Artificial Intelligence Impact Assessment

ECP
Platform voor de InformatieSamenleving

Stappenplan voor het uitvoeren van de AIIA

Organisaties die de AIIA willen gaan doen, kunnen onderstaand stappenplan volgen. Een toelichting hierop is opgenomen in 'Deel 2: Uitvoeren van de AIIA' op pagina 35.

Stap 1 Bepaal de noodzaak voor het doen van een AIIA

1. Wordt de AI toegepast in een nieuw (maatschappelijk) domein?
2. Wordt gebruik gemaakt van een nieuwe vorm van AI-technologie?
3. Heeft de AI een hoge mate van autonomie?
4. Wordt de AI toegepast in een complexe omgeving?
5. Wordt gebruik gemaakt van gevoelige gegevens over personen?
6. Neemt de AI beslissingen die natuurlijke of rechtspersonen in aanzienlijke mate treffen dan wel rechtsgevolgen voor hen hebben?
7. Is de besluitvorming door de AI complex?

Stap 2 Beschrijf de AI-toepassing

1. Beschrijf de toepassing en het doel van de toepassing
2. Beschrijf welke AI-technologie wordt ingezet om het doel te bereiken
3. Beschrijf welke data worden gebruikt in het kader van de toepassing
4. Beschrijf welke actoren een rol spelen in de toepassing

Stap 3 Beschrijf de baten van de AI-toepassing

1. Wat zijn de baten voor de organisatie?
2. Wat zijn de baten voor het individu?
3. Wat zijn de baten voor de maatschappij als geheel?

Stap 4 Is het doel en de wijze waarop het doel wordt bereikt ethisch en juridisch verantwoord?

1. Welke actoren zijn betrokken bij en/of worden geraakt door mijn toepassing van AI?
2. Zijn deze waarden en belangen geconcretiseerd in wet- en regelgeving?
3. Welke waarden en belangen spelen een rol in de context van mijn toepassing van AI?

Stap 5 Is de toepassing betrouwbaar, veilig en transparant?

1. Welke maatregelen zijn genomen om de betrouwbaarheid van het handelen van de AI te borgen?
2. Welke maatregelen zijn genomen om de veiligheid van de AI te borgen?
3. Welke maatregelen zijn genomen om de transparantie van het handelen van de AI te borgen?

Stap 6 Afweging en beoordeling

Stap 7 Vastlegging en verantwoording

Stap 8 Evalueer periodiek

Inhoud

Voorwoord 7

Introductie 13

Noodzaak AIIA 14
Definitie van Artificial Intelligence 14
Voor wie is de Impact Assessment? 15
Hoe ziet het stappenplan eruit? 17
Interdisciplinaire vragen en uitgangspunten 19
Actualisering AIIA 19
Maatschappelijke vragen 19
Ethische overwegingen 20
Transparantie 22

Deel 1 - Achtergrond Artificial Intelligence Impact Assessment 25

Ethische en juridische toetsing 27
Toepassen in ontwerpfase 29
Betrekken Stakeholders 29
Verhouding Privacy Impact Assessment (PIA) 30
Praktische toepassing AIIA en ethiek 30

Deel 2 - Uitvoeren van de AIIA 35

Stap 1 Bepaal de noodzaak voor het doen van een AIIA 39
Stap 2 Beschrijf de AI-toepassing 42
Stap 3 Beschrijf de baten van de AI-toepassing 46
Stap 4 Is het doel en de wijze waarop het doel wordt bereikt ethisch en juridisch verantwoord? 48
Stap 5 Is de toepassing betrouwbaar, veilig en transparant? 51
Stap 6 Afweging en beoordeling 57
Stap 7 Vastlegging en verantwoording 59
Stap 8 Evalueer periodiek 60

Bibliografie 63

Bijlage 1 - Gedragscode Artificiële Intelligentie 67

Bijlage 2 - Stappenplan AIIA 83

Colofon

2018 © ECP | Platform voor de InformatieSamenleving

Met dank aan Turnaround Communicatie

Voorwoord

Het maatschappelijk debat rondom AI heeft zich snel ontwikkeld. Naast de potentiële voordelen van AI, is er daarbij snelgroeiende aandacht voor bedreigingen en risico's (transparantie, privacy, autonomie, cybersecurity *et cetera*) die om een zorgvuldige benadering vragen. Voorbeelden uit het recente verleden (slimme meters, ov-chipkaart) laten zien dat de invoering van IT-toepassingen niet ongevoelig is voor het juridische en ethische debat. Dat geldt ook bij de toepassing van AI. Het vooraf in beeld brengen en adresseren van de impact van AI draagt bij aan een soepele en verantwoorde introductie van AI in de samenleving.

> Wat zijn de relevante juridische en ethische vragen voor onze organisatie als wij besluiten AI in te zetten?

De AIIA helpt bij het beantwoorden van deze vraag en vormt uw gids op weg naar het vinden van het juiste normenkader en het maken van de relevante afwegingen.

De "Gedragscode Artificiële Intelligentie" vormt het vertrekpunt voor deze impact assessment en maakt integraal onderdeel uit van de AIIA. De gedragscode is als bijlage 1 opgenomen in dit document. De gedragscode biedt een set van regels en uitgangspunten die bij de inzet van AI veelal relevant zijn. Omdat zowel het begrip "AI" als het toepassingsgebied zeer ruim zijn, vormt de gedragscode een startpunt voor het ontwikkelen van het juiste juridische en ethische kader waartegen getoetst kan worden. De aard van de AI-toepassing en de context waarin deze wordt gebruikt bepalen in belangrijke mate welke afwegingen in een concreet geval moeten worden gemaakt. Zo zullen toepassingen in de medische sfeer deels tot andere vragen en aandachtspunten leiden dan AI-toepassingen op het terrein van logistiek.

"Kunstmatige intelligentie is geen revolutie. Het is een ontwikkeling die langzaamaan onze samenleving binnentreedt en zich ontwikkelt tot een bouwsteen voor de digitale samenleving. Door steeds hype van realiteit te scheiden, partijen te duiden en te verbinden én de balans te bewaken tussen mogelijkheden, ethiek en rechtsbescherming, zullen we meer en meer de vruchten plukken van AI."

— Daniël Frijters, MT-lid en projectadviseur bij ECP|Platform voor de Informatiesamenleving

De AIIA bevat concrete stappen om u te helpen bij het inzichtelijk maken van de relevante juridische en ethische normen en afwegingen bij besluitvorming inzake de inzet van AI-toepassingen. Ook biedt de AIIA een kader voor het aangaan van de dialoog met belanghebbenden binnen en buiten uw organisatie. De AIIA faciliteert daarmee het gesprek over de inzet van AI.

"AI biedt veel kansen maar leidt ook tot serieuze uitdagingen op het terrein van recht en ethiek. Alleen in samenspraak kunnen daarvoor oplossingen met voldoende draagvlak worden gevonden. De door ECP ontwikkelde gedragscode en daaraan gekoppelde AI Impact Assessment zijn belangrijke hulpmiddelen om ten aanzien van concrete toepassingen de dialoog aan te gaan. Dat draagt bij aan een verantwoorde ontwikkeling en implementatie van AI in de samenleving."

— Prof.dr. Kees Stuurman, Voorzitter ECP-werkgroep Gedragscode AI

AI Impact Assessment steun in de rug

De AIIA is niet bedoeld om organisaties de maat te nemen bij de inzet van AI. Organisaties blijven zelf verantwoordelijk voor de keuzes die zij maken rondom de inzet van AI. Het toepassen van de AIIA is ook niet verplicht en ook geen extra administratieve last. In tegendeel; de AIIA vormt een steun in de rug bij de inzet van AI. Verantwoorde toepassing van AI vermindert immers de risico's en lasten, en helpt de gebruiker en de samenleving vooruit (win-win).

De AIIA is primair gericht op organisaties die AI willen inzetten in de bedrijfsvoering, maar kan ook door ontwikkelaars van AI worden gebruikt om toepassingen te toetsen.

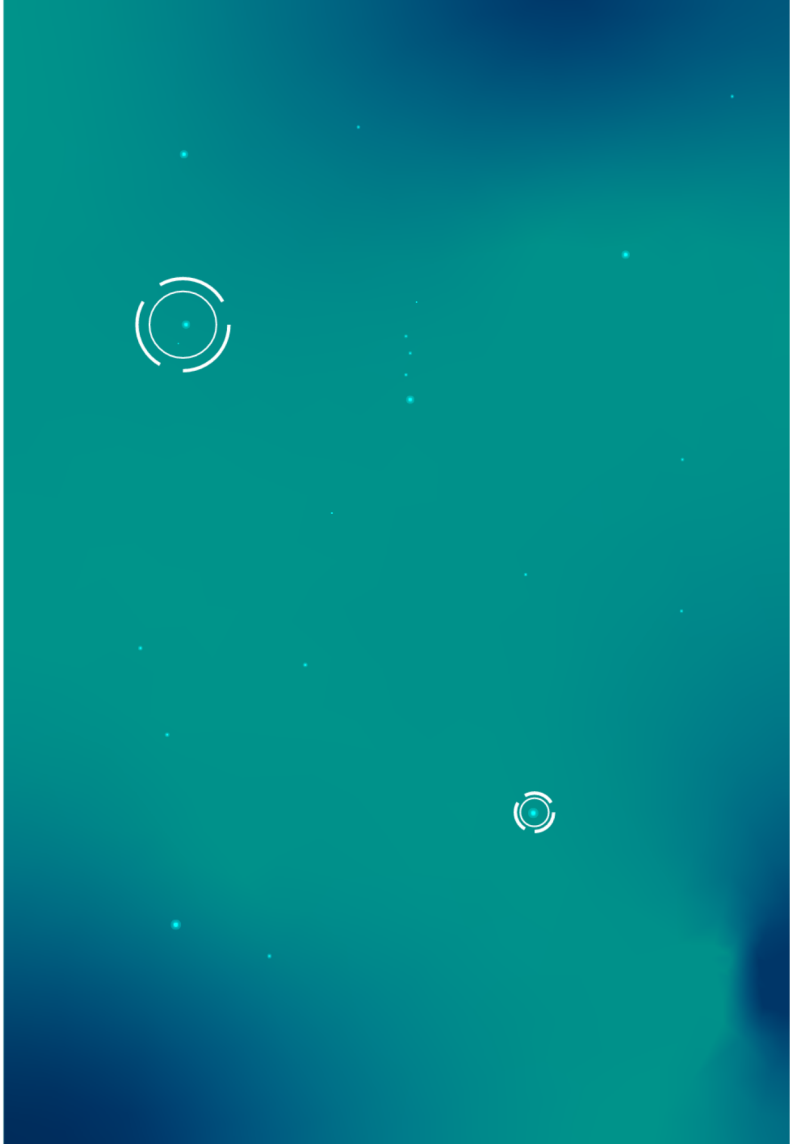
Wij hopen dat de AIIA zijn weg naar de praktijk zal weten te vinden en een effectieve bijdrage zal leveren aan de maatschappelijk verantwoorde introductie van AI in de samenleving.

Prof. dr. Kees Stuurman
Voorzitter ECP Werkgroep Gedragscode AI

Daniël Frijters
MT-lid en projectadviseur ECP

Drs. Jelle Attema
Secretaris

Mr. dr. Bart W. Schermer
Lid werkgroep en CKO Considerati



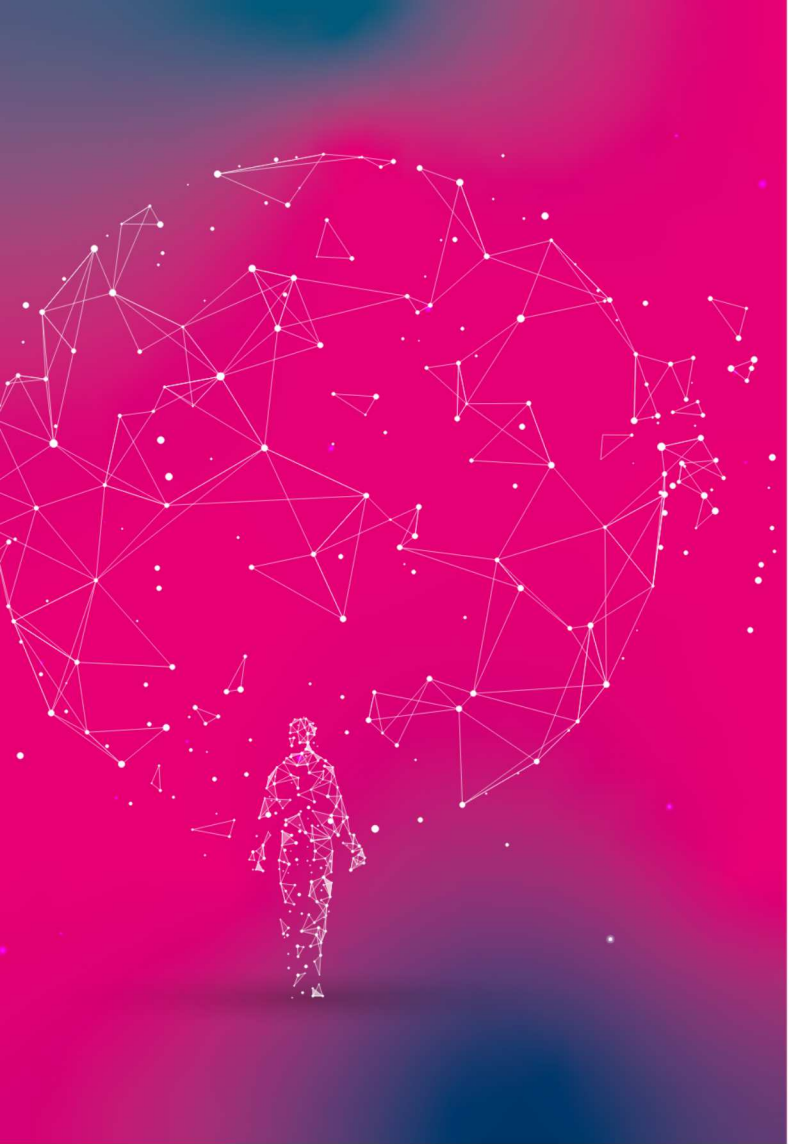
In de werkgroep "Artificial Intelligence Impact Assessment" hadden (op persoonlijke titel) zitting:

Kees Stuurman (voorzitter) *Van Doorne advocaten, Tilburg University* • **Bart Schermer** *Considerati, Universiteit Leiden* • **Daniël Frijters** *ECP | Platform voor de InformatieSamenleving* • **Frances Brazier** *Technische Universiteit Delft* • **Jaap van den Herik** *Universiteit Leiden* • **Joost Heurkens** *IBM* • **Leon Kester** *TNO* • **Maarten de Schipper** *Xomnia* • **Sandra van der Weide** *Ministerie van Economische Zaken en Klimaat* • **Jelle Attema (secretaris)** *ECP | Platform voor de InformatieSamenleving.*

De volgende personen en organisaties hebben waardevol commentaar gegeven op de voorlopige versie (op persoonlijke titel):

Femke Polman en Roxane Daniels *VNG, Data Science Hub* • **Medewerkers van het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties,** *afdeling Informatiesamenleving van de directie Informatiesamenleving en Overheid* • **Marc Welters** *NOREA, EY* • **Marijn Markus, Reinoud Kaasschieter en Martijn van de Ridder** *CAP GEMINI* • **Rob Nijman** *IBM* • **Stefan Leijnen** *Asimov Institute*

Door Considerati is in opdracht van ECP een belangrijke bijdrage geleverd aan het opstellen van de AI Impact Assessment. Met name danken we Joas van Ham, Bendert Zevenbergen en Bart Schermer voor hun inspanningen.



Introductie

De Artificial Intelligence Impact Assessment (AIIA) bouwt voort op de "Handreiking voor gedragsregels Autonome Systemen" (ECP.NL, 2006), die zich richtte op de juridische aspecten van het toepassen van autonome systemen: systemen die handelingen verrichten met juridische gevolgen. De handreiking werd geschreven door een breed samengestelde groep van experts: juristen, bestuurskundigen en technici, afkomstig uit wetenschap, bedrijfsleven en overheid. Het initiatief voor de handreiking werd genomen door ECP. De toenmalige handreiking ontstond op verzoek van ECP-deelnemers, uit bedrijfsleven en overheid, vanwege de grote vlucht die autonome systemen toen leken te nemen en voor zogenaamde "autonomous agents".

De Handreiking richtte zich in 2006 vooral op de juridische aspecten. De AIIA is breder en omvat nu ook de ethische aspecten: een breed gedeelde opvatting in de werkgroep (nog grotendeels bestaande uit dezelfde organisaties en mensen als in 2006) is dat AI welzijn moet verbeteren en menselijke waarden niet alleen moet respecteren maar ook bevorderen.

Noodzaak AIIA

De vraag is gerechtvaardigd, gezien de sterk fluctuerende belangstelling voor AI, of en waarom een Artificial Intelligence Impact Assessment nodig is.

De belangrijkste reden is, dat AI steeds vaker taken overneemt of uitvoert in samenspel met mensen, waarin het ethisch besef van mensen een sturende rol speelt: in het onderwijs, de zorg, bij werk en inkomen en bij de overheid. Bovendien kunnen organisaties, door AI, nieuwe rollen op zich nemen, waar ethische aspecten een rol spelen. Bijvoorbeeld in preventie, controle en fraudedetectie.

Veel van deze voorbeelden van autonomie en intelligentie zijn in technisch opzicht vaak niet spectaculair, maar kunnen desondanks grote impact hebben op degenen die met die systemen te maken krijgen.

De AIIA is zinvol bij AI-toepassingen die handelingen uitvoeren of beslissingen nemen, al dan niet samen met mensen, die vroeger door mensen werden uitgevoerd en waarbij ethische vragen een rol spelen. De Impact Assessment is ook relevant wanneer een organisatie nieuwe doelstellingen nastreeft of activiteiten uitvoert, die door AI mogelijk worden en waar vragen van welzijn, menselijke waarden en juridische kaders relevant zijn.

