen ondertoezichtstaanden (OTS) en de rest van de maatschappij.

Stakeholderparticipatie

f 5. Zijn alle **STAKEHOLDERS** in kaart gebracht middels een stakeholderanalyse en is met hen het gesprek aangegaan?

Bij **STAKEHOLDER** participatie worden verschillende doelgroepen betrokken in het kader van **DIVERSITEIT**, non-discriminatie en rechtvaardigheid. Om rechtvaardige AI te realiseren, moet er goed nagedacht worden over inclusie en diversiteit gedurende de gehele levenscyclus van het AI-systeem. In deze AIIA gaat het in deze context ook vaak over **BETROKKENEN**.

Om te voorkomen dat de **ONTWIKKELAARS** van een AI-systeem te veel in een eigen denkwereld blijven en zich niet bewust zijn van impliciete aannames of gevolgen, is afstemming over AI-systemen essentieel. Je doet dit door bijvoorbeeld af te stemmen met je eigen team, de klant, **EINDGEBRUIKER**, betrokkenen, ervaringsdeskundigen, **DOMEINEXPERTS** als universiteiten, andere overheidsorganisaties etc. Een moreel beraad[5] organiseren is aanbevolen.

fo 9. Met welke mensen en/of groepen is er afgestemd bij het ontwikkelen van **AI-SYSTEEM**?

fo 10. Zijn de stakeholders op de hoogte waarom er gekozen is voor bepaalde input variabelen (waar zij wellicht in staan)?

fo 11. Welke feedback is er verzameld van teams of groepen die verschillende achtergronden en ervaringen representeren? En wat is hier vervolgens mee gedaan?

fo 12. Hoe wordt de invoering van het AI-systeem geïntroduceerd richting collega's van IenW?

fo 13. Hoe wordt de invoering van het AI-systeem geïntroduceerd richting de samenleving?

[5] Overlegorgaan Fysieke Leefomgeving (mei 2021), Moreel Beraad.

Technische robuustheid

Of een **AI-SYSTEEM** werkt waarvoor het bedoeld is wordt afgevangen met technische **ROBUUSTHEID**.

Accuraatheid

t 1. Hoe wordt de doorlopende **ACCURAATHEID** van het systeem gemeten en gewaarborgd?

Een **AI-SYSTEEM** moet over het algemeen goed presteren. Om de kans op verkeerde beoordelingen te minimaliseren, is het belangrijk doorlopend de prestaties van een AI-systeem te meten. Dit omvat het meten van het AI-systeem in zowel de ontwikkel- als productiefase. Ook de kwaliteit van de gebruikte data is van belang. Een AI-systeem is dus nooit af; het blijft noodzakelijk AI-systemen regelmatig te testen en hertrainen. Het is wenselijk een kwantificering te hebben van de kans waarop tóch een verkeerde beoordeling wordt gemaakt.

ACCURAATHEID van het systeem kan je bepalen door **ACCEPTATIECRITERIA** op te stellen voor zowel de data als het systeem en deze te monitoren middels een metriek. Acceptatiecriteria kunnen bijvoorbeeld een hoeveelheid data zijn of bepaalde drempelwaardes van het meetsysteem. Er zijn veel verschillende soorten meetsystemen (vaak 'performance metrics' genoemd door data scientists) beschikbaar om de kwaliteit van **MODELLEN** te kwantificeren, denk bijvoorbeeld aan een accuratesse, precision en recall of F1-score. Hierbij is het van belang dat het meetsysteem en de acceptatiecriteria goed worden afgestemd op de data en het beoogde doel van het AI-systeem.[6] Dit moet samenhangen met onder andere de bevindingen uit de risicoanalyse (zie 'Risicobeheer'), omdat in de loop van tijd nieuwe of andere risico's zich kunnen voordoen met de inzet van een AI-systeem. Ook is het van belang dat de kwaliteit van het systeem doorlopend gemonitord wordt en indien nodig tijdens het hertrainen of doorontwikkelen de acceptatiecriteria en keuze voor meetsystemen opnieuw geëvalueerd worden.

to 1. Wat zijn de opgezette **ACCEPTATIECRITERIA** om de kwaliteit van de **INPUT(DATA)** en **OUTPUT(DATA)** van het **MODEL** aan te toetsen?

to 2. Passen de acceptatiecriteria bij de data en het doel van het AI-systeem?

to 3. Welke evaluatie meetsystemen (performance metrics) ga je gebruiken om de **ACCEPTATIECRITERIA** te waarborgen en waarom?[7] ✶

to 4. Hoe wordt de **OUTPUT(DATA)** (periodiek) steekproefsgewijs en doorlopend getest op juistheid?

to 5. Hoe worden afwijkingen in de output(data) ten opzichte van acceptatiecriteria tijdig geanalyseerd en gecorrigeerd?

to 6. Wat zijn de resultaten als er alternatieve **MODELLEN** zouden worden ingezet?

[6] Het gekozen meetsysteem moet geschikt zijn voor het model en de data die gebruikt wordt om de kwaliteit te meten. Neem bijvoorbeeld één dat in een bepaalde tekst 5 woorden per 100 woorden zou moeten labelen. Als het systeem in deze tekst 0 woorden labelt, dan heeft het model een accuratesse van 95%. Op het moment dat je de kwaliteit van het systeem bepaalt met accuratesse lijkt het model het dus heel erg goed te doen, terwijl de recall 0 is en het dus helemaal niet zo heel goed doet. Daarom is accuratesse niet geschikt om te bepalen hoe goed dit model werkt.

[7] Artikel 15 lid 1 en 2 AI Act.

Betrouwbaarheid

t 2. Is het **AI-SYSTEEM BETROUWBAAR**?

