



Whitepaper

# Verantwoorde AI door het faciliteren van bewuste keuzes

Geleerde lessen in de publieke sector

## Auteurs

Marianne Witte-Schaaphok  
Wouter van der Bij  
Marissa Hoekstra  
Romy van Drie  
Steven Vethman

## Samenvatting

Je wilt verantwoord aan de slag met AI. Eind 2024 kan dat al snel overweldigend zijn door het aantal technische, juridische, ethische en organisatorische aspecten dat daarin moet samenkomen. De doorontwikkeling van de techniek (bijvoorbeeld Generatieve AI) zorgt voor steeds meer kansen. Tegelijkertijd zien we met het explicieter worden van vereisten (AI Act en bijbehorende standaarden) de balans steeds meer verschuiven naar verplichtingen die AI met zich meebrengt. Uiteindelijk willen we AI op een waardevolle manier inzetten, met oog voor waarden en risico's. Hoe zorg je ervoor dat al deze elementen op het juiste moment samenkomen en waardevolle innovatie met AI stimuleren? Hoe zorg je voor de juiste balans tussen de beoogde waarden, benodigde middelen en mogelijke risico's van AI voor jouw organisatie? En hoe zorg je ervoor dat mensen in staat worden gesteld deze keuzes te maken? Op basis van vier jaar ervaring en het analyseren van AI use cases in de praktijk pleiten we ervoor de aandacht te verbreden van techniek en gebruik naar verantwoorde keuzes. We geven daarbij zeven belangrijke aanbevelingen.

1. Formuleer een expliciete doelstelling en beschouw AI enkel als middel
2. Investeer in een interdisciplinaire aanpak
3. Zorg voor een veilige experimenteeromgeving
4. Zorg voor een gestructureerd ontwikkelproces inclusief expliciete keuzemomenten
5. Ontwikkel een AI-visie en maak het eigenaarschap van AI toepassingen duidelijk
6. Investeer in AI kennis en -vaardigheden van medewerkers
7. Begin en leer met laagrisicotoepassingen

## De opmars van AI

De hoeveelheid toepassingen van Artificiële Intelligentie groeit hard, net zoals het aantal hulpmiddelen om dit verantwoord in te zetten. Toch blijft het in de praktijk lastig om verantwoorde AI ontwikkeling goed vorm te geven.

Artificiële Intelligentie (AI) heeft een razendsnelle opmars gemaakt in de laatste jaren met de belofte dat organisaties efficiënter, effectiever en objectiever kunnen werken met AI. Zeker met de opkomst van generatieve AI is deze ontwikkeling in een stroomversnelling geraakt. Ook in de publieke sector creëert AI kansen om effectiever te werken, en heeft het potentie om de publieke dienstverlening te verbeteren en behoeftes van burgers op nieuwe manieren te vervullen<sup>1,2</sup>. Dit is duidelijk terug te zien in de groei van de hoeveelheid AI-experimenten en -toepassingen bij de overheid<sup>3</sup>. Zo worden er vanuit de ambitie om informatiegestuurd te werken experimenten uitgevoerd om de potentie van AI te kunnen benutten. Ook wordt het voor organisaties makkelijker gemaakt om AI in te zetten via veelgebruikte software van commerciële partijen. Gezien de publieke functie van publieke organisaties is het essentieel dat AI verantwoord wordt ingezet. Met verantwoord bedoelen we hier de ontwikkeling en inzet van AI op basis van onderbouwde interdisciplinaire keuzes, waarbij er sprake is van een goede inbedding in de organisatie en een expliciete afweging tussen toegevoegde waarde, publieke waarden, middelen en risico's.

In de praktijk lopen organisaties echter tegen talloze uitdagingen aan die het verantwoord in gebruik nemen van AI uitdagend maken. Dit komt bijvoorbeeld door beperkte middelen in de organisatie, beperkte datakwaliteit en moeite met het vinden van de juiste expertise. Het gevolg is dat innovaties stranden, vertraging oplopen of niet kunnen opschalen. Daarnaast zijn voor veel AI-toepassingen die in gebruik zijn de risico's

niet goed in kaart gebracht<sup>4,5</sup>, met mogelijke gevolgen voor organisatie en burgers. Denk bijvoorbeeld aan berichten over discriminerende algoritmes bij DUO<sup>6</sup>, UWV<sup>7</sup> en gemeente Rotterdam<sup>8</sup>. Ook al besteden organisaties aandacht aan hun algoritme, toch kan de uitwerking ervan onbedoeld discriminerend zijn.

De Europese Unie probeert met de AI-verordening<sup>9</sup> grip te krijgen op verantwoorde AI-ontwikkeling door onder andere AI-geletterdheid te eisen en vangrails te bieden voor hoogrisicotoepassingen. Door de inwerkingtreding van de AI-verordening in augustus 2024 verschuift de focus bij veel organisaties naar het naleven van de AI-verordening. Het voldoen aan de AI-verordening geeft echter nog geen garantie op een verantwoorde en waardevolle AI-innovatie.

