

Tendrapport

Effecten van AI op digitale inclusie en digitaal burgerschap



Alliantie
Digitaal Samenleven

Inhoudsopgave

Managementsamenvatting	4
Introductie	6
Trendrapporten Digitaal Samenleven	7
1. AI - Overzicht van de voornaamste karakteristieken	9
2. AI en Digitale inclusie	11
2.1 Digitale inclusie	12
2.2 Potentieel positieve effecten van AI	14
2.3 Kanttekeningen	16
2.4 Discussie en conclusies	17
3. AI en Digitaal Burgerschap	18
3.1 Digitaal burgerschap	19
3.2 Potentieel positieve effecten van AI	20
3.3 Vier schaduwkanten van AI	23
3.4 Verantwoorde AI	27
3.5 Discussie en conclusies	33
4. Conclusies en een voorstel voor een handelingskader met vijf actierichtingen	35
4.1 Conclusies en trends	36
4.2 Vijf actierichtingen voor een handelings- kader rond AI, digitale inclusie en digitaal burgerschap	38
Totstandkoming van dit rapport	43
Geïnterviewden	44
Deelnemers AI-dialogen	45
Bronnen	46

Managementsamenvatting

Eerste trendrapport met de verkenning van het thema AI

De Alliantie Digitaal Samenleven - een alliantie van publieke, private en maatschappelijke partijen – (hierna ook 'ADS' of 'de Alliantie' genoemd) heeft als doel om digitale inclusie te bevorderen en digitaal burgerschap te stimuleren. Een van de manieren om dat te bewerkstelligen is om kennis bij elkaar te brengen door middel van jaarlijkse Trendrapporten. Dit is de eerste versie met een verkenning van het thema Artificial Intelligence (AI). Het is gebaseerd op 23 interviews, aangevuld met desk research en inhoudelijk gevoed vanuit twee dialoog-sessies met 17 stakeholders.

Doelstelling en doelgroep van dit rapport

De doelstelling van dit rapport is tweeledig:

- Een informatief overzicht bieden van de stand van zaken en ontwikkelingen rond AI vanuit het perspectief van digitale inclusie en digitaal burgerschap en een inzicht te geven in de verschillende benaderingen en openstaande kwesties daarin.
- Met dit informatieve overzicht een kennisbasis leggen onder de dialoog die binnen de Alliantie wordt gevoerd over een te ontwikkelen handelingskader met mogelijke actierichtingen voor de Alliantie.

De doelgroep van dit rapport is de Alliantie (waaronder automatisch ook aan de Alliantie verbonden organisaties vallen) en overige geïnteresseerden.

AI

- Razendsnelle ontwikkelingen: Hoewel bepaalde vormen van AI al jaren worden toegepast, zijn de technologische ontwikkelingen en toepassingen in een stroomversnelling gekomen. De ontwikkeling van zogeheten foundation models (zie voor uitleg hoofdstuk 1) hebben tot de verbeelding sprekende generatieve AI-applicaties zoals ChatGPT mogelijk gemaakt en AI volop in de belangstelling gezet.
- Systeemtechnologie: De impact van AI is te voelen in alle geledingen van de samenleving omdat AI een systeemtechnologie is.
- Kennis en bewustzijn over AI bij het brede publiek blijft achter.

Digitale inclusie

- Omvangrijke en diverse groep mensen: De groep mensen die niet of onvoldoende digitaal kan meekomen, bestaat uit meerdere miljoenen Nederlanders én is zeer divers. Er zijn verschillende redenen waarom deelname aan de digitale wereld voor hen moeilijk is.
- AI-apps die toegang faciliteren: AI-applicaties bieden kansen om een deel van deze groepen te 'empoweren': door andere vormen van interactie (gesproken dialoog) en door vertalingen in simpeler taal geeft AI hun de mogelijkheid om aan de digitale wereld deel te nemen. Voorwaarde is dat deze applicaties op een verantwoorde manier zijn ontwikkeld.
- AI helpt niet iedereen: Een ander deel van deze groepen zal nooit zelf deelnemen aan de digitale wereld omdat zij dit niet kunnen of niet willen. Daarom moet er bij digitale dienstverlening altijd een alternatief zijn in de vorm van menselijk contact.