Een **BETROUWBAAR** AI-systeem geeft in vergelijkbare gevallen dezelfde resultaten. De vraag die centraal staat bij betrouwbaarheid is of de individuele **OUTPUT(DATA)** nogmaals te verkrijgen is met behulp van hetzelfde **MODEL** en dezelfde **INPUT(DATA)**, dezelfde instellingen en dezelfde **PARAMETERS**. Ook is het van belang dat het systeem een betrouwbare indicatie geeft van hoe goed het model gaat presteren in nieuwe situaties.

to 7. Wat zijn de belangrijkste factoren die de prestaties van het **AI-SYSTEEM** beïnvloeden?

to 8. Wordt een deel van de (sub)dataset uitgesloten voor het leren van het model en alleen gebruikt voor het bepalen van de betrouwbaarheid of wordt de betrouwbaarheid van het model berekend met behulp van cross-validatie?

to 9. Hoe is de (hyper)parameter-tuning onderbouwd en getoetst?

Technische implementatie

De **technische implementatie** beschrijft hoe het AI systeem technisch binnen het ICT-landschap van de organisatie is geïntegreerd. De specifieke eisen van het AI-systeem aan hardware en software zijn gedocumenteerd zodat hier rekening mee gehouden kan worden bij het uitrollen en beheer van het systeem. Daarnaast wordt uit de systeemarchitectuur duidelijk hoe de verschillende softwarecomponenten zich tot elkaar verhouden. Een goed doordachte architectuur vermindert de bedrijfsrisico's die gepaard gaan met het bouwen van een technische oplossing en slaat een brug tussen bedrijfs- en technische vereisten.

t 3. Hoe is het AI systeem technisch geïmplementeerd?

to 10. Is er nagedacht hoe het AI-systeem past in de al bestaande technische- en systeeminfrastructuur en zijn hier passende maatregelen voor genomen om deze uit te rollen (indien van toepassing)?

to 11. Hoe ziet de systeemarchitectuur eruit (hoe verhouden de softwarecomponenten zicht tot elkaar)?

to 12. Zijn eventuele specifieke hardware- en software-eisen gedocumenteerd?

Reproduceerbaarheid

t 4. Is het **AI-SYSTEEM REPRODUCEERBAAR**? Is er een proces ingesteld om dit te meten?

Bij **REPRODUCEERBAARHEID** kun je denken aan het vastleggen van welke data gebruikt zijn, hoe het model tot stand is gekomen, of wijzigingen in de data zijn bijgehouden, of uit dezelfde **INPUT(DATA)** dezelfde resultaten voortvloeien, en of er bepaalde situaties of condities zijn waarin de **OUTPUT(DATA)** beïnvloed kunnen worden. Reproduceerbaarheid gaat over trainen, valideren en testen.

Reproduceerbaarheid hangt nauw samen met **TRACEERBAARHEID**. Bij traceerbaarheid gaat het er voornamelijk om dat de datasets en processen goed worden gedocumenteerd. Versiebeheer op de data, het en de training speelt daarin een belangrijke rol.

to 13. Kan je een verkregen **OUTPUT(DATA)** nu of in de toekomst reconstrueren (dus bijvoorbeeld zijn oude versies van het **MODEL**, datasets en omstandigheden opgeslagen middels versiebeheer)?

to 14. Is het mogelijk om gegeven de **PARAMETERS** en een vaste **SEED** het model te reconstrueren?

to 15. Is het **AI-SYSTEEM** aan de hand van documentatie op hoofdlijnen te reproduceren?

to 16. Hoe worden de wijzigingen tijdens de levensduur van het systeem gedocumenteerd?

Uitlegbaarheid

t 5. Is het **AI-SYSTEEM** voldoende **UITLEGBAAR** en te interpreteren voor de **ONTWIKKELAARS**?

Technische **UITLEGBAARHEID** heeft te maken met het vermogen om zowel technische processen als daaraan gerelateerde menselijke beslissingen te kunnen begrijpen. Verder moet helder zijn welke verschillende ontwerpkeuzes zijn gemaakt en wat de rationale is voor het inzetten van het **AI-SYSTEEM**. Zie ook ‘Verantwoordingsplicht’ voor uitlegbaarheid richting **BETROKKENEN**.

to 17. Hoe heb je bij het ontwikkelen van het AI-systeem gekeken naar de uitlegbaarheid van het model?

to 18. In hoeverre is het mogelijk om een verklaring te geven aan een externe AI-expert hoe het AI-systeem op een bepaalde manier werkt (zie ook ‘Uitlegbaarheid’)?

to 19. Is de benodigde deskundigheid voor het beheer van AI-systeem gedocumenteerd?

Data governance

Data **GOVERNANCE** gaat over een (bestuurlijke) werkwijze rondom data met betrekking tot toegang, eigenaarschap, bruikbaarheid, integriteit en veiligheid. Daarnaast is er aandacht voor de kwaliteit van de data die wordt gebruikt.

Onder data governance valt ook privacy. Privacy is een van de fundamentele rechten van de mens, die mogelijk door AI kan worden aangetast. Het is daarom belangrijk dat er adequate data governance en bescherming van persoonsgegevens conform de Algemene Verordening Gegevensbescherming (AVG) is.

Voor AI-systemen is het essentieel om informatiebeveiligings- en privacyrisico's inzichtelijk te maken, deze risico's terug te brengen naar een acceptabel niveau en periodiek (technisch) te laten testen (bv. door het uitvoeren van een pentest). Om dit te realiseren moet het risicomanagementproces voor informatiebeveiliging en privacy van de organisatie worden doorlopen voor de AI-implementatie. Producten die daarbij onder meer moeten worden opgeleverd zijn: BIV-classificatie (BIA), implementatie van en toets aan de BIO, DPIA (in geval van verwerking persoonsgegevens), security testen en indien noodzakelijk een verbeterplan.

Kwaliteit en integriteit van data

d 1. Hoe wordt de kwaliteit van de data gewaarborgd?[8] ✶

Datakwaliteit is essentieel voor de werking van een **AI-SYSTEEM**. Verzamelde gegevens kunnen bijvoorbeeld sociaal geconstrueerde **BIAS**, onjuistheden, fouten en vergissingen bevatten (zie ook 'Bias'). Dit moet worden geadresseerd voordat er verder met deze data wordt gewerkt. De datasets en de werkwijze moeten worden getest en gedocumenteerd bij iedere stap: training, testen, uitrolfase en operationele fase. Dit geldt ook voor AI-systemen die niet intern gebouwd zijn, maar elders zijn verworven. De archiefwet[9] stelt eisen aan de manier van opslaan en bewaartermijn van data.

