

# Mensgerichte Artificiële Intelligentie

*Een oproep voor zinvolle en verantwoorde toepassingen*





## Het vitale belang van verantwoorde AI

Artificiële Intelligentie (AI) is zonder enige twijfel de meest invloedrijke technologie van dit moment. Van de rechtspraak tot het onderwijs, van de zorg tot het openbaar bestuur, en van ons dagelijks internetgebruik tot de journalistiek: overal in de samenleving spelen algoritmes een steeds grotere rol in ons doen en laten. Die invloed wordt zowel bejubeld als gevreesd. AI kan een positieve invloed hebben op de maatschappij en de economie, maar heeft ook een keerzijde. Zitten er geen vooroordelen ingebakken in algoritmes en data, en wordt er niemand buitengesloten? Kunnen mensen nog wel overzien hoe AI-systemen werken, en verantwoordelijkheid nemen voor beslissingen die ze nemen op basis van AI?

Deze grote impact van AI op mens en maatschappij brengt een bijzondere verantwoordelijkheid met zich mee. Zowel voor de ontwerpers en ontwikkelaars van AI-systemen als voor de toepassers en gebruikers ervan. Hoe kunnen we AI-systemen verantwoord ontwerpen? Hoe kunnen we bepalen of deze systemen verantwoorde data gebruiken? En hoe kunnen we ze op een goede manier maatschappelijk inbedden en gebruiken? Het belang van deze vragen wordt breed herkend en in 2019 was er bijna geen branche, regering, bedrijf dat niet een ethische code had opgesteld. Het jaar 2019 is de boeken al ingegaan als het jaar van de AI-ethiek.<sup>1</sup>

Ook in de internationale verhoudingen is het streven naar 'verantwoorde AI' van cruciaal belang. In de mondiale concurrentiestrijd profileert Europa zich immers steeds nadrukkelijker als 'derde weg', naast de kapitalistische benadering van 'Big Tech' in het Westen en de centralistische benadering van staatscontrole in het Oosten. In deze derde weg staat de mens centraal. Niet alleen commerciële belangen van bedrijven of de behoefte van de staat om de samenleving te kunnen sturen vormen hier de drijvende kracht, maar de ambitie om AI te verbinden met menselijke waarden. Mensgerichte AI als de derde weg, een weg die Oost en West met elkaar verbindt, en AI ten dienste van de gehele mensheid wil stellen.

Nederland heeft de potentie om een kernspeler te zijn in deze derde weg. Het wetenschappelijk onderzoek op AI-gebied is van hoog niveau en Nederland kent veel succesvolle AI-gerelateerde bedrijven. Het poldermodel van de Nederlandse economie vormt bovendien een uitstekende basis om overheid, wetenschap, bedrijfsleven en de samenleving met elkaar te verbinden. Daarnaast heeft het Nederlandse onderzoek op het gebied van Maatschappelijk Verantwoord Innoveren (MVI) en de ethiek van technologie een mondiale toppositie.<sup>2</sup> In de Nederlandse aanpak wordt ethiek sterk verbonden met technologisch ontwerp: ethische reflectie mondt niet alleen uit in mooie beschouwingen maar slaat ook neer in de technologie zelf. Deze pragmatische aanpak biedt een solide basis om niet alleen ethisch te *praten* over AI, maar om AI ook ethisch vorm te geven. Daarbij gaat het niet alleen om de algoritmes, maar ook om de vraag hoe, waar en door wie ze worden toegepast. Gebeurt dat rechtvaardig, met respect en met solidariteit? Algoritmes worden bovendien altijd getraind met data. Eerlijke algoritmes vergen dan ook representatieve trainingsdata.<sup>3</sup>

Het wordt dan ook tijd om deze unieke kracht van Nederland optimaal te benutten voor mensgerichte Artificiële Intelligentie. Om dat duidelijk te maken laten we hieronder allereerst zien op welke manier AI de samenleving aan het veranderen is en wat daarin de uitdagingen zijn: we zitten in de 'vierde industriële revolutie', die tot de 'vijfde samenleving' leidt. Wij pleiten vervolgens voor een benadering waarin AI-systemen worden ontwikkeld vanuit een *ethisch verantwoord en maatschappelijk zinvol* perspectief. Tot slot verbinden we die benadering met de rollen die alle partijen in de 'quadrupel helix' kunnen spelen – overheid, bedrijfsleven, wetenschap en de burger – om zodoende een duurzame en verantwoorde basis voor welvaart en welzijn te realiseren.