De groeiende aandacht voor verantwoorde AI en de komst van de AI-verordening leiden tot een overvloed aan technische oplossingen, handreikingen, richtlijnen en middelen die de verantwoorde ontwikkeling en inzet van AI ondersteunen. Vanuit de technische hoek worden er volop nauwkeurigere<sup>10</sup>, eerlijkere<sup>11</sup> en beter uitlegbare<sup>12</sup> methodes ontwikkeld en onderzocht. En vanuit de ethische kant is er de ontwikkeling van hulpmiddelen als de handreiking non-discriminatie<sup>13</sup>, DEDA<sup>14</sup>, IAMA<sup>15</sup>, Aanpak Begeleidingsethiek<sup>16</sup>, ethische richtsnoeren voor betrouwbare AI<sup>17</sup>, ALTAI<sup>18</sup> en het Algoritmekader<sup>19</sup>. Deze middelen zijn van meerwaarde in het ontwikkelproces; ze gaan daarin verder dan de techniek en besteden onder andere aandacht aan doelen, afweging van publieke waarden, verantwoordelijkheden en wat de impact is op stakeholders en de maatschappij. Ondanks het grote aanbod aan hulpmiddelen en ondersteuning blijft het in de praktijk lastig om verantwoorde AI-ontwikkeling goed vorm te geven.

## Verantwoorde keuzes

In de praktijk moet er meer aandacht komen voor het faciliteren van verantwoorde keuzes. De mensen binnen de organisatie moeten in staat worden gesteld om deze afwegingen expliciet te maken, waaronder de doel-waarden-middelen-risico's-balans.

Organisaties ervaren uitdagingen in het aansturen van het ontwikkelproces, het valideren of de innovatie verantwoord is, en de stap maken naar implementatie. Hierbij spelen twee aspecten een essentiële rol. Ten eerste wordt er te weinig aandacht besteed aan het faciliteren van verantwoorde keuzes. Ten tweede is er te weinig expliciete aandacht voor de balans tussen 1) toegevoegde waarde en doel, 2) publieke waarden, 3) benodigde middelen, en 4) bijkomende risico's. We noemen dit de doel-waarden-middelen-risico's-balans, of kortweg de DWMR-balans.

### Faciliteren van verantwoorde keuzes

We zien dat er te weinig aandacht wordt besteed aan het faciliteren van verantwoorde keuzes. De huidige focus van AI-ontwikkeling richt zich met name op twee perspectieven 1) de AI-techniek, en 2) de inzet in de organisatie. Het techniekperspectief legt de focus op de ontwikkeling van technische methodes die aspecten van verantwoorde AI ondersteunen. Denk aan het verhogen van de prestaties, het ontwikkelen van biasdetectie en –mitigatietechnieken, het vergroten van de uitlegbaarheid van algoritmes en het ontwikkelen van hybride AI. Vanuit een breder perspectief wordt er gekeken naar de inzet van AI. Hier draait het om het werkproces waarin het algoritme wordt gebruikt, en de interactie tussen mens en

algoritme. Er wordt onder andere gekeken naar hoe de uitkomsten van de AI worden gebruikt, wat de rol is van de gebruiker, hoe de gebruiker kritisch kan blijven reflecteren, wat de maatregelen zijn om ongewenste bias te voorkomen en hoe er over de inzet van de AI wordt gecommuniceerd

In beide perspectieven zijn het mensen die verschillende keuzes maken om te waarborgen dat de techniek en inzet verantwoord zijn.

Onze overtuiging is dat een derde perspectief, waar het draait om verantwoorde keuzes faciliteren, essentieel is voor het bereiken van waardevolle, verantwoorde AI-innovatie. Dit gaat om keuzes over de beoogde doelen en onderliggende (publieke) waarden, keuzes in het technische ontwerp en de manier van inzetten, de verbinding tussen techniek en organisatie, overkoepelende keuzes over het ontwikkelproces en de verantwoording van deze keuzes. Denk bij overkoepelende keuzes aan de DWMR-balans en de beslissing om een algoritme wel of niet te gebruiken. Veel van deze keuzes zijn zeer afhankelijk van de context van de toepassing en vraagt daarom om de juiste kennis, samenwerking tussen expertises en vakmanschap. Organisaties moeten hun mensen in staat stellen om deze keuzes te maken. Alle genoemde (technische, juridische en ethische) instrumenten zijn behulpzaam, maar voor verantwoorde AI is het essentieel om in de werkwijze voldoende middelen en expertise op te bouwen om deze instrumenten in de praktijk toe te passen.

“Geef meer aandacht voor verantwoorde keuzes in de balans tussen doel, waarden, middelen en risico's.”