De waarde van de AIIA is niet afhankelijk van de mate van autonomie of intelligentie van de ICT. Ook al maken de snelle ontwikkelingen rond AI deze vraag wel concreter en urgenter.

Definitie van Artificial Intelligence

Er is weinig overeenstemming over de definitie van Kunstmatige Intelligentie of Artificial Intelligence (AI).[1] De AIIA volgt de beschrijving en benadering van de IEEE (The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2017).

> "Autonome en/of intelligente systemen (AI/S) zijn systemen die op basis van gedefinieerde principes kunnen redeneren, beslissen, intenties kunnen vormen en handelingen uitvoeren."

De IEEE heeft het initiatief genomen om meer dan tweehonderd experts en wetenschappers van over de hele wereld na te laten denken over de ethische aspecten van autonome en intelligente systemen. Voor de diverse aspecten zijn werkgroepen in het leven geroepen om tot normen en gedragsregels te komen. Het document weerspiegelt de consensus onder een brede groep experts, afkomstig uit vele delen van de wereld en culturen.

Kernelementen uit de benadering van de IEEE, en ook van de AIIA, is dat ethische toepassing van AI betekent dat AI moet bijdragen aan welzijn van mens en planeet. De IEEE volgt de operationalisatie van de OECD van welzijn (OECD, 2018). Deze omvat tal van onderwerpen zoals mensenrechten, economische doelstellingen, onderwijs en gezondheid, maar ook subjectieve aspecten van welzijn. Wat "bijdragen aan welzijn" betekent voor een specifiek project vergt analyse en afweging van vaak vele (soms tegenstrijdige) eisen met oog voor de specifieke culturele context. De AIIA biedt de "Gedragscode Artificiële Intelligentie" (Bijlage 1) als een vertrekpunt bij die analyse. Het derde aspect dat de IEEE benadrukt is dat de toepasser van AI verantwoordelijk is voor de impact van AI en processen moet inrichten om de positieve gevolgen te realiseren en de negatieve te voorkomen en te beheersen.

Voor wie is de Impact Assessment?

De Impact Assessment is voor organisaties die AI willen inzetten in hun (dienstverlenings)processen en een analyse willen doen van de juridische en ethische gevolgen. In de ontwerpfase (omdat dan kostbare fouten voorkomen kunnen worden) maar ook tijdens het gebruik: organisaties zullen vaak de gevolgen van hun dienst willen overzien. Het doorlopen van de Impact Assessment betekent veel werk: een deel kan echter

worden hergebruikt omdat een belangrijk deel van de ethische en juridische uitgangspunten generiek zal zijn voor een bepaalde technologie, voor een bepaalde sector of een bepaalde professie.

De organisatie die AI wil toepassen, voert de Impact Assessment uit. Technologie behoort namelijk te functioneren binnen de juridische en ethische kaders van de organisatie die AI inzet, binnen de kaders van de professionals die samenwerken met AI of delen van hun werk overdragen aan de technologie, de eindgebruikers en de samenleving.

De uitkomsten van de Impact Assessment leiden soms tot bepaalde eisen aan de technologie (bepaalde functionaliteit), organisatorische maatregelen (bijvoorbeeld een fall-back wanneer eindgebruikers menselijk contact willen, of nieuwe taakverdelingen om incidenten te voorkomen en af te handelen), verdere opleiding en training (hoe draagt een arts, accountant, jurist of ambtenaar zijn professionele verantwoordelijkheid wanneer taken door AI worden uitgevoerd; hoe interpreteert een professional de adviezen van AI, wat zijn de zwakke en sterke punten van deze adviezen en hoe komen ze tot stand) en het verzamelen van data over de precieze uitwerking in de praktijk.
De aanbieder en producent van de AI-oplossing moeten zorgen voor een aantal technische randvoorwaarden (bijvoorbeeld rond integriteit van data, veiligheid en continuïteit), maar ook voor voorzieningen waardoor de organisatie die de AI inzet verantwoordelijkheid kan nemen en transparant kan zijn over de gevolgen. De aanbieder van de technologie kan met de Impact Assessment in de hand, organisaties helpen om de juiste vragen te stellen en afwegingen te maken.

> Het uitgangspunt van deze Impact Assessment is dat de organisatie die AI toepast verantwoordelijkheid neemt voor AI.

Voor de werkgroep is dat fundamenteel: de zwarte scenario's rond AI gaan meestal over technologie waarbij de ethische kaders door een externe partij (wellicht de fabrikant, een kwaadwillende of de technologie zelf) worden bepaald.

Dit assessment helpt, met algemene principes en uitgangspunten in de hand, na te gaan wat die principes betekenen voor een specifieke toepassing: voor het ontwerp van de technologie, voor de inrichting van de organisatie die technologie toepast, voor de bestuurders die verantwoording moeten kunnen afleggen, de professionals en vakmensen die met de technologie werken of er taken aan overdragen, voor de eindgebruikers die de gevolgen ondervinden, en de samenleving.

Hoe ziet het stappenplan eruit?

Of het nuttig is de Impact Assessment uit te voeren is vaak afhankelijk van de combinatie van dienst, organisatie, eindgebruikers en samenleving.

1

Stap 1 van de Impact Assessment bestaat uit een aantal screeningsvragen om de vraag te beantwoorden of het zinvol is de assessment uit te voeren. Deze vragen hebben betrekking op:

1. de **maatschappelijke en politieke context van de toepassing** (ervaring met de technologie in dit domein, raakt de technologie gevoelige thema's),
2. **kenmerken van de technologie zelf** (autonomie, complexiteit, begrijpelijkheid, voorspelbaarheid),
3. en de **processen waarvan de technologie deel uitmaakt** (complexiteit van de omgeving en besluitvorming, transparantie, begrijpelijkheid en voorspelbaarheid van de uitkomsten begrijpen, de impact voor mensen).

> Bij één of meerdere positieve antwoorden op de screeningsvragen, kan het zinvol zijn de Impact Assessment uit te voeren.

2

Het uitvoeren van de Impact Assessment begint vervolgens in **stap 2** met het beschrijven van het project: de doelen die met de inzet van AI worden beoogd, de data die worden gebruikt, de actoren zoals de eindgebruikers en andere belanghebbenden. Denk daarbij ook aan de professionals in een organisatie die met AI moeten samenwerken of werk overdragen aan AI.

De doelen van het project worden in **stap 3** niet alleen geformuleerd op het niveau van de eindgebruiker die de gevolgen ondervindt van de dienst, maar ook op het niveau van de organisatie die de dienst aanbiedt en de samenleving. Deze brede benadering van doelen is belangrijk, omdat ethische en juridische aspecten aan de orde komen die betrekking hebben op de relatie tussen organisatie en haar omgeving.

Stap 4 gaat in op de ethische en juridische aspecten van de toepassing. In deze stap worden de relevante ethische en juridische kaders in kaart gebracht en toegepast op de toepassing. Er zijn voor een toepassing veel relevante bronnen voor ethische en juridische kaders: sommige zijn formeel (wetten, besluiten), andere meer informeel: gedragscodes, convenanten of beroepscodes.

Bij **stap 5** maken organisaties zelf strategische en operationele keuzes met een ethische component: hoe ze hun activiteiten willen uitvoeren in relatie tot hun klanten, medewerkers, toeleveranciers, concurrenten en de samenleving.

Een weging van de verschillende facetten, gerelateerd aan ethische en juridische aspecten, wordt in **stap 6** gemaakt. In deze stap worden besluiten genomen over de toepassing van AI.

Deze stappen worden afgesloten door in **stap 7** de vorige stappen goed te documenten en genomen besluiten te verantwoorden.

En door in **stap 8** en in stap 8 de impact van AI te gaan bewaken en evalueren. Omdat de toepassing van AI vaak veranderingen zal veroorzaken in de manier waarop naar ethische en juridische aspecten wordt gekeken, zal dat ook vaak onderwerp van die evaluatie.

Interdisciplinaire vragen en uitgangspunten

De Impact Assessment en de gedragscode zijn uitgewerkt door een breed samengestelde groep experts. Een belangrijke uitdaging was het overbruggen van de verschillende perspectieven. Een jurist kijkt anders naar ethiek dan een aanbieder van deze systemen, een ingenieur, een ambtenaar of een IT-auditor. In de Impact Assessment en de gedragscode is getracht gemeenschappelijke vragen en uitgangspunten te formuleren, die verschillende disciplines vanuit hun eigen invalshoek en expertise adresseren. De handreiking maakt die discipline-specifieke analyses niet overbodig.

Actualisering AIIA

De Impact Assessment en de Gedragscode zijn vastgesteld naar de inzichten van nu. Onder invloed van het maatschappelijke debat en de ervaringen met nieuwe technologie veranderen echter verwachtingen, rollen, normen en waarden. Zo verandert de inhoud van beroepen en de criteria waarop professionals worden beoordeeld. Ook de verwachtingen van eindgebruikers veranderen als bepaalde technologie gemeengoed wordt. Deze veranderingen zijn niet of nauwelijks te voorzien: een belangrijk element in de Impact Assessment is daarom het plannen van nieuwe assessments en het verzamelen van gegevens over de impact van technologie. En dat steeds tegen de actuele stand van zaken op het terrein van toepasselijke (rechts)regels en het maatschappelijk debat.

Maatschappelijke vragen

De Impact Assessment gaat in op de gevolgen van toepassing van AI in organisaties. Veel vraagstukken rond nieuwe technologie worden daarmee niet beantwoord: bijvoorbeeld wat automatisering en robotisering doet met de inhoud van werk en werkgelegenheid, of wat AI betekent voor marktverhoudingen. Vraagstukken als interoperabiliteit van datasets en regie op gegevens worden niet geadresseerd. Het maatschappelijke en

politieke debat rond deze aspecten is erg belangrijk voor de eisen die aan AI worden gesteld. Lezers die zich een beeld willen vormen van deze aspecten wordt aangeraden om publicaties als "Opwaarderen" (Rathenau-instituut) of "Mens en Technologie" (SER) te lezen.[2]

Ethische overwegingen

De Impact Assessment stelt dat ethische vragen niet uitsluitend (gaan) spelen bij vormen van AI die nu nog niet mogelijk zijn: ook de huidige (simpele vormen van) AI en veel oudere ICT systemen raken al aan ethische vragen.

Een belangrijk onderscheid rond ethiek en AI is het onderscheid tussen Artificial Narrow Intelligence (ANI) en Artificial General Intelligence (AGI): doel van AGI is dat intellectuele taken door machines even goed worden uitgevoerd als door mensen. Daarvoor moeten deze systemen informatie hebben over wat ze kunnen, wat hun beperkingen zijn, welke doelen ze na moeten streven en welke strategie daarbij past. Vaak wordt deze informatie "zelfbewustzijn" genoemd.

Het onderscheid tussen AGI en ANI is dat ANI intellectuele taken uitvoert op een beperkt domein.
Ethische principes hebben ook, maar niet uitsluitend, betrekking op systemen met ANI of AGI. Een belangrijk ontwerpprincipe is dat mensen, die ANI of AGI inzetten, controle moeten kunnen uitoefenen: de ethische kaders stellen waarbinnen de systemen handelen. Het expliciet maken van ethische kaders is iets waar organisaties niet aan gewend zijn en misschien gemakkelijk overlaten aan de ontwerpers van systemen.

> Een doel van de Impact Assessment is dat organisaties zelf hun ethische kaders bepalen.

Een tweede belangrijk onderscheid rond ethiek en AI is dat veel systemen die onder de noemer "AI" worden geschaard, niet meer "voorgeprogrammeerd" zijn zoals de ICT-systemen die we kennen, maar zelf leren en hun handelen en oordeel aanpassen. De klassieke systemen hadden meer rekenkracht dan hun scheppers, maar ze konden niet slimmer zijn dan hun bedenkers. De zelflerende systemen kunnen uiteindelijk beslissingen nemen of taken uitvoeren beter dan hun "scheppers". "Zelflerend" betekent dat deze systemen fouten moeten kunnen maken. En dat ze taken soms op nieuwe manieren uitvoeren, onbegrijpelijk en onvoorspelbaar voor mensen. Vraagstukken als controle, transparantie en verantwoording zijn cruciale thema's bij deze zelflerende systemen: hoe kunnen we een systeem controleren dat beter is dan wij, zonder dat we begrijpen hoe. Soms zal dat kunnen betekenen dat systemen niet kunnen worden toegepast: bijvoorbeeld in het domein van de overheid, waar het in heldere taal kunnen uitleggen van een overheidsbeslissing een recht is van burgers.

> De Impact Assessment gaat ervan uit, dat de controle, verantwoording en transparantie niet altijd onderdeel hoeft te zijn van het systeem

Als een systeem beter is dan mensen, zijn er andere maatregelen nodig waardoor mensen toch controle uit kunnen oefenen en verantwoording af kunnen leggen. Bijvoorbeeld door een systeem niet te laten "leren" bij het uitvoeren van taken. Of door een systeem alleen binnen specifieke (ethische) grenzen, geformuleerd door de organisatie die de systemen toepast, handelingen te laten verrichten.

Een derde overweging is dat de meeste AI niet zelfstandig werkt: het maakt onderdeel uit van een dienst of een product. En AI werkt vaak samen met of adviseert mensen. Zo kan een webshop op basis van AI het productaanbod voor een bezoeker aanpassen, de prijs bepalen, toetsen of de informatie die de bezoeker geeft over adres- en betaalgegevens betrouwbaar zijn en voorspellen wanneer het pakketje waarschijnlijk thuis wordt bezorgd. Elk van deze vormen van AI heeft andere ethische en juridische aspecten. Maar ook over de totale webshop kunnen ethische vragen gesteld worden, zoals: helpt de shop bezoekers om duurzame keuzes te maken of richt het zich juist op verleiding en impulsaankopen (of combineert het beide principes). De Impact Assessment gaat dan over de gehele dienst en de afzonderlijke componenten.

Transparantie

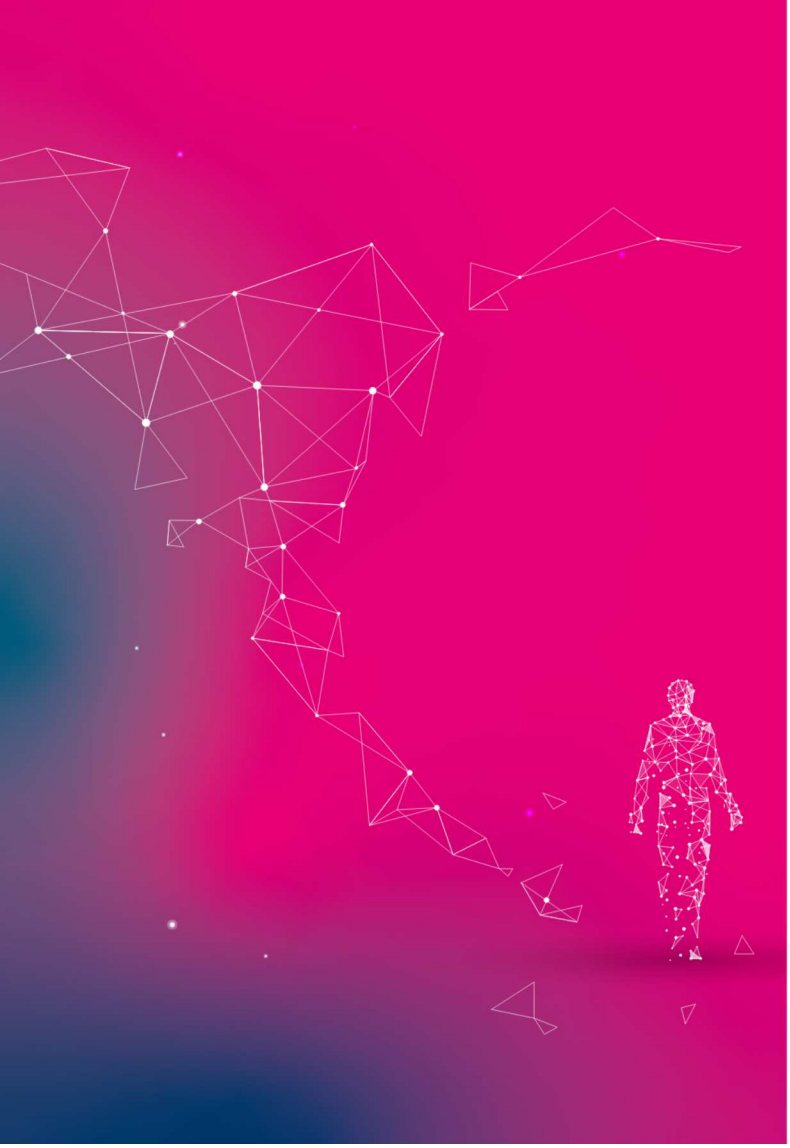
Transparantie over de werking van een AI-toepassing geeft individuen de mogelijkheid om de effecten van de toepassing op de handelingsvrijheid en beslissingsruimte te waarderen.

> Transparantie betekent dat actoren wetenschap hebben van het feit dat AI toegepast wordt, hoe besluitvorming tot stand komt en welke consequenties dit mogelijk voor hen heeft.

Dit kan in de praktijk verschillende dingen betekenen. Het kan betekenen dat er toegang is tot de broncode van een AI-toepassing, dat eindgebruikers in een bepaalde mate betrokken zijn bij het ontwerpproces van de toepassing, of dat op hoofdlijnen uitleg wordt gegeven over de werking en de context van de AI-toepassing. Transparantie over de inzet van AI-toepassingen kan de autonomie van het individu vergroten doordat het de mogelijkheid biedt aan het individu om zich te verhouden tot, bijvoorbeeld, een automatisch genomen besluit.