[8] Artikel 10 lid 2 en 3 AI Act.
[9] Archiefwet 1995 (https://wetten.overheid.nl/BWBR0007376/2020-01-01)

Overkoepelend

do 1. Is de gebruikte data noodzakelijk voor het **AI-SYSTEEM**?

do 2. Hoe voorkom je onbedoelde verdubbelingen van data?

do 3. Is het mogelijk om de trainings- en testgegevens te actualiseren als de situatie daar om vraagt? Wanneer besluit je het AI-systeem te her-trainen, tijdelijk stop te zetten, of door te ontwikkelen?[10] ✶ Input(data)

do 4. Voldoet de data aan de aannames van het **MODEL**?

do 5. Op welke manier is de **INPUT(DATA)** die wordt gebruikt in het AI-systeem verzameld en samengevoegd?

do 6. Hoe wordt de data gelabeld?

do 7. Welke factoren hebben invloed op de kwaliteit van de input(data)? En wat kan je daaraan doen?

do 8. Is de input(data) getoetst op veranderingen die zich voordoen tijdens trainen, testen en evalueren? Ook door de tijd heen tijdens het gebruik van het algoritme?

Output(data)

do 9. Indien output(data) wordt gebruikt als nieuwe input, hoe wordt de output(data) opgeslagen (denk aan een feedbackloop)?

do 10. Hoe zorg je ervoor dat de output(data) tijdig beschikbaar is?

[10] Artikel 14 lid 4.

Privacy en gegevensbescherming

d 2. Hoe wordt er omgegaan met persoonsgegevens of vertrouwelijke gegevens? (Denk aan de DPIA)

Privacy en de bescherming van gegevens moeten gewaarborgd zijn gedurende de hele levenscyclus van het **AI-SYSTEEM**. Met digitaal vastgelegde gegevens van menselijk gedrag kunnen AI-systemen wellicht leeftijd, geslacht en politieke, religieuze of seksuele voorkeuren afleiden. Let erop dat wanneer je persoonsgegevens gebruikt, deze niet gebruikt kunnen worden om te discrimineren, zie ook 'Bias'.

Naast persoonsgegevens kunnen er ook andere vertrouwelijke gegevens gebruikt worden die niet zomaar openbaar gemaakt mogen worden. Dit geldt bijvoorbeeld voor het gebruik van vertrouwelijke informatie als gerubriceerde informatie of bedrijfsgeheimen. Ook deze data moet goed beschermd zijn. De AI Verordening biedt aanvullende regels voor het gebruik van (persoons)gegevens in AI-systemen.

Met betrekking tot persoonsgegevens

do 11. Werkt **HET AI-SYSTEEM** met persoonsgegevens[11] (is de AVG van toepassing)? Zo ja, vul de volgende vragen ook in. Zo nee, ga verder bij 'met betrekking tot vertrouwelijke gegevens'.

do 12. Is de output van het AI-systeem tot personen te herleiden (is de AVG van toepassing)? Zo ja, vul dan de volgende vragen ook in.

do 13. Zijn er verregaande beschermingsmaatregelen genomen om de persoonsgegevens te beveiligen?[12] ✶

do 14. Zijn functionarissen betrokken, zoals de functionaris gegevensbescherming, privacy adviseur, informatiebeveiliger, Chief Information Officer, etc.?

do 15. Hoe vaak wordt de kwaliteit en de noodzakelijkheid van de verwerking van persoonsgegevens geëvalueerd?

do 16. Is er aandacht besteed aan rechten van derden met betrekking tot verspreiding van informatie over het AI-systeem? ✶

Met betrekking tot vertrouwelijke gegevens (niet zijnde persoonsgegevens)

do 17. Worden vertrouwelijke gegevens gebruikt of opgeslagen?

do 18. Hoe wordt de veiligheid van deze informatie gewaarborgd?

[11] Definitie uit artikel 4 lid 1 AVG.
[12] Artikel 10 lid 5 AI Act.

Risicobeheer

Het is van belang dat mogelijke risico's in de gaten worden gehouden. Wanneer risico's niet zijn voorzien, kan een **AI-SYSTEEM** tot onbetrouwbare resultaten komen. Dit kan schade veroorzaken. Het beginsel van preventie moet ervoor zorgen dat schade zoveel mogelijk wordt beperkt. Schade kan opgelopen worden door slecht functioneren van het AI-systeem, of bijvoorbeeld door **HACKAANVALLEN** van buitenaf.

Risicobeheersing

r 1. Hoe is het AI-systeem getest op de passende risicobeheersmaatregelen?[13] ✶

Bij het ontwikkelen en **IN GEBRUIK NEMEN** van een **AI-SYSTEEM** komen gevaren kijken, die in deze AIIA zoveel mogelijk ingekaderd worden. Toch kunnen zich alsnog onvoorziene problemen voordoen. Het is belangrijk om vast te stellen hoe je met deze potentiële gevaren omgaat. Dat betekent ook dat er mechanismes ingesteld moeten worden om risico's te beheersen, en dat deze mechanismen goed zijn getoetst. Denk aan het voorkomen van data vergiftiging, de mate van beheersmaatregelen en de beveiliging van de bewaarplaats van uitkomsten. Daarnaast moet er rekening gehouden worden met het feit dat zich nieuwe risico's kunnen voordoen na invoering van het AI-systeem. De risicobeheersmaat-regelen moeten dus periodiek gecontroleerd worden.

ro 1. Hoe is de toegang tot het AI-systeem en diens componenten ingericht? (Denk aan de Generieke IT-beheersmaatregelen)

ro 2. Hoe is het **AI-SYSTEEM** getest op het beoogde doel voordat het in gebruik wordt genomen?[14] ✶

ro 3. Is het waarschijnlijk dat kwetsbare groepen (zoals kinderen) toegang zullen hebben tot het AI-systeem? In dat geval moeten de risicobeheersmaatregelen extra worden aangescherpt.[15] ✶

ro 4. Zijn er buiten de standaard beveiligingsmaatregelingen van IenW extra maatregelen genomen om het AI-systeem te beveiligen?

ro 5. Hoe wordt het alternatieve plan als er problemen met het **AI-SYSTEEM** zijn in werking gezet?

ro 6. Is de correctheid van de implementatie aangetoond? Denk hierbij bijvoorbeeld aan unit-integratie- en end-to-end tests, indien van toepassing.

ro 7. Hoe kan het AI-systeem interageren met andere hardware of software (indien van toepassing)?