Digitaal burgerschap

- Kansen voor individu, organisatie en samenleving: De kansen van AI voor digitaal burgerschap liggen op individueel niveau in applicaties die mensen helpen (persoonlijke assistent) of hen nieuwe mogelijkheden geven om bijvoorbeeld computercode of beeld te generen. Op het niveau van organisaties, sectoren en de samenleving als geheel liggen de kansen vooral in allerlei applicaties die zaken 'beter', 'meer' en 'toegesneden op specifieke behoeften' kunnen maken. Tenslotte biedt AI ook kansen op participatie in het maatschappelijk debat met applicaties op het gebied van deliberatieve democratie (zie voor toelichting paragraaf 3.2).

Schaduwkanten van AI

- Data in AI: De door AI-applicaties gebruikte datasets kunnen een ongewenste bias hebben, zijn vaak niet transparant, zijn regelmatig verzameld zonder respect voor auteursrechten en bevatten soms privacygevoelige gegevens.
- AI als black box: Met name de generatieve AI-applicaties zoals ChatGPT maken gebruik van miljarden parameters die tot uitkomsten leiden die niet herleidbaar zijn voor zowel gebruikers als ontwerpers; een zogenaamde Black Box. De uitlegbaarheid van de resultaten is daardoor in het geding. Ook werkt dit automation bias in de hand: de gebruiker neemt de resultaten klakkeloos over 'want de computer het zegt'.
- Digitale autonomie in het gedrag: Belangrijke delen van de AI-technologie zijn in handen van grote tech-leveranciers: zij hebben een bijna-monopolie op deze technologie. Daardoor zijn en worden publieke instellingen in toenemende mate afhankelijk van deze tech-leveranciers en komt hun digitale autonomie in het gedrag.
- Kwaadwillenden: Kunnen met AI-applicaties deepfakes en grote hoeveelheden desinformatie produceren. Ook wordt AI ingezet voor cybercriminaliteit.

Verantwoorde AI als oplossingsrichting

De mogelijke gevolgen van de bovengenoemde schaduwkanten van AI tasten onze mensenrechten en publieke waarden aan. Daarom wordt er hard aan gewerkt om bij de ontwikkeling, implementatie en het gebruik van AI-applicaties ELSA-aanpak (Ethical, Legal & Societal Aspects) te hanteren. Indien wordt voldaan aan randvoorwaarden die voortkomen uit de ELSA-aanpak kan men van verantwoorde AI spreken. Zodoende kan een belangrijk deel van de risico's voor de mensenrechten en publieke waarden gemitigeerd worden. Daarbij moet wel worden opgemerkt dat de ELSA-aanpak voor verantwoorde AI nog niet geheel uitgekristalliseerd is. In ELSA-labs wordt deze aanpak in de praktijk gebracht om ervoor te zorgen dat bedrijven, overheden, kennisinstellingen, maatschappelijke organisaties en mensen gezamenlijk verantwoorde toepassingen van AI ontwikkelen.

Belangrijkste conclusies

In dit rapport worden de volgende conclusies over AI in relatie tot digitale inclusie en digitaal burgerschap getrokken:

1. Zeer snelle ontwikkeling van AI-mogelijkheden en AI-applicaties: De gesignaleerde stroomversnelling betekent dat er een grote urgentie is om deze ontwikkelingen in goede banen te leiden.
2. Verwevenheid van AI met de digitale wereld: AI wordt steeds meer onderdeel van software die we dagelijks gebruiken. Ondanks dat, is het een soms weinig zichtbaar onderdeel van de digitale wereld en daarmee des te relevanter voor de Alliantie Digitaal Samenleven.
3. Ontwikkeling AI-applicaties die digitale inclusie kunnen bevorderen: De groep mensen die moeite heeft om digitaal mee te komen is zeer divers. AI heeft het potentieel om een deel van deze mensen beter toegang te geven tot de digitale wereld. Nu al zijn er op AI gebaseerde applicaties die toegang vergemakkelijken voor mensen met fysieke beperkingen en voor mensen die moeite hebben met lezen en schrijven. In de nabije toekomst zijn er ontwikkelingen te verwachten die inspelen op andere manieren van digitale interactie, zoals door middel van spraaktechnologie. Dit geeft een belangrijke kans voor méér digitale inclusie, maar tegelijkertijd is het duidelijk dat AI niet voor iedereen in deze doelgroepen de oplossing is: bij digitale dienstverlening moet er dus altijd een analoog alternatief zijn.
4. Zonder de invulling van de cruciale, strikte randvoorwaarden van verantwoorde AI kunnen de schaduwzijden van AI voor meer uitsluiting zorgen en een negatieve impact op digitaal burgerschap hebben: De schaduwzijden van AI kunnen samengevat worden in vier punten:
(1) bias/vooringenomenheid; (2) black box – gebrek aan transparantie en betrouwbaarheid van AI; (3) bedreiging van de digitale autonomie van burgers en organisaties door de macht van Big Tech bedrijven; (4) misbruik van AI door kwaadwillenden. Als gevolg hiervan kan AI voor meer uitsluiting zorgen en een negatieve impact op digitaal burgerschap hebben. Maar de gevolgen van deze schaduwzijden kunnen beperkt worden door applicaties aan de randvoorwaarden van verantwoorde AI te laten voldoen. Deze randvoorwaarden zijn (1) geen bias die mensen achterstelt; (2) transparantie en betrouwbaarheid van de resultaten; (3) waarborging van de privacy; (4) toegankelijk en inclusief ontwerp. Aan deze randvoorwaarden kan voldaan worden door bij de ontwikkeling, implementatie en het gebruik van de AI-applicaties de ethische, wettelijke en maatschappelijke aspecten ervan te beoordelen.

Daartoe is een ethisch instrumentarium deels ontwikkeld en deels nog in ontwikkeling, bestaande uit codes, instrumenten en/of methoden voor het stellen van ethische vragen. Daarnaast ruimen steeds meer organisaties capaciteit en menskracht in om dit ethisch instrumentarium toe te passen in hun organisatie (ethische infrastructuur). Het toenemend aantal actieve ELSA-labs is een belangrijk illustratie van de ontwikkeling van verantwoorde AI. Tegelijkertijd moet worden vastgesteld dat hier nog een lange weg te gaan is, waarbij de Alliantie Digitaal Samenleven mogelijk een rol kan spelen.

Overige observaties en trends

- Kennis en bewustzijn over AI in het licht van publieke waarden is nog onvoldoende ontwikkeld.
- Digitale autonomie van burgers en publieke organisaties staat onder druk.
- Regulering loopt achter op ontwikkelingen waardoor er in de komende jaren voortdurend een maatschappelijke discussie zal zijn over wat gereguleerd dient te worden en hoe dat moet gebeuren.
- Sneller wetenschappelijke kennis gebruiken door op zoek te gaan naar nieuwe manieren om deze sneller te delen en in de praktijk te brengen op plekken waar dat relevant is.
- In het kielzog van de ontwikkelingen van AI nemen ook nepnieuws, misbruik van deepfake-technologie en cybercriminaliteit door kwaadwillenden toe. Dit maakt kritische vaardigheden belangrijker dan ooit en maakt dat er dringend behoefte is aan een herkenningsmechanisme voor met name door AI gegenereerde audiovisuele content.

Naar een handelingskader voor de Alliantie Digitaal samenleven

Uit de interviews blijkt dat er behoefte is aan een evenwichtig verhaal over AI, digitale inclusie en digitaal burgerschap, waarbij gefocust wordt op handelingen en acties m.b.t. AI-applicaties die digitale inclusie kunnen bevorderen én verantwoorde ontwikkeling, implementatie en gebruik van AI stimuleren. Bij het vervolg - het oppakken en uitwerken van de hieronder genoemde actierichtingen - hanteert de Alliantie de principes van inclusive design. Hierbij is het vanzelfsprekend dat de mensen om wie het gaat gedurende het gehele proces actief betrokken zijn en meedenken.