Bij transparantie is het belangrijk te bedenken dat diensten (maar ook producten zoals de zelfrijdende auto) vaak uit talloze componenten zijn opgebouwd. Sommige componenten kunnen AI worden genoemd. Veel van die componenten zijn niet onder het directe beheer van de organisatie die de dienst aanbiedt: overheden maken gebruik van elkaars data, zelfrijdende auto's vertrouwen op data van wegbeheerders, andere auto's op de weg en aanbieders van navigatiesystemen. Vaak zullen deze diensten gebruik maken van data uit allerlei databronnen die continu veranderen. In veel gevallen is niet meer precies duidelijk welke data op het moment van een beslissing een rol speelden. Vraag is dan welke kennis en organisatorische maatregelen zijn nodig om verantwoordelijkheid te kunnen nemen en ongewenste gevolgen en herhaling te voorkomen: soms kan algoritmische transparantie daarbij belangrijk zijn.

Uitgangspunt van de Impact Assessment is dat bij iedere toepassing van AI, wordt gekeken wat nodig is aan transparantie en wat dat betekent voor de inrichting van de techniek, de organisatie of de mensen die met de techniek werken.



Deel 1 - Achtergrond Artificial Intelligence Impact Assessment (AIIA)

Een Artificial Intelligence Impact Assessment (hierna: AIIA of Impact Assessment) is een gestructureerde methode om:

1. De (maatschappelijke) baten van een AI-toepassing in kaart te brengen.
2. De betrouwbaarheid, veiligheid en transparantie van AI-toepassingen te analyseren.
3. Waarden en belangen te identificeren die door de toepassing van AI geraakt worden.[3]
4. Risico's van de toepassing van AI te identificeren en beperken.
5. Verantwoording af te leggen over de keuzes die zijn gemaakt bij het afwegen van waarden en belangen.

Het doorlopen van een AIIA resulteert in een ethisch en juridisch verantwoorde toepassing van AI. Door in een vroeg stadium na te denken over de kansen en risico's worden problemen voorkomen. Dit zorgt er niet alleen voor dat de toepassing van AI verantwoord is, maar helpt ook bij het beschermen van de reputatie en de investeringen van de gebruiker.[4]

Er bestaat geen wettelijke plicht om een AIIA te doen. De AIIA is een zelfregulerend instrument waarmee een organisatie tot een maatschappelijke verantwoorde toepassing van AI komt.

Ethische en juridische toetsing

Het doorlopen van een AIIA moet resulteren in een ethisch en juridisch verantwoorde toepassing van AI. Wil een AI-toepassing ethisch en juridisch verantwoord zijn dan moet aan twee voorwaarden zijn voldaan:

Is de toepassing van AI betrouwbaar, veilig en transparant?

Betrouwbaarheid, veiligheid en transparantie zijn noodzakelijke randvoorwaarden voor een verantwoorde toepassing van AI. Als een AI niet goed werkt of onveilig is, dan zal de toepassing ervan (ongeacht het concrete doel) niet snel verantwoord zijn. Het betreft hier dus generieke voorwaarden waar een AI-toepassing altijd aan moet voldoen.

Betrouwbaar
Betrouwbaarheid heeft betrekking op de systematisch correcte werking van het systeem: werkt het efficiënt en zijn de uitkomsten technisch en statistisch kloppend. Met andere woorden, doet de AI-toepassing wat het moet doen en zijn de uitkomsten van het systeem correct en valt waar nodig te reconstrueren hoe de AI tot een beslissing is gekomen?

Veilig
Veiligheid van AI speelt op verschillende niveaus een rol. Bovenal moet de AI geen (onacceptabel) gevaar vormen voor de omgeving. Dit is in het bijzonder het geval daar waar het gaat om AI-systemen die gesitueerd zijn in de fysieke wereld (denk bijvoorbeeld aan autonoom rijdende auto's). Daarnaast moet een AI-toepassing als informatie-verwerkend systeem zelf ook veilig zijn (digitale veiligheid). Dit betekent dat de integriteit, vertrouwelijkheid en beschikbaarheid van het systeem en de data die het gebruikt gewaarborgd moeten worden. Dit is niet alleen ter bescherming van de werking van de AI-toepassing, maar ook voor de bescherming van de rechten van (eind)gebruikers, zoals bijvoorbeeld het recht op privacy en gegevensbescherming.

Transparant
Een derde aspect vormt transparantie en in het verlengde daarvan de uitlegbaarheid van het handelen van AI en de (externe) verantwoording over het gebruik. Het individu en/of de maatschappij moet zich een beeld kunnen vormen van de wijze waarop besluiten tot stand komen en wat de gevolgen daarvan zijn voor maatschappelijke actoren. Dit geldt primair voor besluitvorming die een wezenlijke invloed heeft op het individu of de maatschappij. Transparantie impliceert overigens niet noodzakelijkerwijs inzicht in algoritmen en datagebruik.

Is de toepassing ethisch verantwoord en legitiem?

Betrouwbaarheid, veiligheid en transparantie zijn noodzakelijke randvoorwaarden voor een ethisch verantwoorde toepassing van AI. Maar zelfs wanneer deze randvoorwaarden goed zijn ingevuld is de toepassing van AI nog niet per definitie ethisch. Zo kan het doel voor de toepassing van AI zelf illegaal zijn (bijvoorbeeld discriminatie). Andere waarden of belangen zouden zwaarder kunnen wegen dan het doel, of de wijze waarop het doel wordt bereikt is niet ethisch.

Doel
Het doel van de AIIA is niet om te zeggen wat wel en niet mag bij de toepassing van AI. Het is in eerste instantie aan de gebruiker van AI zelf om te bepalen wat hij ethisch verantwoord vindt en welke waarden worden nagestreefd met een AI-toepassing. Uiteraard moet deze afweging wel in lijn zijn met de maatschappelijke opvattingen over wat ethisch is en overeenstemmen met geldende wet- en regelgeving. De "Gedragscode Artificiële Intelligentie" in bijlage 1 biedt een handreiking om invulling te geven aan het ethische kader.

Waarden
Waarden vertalen zich in de maatschappij naar normen, wetten en regels. Het juridisch kader vormt daarom het eerste concrete toetsingskader of een AI-toepassing ethisch verantwoord is. Het gaat dan om wet- en regelgeving, gedragscodes en ethische codes.

Context
Ook de context is relevant, bijvoorbeeld binnen een sector. Zo zijn in de context van de geneeskundige zorg bijvoorbeeld de Wet op de beroepen in de individuele gezondheidszorg, de Wet op de geneeskundige behandelingsovereenkomst en de Wet op de medische hulpmiddelen relevant. Daarnaast zijn er tal van richtlijnen en gedragscodes van toepassing. Deze wetten, regels en gedragscodes vormen het kader waarbinnen de AI-toepassing hoe dan ook moet opereren.

Moreel kompas
Legaal betekent echter niet noodzakelijkerwijs dat een toepassing ook ethisch verantwoord is. Bij de toepassing van geavanceerdere vormen en toepassingen van AI zal het juridisch kader vaak nog niet helder of concreet zijn. Het is dan aan de organisatie om op basis van haar eigen moreel kompas keuzes te maken. Bij het expliciteren van dat kompas is de "Gedragscode Artificiële Intelligentie" behulpzaam (zie bijlage 1).

Toepassen in ontwerpfase

Een AIIA wordt uitgevoerd bij het begin van een project waarin AI technieken toegepast worden. Op deze manier kunnen de ethische overwegingen worden meegenomen in het ontwerp van de toepassing (*value based design of value aligned design*). Ook vanuit kosten- en haalbaarheidsperspectief is dit verstandig, omdat bij een reeds gebouwd product of gerealiseerd project het vaak onmogelijk of heel kostbaar is om nog wijzigingen door te voeren.

Betrekken Stakeholders

Naast de interne stakeholders (de business of het beleid, legal, compliance, IT *et cetera*) is ook het betrekken van de buitenwereld

relevant. Het gesprek met belanghebbenden (politiek, bestuur, maatschappelijk middenveld, wetenschap) en in het bijzonder eindgebruikers die geraakt worden door de toepassing van AI (burgers, patiënten, consumenten, werknemers et cetera) en hun vertegenwoordigers, is essentieel om draagvlak te krijgen voor de resultaten van de AIIA.

Verhouding Privacy Impact Assessment (PIA)

De AIIA en de Privacy Impact Assessment (PIA), ook Data Protection Impact Assessment (DPIA) genoemd, zijn beide risico-inschattingsinstrumenten en hanteren deels dezelfde logica. De beide instrumenten zijn complementair, maar niet onderling uitwisselbaar. Een PIA is enkel gericht op de risico's die verwerking van persoonsgegevens met zich mee kan brengen voor de betrokkene (de persoon wiens gegevens worden verwerkt). De AIIA is een breder instrument dat zich richt op alle mogelijke ethische en juridische vraagstukken die geassocieerd kunnen worden met de toepassing van AI. Voorts kijkt de AIIA niet alleen naar risico's, maar biedt het ook een kader voor het maken van ethische keuzes voor de inzet van kunstmatige intelligentie. Wanneer in het kader van de toepassing reeds een PIA is gedaan, is het sterk aan te raden de resultaten mee te nemen in de AIIA.

Praktische toepassing AIIA en ethiek

Ethiek is een filosofische discipline die zich bezighoudt met de vraag wat juist handelen is. Ethiek biedt geen checklist met wat goed en fout is, maar is veeleer de methodiek om tot een oordeel te komen over wat goed en fout is.

> Ethiek als discipline helpt om een conflict, een probleem, of een dilemma te benaderen, te doorgronden, verschillende oplossingen af te wegen en uitkomsten te analyseren aan de hand van menselijke en maatschappelijke waarden.

Ethiek biedt geen garantie voor een vlekkeloze implementatie. Een ethische analyse kan wel een gesprek over het ontwerp of implementatie van een AI-systeem op een hoger niveau tillen en helpen de juiste keuzes te maken (een ethisch verantwoorde inzet van AI).

De toepassing van AI moet in lijn zijn met de doelstellingen en de ethische richtlijnen van de organisatie zelf. De waarden die de organisatie nastreeft (de relatie met klanten, duurzaamheid, diversiteit et cetera) dienen gereflecteerd te worden in de toepassing van AI. Verder kan de toepassing van AI niet los worden gezien van de bredere inbedding daarvan binnen een organisatie en de interactie tussen de medewerkers en de toepassing. Binnen de organisatie moeten ook keuzes worden gemaakt omtrent de beheersmaatregelen om tot een betrouwbare, veilige en transparante toepassing van AI te komen.

Ethische lenzen

De ethiek kent verschillende redeneermethodes. Zij vormen als het ware de lens waarmee je een vraagstuk beschouwt. Het is van belang te beseffen dat er verschillende lenzen zijn, om dat afhankelijk van de gekozen lens een ander oordeel tot stand kan komen.
De meest typische 'ethische lenzen' zijn:[5]

1. **Gevolgenethiek** (of consequentialisme) legt de nadruk op de gevolgen van een handeling. Een handeling is moreel goed als het resultaat positief is. Wanneer een persoon in een noodsituatie moet kiezen om één persoon te doden, zodat tien personen kunnen overleven, dan is de goede keuze om deze persoon te doden.[6]

2. **Deontologie** betekent letterlijk plichtenleer. In plaats zich te richten op de gevolgen van een handeling is het uitgangspunt het naleven van plichten. Het goede doen betekent doen wat je plicht is. Het effect van het voldoen aan de plicht is dus niet van ethisch belang. Wanneer iemand het moreel onaanvaardbaar vindt om te doden, dan is het voor hem of haar de goede keuze om de persoon niét te doden, ook al is het resultaat dat tien andere personen dan niet gered kunnen worden en sterven.

3. **Deugdethiek** beziet handelingen vanuit de vraag of deze voortkomen of bijdragen aan een bepaalde deugd.[7] Wat deugdelijk is verschilt per actor. Of het een goede keuze is om één persoon te doden om 10 mensen te redden hangt af van wat een deugdzaam mens zou doen. De goede keuze is de keuze die een deugdzaam mens zou maken.

4. **Zorgethiek** heeft als focus zorg voor elkaar en het opbouwen van goede relaties. De nadruk ligt niet algemene principes maar op het individu. Abstracte ethische vragen, bijvoorbeeld wat het goede is, zien volgens zorgethici het individu over het hoofd (waardoor er van moraliteit geen sprake is). De keuze voor het doden van iemand om anderen te redden hangt dus af van welke relatie je hebt met de individuen.

De ethische lenzen bieden u een vertrekpunt voor de analyse van het vraagstuk of uw toepassing van AI ethisch verantwoord is en vormen als het ware uw 'moreel kompas'. Welke waarden stelt u voorop en wat is uw vertrekpunt bij de toepassing van AI? Gaat u voor het grootste geluk voor de grootste groep, of besteed u juist meer aandacht aan kwetsbare groepen? Deze lenzen vertegenwoordigen de hoofdstromingen in de ethiek en zijn daarmee voldoende voor een praktische benadering van ethiek in een AIIA.

Keuzes helder maken

Maatschappelijke actoren kunnen met verschillende ethische lenzen naar hetzelfde ethische dilemma kijken en op basis daarvan tot een andere conclusie komen over wat 'ethisch' is in een gegeven situatie. Door uw keuzes en afwegingen helder te maken en de lens waarmee u kijkt, kunt u de dialoog aangaan met andere maatschappelijke actoren.

Houd er bij de toepassing van deze lenzen rekening mee dat de ene lens niet noodzakelijkerwijs de andere uitsluit. Zo kunnen bijvoorbeeld keuzes wel primair worden ingegeven door de te verwachten resultaten (consequentialisme), maar kan het handelen desondanks beperkt of gestuurd worden door bepaalde principes (deontologie).

Deel 2 - Uitvoeren van de AIIA

Organisaties die een AIIA willen gaan doen, kunnen het volgende stappenplan volgen:

1. Bepaal de noodzaak voor het doen van een AIIA.
2. Beschrijf de toepassing en de context van de toepassing.
3. Stel de baten van de toepassing vast.
4. Stel vast of het doel en de wijze waarop AI wordt ingezet verantwoord is.
5. Stel vast of de toepassing betrouwbaar, veilig en transparant is.
6. Leg de resultaten en afwegingen vast.
7. Evalueer periodiek (creëer een feedback loop).

In de verschillende stappen loont het om de dialoog aan te gaan met de buitenwereld (vertegenwoordigers van eindgebruikers, burgerrechtenorganisaties, klantenpanels *et cetera*) om te toetsen of uw aannames en afwegingen in lijn zijn met de maatschappelijke opvattingen over wat ethisch is.

Stappenplan voor het uitvoeren van de AIIA

Stap 1 Bepaal de noodzaak voor het doen van een AIIA

8. Wordt de AI toegepast in een nieuw (maatschappelijk) domein?
9. Wordt gebruik gemaakt van een nieuwe vorm van AI-technologie?
10. Heeft de AI een hoge mate van autonomie?
11. Wordt de AI toegepast in een complexe omgeving?
12. Wordt gebruik gemaakt van gevoelige gegevens over personen?
13. Neemt de AI beslissingen die natuurlijke of rechtspersonen in aanzienlijke mate treffen dan wel rechtsgevolgen voor hen hebben?
14. Is de besluitvorming door de AI complex?

Stap 2 Beschrijf de AI-toepassing

1. Beschrijf de toepassing en het doel van de toepassing
2. Beschrijf welke AI-technologie wordt ingezet om het doel te bereiken
3. Beschrijf welke data worden gebruikt in het kader van de toepassing
4. Beschrijf welke actoren een rol spelen in de toepassing

Stap 3 Beschrijf de baten van de AI-toepassing

1. Wat zijn de baten voor de organisatie?
2. Wat zijn de baten voor het individu?
3. Wat zijn de baten voor de maatschappij als geheel?

Stap 4 Is het doel en de wijze waarop het doel wordt bereikt ethisch en juridisch verantwoord?

1. Welke actoren zijn betrokken bij en/of worden geraakt door mijn toepassing van AI?
2. Zijn deze waarden en belangen geconcretiseerd in wet- en regelgeving?
3. Welke waarden en belangen spelen een rol in de context van mijn toepassing van AI?

Stap 5 Is de toepassing betrouwbaar, veilig en transparant?

1. Welke maatregelen zijn genomen om de betrouwbaarheid van het handelen van de AI te borgen?
2. Welke maatregelen zijn genomen om de veiligheid van de AI te borgen?
3. Welke maatregelen zijn genomen om de transparantie van het handelen van de AI te borgen?

Stap 6 Afweging en beoordeling

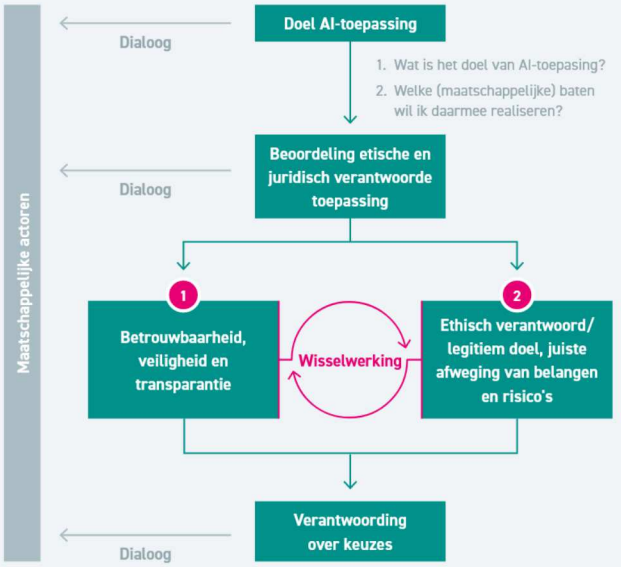
Stap 7 Vastlegging en verantwoording

Stap 8 Evalueer periodiek

Figuur 1. Het stappenplan voor een AIIA

Schematische weergave

Onderstaande figuur is een schematische weergave van de logica van een AIIA.