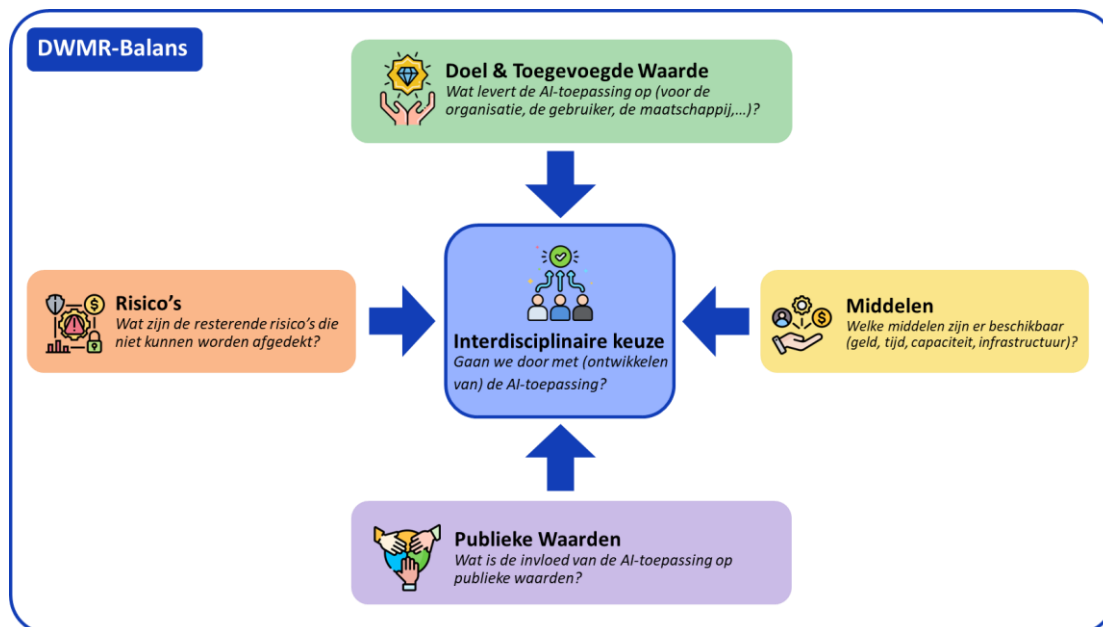
## De DWMR-balans

De DWMR-balans is de expliciete afweging tussen doel (D), publieke waarden (W), beschikbare middelen (M) en de resterende risico's (R). De toegevoegde waarde en de aansluiting bij publieke waarden<sup>20</sup> moet opwegen tegen de inzet van de middelen en de mogelijke risico's. Het uitvoeren van alle vereisten en impact assessments voor een hoogrisicoalgoritme vraagt namelijk om veel middelen, zoals tijd, geld, capaciteit en technische infrastructuur. Deze afweging moet centraal staan in het AI-ontwikkelproces. In elke fase van het ontwikkelproces is het van belang deze balans expliciet op te maken in vastgelegde keuzemomenten. Elke keer wordt opnieuw bepaald of de AI-toepassing het waard is om door te gaan met de ontwikkeling en welke middelen daarvoor kunnen en moeten worden ingezet. Gedurende het ontwikkelproces wordt er door middel van bijvoorbeeld experimenten informatie verzameld om deze keuze te ondersteunen.

Op organisatieniveau wordt het maken van deze balans ondersteunt door een duidelijke AI-visie, die AI als een middel stelt en niet als een doel. Een organisatiecultuur die het mogelijk maakt te experimenteren en te leren versterkt de basis waarop de keuzes worden gemaakt.

Om de balans goed te kunnen opstellen is het voor de teams van belang dat er een gemeenschappelijke taal en een interdisciplinaire manier van werken is. De DWMR-balans overstijgt disciplines. Gezamenlijk moet de keuze worden gemaakt of er door wordt gegaan met de ontwikkeling of inzet van de AI-toepassing.

In de volgende sectie werken we aan de hand van onze ervaringen uit welke uitdagingen er in de praktijk zijn bij het maken van verantwoorde keuzes. In de sectie aanbevelingen stellen we maatregelen voor die organisaties ondersteunen in het maken van de DWMR-balans en het faciliteren van verantwoorde keuzes.



# Uitdagingen rondom verantwoorde keuzes over AI

We zien in de praktijk dat veel organisaties tegen uitdagingen aanlopen bij de verantwoorde ontwikkeling van AI, ondanks de komst van wetgeving, richtlijnen, handboeken en tools. In deze sectie gaan we verder in op de uitdagingen in de praktijk en hoe die bijdragen aan de oproep voor het expliciet maken van de DWMR-balans en het faciliteren van verantwoorde keuzes.

## Beperkte beschikbaarheid van tijd en middelen

Het uitwerken van alle aspecten van verantwoorde AI en de verantwoorde inzet van AI vraagt tijd en middelen, die in de praktijk niet altijd voorhanden zijn. In de ontwikkeling en evaluatie van een AI-toepassing is bijvoorbeeld de inbreng van gebruikers en domeinexperts cruciaal. In de praktijk hebben deze mensen vaak een hoge werkdruk. Daardoor worden ze soms minder of later betrokken dan gewenst is. Ook bij andere stappen in het ontwikkelproces maakt een gebrek aan budget of capaciteit het soms onmogelijk alles tot in detail te onderzoeken of te evalueren.

Om van ontwikkeling naar succesvolle implementatie te komen moeten er in de praktijk dus scherpe keuzes gemaakt worden over wat er wel en niet gedaan wordt. Deze keuzes moeten zo gemaakt worden dat de toegevoegde waarde aangetoond kan worden en de risico's in kaart gebracht kunnen worden. Vanuit management moet er support zijn met voldoende middelen en capaciteit om een verantwoorde keuze te kunnen maken.

## Conflicten tussen verschillende publieke waarden

Elke AI-toepassing raakt meerdere publieke waarden, zoals efficiëntie, effectiviteit, veiligheid, transparantie, accountability, privacy, autonomie en inclusiviteit.<sup>21</sup> Inmiddels zijn er meerdere methodologieën beschikbaar die kunnen helpen met het identificeren en specificeren van publieke waarden, zoals de Aanpak Begeleidingsethiek van ECP<sup>22</sup>, value based sensitive design<sup>23</sup> en de datadialoog<sup>24</sup>. Publieke waarden kunnen ook conflicteren. Denk hierbij aan het wegen van de behoefte aan transparantie en het garanderen van veiligheid bij de inzet van AI-modellen. Of neem bijvoorbeeld de mate van inclusiviteit en efficiëntie tijdens het ontwikkeltraject van een AI-toepassing. Het realiseren van een inclusief ontwikkeltraject voor een AI-toepassing kan tijdrovend zijn en veel middelen vergen om de juiste stakeholders te betrekken, wat ten koste kan gaan van efficiëntie. Het blijft in de praktijk lastig om afwegingen te maken tussen publieke waarden. Daar komt in sommige gevallen bij dat het prioriteren van een publieke waarde een positieve impact kan hebben op één groep maar een (mogelijk) negatieve impact heeft op een andere groep.