Uit deze verkenning zijn vijf mogelijke actierichtingen voor de Alliantie Digitaal Samenleven naar voren gekomen:

1. Organiseren van voorlichting en educatie over AI in relatie tot publieke waarden.
2. Stimuleren van ontwikkeling en gebruik van AI-applicaties die op een verantwoorde manier toegang faciliteren voor mensen met beperkingen.
3. Bevorderen van bewustzijn van het belang van ieders digitale autonomie, van zowel burgers als organisaties.
4. Organiseren van digitaal inclusieve regulering van AI middels participatie van ervaringsdeskundigen.
5. Versterken van de rol van wetenschappelijk onderzoek naar impact en effecten van AI op digitale inclusie en digitaal burgerschap.

Introductie

Trendrapporten digitaal samenleven

De doelstelling van dit Trendrapport is om een informatief overzicht te bieden van de stand van zaken en ontwikkelingen rond AI vanuit het perspectief van digitale inclusie en digitaal burgerschap, en een inzicht te geven in de verschillende benaderingen en openstaande kwesties daarin. Met dit informatieve overzicht wordt een kennisbasis gelegd onder de dialoog binnen de Alliantie over een zich ontwikkelend handelingskader met mogelijke actierichtingen voor de Alliantie.

De doelgroep van dit rapport bestaat uit de aan de Alliantie deelnemende organisaties en overige geïnteresseerden.

Dit Trendrapport over AI is opgesteld in de tweede helft van 2023 en is gebaseerd op 23 interviews met experts van verschillende sectoren, aangevuld met deskresearch. AI is echter een hot topic en de bijeenkomsten, artikelen en rapporten erover buitelen over elkaar. Vanwege deze snelle ontwikkelingen is deze verkenning een momentopname. Deze beperking in acht nemend beoogt de verkenning het volgende:

Doelstelling en opzet

Dit Trendrapport over AI is opgesteld in de tweede helft van 2023 en is gebaseerd op een twintigtal interviews met experts van verschillende sectoren en aanvullend desk research. AI is echter momenteel een hot topic en de bijeenkomsten, artikelen en rapporten erover buitelen over elkaar heen. Vanwege de snelle ontwikkelingen kan een verkenning niet volledig zijn en zal deze op het moment van schrijven reeds verouderd zijn. Daarom beoogt deze verkenning het volgende:

- **Hoofdstuk 1:** Een overzicht van de voornaamste karakteristieken van AI.
- **Hoofdstuk 2:** Een overzicht van de positieve en negatieve effecten van AI op digitale Inclusie.
- **Hoofdstuk 3:** Een overzicht van de positieve en negatieve effecten van AI op digitaal burgerschap.
- **Hoofdstuk 4:** Een afrondend hoofdstuk met conclusies en aanbevelingen voor een handelingskader rond AI, digitale inclusie en digitaal burgerschap met vijf actierichtingen.

Hoofdstuk 1

AI: overzicht van de voornaamste karakteristieken

In dit hoofdstuk wordt kunstmatige intelligentie (artificial intelligence) geïntroduceerd. De verschillende soorten AI worden kort behandeld net als de vraag waarom AI zo belangrijk is en nog belangrijker zal gaan worden.



1. AI: overzicht van de voornaamste karakteristieken

Twee en meer soorten AI

De Wetenschappelijke Raad voor het Regeringsbeleid publiceerde in 2021 'Opgave AI', een uitgebreid rapport over kunstmatige intelligentie.¹ In dat rapport worden AI-systemen gedefinieerd als: 'het soort systemen dat intelligent gedrag vertoont door hun omgeving te analyseren en - met enige graad van autonomie - actie te ondernemen om specifieke doelen te bereiken'.