---

<sup>1</sup> UNESCO heeft een aantal van deze initiatieven verzameld in dit overzicht:  
<https://unesdoc.unesco.org/ark:/48223/pf0000369756>

<sup>2</sup> Zie: Jeroen van den Hoven en Peter-Paul Verbeek, 'Nederland is ethisch gidsland in mondiale hightech'  
(<https://fd.nl/morgen/1251701/nederland-is-ethisch-gidsland-in-mondiale-hightech>).

<sup>3</sup> <https://www.cbs.nl/nl-nl/over-ons/innovatie/project/position-paper-cbs-en-artificial-intelligence>

## De impact van AI: van 'Industry 4.0' naar 'Society 5.0'?

Artificiële Intelligentie is onderdeel van wat wel de 'vierde industriële revolutie' wordt genoemd. In deze revolutie wordt digitale technologie *fysiek*: het internet wordt een internet-der-*dingen*, en de computer krijgt een lichaam als *robot*. Door deze technologische ontwikkelingen beweegt onze samenleving zich volgens sommigen in de richting van 'Society 5.0': de vijfde samenleving na de vierde revolutie.<sup>4</sup> Na de tijd van de jagers-verzamelaars bracht de ploeg een overgang naar de agrarische samenleving, waarna de stoommachine de industriële samenleving mogelijk maakte en de computer tot de informatiesamenleving leidde. De vierde revolutie leidt nu tot 'de digitale samenleving', in Japan ook wel aangeduid als de 'super smart society'. Als volgende stap moeten we antwoord geven op de vraag hoe AI invloed zal hebben op de relatie tussen professionals en hun organisaties, en op het ontwikkelen van nieuwe productiviteitsconcepten zoals 'een leven lang leren' en participatieve arbeidsmarktskills.

De kern van de maatschappelijke impact van AI ligt in haar invloed op het *cognitieve functioneren* van de mens: AI verandert het menselijke denken en waarnemen. Historisch gezien is dat niet voor het eerst. AI eerder had het schrift grote invloed op hoe ons geheugen werkt: Plato was al ongerust dat we niets meer zouden kunnen onthouden als we alles zouden gaan opschrijven, en de boekdrukkunst leidde tot een democratisering van kennis en legde de basis voor de wetenschappelijke revolutie. Wat is de nieuwe stap in ons denken die AI teweeg gaat brengen? Wat gaat er gebeuren met ons begrip van de wereld en van onszelf, onze morele en politieke opvattingen en besluiten, onze existentiële vragen en kwetsbaarheden? En hoe zal AI invloed hebben op het functioneren van onze democratie, op de relatie tussen burgers en overheid, en op mogelijkheden tot participatie aan de samenleving?

Er kunnen drie niveaus van impact van AI worden onderscheiden. Op *individueel* niveau beïnvloeden AI-systemen de keuzes, gedragingen, en interpretatiekaders van individuele gebruikers: zoekmachines kleuren met welke informatie mensen hun wereldbeeld opbouwen, expertsystemen geven artsen en rechters een kader voor een diagnose of een vonnis. Op *sociaal* niveau werkt deze invloed door in een grote variëteit van domeinen waaronder de zorg (zoals het detecteren van kankercellen in afbeeldingen), het onderwijs (adaptief lesmateriaal) en de rechtspraak (expertsystemen), verkeer en vervoer (via zelfrijdende voertuigen), en journalistiek ('*automated journalism*'). Op maatschappelijk niveau, tenslotte, beïnvloedt AI de organisatie van de samenleving: '*predictive policing*' verandert de aard van de handhavende taak van de politie, zoekmachines beïnvloeden politiek engagement en algoritmes zijn betrokken in militaire besluitvorming.<sup>5</sup>

Verantwoord omgaan met deze impact van AI vereist dat we nieuwe denkkaders ontwikkelen. Tot nu toe zagen we technologie vaak als 'passief' en de mens als 'actief', maar AI bezit een technologische vorm van 'autonomie'. Algoritmes kunnen leren, en daardoor ontwikkelen ze zich op manieren die hun programmeurs niet altijd hadden kunnen voorzien. Ze zijn in staat om mensen te adviseren over zaken die het menselijk denken minder goed kan overzien en om zelfstandig beslissingen te nemen in bepaalde situaties.