[13] Artikel 9 lid 5 AI Act.
[14] Artikel 9 lid 6 en 7 AI Act.
[15] Artikel 9 lid 8 AI Act.

Alternatieve werkwijze

r 2. Wat is het plan als er problemen met de werking van het **AI-SYSTEEM** zijn?

Het is wenselijk om een plan te hebben voor wanneer er problemen optreden met het **AI-SYSTEEM**. Dit betekent dat er een alternatieve werkwijze beschikbaar moet zijn in het geval dat er problemen met de werking van het systeem zijn. Denk aan de mogelijkheid om van een machine learning naar een beperkter rule-based **MODEL** terug te schakelen.

Het is goed om er bewust van te zijn dat een mens als expert zich niet op dezelfde manier ontwikkelt als een AI-systeem. Denk hierbij aan het effect van de rekenmachine op onze vaardigheid hoofdrekenen. Een alternatieve werkwijze moet hierop ingespeeld zijn. Wat is de impact in het geval dat het AI-systeem foute resultaten genereert?

ro 8. Wat is de impact als het AI-systeem uitvalt?

ro 9. Zie hierboven het voorbeeld over de rekenmachine. Wat is een equivalent effect wat kan optreden als het AI-systeem in gebruik wordt genomen, en is dit wenselijk?

ro 10. Is het **AI-SYSTEEM** bestand tegen fouten of onregelmatigheden van interactie met natuurlijke personen of andere systemen?[16] ✶

Hackaanvallen en corruptie

r 3. Op welke manier worden informatiebeveiligingsrisico's inzichtelijk gemaakt, teruggebracht naar een acceptabel niveau en (technisch) getest?

Informatiebeveiligingsrisico zoals **HACKAANVALLEN** en **CORRUPTIE** moeten zoveel mogelijk worden beheerst. De te voorziene risico's moeten ingekaderd worden door deze inzichtelijk te maken via het risicomanagementproces van de organisatie. Daaronder vallen onder meer het in kaart brengen van BIV-classificatie, rubriceringsniveau van informatie, implementatie van BIO-maatregelen, security testen en, indien het beveiligingsniveau van de BIO niet voldoet, het ev. uitvoeren van een aanvullende (technische) risico-analyse. Ook is van belang om te kijken of fouten en onregelmatigheden te detecteren en technisch af te vangen.

[16] Artikel 15 lid 1 en 3 AI Act.

ro 11. Hoe wordt er voorkomen dat ongeautoriseerde derden gebruik kunnen maken van kwetsbaarheden van het AI-systeem?[17] ✶

ro 12. Wat is de impact als derden ongewenst toegang hebben tot de broncode, data of uitkomsten van het AI- systeem?

ro 13. Kunnen mensen misbruik maken van het feit dat er een AI-systeem wordt ingezet in plaats van een menselijke beslissing?

ro 14. Hoe wordt er geregistreerd wie er gebruik maakt van het AI-systeem en hoe lang?[18] ✶

[17] Artikel 15 lid 1 en 4 AI Act.
[18] Artikel 12 lid 1 AI Act.

Verantwoordingsplicht

Voor handelen binnen de Rijksoverheid moet verantwoording worden afgelegd binnen de organisatie, naar de Tweede Kamer en naar de samenleving. AI staat op dit moment extra in de belangstelling. De techniek wordt steeds vaker toegepast binnen de Rijksoverheid, maar er zijn ook veel zorgen over de ethische afwegingen bij het inzetten van AI. Daarom moeten er goede mechanismes ingesteld worden om de **VERANTWOORDELIJKHEID** voor AI-systemen en de resultaten daarvan te kunnen garanderen.

Communicatie

v 1. Ben je transparant richting betrokkenen en eindgebruikers over de beperkingen en werking van het AI-systeem? En blijven deze voldoende onder de aandacht zolang ze bestaan

v 2. Worden er mechanismes ingesteld waarin eindgebruikers opmerkingen over het systeem (data, techniek, doelgroep, etc.) kunnen maken? En hoe of wanneer worde deze meldingen gewaarborgd (geanalyseerd en gevolgd)?

Deze sectie gaat over twee vormen van communicatie naar **EINDGEBRUIKERS**. Ten eerste, eindgebruikers moeten ervan op de hoogte worden gesteld dat ze met de resultaten van een **AI-SYSTEEM** te maken hebben. Ten tweede, eindgebruikers hebben te allen tijde het recht om te weten hoe een **ALGORITME** de uitkomsten van een AI-systeem bepaalt. Dat betekent ook dat het doel en beperkingen van het AI-systeem duidelijk en eerlijk moeten worden gecommuniceerd. Zowel technische processen als daaraan gerelateerde menselijke beslissingen moeten begrijpelijk zijn en opgevraagd kunnen worden. Bijvoorbeeld door het aanwijzen van een contactpersoon met inhoudelijke kennis over het AI-systeem. Gezien het zelflerende karakter van AI kan dit niet altijd 100% te herleiden zijn. Wel moet in ieder geval mogelijk zijn om gepaste uitleg te geven over het proces aan eindgebruikers.

Daarnaast is het onder elke vraag binnen deze AIIA belangrijk dat burgers informatie kunnen opvragen over het AI-systeem. Men moet in staat gesteld worden om resultaten van het AI-systeem te kunnen betwisten. Dat betekent ook dat data en de omstandigheden waarin de data ter beschikking zijn gesteld bewaard moeten worden (zie Archivering).