Figuur 2. De logica van een AIIA

Stap 1 **Bepaal de noodzaak voor het doen van een AIIA**

Niet iedere toepassing van AI rechtvaardigt het doen van een volledige AIIA. Zet een AIIA alleen in daar waar dat nuttig en nodig is.
De onderstaande screeningsvragen worden gebruikt om in te schatten of een AIIA noodzakelijk of wenselijk is. Wanneer u één van deze vragen met 'ja' kunt beantwoorden, dan is het verstandig om een AIIA te doen. Wanneer u meerdere vragen met 'ja' beantwoordt, dan is een AIIA zeer sterk aan te raden.[8]

De vragen hebben betrekking op de maatschappelijke en politieke context (vraag 1 en 2), de kenmerken van de technologie (vragen 3, 4 en 5) en de processen waarvan de technologie deel uitmaakt (vragen 6 tot en met 9).

1. Wordt de AI toegepast in een nieuw (maatschappelijk) domein?

Wordt de AI toegepast in een domein waar het daarvoor nog niet gebruikt werd? Bijvoorbeeld een toepassing die voor het eerst in de zorg wordt gebruikt terwijl het daarvoor alleen voor marketingdoeleinden werd gebruikt. Door de verandering van domein bestaat de kans dat de toepassing (nieuwe) ethische vragen oproept.

Wanneer de toepassing op een gevoelig maatschappelijk terrein plaatsvindt zijn de risico's en de ethische vraagstukken potentieel groter. Denk hierbij aan onderwerpen als zorg, veiligheid en terrorismebestrijding of onderwijs. Denk verder aan kwetsbare groepen zoals kinderen, minderheden of gehandicapten.

> Houd er rekening mee dat ook ethische dilemma's kunnen ontstaan in schijnbaar onschuldige gebruikscontexten.

De "Gedragscode Artificiële Intelligentie" (bijlage 1) en andere sectorale of dienstgerelateerde en professionele ethische codes kunnen ook helpen vast te stellen of toepassing plaatsvindt op een gevoelig terrein of onderwerp.

2. Wordt gebruik gemaakt van een nieuwe vorm van AI-technologie?
Risico's van technologie zijn doorgaans groter wanneer zij nieuw en innovatief zijn dan wanneer zij reeds lang gebruikt en beproefd zijn.

3. Heeft de AI een hoge mate van autonomie?
Naarmate een AI zelfstandiger handelt en meer vrije beslissingsruimte heeft, is het van groter belang goed te analyseren wat de consequenties van deze autonomie zijn. Naast de ruimte om beslissingen te nemen kan autonomie bijvoorbeeld ook gelegen zijn in de mogelijkheid om zelf gegevensbronnen te selecteren.

4. Wordt de AI toegepast in een complexe omgeving?
Wanneer de AI gesitueerd is een complexe omgeving zijn de risico's groter dan wanneer de AI zich in een afgebakende omgeving bevindt. De diversiteit van de input en het aantal onverwachte situaties waarop een AI moet anticiperen in een open omgeving is vele malen groter dan in een afgebakende omgeving, hetgeen tot onverwachte of ongewenste uitkomsten kan leiden. Het gebruik van een autonome vrachtenwagen die op een gesloten containerterminal rijdt kent bijvoorbeeld minder risico's dan een autonome vrachtwagen die op de openbare weg rijdt.

5. Wordt gebruik gemaakt van gevoelige gegevens over personen?
Worden bij de toepassing van AI gevoelige persoonsgegevens gebruikt, dan is het risico hoger. Denk hierbij bijvoorbeeld aan medische gegevens, gegevens betreffende etniciteit of gegevens over seksuele voorkeuren.[9]

6. Neemt de AI beslissingen die natuurlijke of rechtspersonen in aanzienlijke mate treffen dan wel rechtsgevolgen voor hen hebben?
Wanneer er sprake is van geautomatiseerde besluitvorming door de AI (zonder menselijke tussenkomst) en het besluit kan ertoe leiden dat iemand rechtsgevolgen daarvan ondervindt of anderszins aanzienlijk wordt getroffen, dan is het risico groter. Denk hierbij aan: het niet kunnen krijgen van een hypotheek, je baan kwijtraken, een verkeerde medische diagnose of schade aan je reputatie door een bepaalde categorisatie.[10]

7. Is de besluitvorming door de AI complex?[11]
Naarmate de besluitvorming door de AI complexer is (bijvoorbeeld meer variabelen of probabilistische inschattingen op basis van profielen) nemen de risico's toe. Eenvoudige toepassingen gebaseerd op een beperkt aantal keuzes en variabelen zijn minder risicovol.

Wanneer de wijze waarop een AI tot zijn besluiten is gekomen niet meer (volledig) te begrijpen of te herleiden is voor mensen, dan is het risico van het handelen of de besluitvorming potentieel groter. Met complexe neurale netwerken is het bijvoorbeeld niet altijd meer terug te redeneren hoe de AI tot de beslissing is gekomen.

Stap 2 **Beschrijf de AI-toepassing**

De analyse begint met het beschrijven van de doelen die een organisatie wil bereiken met het toepassen van AI. Welk beleidsdoel of commercieel doel streeft de organisatie na en hoe gaat de toepassing van AI bijdragen aan het realiseren van dit doel?

Zonder een duidelijke doelomschrijving is het niet mogelijk te beoordelen of de toepassing ethisch verantwoord is.

1. Beschrijf de toepassing en het doel van de toepassing

Bij de toepassing van AI kan het gaan om uiteenlopende verschijningsvormen van relatief eenvoudige beslissingsondersteunende systemen tot en met volledig autonome auto's of zelfs wapensystemen. Geef daarom een beschrijving van het product, de dienst, het systeem of het proces waarbinnen de toepassing van AI een rol speelt, welke vorm de toepassing van AI aanneemt en wat het doel is.

> Naast de algemene beschrijving van het doel is het ook van belang om meer in detail de 'ruimte' te beschrijven waarbinnen de AI opereert en de waarden die worden nagestreefd. Hiertoe moeten de volgende vragen beantwoord kunnen worden:

1. Zijn het specifieke doel van de toepassing van de AI en de gewenste eindstaat (*goal state*) voldoende duidelijk omschreven?
2. Hoe draagt de *output* van de AI bij aan het realiseren van het doel?
3. Is de context waarbinnen dit doel moet worden door de AI bereikt voldoende helder en afgebakend?
4. Is er een hiërarchie van doelen /belangen?
5. Wat zijn de regels /beperkingen waar binnen de AI moet blijven (*constraints*)?
6. Wat is een acceptabele tolerantie /foutmarge?

AI zou een begrip van ethisch handelen moeten hebben. Dit betekent dat de AI binnen de relevante context 'begrijpt' wat als ethisch handelen wordt gezien door de gebruiker /maatschappij.[12]

Wat ethisch is moet daarom zoveel mogelijk expliciet en kwantificeerbaar worden gemaakt, zodat de AI op basis van de gewenste waarden en belangen een optimale oplossing kan zoeken voor het probleem. Dit kan bereikt worden door het definiëren van de gewenste *goal state* en eventuele regels en *constraints* om deze *goal state* te bereiken. Het kan ook, voor complexere situaties, het definiëren van 'doel' of '*utility*' functies zijn. Deze doelfuncties beschrijven het nut (*utility*) van een bepaalde staat voor een AI. De AI baseert zijn keuzes op de gevolgen die dit heeft voor de gedefinieerde doelfuncties, waarbij een maximaal nut wordt nagestreefd door de AI.

Verschillende doelfuncties en de daarmee geassocieerde waarden en belangen kunnen echter botsen. Het is aan de mens om daarom niet alleen expliciet te maken wat de doelfuncties zijn, maar ook hoe zij zich tot elkaar verhouden.[13] Het onderstaande (sterk versimplificeerde voorbeeld) kan dit illustreren:

Het dilemma van de autonome auto

Een autonome auto heeft als doelfunctie meegekregen om zo snel mogelijk van punt A naar punt B te komen. Gegeven deze functie zal de auto waarschijnlijk zo hard mogelijk rijden en geen rekening houden met de veiligheid van andere weggebruikers, omdat dit niet relevant is voor de opdracht. Wanneer dezelfde autonome auto enkel de opdracht heeft gekregen om de verkeersveiligheid te garanderen, dan zal de auto waarschijnlijk niet vertrekken, omdat de meest veilige optie niet bewegen is.

In het voorgaande voorbeeld moeten beide doelfuncties dus worden gecombineerd om tot een optimaal resultaat te komen. Hiertoe moet expliciet worden gemaakt wat verkeersveiligheid in concreto betekent en wat het belang is in relatie tot het bereiken van het andere doel (van A naar B komen). Als dit expliciet (kwantificeerbaar) is, dan kan de AI een optimale strategie uitstippelen om zijn doelen te bereiken.

Ook hier spelen ethische lenzen een rol (zie op pagina 27): wordt een AI ontworpen om keuzes te maken die consequentieel van aard zijn of handelt de AI deontologisch? Met andere woorden, maakt de AI beslissingen op basis van wat het meeste oplevert voor de gedefinieerde waarde, of handelt de AI altijd conform specifieke ethische principes, ook al kan het resultaat hiervan mogelijk minder zijn of zelfs negatief uitpakken voor de gedefinieerde waarde? Hierbij is het dus wederom van belang te beseffen dat de ene lens de andere niet noodzakelijkerwijs uitsluit.

2. Beschrijf welke AI-technologie wordt ingezet om het doel te bereiken

Geef een beschrijving van de gebruikte AI-technologie of technologieën. Hierbij gaat het met name om de functionaliteiten van het systeem, de *input* en *output*, de autonomie die het systeem heeft en hoe het effectief handelt binnen de ruimte die het wordt geboden.

3. Beschrijf welke data worden gebruikt in het kader van de toepassing

Beschrijf de databronnen die gebruikt worden om de AI beslissingen te laten nemen (de *input*) en de oorsprong van deze bronnen. Denk aan de trainingsdata die wordt gebruikt om een algoritme te trainen en de data die het systeem vervolgens gebruikt voor het daadwerkelijk functioneren.

Neem in de beschrijving van de data ook sensordata mee die het systeem gebruikt als input. Houd ook rekening met de kwaliteit van de data en de aard van de data (bijvoorbeeld synthetische data of echte data).[14]

4. Beschrijf welke actoren een rol spelen in de toepassing

Beschrijf welke actoren een rol spelen in of bij de toepassing, wat hun positie is en wat hun verwachtingen of wensen zijn (een stakeholderanalyse). Het gaat hierbij met name om de actoren in de samenleving waarmee de toepassing in aanraking komt. Denk hierbij aan burgers, andere organisaties en de overheid.

Stap 3 **Beschrijf de baten van de AI-toepassing**

Wanneer AI wordt ingezet voor het realiseren van een bepaald doel, dan is dit met de gedachte om baten te realiseren voor de organisatie, het individu en/of de maatschappij als geheel. Bij baten kunt u denken aan zaken als vrijheid, welzijn, welvaart, duurzaamheid, inclusiviteit en diversiteit, gelijkheid, efficiëntie en kostenreductie.[15]

Beschrijf in deze stap de baten van de toepassing van AI voor de organisatie, het individu en de maatschappij als geheel. Deze baten dienen te worden meegewogen in de afweging omtrent de ethische en legitieme toepassing van de AI.

Baten van de toepassing zijn er op verschillende niveaus en voor verschillende actoren. Zo zal de organisatie die de AI toepast dit allereerst doen moet het oog op het realiseren van eigen baten (reduceren kosten, vergroten winst *et cetera*). In het geval van de overheid zullen de baten veelal hand in hand gaan met maatschappelijke baten (realiseren beleidsdoelstellingen). Daarnaast kunnen er maatschappelijke baten zijn naast of in aanvulling op de baten voor de overheidsorganisatie. Zo kan de toepassing van AI in de context van HR bijvoorbeeld zorgen voor de selectie van de beste kandidaat (organisatie baten), maar tegelijkertijd ook discriminatie in het selectieproces voorkomen (individuele en maatschappelijke baten).

1. **Wat zijn de baten voor de organisatie?**
   Hoe wordt de doelstelling beschreven, in stap 2, behaald en welke voordelen heeft dit ten opzichte van andere methoden (kostenreductie, efficiëntie *et cetera*)? Houd bij dit punt ook rekening mee hoe de te realiseren baten zich verhouden tot de normen en waarden van de organisatie. In hoeverre draagt AI bij aan het doel en de wijze waarop dat bereikt wordt en past dit binnen de normen en waarden van de organisatie? Draagt de toepassing bij aan de organisatiedoelstellingen en is het in lijn met de ethische richtlijnen van de organisatie?

2. **Wat zijn de baten voor het individu?**
   Welke baten heeft de nieuwe toepassing voor het individu? Is de toepassing van AI bijvoorbeeld veiliger, objectiever of eerlijker dan bestaande besluitvorming? Of maakt de toepassing van AI een product of dienstverlening mogelijk voor het individu die zonder AI niet mogelijk was?

3. **Wat zijn de baten voor de maatschappij als geheel?**
   Een toepassing van AI heeft mogelijk ook maatschappelijke baten. Stel de volgende vragen om de maatschappelijke baten in kaarten te brengen:

   1. Welk maatschappelijk belang is gediend bij de toepassing van AI?
   2. Hoe draagt het project/systeem bij aan of vergroot het welzijn?
   3. Hoe zal het project/systeem bijdragen aan menselijke waarden?

Stap 4 Is het doel en de wijze waarop het doel wordt bereikt ethisch en juridisch verantwoord?

In deze stap bepaalt u of het doel en meer specifiek de wijze waarop dit doel wordt bereikt ethisch en juridisch verantwoord is.

Vertrekpunt voor uw analyse is het bestaande juridisch kader. Maar dit kader kan onvolledig of ontoereikend zijn voor een goede ethische afweging. Daarom identificeert u de waarden en belangen die in het geding zijn bij de toepassing van AI. In het bijzonder kijkt u naar de mogelijke risico's van uw toepassing. Het identificeren van deze risico's is van belang omdat het u inzichten oplevert waarmee u het ontwerp en de toepassing van de AI kan verbeteren. De keuzes die u maakt (gaan we risico's uitsluiten of beperken, hoeveel restrisico accepteren we, accepteren we überhaupt dat onze toepassing risico's creëert?) vormen de ethische afweging van de organisatie. Hierbij speelt de ethische lens waarmee naar de toepassing wordt gekeken een belangrijke rol.

Om te kunnen beoordelen of de inzet van AI ethisch verantwoord is, moet u bepalen welke waarden en belangen mogelijk in het geding zijn bij uw toepassing van AI. Hiertoe kunt u zichzelf de volgende vragen te stellen:

1. Welke actoren zijn betrokken bij en/of worden geraakt door mijn toepassing van AI?

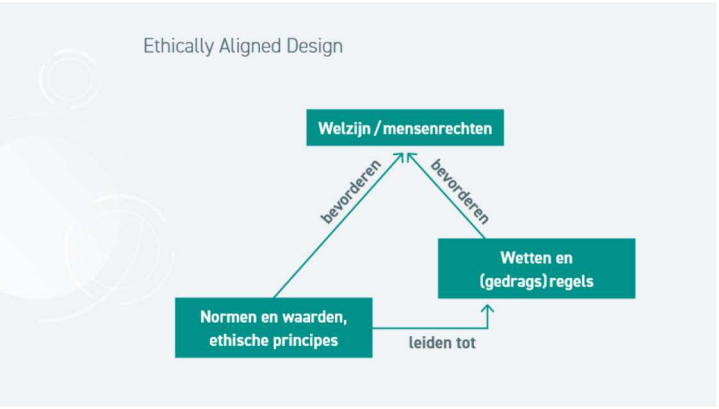
Waarden (eerlijkheid, gelijkheid, vrijheid) zijn idealen en motieven die een samenleving en de actoren daarbinnen nastreven. In Bijlage 1 vindt u de "Gedragscode Artificiële Intelligentie" met daarin de ethische principes van de European Group on Ethics in Science and New Technologies, die een handreiking kunnen bieden bij de analyse van relevante waarden. Doordat waarden abstract zijn, is het vaak lastig te beoordelen of een toepassing van AI in lijn is met de waarden binnen een samenleving. Doorgaans betekent het handelen in strijd met waarden dat de belangen van actoren direct of indirect geschaad worden (zie figuur 3). Zo kan het aantasten van de waarde 'gelijkheid', in concreto betekenen dat een persoon of groep gediscrimineerd wordt. Vandaar dat het vertalen van waarden naar belangen richting kan geven aan de beoordeling of een AI-toepassing ethisch verantwoord is of niet.

Maatschappelijke actoren hebben verschillende belangen. Door AI kunnen bestaande machtsverhoudingen veranderen en belangen van actoren worden geschaad of versterkt. AI-toepassingen kunnen dus belangen raken op verschillende niveaus. Zo kan de toepassing van AI heel concreet het belang van een individu raken (bijvoorbeeld een aantasting van diens privacy) maar kan de toepassing van AI ook belangen en verhoudingen beïnvloeden op het niveau van de samenleving. Denk bijvoorbeeld aan veranderingen op het gebied van werkgelegenheid door de toepassing van AI. In deze AIIA ligt de nadruk op de belangen van het individu. Deze komen in belangrijke mate overeen met de klassieke grondrechten (recht op vrijheid van meningsuiting, privacy *et cetera*) en de sociale grondrechten (recht op onderwijs, werkgelegenheid *et cetera*).

2. Zijn deze waarden en belangen geconcretiseerd in wet- en regelgeving?

Normen, waarden en ethische principes binnen een samenleving zijn (deels) uitgekristalliseerd in wetten en gedragsregels. Deze regels hebben tot doel het bevorderen van welzijn, het beschermen van (mensen)rechten en het ordenen van de samenleving.

Deze wetten en regels vormen het concrete kader waarbinnen uw toepassing moet blijven. In zoverre de kaders onduidelijk of onvolledig zijn dient het ontwerp van uw toepassing in lijn te zijn met de waarden zoals die gelden in de samenleving. In zoverre uw toepassing de belangen van derden raakt moet u kunnen onderbouwen waarom dit gerechtvaardigd is.