Computers kunnen inmiddels dingen die vroeger alleen mensen konden, en omdat ze dat vaak beter en sneller kunnen, ervaren sommigen deze ontwikkeling als een bedreiging voor de mensheid. Vooral nog geldt dat mensen eindverantwoordelijkheid dragen. Maar de relatie tussen mens en technologie verandert snel, en er worden steeds nieuwe vaardigheden van mensen gevraagd. Als computers sommige vormen van kanker al beter kunnen herkennen dan pathologen, wat gaat dat dan betekenen voor de rol van de arts? En wat is de toekomst van de journalistiek als nieuwssites niet alleen gevuld worden door journalisten maar ook door algoritmes? Hoe kunnen we controleren of het systeem de juiste analyse maakt en niet bevooroordeeld is, en kunnen we überhaupt wel verantwoordelijkheid nemen voor adviezen die tot stand komen op basis van AI?

---

<sup>4</sup> Zie: [https://www8.cao.go.jp/cstp/english/society5\\_0/index.html](https://www8.cao.go.jp/cstp/english/society5_0/index.html)

<sup>5</sup> Een korte video met daarin een kernachtig beeld van de ethische uitdagingen van AI is te vinden op: <https://www.youtube.com/watch?v=TlihPx4z0Nw&t=14s>.



De praktijk ligt echter aanmerkelijk genuanceerder dan een simpel streven naar vervanging van de mens door AI. Uiteindelijk gaat het steeds om het ontstaan van nieuwe *verbindingen* van de mens met AI. Artsen stellen diagnoses *in interactie met AI*, net zoals journalisten het nieuws maken *in samenwerking met* algoritmes. De cruciale vraag is niet of AI de mens wel mag vervangen, maar hoe we de mens en AI op een goede manier kunnen verbinden en met elkaar laten samenwerken.

### Verantwoorde AI: van principes naar praktijk

In de veelheid aan ethische analyses en kaders rondom Artificiële Intelligentie<sup>6</sup> komt steevast een aantal kernthema's terug. Allereerst draait de discussie vaak om *eerlijkheid en rechtvaardigheid*: algoritmes kunnen bevooroordeeld zijn, door hun eigen ontwerp of door de data waarmee ze getraind zijn, en een onrechtvaardige impact op de samenleving hebben. Een tweede centraal thema is *aansprakelijkheid*: in hoeverre is het mogelijk mensen aansprakelijk te houden voor het functioneren en de impact van AI-systemen, bijvoorbeeld voor besluiten die mede genomen zijn op basis van AI? Daarnaast zijn *transparantie en uitlegbaarheid* belangrijke ethische issues: in welke mate kan een systeem uitleggen hoe het tot zijn conclusies is gekomen, zodat mensen kunnen overzien op grond waarvan ze het systeem en de onderliggende data kunnen vertrouwen, ook als die data zelf onvolledig en onbetrouwbaar zijn? En tenslotte wordt ook de *veiligheid* van de systemen vaak benoemd: in hoeverre zijn ze betrouwbaar en bestand tegen vijandig gebruik ('dual use') en is de mens vaardig genoeg om goed met AI-systemen om te gaan?

Ethische issues als deze zijn sterk gericht op het beschermen van de menselijke autonomie tegen de nieuwe vorm van 'autonomie' die de techniek ontwikkelt: is er nog voldoende menselijke controle mogelijk, en sluit het systeem de mens of groepen mensen niet teveel buiten in bepaalde gevallen? Er is echter ook een hele categorie ethische vragen die hier buiten valt. Die richten zich minder op het individu, en meer op de gemeenschap. *Inclusie* is zo'n thema: hoe waarborgen we participatie en diversiteit in onze samenleving? *Welzijn* is een ander voorbeeld: hoe ontwikkelen we AI die een positieve bijdrage levert aan het floreren van de gehele samenleving? Ook deze thema's verdienen expliciete ruimte in het debat.

Na een periode van het formuleren van kaders, principes en uitgangspunten is het nu tijd om de ethiek concreet te verbinden met de praktijk. Dat vereist een andere invulling van ethiek dan het klassieke beeld van de 'ethische toetsing'. Al te vaak wordt ethiek geassocieerd met een soort externe beoordeling, zoals een medisch-ethische toetsingscommissie beoordeelt of een bepaald onderzoek wel of niet uitgevoerd mag worden. Maar de ethiek van Artificiële Intelligentie moet zich juist met de techniek *verbinden*. Het gaat erom het samenspel tussen AI, data, mens en maatschappij op een goede manier te begeleiden, en in een participatief proces te zoeken naar zinvolle en verantwoorde toepassingen. Door te identificeren welke impact concrete AI-systemen en de onderliggende data kunnen hebben op mens en maatschappij, en de waarden die daarbij in het geding zijn een plek te geven in het ontwerp, de maatschappelijke inbedding en het gebruik van die systemen.<sup>7</sup>

Op deze manier blijft het ontwikkelen van mensgerichte en maatschappelijk verantwoorde AI op basis van eerlijke data niet bij het formuleren van algemene principes, maar wordt ethiek concreet verbonden met de technologische praktijk: het leidt tot een ethiek 'van binnenuit' en niet alleen 'van buitenaf'. Bovendien krijgt ethiek hierdoor niet alleen een negatieve, maar ook een positieve rol: naast het formuleren van de grenzen van wat we *niet* willen, richt ze zich op het formuleren van de voorwaarden/handelingsopties om te bereiken wat we *wel* willen.