We observeren dat organisaties vaak met AI aan de slag willen omdat ze hopen dat het hun werkprocessen efficiënter en effectiever kan maken. Ze zetten daarmee, vaak zonder expliciete afweging, efficiëntie en effectiviteit als centrale publieke waarden. Het is van belang om te beoordelen of efficiëntie en effectiviteit de centrale waarden zijn voor de beoogde toepassing en of deze binnen de DWMR-balans kunnen worden gerealiseerd.

## **Beperkte AI-geletterdheid binnen organisaties**

We zien dat er binnen organisaties vaak nog beperkte AI-geletterdheid is. AI-geletterdheid betekent dat gebruikers kennis en kunde hebben om verantwoord met de AI-toepassing te kunnen werken en de AI-uitkomsten kritisch kunnen evalueren<sup>25</sup>. AI-geletterdheid is nodig om medewerkers op de hoogte te stellen van de mogelijkheden en beperkingen van AI. Zowel bij de ontwikkeling als in het gebruik van AI is het van belang dat werknemers de kennis en kunde hebben om verantwoorde keuzes te maken. Bij de ontwikkeling gaat dit om keuzes in het ontwerp van het algoritme, het ontwerp van het werkproces en mogelijkheden tot passende menselijke controle, maar ook de ethische afwegingen, en de invloed van AI-gebruik op de organisatie. Er moet voldoende kennis zijn binnen de organisatie op deze vlakken.

## **Moeizame interactie tussen verschillende expertises**

De benodigde kennis is vaak verspreid over verschillende personen met verschillende expertises. In de praktijk merken wij op dat de interactie tussen verschillende expertises in veel gevallen moeizaam verloopt. Vaak wordt de ontwikkeling van AI gestart met een focus op technische ontwerpkeuzes. Sommige van die ontwerpkeuzes nemen een ethisch of organisatorisch vraagstuk met zich mee. Door een beperkte interactie van expertises worden niet de juiste mensen op tijd betrokken bij het maken van die keuzes, waardoor deze keuzes bijvoorbeeld (onbewust) bij de data scientist komen te liggen. Wanneer er wel interactie is tussen mensen met verschillende expertises, blijkt dat wederzijds begrip een uitdaging vormt. Ten slotte is het vaak nog onduidelijk wie de uiteindelijke keuzes moet maken tijdens de ontwikkeling. Dit maakt het lastiger om de mensen met

de relevante expertises om tafel te krijgen. Verantwoorde keuzes vragen om een interdisciplinaire aanpak. Mensen met verschillende perspectieven en expertises moeten met elkaar samenwerken. Hiervoor is het nodig dat mensen kennis hebben van elkaars expertise, elkaar begrijpen en op de hoogte zijn van elkaars verantwoordelijkheden.

## **(Bestuurlijke) focus op AI verordening waarborgt geen verantwoorde inzet**

Veel van de bestuurlijke aandacht is gericht op het voldoen aan de AI-verordening, wat voor blinde vlekken kan zorgen bij de verantwoorde inzet van AI. De komst van de AI-verordening brengt voor organisaties veel uitdagingen met zich mee. Bovendien is het voldoen aan de AI-verordening geen garantie om een verantwoorde inzet te waarborgen.

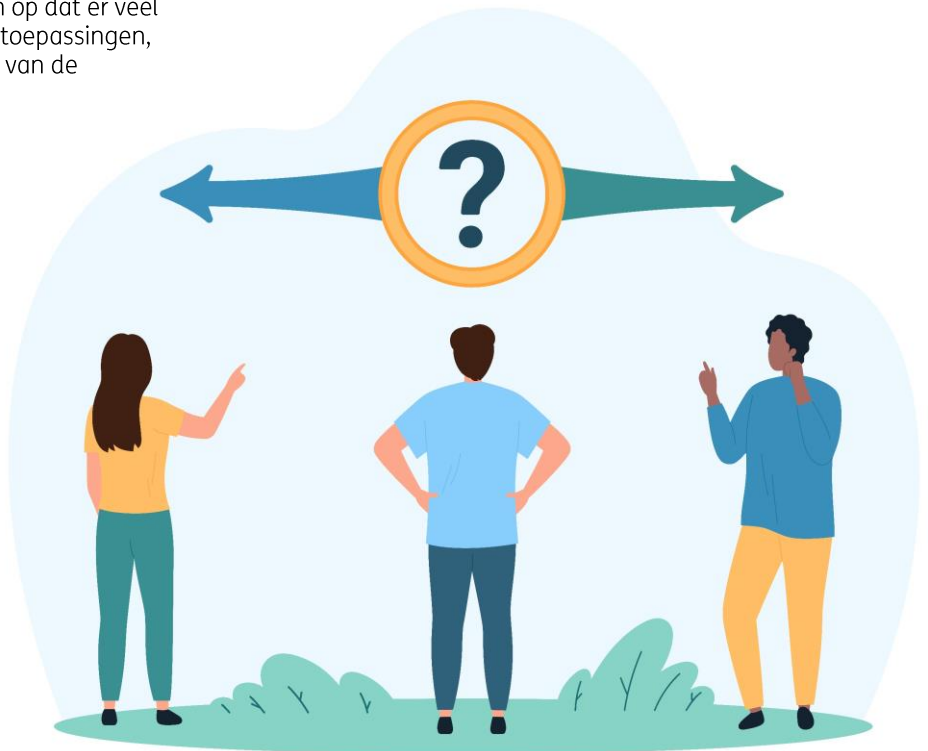
De AI-verordening heeft een productperspectief en focust zich met name op technische aspecten van AI. Er zijn veel vereisten voor de technische ontwikkeling en documentatie met betrekking tot prestaties, robuustheid en bias, risicomanagement, en menselijke controle van het systeem. De uitdaging blijft hoe hier in de praktijk invulling aan moet worden gegeven en hoe mensen in staat kunnen worden gesteld om deze ontwikkeling en evaluatie verantwoord uit te voeren. Veel organisaties kijken daarbij hoopvol naar de komst van geharmoniseerde normen en standaarden vanuit de EU en standaardisatieorganisaties die hen daarbij een eind op weg zullen helpen. Tegelijkertijd zullen deze niet volledig kunnen voorzien in het maken van afwegingen bij specifieke use cases. Deze afwegingen zullen nog steeds een beroep doen op het vermogen van de medewerkers die betrokken zijn bij de ontwikkeling om verantwoorde keuzes te maken.