Figuur 3. De relatie tussen normen en waarden, wet- en regelgeving, welzijn/mensenrechten (IEEE, 2017)

3. Welke waarden en belangen spelen een rol in de context van mijn toepassing van AI?

De discussie over waarden is op dit moment volop bezig en tal van partijen ontwikkelen gedragscodes en normenkaders. De Impact Assessment steunt op deze kaders, zonder een enkele te kiezen: in Bijlage I zijn een gedragscode (afkomstig van de European Group on Ethics in Science and New Technologie) en praktijkregels (een update van de praktijkregels van de in 2006 door ECP gepubliceerde handreiking Autonome Systemen) opgenomen, die behulpzaam kunnen zijn om waarden en belangen te beschrijven die mogelijk geraakt worden door de toepassing van AI.[16] Bij elk van de waarden en belangen zijn een aantal (voorbeeld) vragen geformuleerd die u helpen om richting te geven aan het denken over de ethische aspecten en risico's bij de betreffende waarde.

Stap 5 Is de toepassing betrouwbaar, veilig en transparant?

Betrouwbaarheid, veiligheid en transparantie zijn noodzakelijke randvoorwaarden voor een verantwoorde toepassing van AI. Veel van de risico's van AI vloeien voort uit gebrekkige betrouwbaarheid, veiligheid of transparantie van AI. De vragen in deze stap helpen veelvoorkomende valkuilen bij de toepassing van AI te vermijden en zorg te dragen voor een verantwoorde toepassing.

In deze stap gaat het niet enkel om de betrouwbaarheid, veiligheid en transparantie van de AI-toepassing zelf, maar ook om de bredere inbedding van de AI en deze randvoorwaarden binnen de organisatie. Met andere woorden het gaat om het hele stelsel van organisatorische en technische beheersmaatregelen die ervoor zorgen dat een AI betrouwbaar, veilig en transparant wordt ingezet door een organisatie.

1. Welke maatregelen zijn genomen om de betrouwbaarheid van het handelen van de AI te borgen?

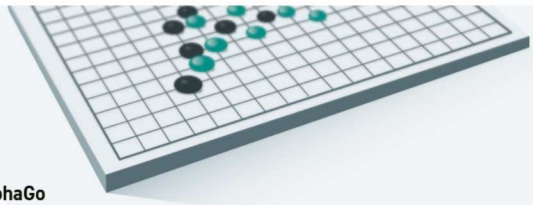
De eerste randvoorwaarde is dat een AI-toepassing betrouwbaar is. Kort gezegd komt dit er op neer dat gegeven de doelfunctie van het systeem, het systeem consequent de juiste beslissingen neemt. Om de betrouwbaarheid vast te kunnen stellen moet rekening worden gehouden met op zijn minst de onderstaande punten.

Zijn duidelijke criteria / parameters vastgesteld voor het correct functioneren van de AI?

Op basis van de doelomschrijving uit stap 3 moeten duidelijke parameters worden gedefinieerd voor het correct functioneren van de AI. Wat is het doel van de AI? Hoe moet het doel worden bereikt? Wat zijn eventuele beperkingen die aan het handelen worden gesteld of waaraan het handelen is onderworpen (*constraints*)? Op basis van deze parameters moet worden getest of de AI in lijn met de vastgestelde parameters handelt. Bij het vaststellen van de parameters moeten ook de waarden en belangen zoals beschreven in stap 5 in ogenschouw worden genomen.

Is de AI consequent in het handelen?
Het handelen van een AI moet consequent zijn. Dit betekent dat in vergelijkbare situaties de AI niet ineens tot totaal andere uitkomsten moet komen. Om vast te kunnen stellen of een AI consequent handelt moet op basis van de vastgestelde parameters getest worden of de AI consequent handelt. Bij lerende AI systemen moet rekening worden gehouden dat het handelen in de tijd kan veranderen.

Hoe wordt omgegaan met de onvoorspelbaarheid van het handelen van de AI?
Gezien de complexiteit van de besluitvorming door AI is in steeds meer situaties niet (volledig) duidelijk en/of reproduceerbaar waarom een AI tot een bepaald besluit is gekomen.



AlphaGo

Een bekend voorbeeld is de AI AlphaGo die in ronde 37 van een spelletje Go tegen de menselijke wereldkampioen Lee Sedol een volledig onvoorspelbare en onnavolgbare zet deed die het uiteindelijk de winst opleverde.[17] Binnen de context van een afgebakend en begrensd spel met heldere regels is dit geen probleem. Maar als een AI gesitueerd is in de fysieke wereld, waar de complexiteit eindeloos veel groter is waardoor 'juist' handelen daarmee moeilijker te definiëren valt, is onvoorspelbaar gedrag mogelijk risicovol.

Het feit dat AI inmiddels te complex is om volledig te kunnen doorgronden moet niet een excuus zijn voor het ongecontroleerd introduceren van een AI in een 'live' omgeving waar het belangrijke of risicovolle besluiten moet nemen. Bij de toepassing moet worden bepaald hoe omgegaan wordt met de onvoorspelbaarheid van het systeem en in hoeverre het niet kunnen reconstrueren van resultaten problematisch is voor de toepassing van de AI. Ook dient aangegeven te worden hoe gereageerd wordt op onverwachte uitkomsten en welke mitigerende maatregelen er zijn om de negatieve gevolgen te beperken. Hierbij is het van belang of het gaat om besluitvorming die voor de mens in real time relevant is om te begrijpen en te doorgronden, of dat het voldoende is dat achteraf indien nodig het proces reconstrueerbaar is (bijvoorbeeld voor een rechterlijke toetsing).

Is de juiste data beschikbaar en gekozen voor de toepassing van de AI?
AI is voor de correcte werking afhankelijk van de gebruikte data. Het is daarom van belang om te beoordelen of de juiste data worden gebruikt en hoe deze data aangeboden worden door de AI. Dit geldt zowel voor de fase waarin de AI leert en getraind wordt, als in de fase van de daadwerkelijke toepassing.

In zoverre het systeem getraind wordt met behulp van trainingsdata moet worden beoordeeld of deze data een accurate afspiegeling zijn van de daadwerkelijke omgeving en het probleemgebied waarin de AI gaat opereren. In het bijzonder moet rekening worden gehouden met het aanleren van verkeerd gedrag door de selectie en het gebruik van data (denk bijvoorbeeld aan sample selection bias). Dit om zaken als vooringenomenheid en discriminatie te voorkomen.

> Bij de toepassing in een 'live' omgeving moeten maatregelen worden genomen om te zorgen dat de beschikbaarheid van de juiste bronnen en de integriteit van deze bronnen gewaarborgd is.

Zijn de juiste algoritmen gekozen die de kunstmatige intelligentie in staat stellen effectief te handelen en het doel te bereiken?
Om de correcte werking van het systeem te kunnen borgen moeten de juiste componenten worden gebruikt, in het bijzonder de algoritmen die worden gebruikt voor de besluitvorming.

Welke methoden worden gebruikt om te verifiëren of de AI binnen de gestelde parameters blijft?
Om te kunnen controleren of de AI correct handelt en betrouwbaar is, moet het functioneren van een AI worden getoetst. Hierbij gaat het om technische en organisatorische maatregelen om achteraf na te gaan of de keuzes van de AI hebben geleid tot het juiste resultaat en er geen sprake is van nadelige gevolgen voor het individu of de maatschappij. Het testen van een AI-toepassing speelt hierbij een belangrijke rol.

2. Welke maatregelen zijn genomen om de veiligheid van de AI te borgen?

Een AI mag geen gevaar vormen voor zijn omgeving. Bij de toepassing van AI moet daarom rekening worden gehouden met de veiligheidsaspecten van de toepassing. Dit is in het bijzonder van belang wanneer het AI-systeem gesitueerd is in de fysieke wereld en daar ook fysieke schade kan aanrichten.

> Een groot deel van de risico-beperkende maatregelen die gericht zijn op het betrouwbaar handelen van de AI zullen ook relevant zijn voor het borgen van de veiligheid.

Welke veiligheidsmaatregelen getroffen moeten worden is (mede-) afhankelijk van de risico's geïdentificeerd in stap 5.

Naast het feit dat een AI veilig in een bepaalde omgeving moet opereren, is ook de (digitale) veiligheid van de AI zelf van belang. Hierbij gaat het in het bijzonder om de informatiebeveiliging en de daarbij behorende belangen vertrouwelijkheid, integriteit en beschikbaarheid.

Welke maatregelen zijn genomen om de vertrouwelijkheid te borgen?
In veel gevallen moeten de gegevens die door een AI worden verwerkt vertrouwelijk blijven. Maatregelen die genomen worden om de vertrouwelijkheid te borgen zijn gericht op het voorkomen van ongeautoriseerde kennisname van de gegevens die een AI verwerkt ten behoeve van het functioneren.

Welke maatregelen zijn genomen om de integriteit te borgen?
Voor het goed functioneren van de AI zelf en voor het beschermen van de rechten en vrijheden van derden moeten de data die de AI verwerkt worden beschermd tegen manipulatie en beschadiging. Maatregelen die genomen worden om de integriteit te borgen zijn gericht op het voorkomen van ongeautoriseerde toegang en aanpassing van de gegevens die een AI verwerkt ten behoeve van het functioneren.

Welke maatregelen zijn genomen om de beschikbaarheid te borgen?
Voor het goed functioneren van de AI zelf en voor het beschermen van de rechten en vrijheden van derden moeten de data die de AI verwerkt beschikbaar zijn en blijven. Zonder data kan de AI niet (correct) functioneren. Ditzelfde geldt voor gebruikte modellen en algoritmen. Maatregelen die genomen worden om de beschikbaarheid te borgen zijn gericht op het wegnemen of verminderen van dreigingen die er voor zorgen dat gegevens niet langer beschikbaar zijn (bijvoorbeeld maatregelen nemen om DDoS aanvallen af te slaan).

3. Welke maatregelen zijn genomen om de transparantie van het handelen van de AI te borgen?

Transparantie kan een belangrijke bijdrage leveren aan de legitieme en ethische toepassing van AI. Het gaat hierbij om de openbaarheid van het gebruik en de werking van AI. Transparantie kan op verschillende niveaus worden geboden (openbaarheid van gebruik en werking, inzicht in de gevolgen en de mogelijkheid tot verantwoording). De mate van transparantie is mede afhankelijk van de potentiële impact die de toepassing heeft op de eindgebruiker. Naarmate deze groter is past een hogere mate van transparantie.

In welke mate is AI transparant?
De organisatie moet beoordelen in welke mate zij transparant is over de toepassing van AI. Het doel van transparantie is uitlegbaarheid van het gebruik en de werking van AI. Zo kan ex post de werking van AI nagegaan worden (bijvoorbeeld bij een audit of na een incident). De meest volledige vorm van transparantie is de publicatie van alle algoritmen, de gebruikte datasets en de resultaten. Op deze wijze kan eenieder verifiëren of de AI-toepassing correct is. Deze vorm van transparantie vraagt echter

technische kennis en geeft niet noodzakelijk inzicht in het gebruik en de werking van de AI-toepassing.[18] Daarnaast kunnen echter voor organisaties overwegingen zijn om hun algoritmen en datasets geheim te houden. Naast commerciële overwegingen zoals het niet willen openbaren van intellectueel eigendom, kunnen ook andere zaken als het beschermen van de privacy van personen of nationale veiligheid een rol spelen.

Naast volledige openbaarheid kan daarom ook gedacht worden aan meer beperkte vormen van transparantie, zoals bijvoorbeeld te vinden is in artikel 13 of artikel 22 van de Algemene Verordening gegevensbescherming. In artikel 13 staat dat wanneer er sprake is van 'geautomatiseerde besluitvorming zonder menselijke tussenkomst' de logica van de besluitvorming en de mogelijke gevolgen daarvan voor de betrokkene helder moeten worden gecommuniceerd.

Is inzicht in de gevolgen van AI mogelijk?
Naast het bestaan en de werking van de AI is het ook relevant om de gevolgen van de toepassing van de AI inzichtelijk te maken. Het kan hierbij gaan om het inzicht bieden aan individuen hoe de besluitvorming door een AI hun (rechts)positie beïnvloedt, maar ook om de bredere maatschappelijke gevolgen van de toepassing van AI, bijvoorbeeld op zaken als werkgelegenheid.

Welke beheersmaatregelen worden toegepast?
Voor een zorgvuldige toepassing van AI moeten technische en organisatorische (beheers)maatregelen worden genomen. Een organisatie moet intern en waar relevant extern verantwoording kunnen afleggen (*accountability*) over deze beheersmaatregelen. Dit kan bijvoorbeeld door middel van audits.

Om de genomen maatregelen en de toepassing van AI in zijn algemeenheid te toetsen wordt een 'algoritmische audit' aanbevolen. Hiermee wordt door een onafhankelijke derde partij het gebruik van de algoritmen en data getoetst

Stap 6 **Afweging en beoordeling**

Bij de beoordeling of de AI-toepassing ethisch verantwoord is, moeten de baten (stap 3) en de geïdentificeerde risico's (stap 4) in gezamenlijkheid worden beschouwd. Bij de afweging moeten tenminste de onderstaande elementen in het oog worden gehouden.

Is de toepassing proportioneel?
Om een oordeel te kunnen vormen over de legitimiteit van de toepassing dient beoordeeld te worden of het toepassen van AI evenredig is. De vraag die gesteld moet worden is: wat is het doel dat wordt nagestreefd met de AI-toepassing en hoe verhoudt dit doel zich tot de impact van de AI-toepassing op het individu en/of de samenleving als geheel? Hoewel het doel de middelen niet heiligt, zal in zijn algemeenheid voor een gewichtiger doel meer geoorloofd zijn.

Is met minder ingrijpende middelen hetzelfde doel te verwezenlijken?
Samen met de proportionaliteit van de toepassing moet de subsidiariteit worden beoordeeld. Dit betekent dat er geen minder ingrijpende manier moet zijn om hetzelfde doel te bereiken. In de context van AI moet dan onder andere worden gekeken naar de noodzaak van de verwerking van gegevens, de mate van autonomie en de complexiteit van de AI.

Gaat het om *positive sum* in plaats van *zero sum*?
Het is van belang dat de beoordeling of een toepassing legitiem/ethisch geen *zero sum game* is. Dat wil zeggen dat het niet puur een keuze moet zijn voor het ene belang dat zwaarder weegt dan het andere belang. Alle waarden en belangen dienen maximaal gediend te zijn bij een toepassing. Slechts daar waar belangen niet langer met elkaar verenigd kunnen worden of ten koste gaan van elkaar dient een afweging te worden gemaakt welk belang uiteindelijk zwaarder weegt.

Zijn er rest-risico's?
Bij de beoordeling of een toepassing legitiem is zal zelfs bij een *positive sum* aanpak er een afweging moeten worden gemaakt welk risico acceptabel is. Bepaal of de risico's die u heeft geconstateerd wegenomen zijn of worden door risico-beperkende maatregelen of dat er nog rest-

risico bestaat. Daar waar er rest-risico blijft bestaan moet worden onderbouwd waarom het rest-risico wordt geaccepteerd en wat eventuele maatregelen zijn om schade te beperken en te herstellen als het risico zich manifesteert.

Hoe wordt verantwoordelijkheid genomen voor verder gebruik?
Een laatste punt dat overwogen moet worden is de verantwoordelijk voor het verdere gebruik (*downstream responsibility*). AI-toepassingen werken doorgaan niet in isolatie, maar zijn gekoppeld aan tal van andere systemen en processen. Er dient rekening te worden gehouden met de vraag hoe door AI gegenereerde data, beslissingen en observaties door kunnen werken in andere systemen en gebruikt worden door andere actoren.

Stap 7 **Vastlegging en verantwoording**

Leg de uitkomsten van de stappen 2 tot en met 5 vast. Besteed in het bijzonder aandacht aan de verantwoording van de legitimiteit van de toepassing. Een goede vastlegging geeft richting voor de daadwerkelijke bouw en inrichting van de toepassing, maar stelt u ook in staat om de maatschappelijke discussie aan te gaan en waar nodig verantwoording af te leggen over uw keuzes.

Of een AI-toepassing ethisch en juridisch verantwoord is, is afhankelijk van de persoon of organisatie die het oordeel velt en de ethische lens waarmee naar de toepassing wordt gekeken. Het oordeel van de organisatie die AI toepast kan afwijken van het oordeel van dat van andere maatschappelijke actoren en/of de samenleving als geheel. Door het doen van een AIIA kunt u uw keuzes en afwegingen onderbouwen en verantwoorden en gestructureerd de maatschappelijke discussie aangaan.

Stap 8 **Evalueer periodiek**

Het beoordelen of een toepassing van AI ethisch verantwoord is, is geen eenmalig proces. De organisatie en de buitenwereld veranderen. Dat heeft wellicht invloed op de ethische en juridische kaders van de AI-toepassing en daarmee op de legitimiteit. Daarom is het van belang periodiek te evalueren of de toepassing nog steeds verantwoord is. Zeker voor lerende AI is het van belang te volgen hoe de AI zich ontwikkelt en de omgeving beïnvloedt. Door periodiek te evalueren worden nieuwe risico's op tijd ontdekt, maar kan ook een *feedback loop* worden gecreëerd die de toepassing van AI beter en effectiever maakt.

Een evaluatie kan plaatsvinden met een zekere tijdsinterval (bijvoorbeeld elk jaar), maar het is verstandig om ook situaties te definiëren die nopen tot een herevaluatie. Denk bijvoorbeeld aan:

1. De AI-toepassing wordt ingezet voor een ander doel dan waarvoor het oorspronkelijk is bedoeld;
2. De beslissingsruimte van de AI wordt uitgebreid of anderszins aangepast;
3. Er worden nieuwe databronnen gebruikt;
4. Bestaande databronnen worden aangepast of niet meer gebruikt.