Communicatie met het AI-SYSTEEM

vo 1. Wordt er aan de **EINDGEBRUIKER** en **BETROKKENEN** van het AI-systeem gecommuniceerd dat de resultaten gegenereerd worden door een AI-systeem en wat dat voor hen betekent

vo 2. Zijn er eindgebruiksinstructies opgesteld? Deze moeten minstens het volgende bevatten:[19] ✶
- De naam en contactgegevens van de aanbieder;
- Kenmerken, capaciteiten en beperkingen;
- Mogelijke toekomstige wijzigingen;
- Menselijk toezicht;
- Verwachte levensduur.

vo 3. Wat zijn de potentiële (psychologische) bijwerkingen zoals het risico op verwarring, voorkeur of cognitieve vermoeidheid van de **EINDGEBRUIKER** bij het gebruik maken van het AI-systeem?

Communicatie over de uitkomsten AI-systeem

vo 4. In hoeverre is het mogelijk om een verklaring te geven aan een **BETROKKENE** waarom het AI-systeem op een bepaalde manier werkt?

vo 5. Is het systeem voldoende **TRANSPARANT** om **EINDGEBRUIKERS** in staat te stellen de output(data) van het systeem te interpreteren en op passende wijze te gebruiken?[20] ✶

vo 6. Is er iets ingericht om eindgebruikers eventuele bijscholing te verlenen?

Communicatie naar aanleiding van het AI-systeem

vo 7. Hoe wordt ervoor gezorgd dat commentaar van betrokkenen en eindgebruikers intern goed wordt behandeld?

vo 8. Als een betrokkene bezwaar wil aantekenen,[21] of een klacht wil indienen tegen een besluit van het AI-systeem,[22] is het dan duidelijk welke stappen hij/zij kan nemen? Hetzelfde geldt voor beroep instellen.[23] ✶

Controleerbaarheid

v 3. Hoe wordt het **AI-SYSTEEM** gecontroleerd?

v 4. Hoe is menselijke controle en toezicht gewaarborgd?

[19] Artikel 13 lid 2 en 3 AI Act.
[20] Artikel 13 lid 1 AI Act.
[21] Artikel 7:1 Algemene Wet Bestuursrecht.
[22] Artikel 9 Algemene Wet Bestuursrecht.
[23] Artikel 8:1 Algemene Wet Bestuursrecht.

Met controleerbaarheid kijken we naar op welke manier de processen voor de evaluatie van de data en het **MODEL** en de resultaten gecontroleerd kunnen worden. Deze controle kan in de vorm van audits kan intern of extern plaatsvinden. Naarmate toepassing plaatsvindt op meer kritische gebieden moeten er strengere eisen worden gesteld.

Het is van belang dat er inzicht in de bronnen, het systeem en de uitkomst is. Deze **VERANTWOORDELIJKHEID** zal doorgaans bij de **GEBRUIKER** liggen.

Om autonoom te kunnen zijn in het gebruik van **AI-SYSTEMEN**, moet de **EINDGEBRUIKER** voldoende begrip hebben van het systeem, of de werking ervan kunnen opvragen. Ook is het belangrijk dat begrip over het AI-systeem makkelijk overgedragen kan worden wanneer er een nieuwe eindgebruiker met het systeem gaat werken die niet bij de ontwikkeling betrokken was. Daarom moeten AI-systemen zoveel mogelijk opgesteld worden in samenspraak met de beoogde eindgebruiker. Toezicht kan worden verwezenlijkt door middel van **GOVERNANCE** mechanismen.

vo 9. Hoe wordt rekening gehouden met het ingaan van aangekondigde nieuwe wet- en regelgeving tijdens de levensduur van dit AI-systeem?

vo 10. Hoe wordt ervoor gezorgd dat het AI-systeem onafhankelijk kan worden gecontroleerd?

vo 11. Hoe wordt de correctheid van de **INPUT(DATA)** gecontroleerd en geïnterpreteerd?

vo 12. Hoe wordt de correctheid van het **MODEL** gecontroleerd en geïnterpreteerd?

vo 13. Hoe wordt de correctheid van de **OUTPUT(DATA)** gecontroleerd en geïnterpreteerd?[24] ✶

Archivering

Archivering is het bewaren van informatie zodat je deze informatie in de toekomst kunt hergebruiken of voor andere doelen in kan zetten. Denk bijvoorbeeld aan het herconstrueren van het model (zie 'Reproduceerbaarheid'), of een nieuwe medewerker kunnen uitleggen hoe het systeem in elkaar zit (zie 'Uitlegbaarheid'), of om verantwoording af te leggen naar een **BETROKKENE** (zie 'Verantwoordingsplicht').

Input(data)

vo 14. Hoe wordt de **INPUT(DATA)** opgeslagen?

vo 15. Wat is de bewaartermijn van de input(data)?

Model

vo 16. Hoe wordt het **MODEL** opgeslagen?

Output(data)

vo 17. Kunnen de gebruikers de output(data) op de juiste manier interpreteren?

vo 18. Wat is de bewaartermijn van de output(data)?

[24] Artikel 14 lid 4.

Klimaatadaptie

AI-systemen kunnen bijdragen aan oplossingen voor de meest urgente maatschappelijke zorgen, tegelijkertijd is het van belang dat dit zo milieuvriendelijk mogelijk gebeurt. Het is van belang om de milieuvriendelijkheid van de volledige toeleveringsketen van het **AI-SYSTEEM** te waarborgen.

Aan de andere kant kan het natuurlijk ook zo zijn dat het AI-systeem juist wordt ingezet om milieuwinst te behalen. Die impact moet afgewogen tegen de milieukosten van bijvoorbeeld het laten draaien van het systeem.

Hierbij is het natuurlijk wel van belang om proportionaliteit aan te houden. Als het veel tijd en energie kost om de milieu-impact van een systeem te meten dat maar een hele kleine ecologische voetafdruk heeft, kan je hiertussen een afweging maken.

vo 19. Is er impact op het milieu door het invoeren van het **AI-SYSTEEM** (ontwikkeling, installatie en gebruik), en hoe wordt dit gemeten?

vo 20. Hoe wordt de impact van het AI-systeem afgewogen tegen de milieukosten van het laten draaien van het AI-systeem?

vo 21. Wat voor maatregelen zijn er genomen om de milieu-impact van het AI-systeem te minimaliseren?