Naast voldoen aan de AI-verordening, moet er ook verder worden gekeken. Vooral organisaties met een publieke taak moeten kritisch en zorgvuldig zijn in hun processen en daarmee ook de inzet van AI. Als een AI-systeem niet kwalificeert voor hoog-risico, betekent dat niet dat alle vereisten uit de AI Act niet meer noodzakelijk zijn. Het is daarom van belang om verantwoorde keuzes te kunnen maken over alle typen AI-toepassingen, ongeacht het type risicoclassificatie dat de AI Act voorschrijft.

### **Onvoldoende voorbereiding van de organisatie om AI in gebruik te nemen**

Er wordt op veel plekken in de publieke sector geëxperimenteerd met de ontwikkeling van AI, maar slechts een klein aandeel van de AI-toepassingen leidt tot een succesvolle implementatie. Wij merken op dat er veel wordt gewerkt aan de ontwikkeling van AI-toepassingen, maar niet voldoende aan het voorbereiden van de organisatie om AI in gebruik te nemen.

Zo is er bijvoorbeeld in veel organisaties nog geen AI-visie vastgelegd waardoor het niet duidelijk is wat de organisatie wel of juist niet wil met AI. Onvoldoende voorbereiding uit zich ook in het ontbreken van de juiste kennis en vaardigheden bij medewerkers om verantwoord AI te kunnen ontwikkelen en gebruiken. Als laatste hebben wij opgemerkt dat er in veel gevallen geen duidelijke afspraken zijn over verantwoordelijkheden van medewerkers. Dit alles zorgt ervoor dat medewerkers niet in staat worden gesteld om de juiste keuzes te maken op het gebied van AI.





# Aanbevelingen voor verantwoorde keuzes over AI

In deze sectie geven we zeven aanbevelingen die organisaties helpen om verantwoorde keuzes te maken over AI. Als eerste bevelen wij aan om een expliciete doelstelling te formuleren en AI niet als doel maar als middel te zien. Verder versterken interdisciplinaire teams het bij elkaar brengen van de verschillende perspectieven voor beter onderbouwde keuzes. Een veilige experimenteeromgeving biedt de ruimte om evaluaties te doen, onderbouwde informatie te verzamelen, en lessen te leren over de inzet. Starten met laag-risico toepassingen helpt de organisatie om de juiste expertise en ervaring op te bouwen. Bij de ontwikkeling helpt een duidelijk proces waarin expliciete keuzemomenten zijn opgenomen. Het vastleggen van een concrete doelbeschrijving en werkproces helpt om de evaluatie en inzet vorm te geven. Als laatste is het essentieel om een AI-visie en eigenaarschap vast te leggen binnen de organisatie, waardoor beoogde doelstellingen, relevante publieke waarden, rollen en verantwoordelijkheden duidelijk zijn. In onderstaande alinea's lichten we de aanbevelingen verder toe.

## Formuleer een expliciete doelstelling en beschouw AI enkel als middel

Het is essentieel om het doel van de AI-toepassing scherp voor ogen te hebben. Het inzetten van AI is geen doel op zich, maar altijd een middel om een doel te behalen. Dit doel moet expliciet worden vastgelegd: voor welk organisatorisch of maatschappelijk probleem biedt de AI-toepassing een mogelijke (deel)oplossing. AI-toepassingen worden vaak ontwikkeld vanuit een technische analyse van de mogelijkheden van AI. Dat kan

ertoe leiden dat niet voldoende duidelijk is wat het doel van het AI-systeem is of hoe het in het werkproces wordt ingepast. Het doel heeft bijvoorbeeld gevolgen voor de technische optimalisatie van het systeem, het ontwerp van de interface met de mens, en de keuze voor evaluatiemethoden. Bovendien is een duidelijk doel noodzakelijk bij het opstellen van de DWMR-balans en de uiteindelijke keuze of de AI-toepassing wel of niet wordt ingezet. Een gebrek aan een duidelijke doelbeschrijving kan leiden tot verschillende meningen over het doel van het systeem, ontwerpkeuzes die niet met elkaar in lijn zijn en een ontoereikende evaluatie. Stel gezamenlijk een duidelijke beschrijving op van het beoogde doel. Een AI-toepassing kan meerdere doelen hebben. Maak expliciet wat de doelen zijn en maak de verschillende afwegingen, zoals de DWMR-balans, voor elk doel. Dit helpt de verschillende stakeholders om de juiste keuzes te maken met betrekking tot de ontwikkeling, evaluatie en implementatie van de AI-toepassing.