Bibliografie

Adviesraad voor Internationale Vraagstukken (2017), *De wil van het volk? Erosie van de democratische rechtstaat in Europa*

Committee on Technology National Science and Technology Council. (2016). *Preparing for the Future of Artificial Intelligence*. CreateSpace Independent Publishing Platform. Retrieved 05 01, 2018, from https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF

Eck, M. v. (2018). *Geautomatiseerde ketenbesluiten & rechtsbescherming: Een onderzoek naar de praktijk*. Retrieved 5 1, 2018, from https://pure.uvt.nl/portal/files/20399771/Van_Eck_Geautomatiseerde_ketenbesluiten.pdf

ECP. (2018). *Artificial Intelligence. Gespreksstof en handvatten voor een evenwichtige inbedding in de samenleving*. Retrieved from http://www.ecp.nl/AI

ECP. (2018). *Het verhaal van digitaal. Samen vormgeven aan onze digitale samenleving*. Retrieved from http://ecp.nl/publicaties/het-verhaal-van-digitaal

ECP.NL. (2006, mei 15). *Handreiking voor gedragsregels Autonome Systemen. Juridische aandachtspunten voor de bouw en het gebruik van autonome systemen*. Leidschendam: ECP.NL. Retrieved mei 1, 2018, from ecp.nl: https://ecp.nl/wp-content/uploads/2017/04/Handreiking-voor-gedragsregels-autonome-systemen-2006.pdf

European group on ethics in science and new technologies. (2018). *Statement on artificial intelligence, robotics and autonomous systems. Brussel: European Commission*. Retrieved 05 01, 2018, from https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf

Future of Life. (2018, 05 01). *AI principles*. Retrieved from Future of Life: https://futureoflife.org/ai-principles/

ISACA. (2016). *Cisa Review Manual*, 26th edition.

Kiran, A., Oudshoorn, N., & Verbeek, P. (2015). *Beyond checklists: toward an ethical-constructive technology assessment*. Journal of Responsible Innovation, 2(1), 5-19.

KNMG. (2013). *Gedragsregels voor artsen*, versie 3.1.

Kool, L., Timmer, J., Royakkers, L., & Est, R. v. (2017). *Opwaarderen - Borgen van publieke waarden in de digitale samenleving*. den Haag: Rathenau Instituut.

Motivaction. (n.d.). *Burgerschapsstijlen*. Retrieved 01-05-2018, from https://www.motivaction.nl/onderzoeksmethoden/burgerschapsstijlen

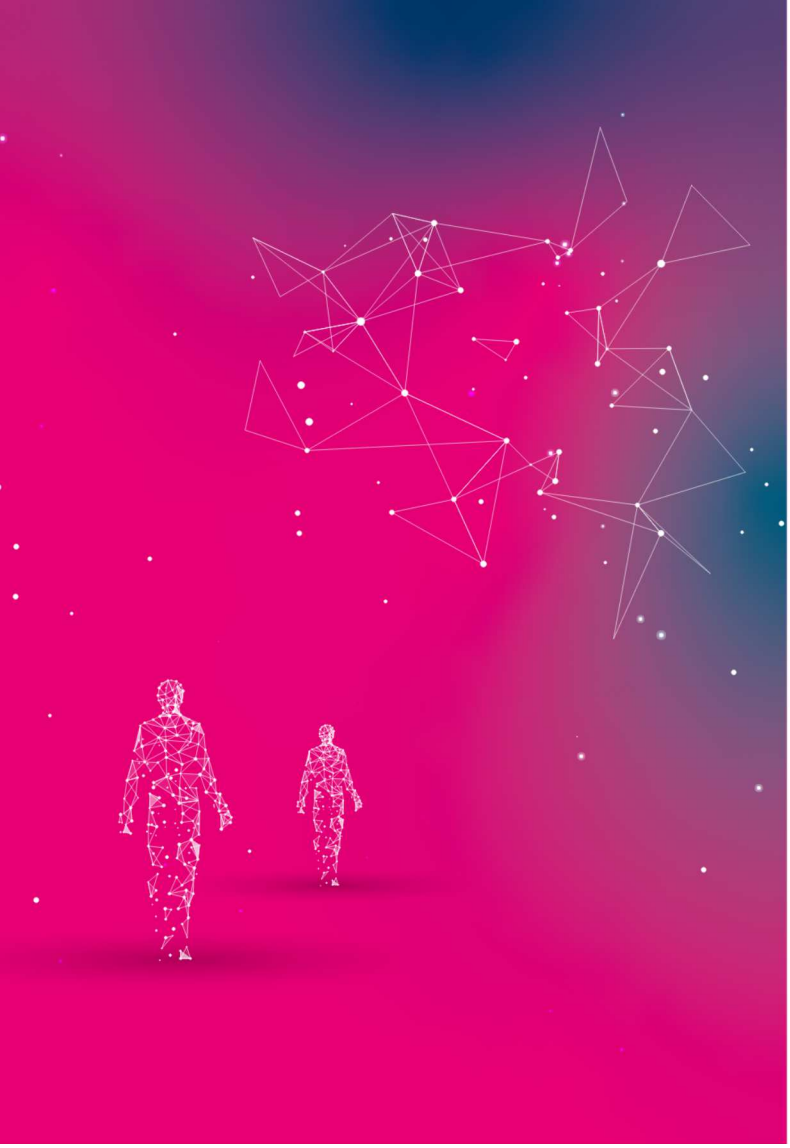
OECD. (2018, 05 09). *measuring-well-being-and-progress.htm*. Retrieved from oecd.org: http://www.oecd.org/statistics/measuring-well-being-and-progress.htm

Tewari, W. (2011). *A structured approach to IT auditing : model based development of audit terms of reference*. Amsterdam: VU University Press.

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2017). *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*. Version 2. IEEE. Retrieved from http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

Wildlak, A., & Peeters, R. (2018). *De digitale kooi. (On)behoorlijk bestuur door informatiearchitectuur of: hoe we de burger weer centraal zetten in een digitaliserende overheid*. Boom.

WRR. (2011). *IOverheid. WRR-Rapport nr. 86*. den Haag: Wetenschappelijke Raad voor het Regeringsbeleid.



Bijlage 1 - Gedragscode Artificiële Intelligentie

De Gedragscode Artificiële Intelligentie biedt een handreiking voor het vaststellen van het normenkader waaraan een concrete AI-toepassing wordt getoetst bij het uitvoeren van een Artifical Intelligence Impact Assessment (AIIA). Deze handreiking is qua aard en context van de toepassing generiek. De gedragscode is in zekere zin ook een momentopname. Het debat over de kaders waarbinnen AI wordt ontwikkeld en toegepast is erg dynamisch en kent een breed spectrum aan meningen en visies. De verwachting is dat er in de nabije toekomst verdere stappen zullen worden gezet om tot Europese, en zo mogelijk ook internationale kaders voor ontwikkeling en toepassing van AI te komen. Als er in dat proces verdere resultaten worden geboekt, ligt het voor de hand om in deze gedragscode daarbij aan te sluiten.

Gedragscode Artificiële Intelligentie

De Gedragscode Artificiële Intelligentie maakt onlosmakelijk deel uit van de Artificial Intelligence Impact Assessment (AIIA). Deze set gedragsregels is het fundament onder de AIIA.

Deel 1 Ethische principes

Toepassing van AI moet voldoen aan de volgende algemene ethische principes, gebaseerd op de European Group on Ethics in Science and New Technologies.[19]

1. Wij maken geen inbreuk op de menselijke waardigheid
2. Wij respecteren de menselijke autonomie
3. Wij onderzoeken en ontwikkelen AI in overeenstemming met de mensrechten en universele waarden
4. Wij dragen bij aan rechtvaardigheid, gelijke kansen en solidariteit
5. Wij respecteren de uitkomsten van democratische besluitvorming
6. Wij passen AI toe conform de principes van de rechtsstaat
7. Wij waarborgen de veiligheid en integriteit van gebruikers
8. Wij voldoen aan wet- en regelgeving omtrent databescherming en privacy
9. Wij voorkomen schadelijke invloed op het milieu

Deel 2 Praktijkregels

De praktijkregels zijn concrete handvatten om AI verantwoord toe te passen in de praktijk. Deze set van regels is gebaseerd op en een update van de "Handreiking voor gedragsregels autonome systemen" van ECP.NL.

10. Wij maken de gebruiker waar noodzakelijk identificeerbaar
11. Wij geven inzicht in de werking en handelingsgeschiedenis van AI-systemen
12. Wij dragen zorg voor de integriteit van AI-systemen, opgeslagen informatie en overdracht daarvan
13. Wij dragen zorg voor vertrouwelijkheid van informatie
14. Wij dragen zorg voor continuïteit
15. Wij dragen zorg voor traceerbaarheid, toetsbaarheid en voorspelbaarheid van AI-handelingen
16. Wij maken geen inbreuk op intellectuele eigendommen
17. Wij respecteren de privacy van mensen, en de wet- en regelgeving op dat gebied
18. Wij maken verantwoordelijkheden in de keten helder
19. Wij laten de informatieverwerking door AI-systemen auditen

Figuur 4. Gedragscode Artificiële Intelligentie

Terminologie

Wanneer de AIIA spreekt over de 'gebruiker' dan doelen wij op de organisatie die AI inzet. Dat kan ook de medewerker zijn die in een organisatie samenwerkt met AI. Wanneer de assessment spreekt over het 'individu' of de 'eindgebruiker' dan doelen wij op de natuurlijke persoon die de AI van een organisatie gebruikt (bijvoorbeeld de bestuurder van een autonome auto) of onderworpen is aan de besluitvorming van de AI (bijvoorbeeld een sollicitant die door een AI beoordeeld wordt). Onder 'belanghebbenden' verstaat de assessment alle individuen en partijen die een belang hebben bij AI-toepassing en direct of indirect gevolgen ondervinden van AI en daaropvolgende besluitvorming. Onder 'bouwers en aanbieders' gaat het om de partijen die AI systemen ontwikkelen. Veel AI-toepassingen worden aangeboden via de cloud.

Onderdelen

De gedragscode bestaat uit twee onderdelen:

1. Ethische principes en democratische randvoorwaarden zoals geformuleerd door de European Group on Ethics in Science and New Technologies;
2. Praktijkregels voor het omgaan met AI-toepassingen.

Deel 1
Ethische principes

Toepassing van AI moet voldoen aan de volgende algemene ethische principes, gebaseerd op de European Group on Ethics in Science and New Technologies.[19]

Deze negen, op EU-initiatief gepubliceerde, basisprincipes en democratische randvoorwaarden zijn een eerste stap op weg naar het vaststellen van een mondiaal ethisch kader. De principes zijn vastgelegd in de "Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems" van de European Group on Ethics in Science and New Technologies. Deze principes zijn gebaseerd op de fundamentele waarden welke zijn neergelegd in de EU-verdragen en het Handvest van de grondrechten van de Europese Unie.

1. **Menselijke waardigheid: AI mag geen inbreuk maken op menselijke waardigheid** [20]
   Iedere mens heeft een op zichzelf staande en intrinsieke waarde als mens waaraan geen afbreuk mag worden gedaan. Vernedering, dehumanisering, instrumentalisering en objectivering (mensen gebruiken als instrument voor een doel, zonder ze als doel op zichzelf te zien) en andere vormen van inhumane behandeling tasten deze waardigheid aan. Bij AI-toepassingen moet rekening gehouden worden met menselijke waardigheid en worden nagegaan op welke wijze een voorgenomen toepassing deze waardigheid aantast. Respect voor menselijke waardigheid betekent bovenal dat de toepassing in lijn moet zijn met mensenrechten. Daarnaast moet waar nodig duidelijk gemaakt worden aan het individu dat deze met een AI-toepassing interacteert. Respect voor menselijke waardigheid kan ook dwingen tot het afzien van de inzet van een AI-toepassing omdat een menselijke tussenkomst of interactie passender is.

- In welke mate wordt menselijke beraadslaging vervangen door geautomatiseerde systemen?
- Kunnen mensen het geautomatiseerde besluitvormingsproces overnemen?
- Is er een sterke prikkel voor de mens om de geautomatiseerde besluiten te volgen?
- Zijn individuen die met de AI-toepassing in aanraking komen zich daarvan bewust?
- Worden mensen geobjectiveerd en mogelijk gedehumaniseerd door de inzet van het systeem?

2. Autonomie: AI moet menselijke autonomie respecteren[21]

Autonomie is het vermogen van een individu om zelfstandig te handelen en beslissen. AI-toepassingen kunnen de handelingsvrijheid en beslissingsruimte van mensen inperken. Ook stelt het actoren in staat mensen onbewust te beïnvloeden (nudging) of zelfs te manipuleren. Paternalisme is een specifieke vorm van het inperken van de autonomie van het individu vanuit het oogpunt van bescherming. Het idee is dat de besluitvorming beter door de organisatie (of het algoritme) kan worden gedaan, omdat deze betere keuzes maakt dan het individu zelf. Een AI-toepassing kan dus voor het individu, zowel bewust als onbewust, de autonomie beperken (of juist vergroten).

Transparantie over de werking van een AI-toepassing geeft individuen de mogelijkheid om de effecten van de toepassing op de handelingsvrijheid en beslissingsruimte te waarderen. Transparantie betekent dat actoren wetenschap hebben van het feit dat AI toegepast wordt, hoe besluitvorming tot stand komt en welke consequenties dit mogelijk voor hen heeft. Dit kan in de praktijk verschillende dingen betekenen. Het kan betekenen dat er toegang is tot de broncode van een AI-toepassing, dat eindgebruikers in een bepaalde mate betrokken zijn bij het ontwerpproces van de toepassing, of dat op hoofdlijnen uitleg wordt gegeven over de werking en de context van de AI-toepassing. Transparantie over de inzet van AI-toepassingen kan de autonomie van het individu vergroten doordat het de mogelijkheid biedt aan het individu om zich te verhouden tot, bijvoorbeeld, een automatisch genomen besluit.

- In welke mate is er toegang tot de broncode van de AI-toepassing (openbaarheid van algoritmen) en is deze kennis bruikbaar voor buitenstaanders?
- In welke mate kan de werking van de toepassing /het algoritme uitgelegd worden aan eindgebruikers en betrokkenen?
- Is inzichtelijk voor eindgebruikers (en andere relevante actoren) wat de consequenties zijn van de besluitvorming door de AI?
- Kunnen de gebruikte datasets openbaar worden gemaakt?
- Kunnen de bronnen van gebruikte gegevens openbaar gemaakt worden?
- Kan de organisatie op een andere wijze transparant zijn voor gebruikers en betrokkenen?
- Vraagt het domein waarin de AI-toepassing wordt ingezet om een hogere mate van transparantie voor gebruikers en betrokkenen (bijvoorbeeld zorg of justitie)?
- In hoeverre neemt de organisatie of AI-toepassing besluiten over of voor het individu?
- Is een balans gevonden tussen de voordelen van het doel en de vrijheid van het individu?
- Is er een moment waarop het individu invloed kan uitoefenen op de besluitvorming door de AI? Zou deze functionaliteit beschikbaar gemaakt moeten worden?
- In hoeverre stuurt de AI de gebruiker in een door de organisatie gewenste richting (nudging)?
- In hoeverre kan een individu zich onttrekken aan (onbewuste) beïnvloeding?

3. Verantwoordelijkheid: het verantwoordelijkheidsprincipe moet ten grondslag liggen aan ieder onderzoek naar en iedere toepassing van AI[22]

Verantwoordelijkheid betekent dat AI-toepassingen alleen ontwikkeld worden in overeenstemming met de mensenrechten en andere universele waarden. Dit betekent dat tijdens het hele traject een AI-toepassing doorlopend oog moet hebben voor (onderzoeks)ethiek en individuele en de effecten die de toepassing van AI heeft op het individu en de maatschappij. Omdat de negatieve effecten van AI-toepassingen potentieel groot zijn is risicobewustzijn en een goed doordachte toepassing van belang.

- Welke technische en organisatorische maatregelen zijn genomen om eventuele negatieve effecten van AI te voorkomen of in te perken (risicobeperking)?
  Op welke manier kunnen eventuele onvoorziene effecten, na inzet van de AI-toepassing, gemitigeerd worden?
- Is het helder bij wie de juridische verantwoordelijke rust voor de inzet van de AI-toepassing?
- Kan de organisatie verantwoording afleggen over de toepassing (accountability)?

4. Rechtvaardigheid, gelijke toegang en solidariteit: AI moet bijdragen aan rechtvaardigheid, gelijke kansen en solidariteit[23]

Rechtvaardigheid (fairness) kent verschillende definities. Rechtvaardigheid kan betekenen dat mensen krijgen wat ze verdienen volgens relevante criteria. Rechtvaardigheid kan ook betekenen dat gelijke gevallen gelijk behandeld worden (gelijkheid). Rechtvaardigheid kan ook verwijzen naar het concept van sociale gelijkheid, het idee dat de zwakkeren prioriteit moeten krijgen boven degenen die voordeel hebben van instituties die ongelijkheid voortbrengen. Bij de toepassing van AI moet de gebruiker beoordelen of de toepassing van AI en de besluiten die genomen worden leiden tot rechtvaardige uitkomsten. Hierbij dient in het oog gehouden te worden dat informatiesystemen nooit helemaal waarde-neutraal zijn. In het ontwerp van het systeem liggen veelal (impliciete) keuzes voor bepaalde waarden besloten (bijvoorbeeld efficiëntie versus accuraatheid). Toepassingen van AI kunnen ongewenste vooringenomenheid (bias) vertonen, wanneer bij het systeemontwerp geen rekening is gehouden met bewuste of onbewuste vooringenomenheid (denk bijvoorbeeld aan een vooringenomenheid in de selectie van data waarmee een AI wordt getraind). Dit kan niet alleen leiden tot verkeerde of discriminerende besluiten, maar bijvoorbeeld ook dat groepen, gedrag of informatie afwijken van de heersende norm (of de norm van de ontwikkelaars/gebruikers).

Het is van belang na te gaan welke effecten de AI-toepassing, naast de rechtvaardigheid van individuele beslissingen, heeft op abstractere normen als rechtszekerheid, gelijke kansen en gelijke toegang.