---

<sup>6</sup> Zie: Roos de Jong, 'Het debat over de betekenis van AI begint pas net': Rathenau Instituut, 2019 (<https://www.rathenau.nl/nl/maakbare-levens/het-debat-over-de-betekenis-van-ai-begint-pas-net>)

<sup>7</sup> Zie: P.P. Verbeek en D. Tijink, Aanpak Begeleidingsethiek. Den Haag: ECP, 2019: <https://ecp.nl/wp-content/uploads/2019/11/060-001-Boek-Aanpak-begeleidingsethiek-240165-binnenwerk-digitaal.pdf>

## **Naar ethisch verantwoorde en maatschappelijk zinvolle toepassingen: een oproep**

Om AI en de onderliggende data te verankeren in ethisch verantwoorde en maatschappelijk zinvolle toepassingen is een nieuwe aanpak nodig, en nauwe afstemming tussen alle betrokken partijen. We stellen daarom de volgende acties voor:

### *1. Stel het streven naar maatschappelijke waarden centraal*

Om AI-toepassingen ethisch verantwoord en maatschappelijk zinvol te maken is het belangrijk om uit te gaan van een goede identificatie van de sociaal-maatschappelijke problemen en uitdagingen die een brede welvaart in de weg staan. Voorbeelden hiervan zijn armoede, onveiligheid, kans-ongelijkheid, arbeidsmarkt-uitsluiting, eenzaamheid, lichamelijk ongemak en onderwijs-barrières. Verzamel data, kennis en inzichten in deze onderliggende problematiek en valideer deze zodat ze erkend worden. Stel de inzichten ter beschikking in een open context om partijen die aan innovaties werken de kans te geven zinvolle oplossingsrichtingen te exploreren en te exploiteren.

### *2. Het bedrijfsleven neemt medeverantwoordelijkheid voor maatschappelijke innovatie*

Om deze gezamenlijke agenda te realiseren is het van belang dat het bedrijfsleven zich committeert aan 'Responsible Innovation'. Innovatie is dan niet alleen technologie-gedreven maar ook maatschappij-gedreven. Daarin verdienen juist publieke waarden een centrale rol: digitale technologie is gaan behoren tot de vitale infrastructuur van samenlevingen overal ter wereld, en kan daarom niet uitsluitend vanuit private waarden worden vormgegeven. AI in het ontwerp en de ontwikkeling van AI-systemen dient ethiek een rol te spelen. Om dit te kunnen realiseren is nauwe samenwerking nodig met wetenschappers op het gebied van AI, data en ethiek en met maatschappelijke organisaties uit de sectoren waarvoor AI-systemen ontwikkeld worden. Ethiek van binnenuit, niet van buitenaf.

### *3. Kennisinstellingen zorgen voor verantwoord AI onderwijs en onderzoek*

De wetenschap en de kennisinstellingen zouden zich moeten richten op verantwoorde onderwijs en onderzoekspraktijken: omdat AI zowel is gebaseerd op wetenschappelijk onderzoek als ook een belangrijke rol speelt in de hedendaagse wetenschappelijke praktijk, dienen de kaders voor wetenschappelijke integriteit en kwaliteit te worden bijgesteld in interactie met de nieuwe mogelijkheden die AI biedt. Datasets moeten op een eerlijke en transparante manier worden samengesteld en algoritmes moeten ze op een uitlegbare manier bewerken. Van waarde-vrije wetenschap naar waardevolle wetenschap.

### *4. Introduceer ethische begeleiding en regulering*

Voor het verantwoord inbedden van AI-systemen in de samenleving draagt de overheid een bijzondere verantwoordelijkheid. Wet- en regelgeving dient afgestemd te zijn op de nieuwe uitdagingen waar AI de samenleving voor stelt, en dat vergt ruimte voor experiment en dynamiek. Daarnaast vraagt mensgerichte AI om het stimuleren van normatieve analyse en reflectie in de praktijken zelf waar AI-systemen en de onderliggende data worden ingebed, zodat de ethiek concreet wordt ingezet voor herontwerp, maatschappelijke implementatie en verantwoord gebruik. AI en data hebben bovendien invloed op het functioneren van de overheid zelf: AI-systemen helpen de overheid de samenleving te analyseren en beslissingen te nemen. Dat vergt bijzondere aandacht voor de impact van AI en de kwaliteit van data op het functioneren van de democratie, en op de manier waarop politieke kwesties worden benaderd, gerepresenteerd en geanalyseerd. Een ethiek van begeleiding, niet van beoordeling.