## Investeer in een interdisciplinaire aanpak

Interdisciplinaire teams zijn van grote meerwaarde in het ontwikkel- en keuzeprocess van AI. Bij de ontwikkeling en inzet van AI moeten veel afwegingen en keuzes gemaakt worden. Dit kunnen technische afwegingen zijn, maar bijvoorbeeld ook keuzes over de implementatie in het beleidsproces of de risico's die de inzet van de AI met zich meebrengt. Wanneer bij het keuzeprocess niet de benodigde disciplines zijn betrokken, kan dit er bijvoorbeeld voor zorgen dat kritieke punten pas in een later stadium signaleerd worden of dat ethische afwegingen in de ontwikkeling bij het technische team komen te liggen.

In interdisciplinaire teams wordt een vraagstuk geanalyseerd vanuit de gecombineerde technische, ethische, organisatorische en juridische expertise, waarbij er een gemeenschappelijke taal en wederzijds begrip is.

Door van elkaars expertise en zienswijze te leren ontstaat er een beter begrip van AI in de organisatie. Het integreren van verschillende perspectieven in een interdisciplinair team verbetert de keuzes die in het ontwikkel- en evaluatieproces worden genomen. Dit draagt bij aan een soepele en verantwoorde ontwikkeling van nieuwe AI-toepassingen. Zorg dat de organisatie de verschillende expertises in huis heeft. Investeer in het samenbrengen van deze mensen en het ontwikkelen van onderling begrip.

### **Zorg voor een veilige experimenteeromgeving**

Om verantwoorde keuzes te kunnen maken over de inzet van AI in de praktijk is het van belang om een veilige experimenteeromgeving te gebruiken. Met zo'n omgeving kun je in de testfase de realiteit zo goed mogelijk benaderen, zonder dat burgers en professionals hiermee risico lopen. Met een experimenteeromgeving kun je een AI-toepassing ontwikkelen en evalueren, en kun je de impact op de organisatie, gebruiker en burgers evalueren. Een veilige experimenteeromgeving en cultuur waarin leren, evalueren en reflecteren centraal staat is tevens cruciaal om het juiste kennis- en ervaringsniveau op te bouwen in de organisatie. Met een veilige experimenteeromgeving bedoelen we ook dat er een plek is waar mensen kritische vragen kunnen stellen over de AI-toepassing. Hierbij zijn het werken vanuit de basishouding dat innovatie ook kan "mislukken", het werken met verschillende implementatievormen (AI of alternatieve oplossing) en het constant expliciteren van hypothesen en ontwerpkeuzes essentiële voorwaarden. Tot slot moet de veilige experimenteeromgeving het mogelijk maken om de benodigde informatie te verzamelen voor de uiteindelijke afweging in de DWMR-balans en de keuze om een oplossing wel/niet in te zetten. De DWMR-balans is de afweging die herhaaldelijk gemaakt moet worden tussen 1) toegevoegde waarde en doel, 2) publieke waarden, 3) middelen en 4) risico's.

De volgende vragen helpen om een veilige experimenteeromgeving op te zetten. Bepaal welke aspecten van de beoogde implementatie essentieel zijn om na te bootsen tijdens het experiment. Denk hierbij aan het opstellen van een representatieve steekproef of het toevoegen van extra ruis, uitzonderlijke en specifieke gebruiksscenario's aan het experiment. Bepaal waar het in de normale werkwijze kan worden ingezet of waar de werkwijze een herontwerp vraagt en waar het moet worden losgekoppeld om burgers/professionals te beschermen. Bepaal welke informatie moet worden opgehaald in de experimenteeromgeving en welke experimenten er nodig zijn om tot deze informatie te komen. Creëer een open en kritische werkomgeving met reflectiemomenten.

### **Zorg voor een gestructureerd ontwikkelproces inclusief expliciete keuzemomenten**

Stel een gestructureerd ontwikkelproces op waarin de keuzemomenten expliciet zijn opgenomen. Alhoewel er veel gestructureerde ontwikkelprocessen<sup>26</sup> bestaan, wordt er vaak niet voldoende aandacht besteed aan expliciete keuzemomenten. Denk aan een zogenaamde go/no-go die zich tussen de verschillende fasen van het ontwikkelproces bevindt.

Voor elk keuzemoment moet duidelijk zijn wat er op dit punt besloten wordt, wie deze keuze gaat maken, welke informatie er nodig is om deze keuze te kunnen maken en waar de keuze wordt vastgelegd. In dit keuzemoment is er ruimte voor inzichten en afwegingen vanuit verschillende disciplines maar ook de gezamenlijke verantwoordelijkheid om de DWMR-balans opnieuw te bepalen. Dit helpt in het ontwikkelproces om te prioriteren wat er onderzocht moet worden en voorkomt dat deze belangrijke keuzes niet expliciet worden geëvalueerd, door mensen zonder mandaat worden gemaakt of niet goed worden gedocumenteerd.

## **Ontwikkel een AI-visie en maak het eigenaarschap van AI-toepassingen duidelijk**

Stel een AI-visie en bijbehorend beleid op dat duidelijk maakt wat de organisatie wil bereiken met AI-toepassingen, waar het wel en niet voor mag worden ingezet, wie welke verantwoordelijkheden heeft en welke vaardigheden er nodig zijn binnen de organisatie om AI verantwoord in te zetten. Maak daarin expliciet wie er eigenaarschap heeft voor de ontwikkeling en implementatie van AI-toepassingen. In de visie moet er aandacht zijn voor deze aspecten op zowel strategisch, tactisch als operationeel niveau. Op deze manier kunnen de juiste rollen in de organisatie in staat worden gesteld om de juiste keuzes te maken en is duidelijk hoe verantwoording over de keuzes wordt afgelegd. Daarnaast stelt een AI-visie de organisatie in staat om zich voor te bereiden op de ontwikkeling van AI, waardoor experimenten succesvol kunnen landen in de organisatie. Eigenaarschap in de organisatie zorgt ervoor dat de ontwikkeling en implementatie aansluit bij de AI-visie en de werkvloer en dat er duidelijkheid is over verantwoordelijkheden en mandaat.