- Welke waarden heeft de organisatie besloten te bevorderen, en hoe?
- Zijn er specifieke groepen die bevoordeeld of benadeeld zijn in de context waar de AI-toepassing ingezet wordt?
- Wat is het mogelijke schadelijke effect van onzekerheid en foutmarges voor verschillende groepen?
- Welke keuzes liggen impliciet besloten in de architectuur van het systeem? Zijn deze keuzes gemaakt door de organisatie die de AI gaat gebruiken, of door de ontwikkelaar?
- Neemt de AI-toepassing minder vooringenomen besluiten dan het menselijke besluitvormingsproces?
- In welke mate is de AI-toepassing een voortzetting van menselijke vooringenomenheid?
- Worden heersende beelden en stereotyperingen versterkt door de toepassing van AI?
- Zijn waarden als inclusiviteit en diversiteit actief meegenomen als functionele vereisten voor de AI-toepassing?

5. Democratie: AI moet de uitkomsten van democratische besluitvorming respecteren[24]

Een democratische rechtstaat kent een electorale dimensie en een constitutionele dimensie. Tot de electorale dimensie behoren aspecten als vrije en eerlijke verkiezingen, een pluriform aanbod van partijen en ruimte voor debat en overleg. De constitutionele dimensie omvat aspecten als gelijkheid voor de wet, het recht om verhaal te halen, rechtszekerheid, bescherming van burgerlijke vrijheden, een vrije en pluriforme pers.[25] Zoals het schandaal met Cambridge Analytica duidelijk heeft gemaakt kan de toepassing van AI het verkiezingsproces beïnvloeden.[26]

Overheden in het bijzonder dienen bij de toepassing van AI rekening te houden met de invloed die deze toepassing heeft op de democratische rechtsstaat, met name daar waar het gaat om de constitutionele dimensie. Ook bij toepassingen die verder van de rechtsstaat afstaan kan de democratische dimensie relevant zijn doordat democratische waarden als diversiteit, moreel pluralisme, en gelijke toegang tot informatie aangetast kunnen worden.

- In welke mate ondermijnt de AI-toepassing principes van democratie, bijvoorbeeld doordat de technologie beleid afdwingt zonder openbare beraadslaging?
- In hoeverre beïnvloedt de toepassing van AI de rechtszekerheid en burgerlijke vrijheden? Is deze invloed duidelijk voor eindgebruikers, betrokkenen, en (volks)vertegenwoordigers?
- In hoeverre beïnvloedt de AI-toepassing de vrije meningsuiting en het forum voor publiek debat?
- In hoeverre beïnvloedt de AI-toepassing democratische waarden zoals moreel pluralisme en diversiteit?
- In hoeverre filtert de AI-toepassing informatie van of voor de gebruiker (curatie)?
- In hoeverre blokkeert AI de toegang tot informatie?
- Wat zijn de criteria op basis waarvan informatie wordt gefilterd, geblokkeerd en gecureerd?
- Heeft de AI een vooringenomenheid met betrekking tot de te filteren informatie?

6. Rechtsstaat, verantwoording en aansprakelijkheid: toepassingen van AI moeten zich voegen naar en onderwerpen aan de principes van de rechtsstaat[27]

7. Veiligheid, lichamelijke en mentale integriteit: AI-systemen moeten veiligheid en integriteit van gebruikers respecteren[28]

Veiligheid in de context van AI-toepassingen gaat over meer dan de fysieke veiligheid voor de gebruiker of de omgeving waarin de AI-toepassing ingezet wordt.[29] Ook de interne veiligheid en betrouwbaarheid (cybersecurity) en de emotionele veiligheid bij mens-machine interactie moeten geborgd worden. Speciale aandacht moet hierbij besteed worden aan kwetsbare groepen die met de AI-toepassing in aanraking kunnen komen.

- Wat is het het effect op de fysieke veiligheid van de gebruikers en omgeving van de AI-toepassing?
- In welke mate is de cyberveiligheid van de toepassing geborgd?
- Welk effect heeft de AI-toepassing op de emotionele veiligheid van gebruikers en betrokkenen?
- Welke kwetsbare groepen kunnen in aanraking komen met de AI-toepassing? Op welke wijze is er zorg gedragen dat deze groepen geen nadelige effecten ondervinden van de toepassing?

8. Bescherming van data en privacy: AI moet voldoen aan de wet- en regelgeving rond databescherming en privacy[30]

Het recht op privacy is het recht op de bescherming van de persoonlijke levenssfeer. Wat privacy in de praktijk betekent is sterk afhankelijk van de context. Bij AI-toepassingen speelt met name de informationele dimensie van privacy een rol (het recht op de bescherming van persoonsgegevens). Concreet kan worden aangehaakt bij de uitgangspunten en regels van de Algemene Verordening Gegevensbescherming.

- Heeft de organisatie bepaald hoe de privacy van betrokkenen beschermd wordt?
- Verzamelt en verwerkt de toepassing alleen de voor de toepassing noodzakelijke gegevens?
- Zijn eindgebruikers in staat om te bepalen welke gegevens van/over hen worden verzameld en welke conclusies daaruit getrokken worden?
- Kan de gebruiker zijn gegevens verwijderen uit het systeem?

9. Duurzaamheid: AI mag geen schadelijke invloed hebben op het milieu.[31]

AI-toepassingen hebben net als andere technologieën een invloed op de leefbaarheid van onze planeet en de toekomstige welvaart van de mensheid en de leefomgeving voor volgende generaties. AI heeft een directe invloed op de leefomgeving (denk aan het vergroten of verkleinen van het energieverbruik en e-waste) en een indirecte invloed, bijvoorbeeld door het stimuleren van milieubewust gedrag (bijvoorbeeld via beslisondersteuning of nudging).

- Wat zijn de milieu-effecten van de AI-toepassing?
- Vergroot of verkleint de toepassing van AI het gebruik van grondstoffen en natuurlijke hulpbronnen?
- Welke invloed heeft de AI-toepassing op de leefwereld van volgende generaties?

Deel 2
Praktijkregels

Deze regels vormen een update van de "Handreiking voor gedragsregels autonome systemen"(2006) van ECP.NL. Beide componenten van de gedragscode (ethische principes en praktijkregels) hebben elk een eigen waarde en functie. De ethische principes bieden op een wat hoger abstractieniveau een breed kader voor AI. De praktijkregels zijn over het geheel genomen iets concreter. Zij zijn echter niet ontworpen als een (sluitende) uitwerking van de genoemde ethische principes, maar sluiten daar wel goed bij aan en geven richting voor het toepassen van AI in de praktijk

1. Identificatie

Waar noodzakelijk dient de gebruiker van een AI-systeem identificeerbaar te zijn. Deze identiteit moet aan het AI-systeem gekoppeld kunnen worden.

2. Transparantie

Partijen dienen na te gaan of zij een overeenstemmend beeld hebben van de mogelijkheden en onmogelijkheden van het gebruikte AI-systeem.

Bouwers en gebruikers van AI-systemen geven indien mogelijk duidelijk inzicht in de werking van de door hen gebouwde of aangeboden AI-systemen

Bouwers en gebruikers geven de eindgebruiker steeds inzicht in de handelingsgeschiedenis van de door hen gebouwde of aangeboden AI-systemen. Dit beginsel kent slechts uitzondering in die gevallen waarin het genereren van een handelingsgeschiedenis wettelijk niet verplicht is en redelijkerwijs niet mogelijk is.

3. Integriteit

Partijen dragen zorg voor de integriteit van het AI-systeem, de daarin opgeslagen informatie en de overdracht daarvan.

Partijen nemen passende maatregelen om schendingen van de integriteit van een AI-systeem te kunnen detecteren, en maken afspraken over de acties die ondernomen dienen te worden wanneer een schending wordt geconstateerd.

4. Vertrouwelijkheid

Partijen dragen zorg voor de vertrouwelijkheid van de opgeslagen informatie in door hen gebouwde of gebruikte AI-systemen.

Partijen nemen passende maatregelen om onrechtmatige openbaringen van vertrouwelijke informatie te kunnen detecteren, en maken afspraken over de acties die ondernomen dienen te worden wanneer een onrechtmatige openbaring wordt geconstateerd.

5. Continuïteit

Partijen dragen zorg voor de continuïteit van de door hen aangeboden of gebruikte AI-systemen.

Partijen nemen passende maatregelen om te voorkomen dat een fout in een AI-systeem of het platform waarop deze draait, leidt tot het volledig verloren gaan van een AI-systeem.

6. Toetsbaarheid, voorspelbaarheid en traceerbaarheid

Partijen dragen zorg voor de traceerbaarheid, toetsbaarheid en voorspelbaarheid van de door een AI-systeem verrichte handelingen.

Partijen dragen zorg voor de integriteit van de door AI-systemen gegenereerde logs.

Partijen dragen zorg voor de vertrouwelijkheid van gegenereerde logs.

7. Intellectuele eigendom

Partijen (bouwer, gebruiker en andere belanghebbenden en /of eindgebruiker) zullen voorafgaand onderling duidelijke afspraken maken over de intellectuele eigendomsrechten en bedrijfsgeheimen met betrekking tot het systeem. Hieronder vallen in ieder geval: eigendom /gebruik van al bestaande intellectuele eigendomsrechten en

bedrijfsgeheimen van een of meer partijen, en eigendom /registratie / handhaving van intellectuele eigendomsrechten en bedrijfsgeheimen voortkomend uit de ontwikkeling en/of gebruik van het systeem

Vóór gebruik dient gekeken te worden of en zo ja welke intellectuele eigendomsrechten van derden een rol spelen bij het systeem. Vervolgens moeten partijen er voor zorgen dat er geen inbreuk gemaakt wordt op een dergelijk intellectueel eigendomsrecht.

8. Privacy

De verwerking van persoonsgegevens moet rechtmatig, behoorlijk en transparant zijn.[32]

De verzameling en verdere verwerking van persoonsgegevens moet gebonden zijn aan specifieke doelen.[33]

De gegevens moeten toereikend, ter zake dienend en beperkt tot het noodzakelijke zijn).[34] De gegevens moeten juist zijn.[35] De gegevens mogen niet langer worden bewaard dan nodig.[36] De gegevens moeten goed beveiligd zijn.[37]

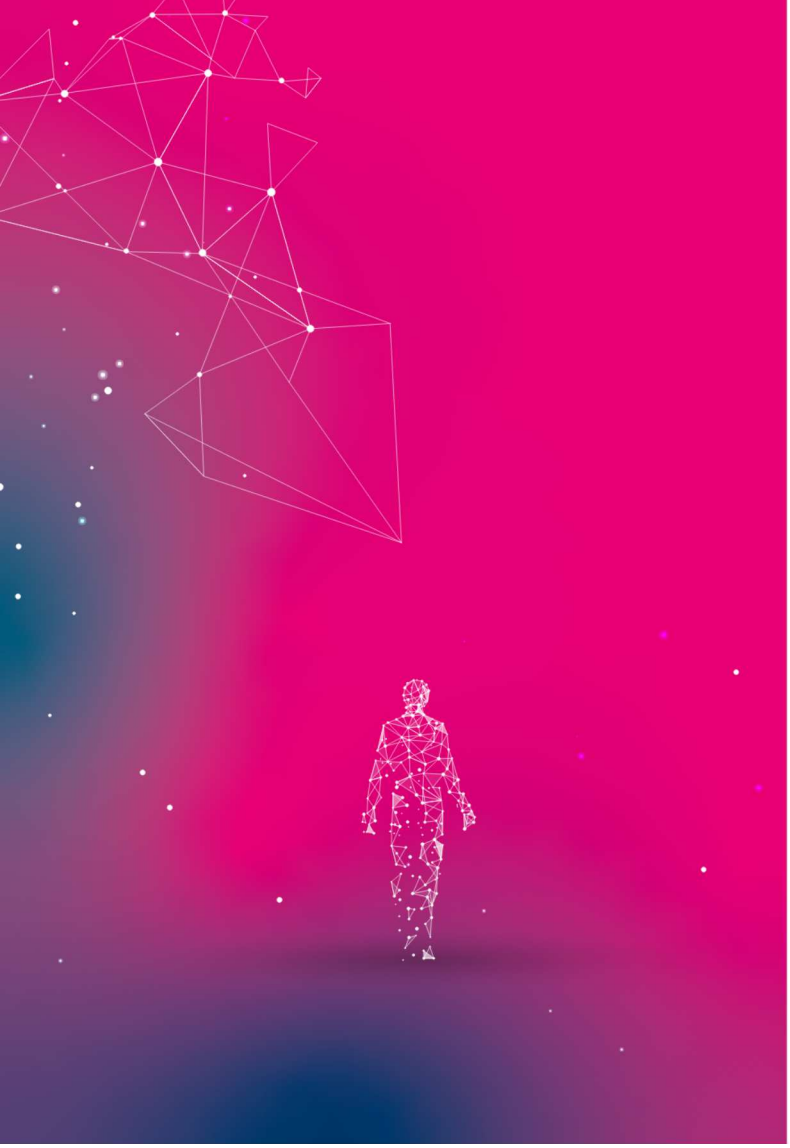
Betrokkenen hebben het recht om niet te worden onderworpen aan geautomatiseerde besluitvorming die rechtsgevolgen heeft dan wel de betrokkene anderszins in aanzienlijke mate treft.[38]

9. Verantwoordelijkheid

Bij het ontwikkelen en toepassen van complexe AI-systemen waar vele componenten en (deel)dienstverleners een rol spelen en waar gedrag niet altijd meer te herleiden is tot specifieke componenten of dienstverleners, dienen maatregelen getroffen te worden waardoor afbakening van verantwoordelijkheden helder is.

10. Audit

Voordat gebruik gemaakt wordt van een AI-systeem dient te worden vastgesteld op welke wijze (middelen, proces) de relevante aspecten van de betreffende informatieverwerking kunnen worden geverifieerd door middel van een audit.



Bijlage 2 - Stappenplan AIIA

Stap 1 **Is het zinvol om een AIIA te doen?**

Bepaal aan de hand van de volgende screeningsvragen of het zinvol is om een AIIA te doen.

Wordt de AI toegepast in een nieuw (maatschappelijk) domein?	☐ Ja	☐ Nee
Vindt de toepassing plaats op een gevoelig (maatschappelijk) terrein of onderwerp?	☐ Ja	☐ Nee
Wordt gebruik gemaakt van een nieuwe vorm van AI-technologie?	☐ Ja	☐ Nee
Heeft de AI een hoge mate van autonomie?	☐ Ja	☐ Nee
Is de besluitvorming door de AI complex?	☐ Ja	☐ Nee
Wordt de AI toegepast in een complexe omgeving?	☐ Ja	☐ Nee
Wordt gebruik gemaakt van gevoelige gegevens over personen?	☐ Ja	☐ Nee
Neemt de AI beslissingen die natuurlijke of rechtspersonen in aanzienlijke mate treffen dan wel rechtsgevolgen voor hen hebben?	☐ Ja	☐ Nee
Zijn de uitkomsten van de AI-toepassing niet meer (volledig) te begrijpen / herleiden?	☐ Ja	☐ Nee

Is het antwoord op één of meer van deze vragen 'Ja' dan is het zinvol om een AIIA te doen. Ga naar Stap 2.

Stap 2 **Beschrijf de AI-toepassing**

Beantwoord de volgende vragen over de voorgenomen inzet van AI.

1. Wat is het doel van de toepassing?
2. Welke AI-technologie(ë)n worden in gezet om het doel te bereiken?
3. Welke data worden gebruikt om het doel te bereiken?
4. Welke actoren (leveranciers, eindgebruikers andere belanghebbenden) spelen een rol bij de toepassing?

Stap 3 **Beschrijf de baten van de AI-toepassing**

Beschrijf de positieve aspecten /opbrengst (baten) van de toepassing door de volgende vragen te beantwoorden:

1. Wat zijn de baten voor de organisatie?
2. Wat zijn de baten voor het individu?
3. Wat zijn de baten voor de maatschappij als geheel?

Stap 4 **Is het doel en de wijze waarop het doel wordt bereikt ethisch en juridisch verantwoord?**

Beschrijf de invloed die de toepassing heeft op menselijke en maatschappelijke waarden. Wanneer waarden negatief beïnvloed worden door de toepassing (bijvoorbeeld privacyrisico's of negatieve milieu-effecten) moet onderbouwd worden hoe deze risico's worden beperkt en als er sprake is van restrisico, waarom dit geaccepteerd is. Denk bij waarden aan onder andere:

1. Menselijke waardigheid
2. Autonomie (vrijheid)
3. Verantwoordelijkheid
4. Transparantie

5. Rechtvaardigheid (*fairness*)
6. Democratie en rechtsstatelijkheid
7. Veiligheid
8. Privacy en gegevensbescherming
9. Duurzaamheid

Nota bene: of een toepassing ethisch verantwoord is, is naast het doel ook sterk afhankelijk van de inrichting van de randvoorwaarden (Stap 5).

Stap 5 **Is de toepassing betrouwbaar, veilig en transparant?**

Beschrijf de randvoorwaarden voor de ethische toepassing van AI (betrouwbaarheid, veiligheid, transparantie) door de volgende vragen te beantwoorden:

1. Welke maatregelen zijn genomen om de betrouwbaarheid van het handelen van de AI te borgen?

2. Welke maatregelen zijn genomen om de veiligheid van de AI te borgen
   - Hoe is de veiligheid van de AI ten opzichte van de buitenwereld geborgd?
   - Hoe is de (digitale) veiligheid van de AI zelf geborgd?

3. Welke maatregelen zijn genomen om de transparantie van het handelen van de AI te borgen?
   - Is de werking van de AI (de logica van de besluitvorming) inzichtelijk / openbaar?
   - Is er duidelijk wat de gevolgen zijn van de toepassing van AI (in het bijzonder de gevolgen voor eindgebruikers)?
   - Zijn er maatregelen genomen om verantwoording af te kunnen leggen over de toepassing (*accountability*)?