Bijlagen

Begrippenlijst

In dit document worden begrippen gebruikt die in de literatuur vaak verschillend gedefinieerd worden. Hieronder volgen eenduidige definities die gebruikt worden in dit document.

ACCEPTATIECRITERIA	Op het beoogde doel en data afgestemde voorwaarden, waaraan het **AI-SYSTEEM** moet voldoen. Dit kan bijvoorbeeld de hoeveelheid data zijn, een accuratesse maatstaaf voor de **OUTPUT(DATA)** of het inrichten van een onafhankelijke controle van output. Acceptatiecriteria moeten waar mogelijk meetbaar gemaakt worden zodat deze gemonitord kunnen worden met een geschikt meetsysteem. Goede acceptatiecriteria zijn SMART en voldoende verschillend zodat alle relevante aspecten van het AI-systeem goed gemonitord worden
ACCURAATHEID	Zeer nauwgezet, precies of zorgvuldig; als een systeem in staat is om juiste én accuratesse beoordelingen te maken. In een formule: TP+TN/(TP+TN+FP+FN). TP= werkelijke positief, TN=Werkelijk negatief, FP=Verkeerde positief, FN= Verkeerd negatief. Hoe meer werkelijke resultaten t.o.v. verkeerde resultaten hoe hoger de accuraatheid.
AI MET BEPERKT RISICO	De AI Verordening stelt vast wat beperkt risico AI is. AI ingericht op interactie met mensen, emoties herkennen, of gemanipuleerde beelden produceren. Denk aan spamfilters, het samenvatten van teksten, het classificeren van onderwerpen van luchtvaartvoorvallen, of bijvoorbeeld AI-systemen die kantoorverlichting regelen.
AI MET HOOG RISICO	De AI Verordening stelt vast wat hoog risico AI is. Dit zijn vaak producten die nauw te maken hebben met fundamentele rechten en/of productveiligheid. Denk hierbij bijvoorbeeld aan AI in vliegtuigen, vaartuigen, voertuigen, rails, wegverkeer, vliegnavigatie en drinkwatertoevoer. Zo lang de AI Verordening nog niet in werking is gesteld, gaat het erom dat we bewust met AI omgaan. Dat betekent dat we ons moeten realiseren wanneer AI een hoog risico heeft.
AI MET MINIMAAL RISICO	Alle AI-systemen die niet verboden zijn of onder **AI MET HOOG RISICO** of **AI MET BEPERKT RISICO** vallen.
AI-SYSTEEM	Een systeem welke (deels) tot stand is gekomen door middel van het toepassen van zelflerende algoritmes (machine learning, statistiek of logica) op historische data, met het doel om voorspellingen of aanbevelingen te doen, of om zelfstandig beslissingen te nemen.
ALGORITME	Een 'recept', of eindige reeks van wiskundige instructies die vanuit een gegeven begintoestand naar een vooraf gesteld doel leiden. Doorgaans zijn deze **ALGORITMES** geïmplementeerd in een computerprogramma.

ARTIFICIAL INTELLIGENCE	AI kent geen eenduidige definitie. Wij hanteren de omschrijving van AI door de Algemene Rekenkamer: “het vermogen [..] om externe gegevens correct te interpreteren, om te leren van deze gegevens en om deze lessen te gebruiken om specifieke doelen en taken te verwezenlijken via flexibele aanpassing”. Ook belichten wij graag al die van de Europese Commissie alhoewel deze nog niet gehanteerd wordt in dit document: AI omvat systemen die intelligent gedrag vertonen door hun omgeving te analyseren en – met een zekere mate van zelfstandigheid – actie ondernemen om specifieke doelen te bereiken.
BEHEERORGANISATIE	Een organisatie die applicatiebeheer van het **AI-SYSTEEM** inricht en optimaliseert.
BELANGENGROEP	Samenstelling van **STAKEHOLDERS** om **DIVERSITEIT** te meten. Dit kan zowel een groep van **EINDGEBRUIKERS** zijn als een groep van mensen die impact ervaren door het systeem.
BETROKKENEN	Natuurlijk persoon of organisatie die bij het gebruik of de uitkomsten van het systeem belang heeft, of belang denkt te hebben. Hier wordt bewust niet het woord ‘belanghebbende’ gebruikt, omdat het meer omvat dan het in het bestuursrecht gedefinieerde ‘belanghebbende’. Denk aan burgers, een onder toezicht staande, maar ook de **EINDGEBRUIKER** zelf.
BETROUWBAAR	De eigenschap beschikken van consistent gedrag en consistente resultaten.
BIAS	Vooringenomenheid. Het doen van aannames over dingen, mensen of groepen die vaak niet gebaseerd zijn op werkelijke metingen.
BIAS IN DE INPUT	Kwaliteit, consistentie en integriteit van data is een belangrijke voorwaarde voor een unbiased analyse.
BIAS IN DE OUTPUT	De manier waarop de **OUTPUT(DATA)** wordt gebruikt kan invloed hebben op de levens van mensen. Het is belangrijk dat hierbij geen onterechte correlatie gaat leiden tot causaliteit.
BIAS IN HET MODEL	Hoe correct zijn de **MODELLEN**; in hoeverre corrigeren ze voor bekende gebreken in representativiteit van de data? Dit kan bijvoorbeeld ook gaan over wat het **AI-SYSTEEM** leert en wat ongewenste leereffecten zijn.
CIO	Chief Information Officer.
CISO	Chief Information Security Officer.
CORRUPTIE	Het misbruiken of uitbuiten van fouten van het systeem, of het uitbuiten van ogenschijnlijke neutrale eigenschappen van het systeem.[25] We maken onderscheid met **ONBEDOELDE CORRUPTIE**.
DATA BIAS	Wanneer de steekproef niet representatief is voor de gehele populatie.
DATA PIPELINE	Hoe de data vanuit het veld naar het model komt; het proces dat de data doorloopt.

[25] Voorbeeld: Een formulier waarvan de resultaten worden gebruikt voor input(data) van het model kan door derden per ongeluk (onbedoelde corruptie) of expres (bedoelde corruptie) met foutieve informatie ingevuld worden zodat het algoritme op verkeerde data kan gaan draaien.