## **Investeer in AI-kennis en -vaardigheden van medewerkers**

Zorg ervoor dat medewerkers een breed begrip van AI hebben en over de vaardigheden beschikken om met AI te werken. Dit is essentieel om medewerkers in staat te stellen om verantwoorde keuzes te kunnen maken over AI. Verschillende rollen hebben daarin verschillende vaardigheden nodig. Gebruikers hebben een basisbegrip van AI nodig om verantwoord met de AI-toepassing te kunnen omgaan. Medewerkers die betrokken zijn bij de ontwikkeling moeten bredere kennis en vaardigheden hebben om de keuzes in de ontwikkeling en DWMR-balans te kunnen maken. Dit betekent dat de organisatie voldoende technische kennis in huis moet hebben, maar

ook ethische, juridische en organisatorische kennis. Daarnaast moeten de vaardigheden van medewerkers worden ontwikkeld om complexe afwegingen te maken, bijvoorbeeld bij conflicterende publieke waarden of in de DWMR-balans. Zorg dat de medewerkers die betrokken zijn bij de ontwikkeling naast hun eigen expertise ook het basisbegrip van de andere expertises hebben om interdisciplinair werken te bevorderen. Zorg dat alle medewerkers, waaronder juridische medewerkers en beleidsmedewerkers, een realistisch beeld hebben van wat AI is en kan. Besef dat het vanuit de AI-verordening verplicht is om bij het gebruik van AI te zorgen voor AI-geletterdheid binnen de organisatie.

## **Begin en leer met laagrisicotoeepassingen**

Begin klein om ervaring op te doen, te leren en vertrouwen te creëren. Met klein bedoelen we o.a. toepassingen met laag risico voor burgers en gebruikers en met beperkte wijzigingen in organisatorische en technische processen. Verantwoorde keuzes maken vraagt namelijk om kennis, ervaring en vakmanschap. Deze ervaring bouw je op binnen je organisatie door te starten met ontwikkelen bij laag risico. Op deze manier kan er bijvoorbeeld geleerd worden welke mensen betrokken moeten worden, wanneer welke keuzes moeten worden gemaakt en wat voor informatie er nodig is voor deze keuzes. Bij eenvoudige toepassingen hoeven nog niet direct alle complexe ethische en organisatorische afwegingen gemaakt te worden en zijn de risico's lager als er fouten worden gemaakt. Tegelijkertijd kunnen laagrisicotoeepassingen nog steeds een positieve impact hebben op de organisatie. Documenteer en deel de geleerde lessen en neem ze mee voor vervolgotrajecten. Dit ondersteunt het leerproces en maakt het sneller mogelijk om succesvolle innovaties te realiseren en op te schalen. Succesvolle innovaties stimuleren het enthousiasme en het vertrouwen binnen en buiten de organisatie.

## Conclusie

Het gebruik van AI wordt steeds gewoner en de potentie van de technologie is groot. In de publieke sector worden er stappen gezet om deze technologie om te zetten naar toepassingen om deze potentie te benutten. We zien echter dat organisaties in veel gevallen tegen uitdagingen aanlopen bij het faciliteren van de verantwoorde inzet en ontwikkeling van AI.

Wij pleiten ervoor dat er meer expliciete aandacht komt voor de balans tussen 1) toegevoegde waarde en doel, 2) publieke waarden, 3) benodigde middelen, en 4) bijkomende risico's. De toegevoegde waarde en de aansluiting bij publieke waarden moet opwegen tegen de inzet van de middelen en de mogelijke risico's. Daarnaast moeten medewerkers beter gefaciliteerd worden in het maken van verantwoorde keuzes. In het maken van deze keuzes is interdisciplinariteit en onderling begrip essentieel. De verantwoorde ontwikkeling en inzet van AI valt of staat bij de keuzes die gemaakt worden door de betrokken mensen.

We sporen organisaties daarom aan om aan de slag te gaan met zeven aanbevelingen die in de praktijk bijdragen aan het maken van verantwoorde keuzes over AI. Dit vraagt om actie vanuit bestuur, beleid en de werkvloer, een open blik en een goede samenwerking.

1. **Formuleer een expliciete doelstelling en beschouw AI enkel als middel**
2. **Investeer in een interdisciplinaire aanpak**
3. **Zorg voor een veilige experimenteeromgeving**
4. **Zorg voor een gestructureerd ontwikkelproces inclusief expliciete keuzemomenten**
5. **Ontwikkel een AI-visie en maak het eigenaarschap van AI-toepassingen duidelijk**
6. **Investeer in AI-kennis en -vaardigheden van medewerkers**
7. **Begin en leer met laagrisicotoepassingen**

Wanneer organisaties met deze punten aan de slag gaan, stellen ze hun medewerkers in staat op verantwoorde keuzes te maken in de ontwikkeling en inzet van AI. Dit stimuleert waardevolle AI-innovaties met oog voor de risico's en ondersteunt in een succesvolle landing in de praktijk.

# Het AI Oversight Lab

Het AI Oversight Lab ondersteunt publieke organisaties bij het nemen van verantwoorde en legitieme beslissingen over de ontwikkeling en het gebruik van algoritmen en AI in hun werkprocessen. In het AI Oversight Lab werken we in een interdisciplinair team al vier jaar samen met gemeenten, rijksorganisaties en uitvoeringsorganisaties. We focussen ons daarbij op het creëren van handelingsperspectief in deze samenwerkingen en het dichtn van kennishiaten in de wetenschap en beleid. Het delen van ervaringen en lessen binnen de learning community speelt hierbij een essentiële rol.