Stap 6 **Afweging en beoordeling**

Weeg op basis van het bovenstaande (stappen 3, 4 en 5 in het bijzonder) af of de toepassing in zijn geheel ethisch verantwoord is. De volgende aspecten kunnen worden meegenomen in deze beoordeling:

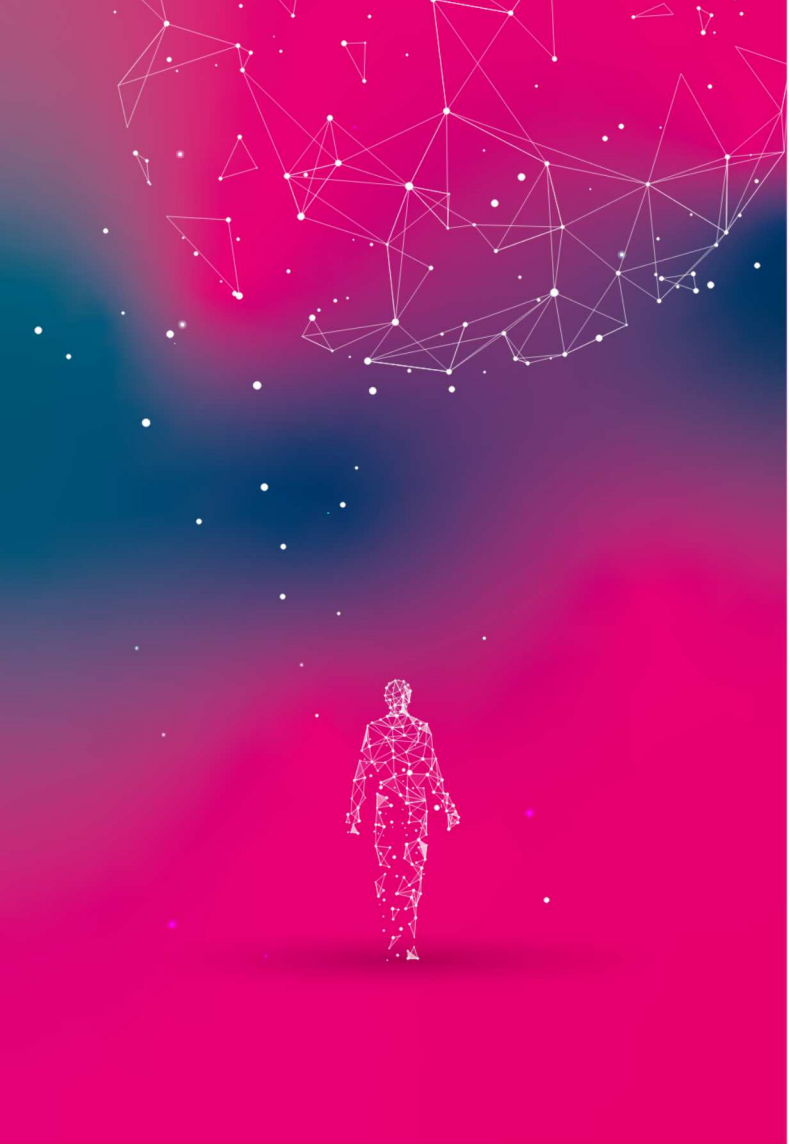
1. Is de toepassing proportioneel?
2. Is met minder ingrijpende middelen hetzelfde doel te verwezenlijken (subsidiariteit)?
3. Is de keuze voor de toepassing *positive sum* of *zero sum*
4. Wat zijn rest-risico's en waarom zijn deze acceptabel?
5. Wordt rekening gehouden met verder gebruik (*downstream responsibility*)?

Stap 7 **Vastlegging en verantwoording**

Leg de antwoorden op de bovenstaande vragen vast, zodat intern en extern verantwoording kan worden afgelegd over de keuzes.

Stap 8 **Evalueer periodiek**

Evalueer bij veranderingen in de toepassing en/of periodiek of de bovenstaande conclusies nog steeds gelden.



Noten

1 https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF(OECD, 2018)

2 Praktische hulpmiddelen zijn onder andere de Ethische Data Assistent (https://dataschool.nl/deda/), het AI Ethics Framework (https://www.migarage.ai/ethics-framework/), de AI NOW Algorithmic Impact Assessment (https://ainowinstitute.org/aiareport2018.pdf) en de Princeton Dialogues on AI and Ethics (https://www.migarage.ai/ethics-framework/).

3 Waarden hebben geen vaste, in code omzetbare definitie, maar zijn afhankelijk van diverse culturele, historische en maatschappelijke factoren. Binnen deze AIIA wordt daar waar wordt gesproken over waarden zoveel mogelijk concreet gemaakt wat wordt verstaan onder een bepaalde waarde in de context van de AIIA.

4 De assessment maakt onderscheid tussen de gebruiker van AI (de organisatie die AI toepast bij dienstverlening, de medewerker die samenwerkt met AI bij het uitvoeren van werkzaamheden), de ontwikkelaar (technologie- en platformpartijen, clouddienstverleners), de eindgebruiker (die direct de gevolgen ondervindt van beslissingen en handelingen van het AI systeem, zoals de bestuurder van een zelfrijdende auto of de burger die te maken krijgt met een door AI genomen beslissing) en de belanghebbenden (de bredere kring van partijen die gevolgen ondervindt van toepassing van AI: zoals maatschappelijke en politieke organisaties, beroeps- en brancheorganisaties).

5 De onderstaande voorbeelden zijn extreme voorbeelden en vormen een versimpeling van het denken binnen deze ethische strominge

6 In veel gevallen is de afweging over de toepassing van AI binnen een organisatie en het maatschappelijk debat consequentieel van aard: als het resultaat van de inzet van AI een positief effect heeft op de in de AIIA genoemde belangen, of in een belangenafweging het gerechtvaardigd blijkt om bepaalde risico's van AI te accepteren, dan wordt de toepassing als ethisch en legitiem verantwoord gezien.

7 Deugden zijn eigenschappen van een persoon die als moreel goed worden beschouwd. De vier kardinale deugden zijn voorzichtigheid, rechtvaardigheid, gematigdheid en moed.

8 Wanneer het bij het gebruik van een AI te verwachten valt dat u persoonsgegevens gaat verwerken is het verstandig deze stap te combineren met de afweging of een DPIA noodzakelijk is.

9 Zie ook artikel 9 Algemene Verordening Gegevensbescherming

10 Zie ook artikel 22 Algemene Verordening Gegevensbescherming

11 Wanneer wij spreken over besluitvorming door AI dan bedoelen wij

het handelen van de AI om tot de optimale uitkomst te komen voor de doelen en waarden zoals door de mens gedefinieerd. Hoewel de AI dus beslissingen neemt om tot een optimale uitkomst te komen, is dit op basis van de doelen en de bijbehorende doelfuncties zoals door de gebruiker gedefinieerd. De uitkomst kan ook een advies zijn, waarbij en de mens uiteindelijk het daadwerkelijke besluit neemt.

12 De vraag of een AI ethisch besef moet hebben is uiteraard sterk afhankelijk van de context en de complexiteit van de toepassing van AI.

13 Asimov's 'Three Laws of Robotics' zijn een populair voorbeeld van een dergelijke hiërarchie.

14 Kwaliteit kan betrekking hebben op de data zelf (zijn de data bijvoorbeeld consistent en compleet) maar kan ook betrekking hebben op inhoudelijke kwaliteiten zoals waarheidsgetrouwheid. Synthetische data zijn data die gegenereerd zijn door de computer. Het betreffen gegevens die niet 'echt' zijn, maar wel zo dicht mogelijk een dataset met 'echte' data weerspiegelen. Synthetische datasets worden onder andere gebruikt om te voorkomen dat met echte persoonsgegevens wordt getest.

15 Zie onder andere: IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (2017). Voor het doen van maatschappelijke kosten-baten analyses kan worden aangesloten bij de *Algemene Leidraad voormaatschappelijke kosten-batenanalyse van het CPB en PBL en/of de Werkwijzer voor maatschappelijke kosten-batenanalyse bij de digitale overheid.*

16 Dit is geen uitputtende lijst. Welke waarden en belangen geraakt worden verschilt uiteraard per toepassing. Het is aan de organisatie zelf om te bepalen of er nog andere waarden en belangen in het geding zijn. Ook hangen de genoemde waarden en belangen sterk met elkaar samen. Het zijn dus niet belangen die in isolatie moeten worden gewogen.

17 https://www.wired.com/2016/03/two-moves-alphago-lee-sedol-redefined-future/

18 Net zoals als kennis van de anatomie van een organisme of de werking van cellen maar beperkte voorspellende waarde heeft voor het voorspellen van gedrag.

19 De ethische principes zijn gebaseerd op: European group on ethics in science and new technologies. (2018). Statement on artificial intelligence, robotics and autonomous systems. Brussel: European Commission. Retrieved 05 01, 2018, from https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf

20 **Toelichting EGE (2018):** The principle of human dignity, understood as the recognition of the inherent human state of being worthy of respect, must not be violated by 'autonomous' technologies. This means, for instance, that there are limits to determinations and classifications concerning persons, made on the basis of algorithms and 'autonomous' systems, especially when those affected by them are not informed about them. It also implies that there have to be (legal) limits to the ways in which people can be led to believe that they are dealing with human beings while in fact they are dealing with algorithms and smart machines. A relational conception of human dignity which is characterised by our social relations, requires that we are aware of whether and when we are interacting with a machine or another human being, and that we reserve the right to vest certain tasks to the human or the machine.

21 **Toelichting EGE (2018):** The principle of autonomy implies the freedom of the human being. This translates into human responsibility and thus control over and knowledge about 'autonomous' systems as they must not impair freedom of human beings to set their own standards and norms and be able to live according to them. All 'autonomous' technologies must, hence, honour the human ability to choose whether, when and how to delegate decisions and actions to them. This also involves the transparency and predictability of 'autonomous' systems, without which users would not be able to intervene or terminate them if they would consider this morally required.

22 **Toelichting EGE (2018):** The principle of responsibility must be fundamental to AI research and application. 'Autonomous' systems should only be developed and used in ways that serve the global social and environmental good, as determined by outcomes of deliberative democratic processes. This implies that they should be designed so that their effects align with a plurality of fundamental human values and rights. As the potential misuse of 'autonomous' technologies poses a major challenge, risk awareness and a precautionary European Group on Ethics in Science and New Technologies 17 approach are crucial. Applications of AI and robotics should not pose unacceptable risks of harm to human beings, and not compromise human freedom and autonomy by illegitimately and surreptitiously reducing options for and knowledge of citizens. They should be geared instead in their development and use towards augmenting access to knowledge and access to opportunities for individuals. Research, design and development of AI, robotics and 'autonomous' systems should be guided by an authentic concern for research ethics, social accountability of developers, and global academic cooperation to protect fundamental rights and values and aim at designing technologies that support these, and not detract from them.

23 **Toelichting EGE (2018):** AI should contribute to global justice and equal access to the benefits and advantages that AI, robotics and 'autonomous' systems can bring. Discriminatory biases in data sets used to train and run AI systems should be prevented or detected, reported and neutralised at the earliest stage possible. We need a concerted global effort towards equal access to 'autonomous' technologies and fair distribution of benefits and equal opportunities across and within societies. This includes the formulating of new models of fair distribution and benefit sharing apt

to respond to the economic transformations caused by automation, digitalisation and AI, ensuring accessibility to core AI-technologies, and facilitating training in STEM and digital disciplines, particularly with respect to disadvantaged regions and societal groups. Vigilance is required with respect to the downside of the detailed and massive data on individuals that accumulates and that will put pressure on the idea of solidarity, e.g. systems of mutual assistance such as in social insurance and healthcare. These processes may undermine social cohesion and give rise to radical individualism.

24 **Toelichting EGE (2018):** Key decisions on the regulation of AI development and application should be the result of democratic debate and public engagement. A spirit of global cooperation and public dialogue on the issue will ensure that they are taken in an inclusive, informed, and farsighted manner. The right to receive. education or access information on new technologies and their ethical implications will facilitate that everyone understands risks and opportunities and is empowered to participate in decisional processes that crucially shape our future. The principles of human dignity and autonomy centrally involve the human right to self-determination through the means of democracy. Of key importance to our democratic political systems are value pluralism, diversity and accommodation of a variety of conceptions of the good life of citizens. They must not be jeopardised, subverted or equalised by new technologies that inhibit or influence political decision making and infringe on the freedom of expression and the right to receive and impart information without interference. Digital technologies should rather be used to harness collective intelligence and support and improve the civic processes on which our democratic societies depend.

25 Zie Adviesraad voor Internationale Vraagstukken (2017), De wil van het volk? Erosie van de democratische rechtsstaat in Europa

26 Zie: https://www.theguardian.com/news/series/cambridge-analytica-files

27 **Toelichting EGE (2018):** Rule of law, access to justice and the right to redress and a fair trial provide the necessary framework for ensuring the observance of human rights standards and potential AI specific regulations. This includes protections against risks stemming from 'autonomous' systems that could infringe human rights, such as safety and privacy. The whole range of legal challenges arising in the field should be addressed with timely investment in the development of robust solutions that provide a fair and clear allocation of responsibilities and efficient mechanisms of binding law. In this regard, governments and international organisations ought to increase their efforts in clarifying with whom liabilities lie for damages caused by undesired behaviour of 'autonomous' systems. Moreover, effective harm mitigation systems should be in place.

28 **Toelichting EGE (2018):** Security, safety, bodily and mental integrity: Safety and security of 'autonomous' systems materialises in three forms: (1) external safety for their environment and users, (2) reliability and internal robustness, e.g. against hacking, and (3) emotional safety with respect to human-machine interaction. All dimensions of safety must be taken into account by AI developers and strictly tested before release in order to ensure that 'autonomous' systems do not infringe on the human right to bodily and mental integrity and a safe and secure environment. Special attention should hereby be paid to persons who find themselves in a vulnerable position. Special attention should also be paid to potential dual use and weaponisation of AI, e.g. in cybersecurity, finance, infrastructure and armed conflict.

29 Zie ook onder 'Stap 4', punt 2: Welke maatregelen zijn genomen om de veiligheid van de AI te borgen

30 **Toelichting EGE (2018):** Data Protection and Privacy: In an age of ubiquitous and massive collection of data through digital communication technologies, the right to protection of personal information and the right to respect for privacy are crucially challenged. Both physical AI robots as part of the Internet of Things, as well as AI softbots that operate via the World Wide Web must comply with data protection regulations and not collect and spread data or be run on sets of data for whose use and dissemination no informed consent has been given. 'Autonomous' systems must not interfere with the right to private life which comprises the right to be free from technologies that influence personal development and opinions, the right to establish and develop relationships with other human beings, and the right to be free from surveillance. Also in this regard, exact criteria should be defined and mechanisms established that ensure ethical development and ethically correct application of 'autonomous' systems. In light of concerns with regard to the implications of 'autonomous' systems on private life and privacy, consideration may be given to the ongoing debate about the introduction of two new rights: the right to meaningful human contact and the right to not be profiled, measured, analysed, coached or nudged.

31 **Toelichting EGE (2018):** Sustainability: AI technology must be in line with the human responsibility to ensure the basic preconditions for life on our planet, continued prospering for mankind and preservation of a good environment for future generations. Strategies to prevent future technologies from detrimentally affecting human life and nature are to be based on policies that ensure the priority of environmental protection and sustainability.

32 Persoonsgegevens mogen alleen worden verwerkt voor gerechtvaardigde doeleinden. Dit betekent dat bij de inzet van kunstmatige intelligentie vooraf moeten worden bepaald met welk doel de gegevens voor/door de kunstmatige intelligentie worden verwerkt. Dit moet ook transparant zijn voor de buitenwereld, meer specifiek de betrokkenen.

33 Wanneer gegevens eenmaal voor het hierboven beschreven gerechtvaardigde doel zijn verzameld, dan mogen de gegevens ook alleen voor dit doel worden verwerkt. Enige uitzondering op deze regel is wanneer het nieuwe doel verenigbaar is met het oorspronkelijke verzameldoel.

34 Er mogen niet meer gegevens worden verwerkt dan voor het doel van de verwerking noodzakelijk zijn (data minimalisatie). Het gebruik van datasets door/ten behoeve van kunstmatige intelligentie moet dus beperkt blijven tot hetgeen noodzakelijk is voor het goed functioneren van de kunstmatige intelligentie ten behoeve van het gespecificeerde doel waartoe de kunstmatige intelligentie wordt ingezet. Data minimalisatie betekent overigens niet altijd 'zo min mogelijk gegevens'. De kunstmatige intelligentie moet wel voldoende data hebben om correct te kunnen functioneren.

35 De gegevens moeten juist en actueel zijn. Onjuiste of verouderde gegevens moeten worden aangepast of verwijderd.

36 Persoonsgegevens mogen niet langer bewaard worden dan noodzakelijk voor het doel van de verwerking. Gegevens die niet langer het verwerkingsdoel dienen moeten worden geanonimiseerd of gewist

37 De vertrouwelijkheid, integriteit en beschikbaarheid van persoonsgegevens bij het gebruik van persoonsgegevens voor/door kunstmatige intelligentie moet geborgd worden met passende technische en organisatorische maatregelen. Naast deze algemene beginselen is in het kader van kunstmatige intelligentie ook specifiek artikel 22 AVG relevant.

38 Wanneer een kunstmatige intelligentie zonder menselijke tussenkomst beslissingen neemt (algoritmische besluitvorming), dan is dit niet toegestaan wanneer dit rechtsgevolgen heeft of de betrokkenen anderszins significant raakt in diens rechten. De AVG en de Nederlandse Uitvoeringswet AVG maakt enkele specifieke uitzonderingen op dit algemene verbod. Wanneer er bijvoorbeeld uitdrukkelijke toestemming is van de betrokkene, of de besluitvorming is noodzakelijk voor de totstandkoming van een overeenkomst, dan is de besluitvorming toegestaan. Daarnaast kunnen in nationaal of Europees recht specifieke uitzonderingen worden gecreëerd.