DESIGN BIAS	Problemen in het technisch ontwerp, inclusief beperkingen van computerhulpmiddelen zoals hardware en software.
DIVERSITEIT	Hieronder verstaan we het herkennen van verschillende typen 'subjecten' in onze analyses. Wij proberen hierbij te voorkomen dat groepen van relevante subjecten onterecht niet worden meegenomen in ontwikkelen van een **AI-SYSTEEM**, waardoor het systeem niet op hen aansluit.
DOEL 1	De Algemene Rekenkamer noemt 3 doelen die AI kunnen hebben. Doel 1 is gericht op het automatiseren van eenvoudige menselijke handelingen. Kenmerkend aan dit soort **ALGORITMES** is dat ze vaak voorschrijvend zijn en automatisch een handeling uitvoeren, zonder tussenkomst van een mens. Het risico op fouten met impact op de burger is hierbij laag, gezien de hoge technische **TRANSPARANTIE** en eenvoudig toepassingsgebied.
DOEL 2	De Algemene Rekenkamer noemt 3 doelen die AI kunnen hebben. Doel 2 is gericht op het faciliteren van bedrijfsvoering. Hierbij wordt vaak complexere data gebruikt dan bij **DOEL 1**. Het zijn vaak voorspellende **ALGORITMES** zonder automatische besluitvorming. Het risico op fouten met impact op de burger is aanwezig maar is beperkt. Het algoritme doet namelijk alleen voorbereidend 'werk'.
DOEL 3	De Algemene Rekenkamer noemt 3 doelen die AI kunnen hebben. Doel 3 is gericht op (risico-)voorspellingen. Er is geen sprake van automatische besluitvorming. Het risico op fouten met impact op de burger is hoog. Bijvoorbeeld dat de resultaten in strijd zijn met de wet, of (ongewenste) afwijking vertonen op basis van de verbogen beperkingen van **INPUT(DATA)**. Verklaarbaarheid komt hierbij in gevaar. Daarnaast bestaat de kans dat het advies van het **ALGORITME** de uiteindelijke beslissing van de medewerker beïnvloedt.
DOMEINEXPERT	Iemand die veel kennis heeft over het probleemgebied waarin het **AI-SYSTEEM** gebouwd wordt.
EERLIJKHEID	Als niet elk subject een gelijke behandeling krijgt, moet dat verklaard kunnen worden. Hierbij is het van belang dat we zoveel mogelijk onderscheidende subjectkenmerken in beeld hebben. Zowel om aan te kunnen tonen welke kenmerken daadwerkelijk een rol spelen (en partij A een lager risico toebedelen dan partij B) en welke kenmerken dit juist niet zijn (waardoor partij A en B een onderbouwd gelijkwaardig risico hebben).
EINDGEBRUIKER	Eindgebruikers zijn de personen die het **AI-SYSTEEM** in de praktijk toepassen binnen de '**GEBRUIKER**' als organisatie. Het gaat hierbij om een natuurlijk persoon. Wie zitten er met de handen aan de knoppen? Wie vergaart binnen de organisatie informatie uit het AI-systeem? Denk aan een inspecteur of een wegverkeersleider.
EINDVERANTWOORDELIJKE	Een rol binnen de organisatie die de **VERANTWOORDELIJKHEID** over het **AI-SYSTEEM** draagt. Dat betekent bijvoorbeeld de verantwoordelijkheid over dat de juiste resultaten van het AI-systeem bereikt worden.
ENTITEIT	Een functie binnen een afdeling van een organisatie.

GEBRUIKER	Volgens de AI Verordening "een (...) overheidsinstantie, agentschap of ander orgaan die/dat een **AI-SYSTEEM** onder eigen verantwoordelijkheid gebruikt (...)". De gebruiker zet het systeem in. Dit is nooit een natuurlijk persoon. Bijvoorbeeld de ILT of RWS.
GEEN POSITIEVE IMPACT	**BETROKKENEN** die niet per definitie negatieve impact ervaren van de inzet van het AI-systeem, maar bijvoorbeeld in dezelfde situatie blijven als daarvoor. Daarbij kan een gevaar zijn dat deze betrokkenen niet dezelfde 'positieve impact' van het AI-systeem ervaren dat andere betrokkenen wel krijgen.
GELIJKHEID	Hieronder verstaan we de gedachte dat elk gelijksoortig subject een gelijke behandeling krijgt.
GOVERNANCE	De handeling of de wijze van besturen, de gedragscode en het toezicht op organisaties. Het betreft beslissingen die verwachtingen bepalen, macht verlenen of prestaties verifiëren. Het bestaat ofwel uit een afzonderlijk proces ofwel uit een specifiek deel van management- of leiderschapsprocessen.
HACKAANVAL	Inbreken in het **AI-SYSTEEM**. Met als gevolg bijvoorbeeld vervuiling van data, het ongewenst uitlekken van (de werking van) een AI-systeem, of aantasting van software of hardware.
IN GEBRUIK NEMEN	Het moment dat een AI-systeem 'in gebruik wordt genomen' betekent het moment waarop deze voor het eerst buiten de deur wordt gebruikt. In de praktijk betekent dit dus ook een externe test of pilot. Op het moment van in gebruik name moet de AIIA af zijn.
INPUT(DATA)	Die gegevens welke worden verwerkt met een vooropgesteld doel. In de context van een **AI-SYSTEEM** kunnen dit ruwe data zijn, bijvoorbeeld de waarnemingen uit de werkelijkheid. In de context van het **MODEL** zijn dit normaal gesproken voorbewerkte data.
MODEL	Een (versimpelde) wiskundige vertegenwoordiging van de werkelijkheid, welke wordt gebruikt om informatie te verwerken. In een **AI-SYSTEEM** wordt de wiskundige vertegenwoordiging vaak deels of in zijn geheel volgens een **ALGORITME** 'geleerd', waardoor zelfs door de **ONTWIKKELAARS** niet volledig uit te leggen is hoe het model aan diens uitkomsten komt.
MOREEL BERAAD	Overlegorgaan Fysieke Leefomgeving (mei 2021), Moreel Beraad.
NEGATIEVE IMPACT	**BETROKKENEN** die nadelige gevolgen ervaren door de toepassing van het **AI-SYSTEEM**, bijvoorbeeld omdat ze gediscrimineerd worden op basis van een **BIAS** in het AI-systeem.
ONBEDOELDE CORRUPTIE	Zonder kwaadwillende bedoelingen de werking van het **AI-SYSTEEM** beïnvloeden door bijvoorbeeld verkeerde input te voeden, of de verkeerde knoppen in te drukken. Onbedoelde corruptie valt onder **BETROUWBAARHEID**. We maken onderscheid met (bedoelde) corruptie.
ONTWIKKELAAR	Een organisatie of een persoon die een **AI-SYSTEEM** ontwerpt, ontwikkelt en/of traint.