In de vier jaar dat het AI Oversight Lab bestaat hebben we samen met organisaties in de publieke sector gewerkt. Dit heeft geleid tot brede praktijkkennis en wetenschappelijke verdieping van de verantwoorde inzet van AI.

Voorbeelden van de uitdagingen die centraal staan in de samenwerkingen zijn:

- technische evaluaties van nauwkeurigheid, robuustheid en bias/non-discriminatie;
- verkenning van randvoorwaarden voor organisatorische inbedding van AI binnen een organisatie;

- onderzoek naar hoe publieke waarden kunnen worden meegenomen en geëvalueerd in het ontwikkelproces;
- het ontwikkelen van een evaluatiemethode voor LLMs, zowel op technisch vlak als op ethisch, juridisch en organisatorisch vlak;
- vormgeven van het betrekken van gebruikers in het ontwikkelproces.

De lessen en aanbevelingen in dit whitepaper zijn gebaseerd op onze samenwerkingen en wetenschappelijke ontwikkelingen.

## Aan de slag met verantwoorde AI?

Wil je beter kunnen omgaan met uitdagingen in de praktijk? De interdisciplinariteit in jouw organisatie vergroten? Of je organisatie beter voorbereiden op de implementatie van AI? Kortom, wil je met jouw organisatie de volgende stap zetten in de verantwoorde ontwikkeling van AI?

Zoek contact om te bespreken hoe we samen jouw organisatie kunnen faciliteren om bewustere keuzes te maken. Met ons interdisciplinaire team kunnen we ondersteunen op technische, ethische, organisatorische en interdisciplinaire vraagstukken. We zetten ons daarbij in om het vakmanschap binnen jouw organisatie te vergroten.

## Met dank aan

- TNO Appl.AI programma voor de financiële bijdrage aan ons onderzoek.
- Onze partners in het AI Oversight Lab, die we bedanken voor hun openheid, inzet en enthousiasme om verantwoord te innoveren met AI. Mede dankzij deze samenwerkingen zijn deze inzichten tot stand gekomen.
- Alle teamleden van het AI Oversight Lab voor hun bijdragen in de onderzoeken, en specifiek Cor Veenman, Anne Fleur van Veenstra en Marianne Schoenmakers voor het aanscherpen van onze inzichten.

## Bronvermelding

1. [World Economic Forum. Unlocking Public Sector Artificial Intelligence. cking Public Sector Artificial Intelligence | World Economic Forum](#)
2. [European Commission. \(2024\). Adopt AI Study.](#)
3. [TNO \(2024\). Quicksan AI in de publieke dienstverlening III.   
https://publications.tno.nl/publication/34642601/SA\\_SNc3ZW/TNO-2024-R11005.pdf](#)
4. [Algemene Rekenkamer. \(2024\). Focus op AI bij de rijksoverheid.](#)
5. [TNO \(2021\). Op zoek naar de mens in AI. Betrek de burger en experimenteer op een verantwoorde wijze.](#)
6. [NOS. \(2023\). DUO mag algoritme niet gebruiken totdat meer bekend is over mogelijke discriminatie](#)
7. [Autoriteit Persoonsgegevens. \(2023\). AP ziet toe op hersteloperatie UWV na illegale inzet algoritme.](#)
8. [Wired. \(2023\). Inside the Suspicion Machine.](#)
9. [European Parliament. \(2024\). EU AI Act: first regulation on artificial intelligence](#)
10. Denk bijvoorbeeld aan de recente ontwikkelingen in LLMs.
11. [Bijvoorbeeld: Gemeente Amsterdam. \(2022\). The Fairness Handbook.](#)
12. [Bijvoorbeeld: UXAI: Design Strategy.](#)
13. [Handreiking non-discriminatie by design | Rapport | Rijksoverheid.nl](#)
14. [Universiteit Utrecht. De Ethische Data Assistent \(DEDA\).](#)
15. [Universiteit Utrecht. Impact Assessment Mensenrechten en Algoritmes](#)
16. [ECP. Aanpak Begeleidingsethiek.](#)
17. [High Level Expert Group. \(2019\). Ethische richtsnoeren voor betrouwbare KI](#)
18. [High Level Expert Group. \(2020\). Assessment List for Trustworthy Artificial Intelligence \(ALTAI\) for self-assessment.](#)
19. [Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. \(2024\). Algoritmekader.](#)
20. Bijvoorbeeld transparantie, privacy, duurzaamheid, rechtvaardigheid
21. [WRR \(2011\). iOverheid. Opgevraagd van: iOverheid | Rapport | WRR](#)
22. [ECP. Aanpak Begeleidingsethiek.](#)
23. [M. Steen and I. van de Poel, "Making Values Explicit During the Design Process," in IEEE Technology and Society Magazine, vol. 31, no. 4, pp. 63-72, winter 2012, doi: 10.1109/MTS.2012.2225671.](#)
24. [IBDS. \(2024\). Datadialogen: doordacht van start met data](#)
25. [Long, D. & Magerko, B. \(2020\). What is AI literacy? Competencies and Design Considerations.](#)
26. [Bijvoorbeeld CRISP-DM, SEMMA of NIST](#)



Benieuwd hoe TNO jouw organisatie kan helpen bij de  
verantwoorde ontwikkeling en inzet van AI?  
Neem contact met ons op via [marianne.schaaphok@tno.nl](mailto:marianne.schaaphok@tno.nl)  
of [wouter.vanderbij@tno.nl](mailto:wouter.vanderbij@tno.nl)

**TNO** innovation  
for life