### *5. Organiseer ethiek vanuit burgers en betrokkenen*

De stem van de burger is onvoldoende aanwezig in debatten over AI en data. In het maatschappelijke debat zijn het vooral maatschappelijke organisaties en intellectuelen die het woord nemen, en niet de burger en de professional die de impact van AI in het dagelijks leven en werken heel direct en concreet ervaart. Deze stem van burgers en betrokkenen zou expliciet moeten worden verbonden met die van wetenschap, overheid en bedrijfsleven. Zoals 'co-design' of 'design thinking' heel gebruikelijk is geworden in het ontwerpveld (waarom zou de ontwerper beter weten wat de gebruiker wil dan de gebruiker zelf?), en zoals in de wetenschap 'burgerwetenschap' sterk in opkomst is (waarom zouden wetenschappers een monopolie hebben op het doen van wetenschappelijk onderzoek?), zo zou de ethiek zich expliciet moeten richten op 'burgerethiek', waar het AI betreft. In de aanpak

begeleidingsethiek zijn mooie eerste resultaten bereikt doordat zorgprofessionals en patiënten een ethische dialoog over AI en data in de zorg zijn aangegaan. Welke concrete ethische vragen en zorgen hebben burgers ten aanzien van concrete toepassingen van AI en data? En hoe kunnen we die verbinden met onderzoek, innovatie, beleid en hun/ons eigen gedrag? Een ethiek van onderop, niet van bovenaf.

#### *6. Verbind AI met Duurzaamheid*

Steeds vaker wordt een vijfde element toegevoegd aan de quadrupel helix: de natuur en de toekomst van onze planeet. Om op een goede manier vorm te geven aan de samenleving dient AI niet alleen mensgericht maar ook natuurgericht te zijn. Ecosystemen zijn kwetsbaar en de mensheid met hen. AI en de onderliggende data hebben een substantiële impact op die systemen, vanwege de grondstoffen waarmee computers worden gebouwd en de energie die AI-systemen en dataopslag consumeren, maar ook vanuit de manier waarop AI-systemen en data het mogelijk maken klimaatsystemen beter te begrijpen en klimaatproblemen te voorspellen, analyseren en tegemoet te treden. Van quadrupel helix naar quintupel helix, en naar duurzame, mensgerichte AI.

#### *7. Ondersteun en stimuleer brede toegankelijkheid van verantwoorde en zinvolle oplossingen*

Om de doelstelling van brede welvaart te realiseren, is het belangrijk dat de AI-oplossingen die bedacht en ontwikkeld worden leiden tot een verantwoord gebruik van AI ook in commerciële toepassingen en dat deze kunnen worden opgeschaald tot proporties die alle inwoners bereiken en insluiten. Iedereen moet erop kunnen vertrouwen dat AI-systemen verantwoord worden ontworpen en gereguleerd, en kunnen profiteren van de positieve bijdrage die AI-systemen kunnen leveren aan ons welzijn en onze welvaart.

Met de bovenstaande zeven aanbevelingen zetten we in op Mensgerichte Artificiële Intelligentie voor een brede samenleving en welvaart. We nodigen u uit om ze kritisch onder de loep te nemen in het licht van snelle ontwikkelingen rond het gebruik van AI. Belangrijke vragen die u zich daarbij kunt stellen zijn: wat is AI en wat is er goed aan? Hoe krijgen we het beter met AI en waar trekken we de grens? Hoe benutten we de mogelijkheden van AI zonder de balans te verstoren tussen individuele vrijheid en collectieve waarden en rechten? Hoe reguleren we het gebruik van AI in toepassingen met de juiste machtsverhoudingen tussen bedrijven, overheden en burger?

Omdat AI een nieuw interdisciplinair innovatiedomein is, dat ons dagelijks leven, zorgen, leren en werken in de samenleving ingrijpend zal veranderen, willen wij als Nederlanders graag zelf bepalen vanuit welke Nederlandse cultuur en waarden we willen omgaan met AI-ontwikkelingen en dit vraagt om een brede maatschappelijke dialoog.