OPDRACHTGEVER	Een persoon of organisatieonderdeel die een opdracht verstrekt aan een opdrachtnemer. Deze is ook (samen met de projectleider) eindverantwoordelijk voor het maken van een AIIA.
OUTPUT(DATA)	De gegevens die een **AI-SYSTEEM** oplevert. Dit zijn de resultaten van het **MODEL**.
PARAMETER	Een variabele binnen het **MODEL**. Wanneer deze variabele gewijzigd wordt, wordt ook de resulterende grootheid van het model of van de berekening gewijzigd.
POSITIEVE IMPACT	**BETROKKENEN** die gunstige gevolgen ervaren van de inzet van het **AI-SYSTEEM**. Denk aan een minderheidsgroep die wordt bevoordeeld. Daarbij kan het gevaar zijn dat deze positieve bias te optimistisch is, en dus niet waarheidsgetrouw. Ook kan de keerzijde hiervan een 'negatieve impact' voor andere betrokkenen zijn.
PROJECTLEIDER	De eindverantwoordelijke voor het project waarbinnen het **AI-SYSTEEM** valt. Deze is ook (samen met de **OPDRACHTGEVER**) **EINDVERANTWOORDELIJK** voor het maken van een AIIA.
PROPORTIONEEL	AI is een ingrijpende techniek, met verklaarbaarheidsproblemen. Staat het gebruik van AI in verhouding tot het probleem wat er met het **ALGORITME** opgelost gaat worden? Het verwachte voordeel moet groter zijn dan het risico dat AI met zich meebrengt.
REPRODUCEERBAAR	Het steeds opnieuw kunnen bereiken van een vergelijkbaar resultaat wanneer een beschreven procedure wordt uitgevoerd.
ROBUUSTHEID	Met een preventieve benadering ontwikkeld zijn; zich gedragen zoals voorzien en van tevoren beschreven. Onaanvaardbare schade vermijden.
SEED	Een "seed" is het uitgangspunt van een willekeurig getal generator. Deze generator maakt altijd vanuit dit uitgangspunt volgens dezelfde "route" nieuwe (pseudo) willekeurige getallen. Door de "seed" te documenteren kan de "route" van (pseudo) willekeurige getallen worden herhaald. Dit betekent dat deze seed nodig is om reconstructie van een **MODEL** te controleren wanneer het model ergens gebruik maakt van willekeurige getallen.
DE SEED ZELF IS OOK EEN GETAL	Er zijn geen specifieke eisen aan dit getal, dus vaak wordt er voor iets "herkenbaars" gekozen (bijvoorbeeld "123456", of "0, 42, 1234" of de geboortedatum van een **ONTWIKKELAAR**).
STAKEHOLDER	Persoon of organisatie die een beslissing of activiteit kan beïnvloeden, erdoor kan worden beïnvloed, of zichzelf als beïnvloed beschouwd. Een stakeholder kan bijvoorbeeld ook de eigenaar van gebruikte data zijn.
SUBSIDIAIR	AI is een ingrijpende techniek, met verklaarbaarheidsproblemen. Kan het probleem ook met minder vergaande middelen opgelost worden?
TRACEERBAARHEID	Wanneer processen en resultaten te controleren zijn.
TRANSPARANT	Wanneer de werking en doelen van het **AI-SYSTEEM** duidelijk worden gecommuniceerd en resultaten van het AI-systeem **UITLEGBAAR** zijn.

TYPE ALGORITMES	Verschillende technieken kunnen gebruikt worden om AI te maken, zoals neurale netwerken, random forests of andere vormen van machine learning. Maar ook minder complexe **ALGORITMES** zoals business-rules of beslisbomen kunnen gebruikt worden.
UITLEGBAAR	Een verklaring van hoe input variabelen bijdragen aan een output van het algoritme die uitgelegd moet worden.
VERANTWOORDELIJK	Handelingen van een **ENTITEIT** kunnen op unieke wijze worden herleid tot die entiteit, en deze entiteit is voor deze handelingen aansprakelijk. Wie is wie Vul in welke personen een rol hebben gespeeld bij het beantwoorden van deze AIIA.

Wie is wie

Vul in welke personen een rol hebben gespeeld bij het beantwoorden van deze AIIA.

BELANGENGROEP: .....

CISO of **CIO**: .....

Communicatieadviseur: .....

Data scientists: .....

Databeheerder of bronhouder: .....

DOMEINEXPERT: .....

Functionaris Gegevensbescherming: .....

Jurist: .....

OPDRACHTGEVER: .....

Overige leden projectteam: .....

PROJECTLEIDER: .....

Strategisch adviseur ethiek: .....

Wie doet wat

	H1	H2.	H3.	H4.	H5.	H6.
BELANGENGROEP:						
CISO of **CIO**:					X	
Communicatieadviseur:		X				X
Data scientists:		X	X	X	X	X
Databeheerder of bronhouder:						X
DOMEINEXPERT:		X	X	X	X	X
Functionaris Gegevensbescherming:				X		
Jurist:		X				
OPDRACHTGEVER:	X	X	X	X	X	X
Overige leden projectteam:						
PROJECTLEIDER:	X	X	X	X	X	X
Strategisch adviseur ethiek:						

Dit is een publicatie van:

Ministerie van Infrastructuur en Waterstaat

Postbus 20901

2500 EX Den Haag

November 2022 | 73263