AI is op verschillende manieren in te delen, maar grofweg zijn er momenteel twee soorten AI-systemen op de markt: (1) AI-toepassingen om één taak uit te voeren, zoals applicaties voor beeldherkenning of spraakherkenning, aangeduid met narrow AI; (2) AI-toepassingen die tekst, beelden of andere media kunnen genereren als reactie op aanwijzingen of vragen (ook wel 'prompts' genoemd), aangeduid met generatieve AI. Generatieve AI-applicaties zijn gebaseerd op zogenaamde foundation models. Dit zijn modellen die door middel van machine learning getraind zijn middels grote datasets. Deze foundation models hebben patronen en structuur van hun ingevoerde trainingsgegevens geleerd, die vervolgens als 'fundament' ingezet worden om nieuwe gegevens met vergelijkbare kenmerken te genereren, zoals teksten, beeldmateriaal, video, audio en computercode. Er zijn allerlei vormen van foundation models. Toepassingen als ChatGPT en Bard gebruiken de zogenaamde Large Language Models, de beeldgenerator DALL-E gebruikt een Text-to-Language model. Deze modellen genereren o.a. teksten en computercode op basis van vragen van de gebruiker (zogenaamde 'prompts'). Deze generatieve AI-systemen maken sinds een jaar opgang met de lancering van ChatGPT. Het gaat onder meer om tekst- en code-genererende systemen (zoals ChatGPT, de variant Bing Chat en Bard van Google) en beeld-genererende systemen (zoals Stable Diffusion, Midjourney, and DALL-E).

Kenner van AI-systemen geven overigens aan dat (1) er allerlei andere vormen van AI zijn of in ontwikkeling zijn, (2) de technische ontwikkelingen heel snel gaan en (3) zij het meest verwachten van combinaties van diverse vormen van AI.

AI toepassingen

Een in november 2023 te verschijnen Rathenau rapport² maakt de volgende indeling in AI-toepassingen:

- Instrumenten om te leren
- Productie-tools
- Complexe probleemoplossers
- Interacteren en ervaren

Systeemtechnologie in een stroomversnelling

AI-applicaties gebaseerd op narrow AI zijn al vele jaren met ons en worden breed toegepast. Echter, de lancering van ChatGPT in november 2022 en de verbazend snelle ontwikkelingen daarin heeft inmiddels velen van de impact van AI op de samenleving doordrongen.

Net als de verbrandingsmotor en elektriciteit is AI een systeemtechnologie: een technologie die alomtegenwoordig is, voortdurend technisch verbeterd wordt en leidt tot complementaire innovaties¹. Dit betekent dat AI brede en veelomvattende effecten heeft en zal hebben op vrijwel alle aspecten van de samenleving. En die effecten zullen naar verwachting erg snel zichtbaarder worden.

Moeten we er blij mee of juist bang voor zijn?

Het gebruik van het woord intelligentie vermenschlijkt AI-applicaties waardoor er gemakkelijk verkeerde conclusies getrokken kunnen worden. Daarnaast doemen hypothetische toekomstbeelden op van een AI-systeem dat elke intellectuele taak van mensen zou kunnen uitvoeren (Artificial general intelligence). Met andere woorden: een autonoom systeem dat de menselijke vermogens evenaart. En nog verder weg zelfs het concept van Artificial Super Intelligence, waarin AI als systeem zoveel intelligenter is dan de mens dat het de leiding zou kunnen nemen en daardoor een risico is voor de mensheid.¹

In het maatschappelijk debat zien we optimisten die verwachten dat AI kan helpen bij het oplossen van grote problemen in de wereld doordat AI-systemen vele data - gestructureerd en ongestructureerd - kunnen verwerken. Een andere groep, meer pessimistisch gestemde personen, wijst vooral op de Artificial Super Intelligence en voorspelt dat AI de ondergang van de mens betekent. Er zijn ook meer genuanceerde stemmen te horen van mensen die de onvermijdelijkheid van AI onderkennen, maar kritisch zijn op sommige ontwikkelingen en toepassingen.

De discussie over deze vooralsnog hypothetische AI-applicaties en over de (on-)mogelijkheden en risico's ervan staat helaas vaak een nuchtere en feitelijke discussie over AI en de effecten die AI nu al heeft in de weg.

¹Algoritmes en AI worden soms door elkaar gebruikt. Het verschil is subtiel: algoritmen zijn stapsgewijze procedures gebruikt om problemen op te lossen of taken uit te voeren. Sommige meer traditionele algoritmes kunnen wel dynamisch zijn door hun output te veranderen als er meer data worden ontvangen (o.a. d.m.v. Bayesiaanse statistiek). Geavanceerde zelflerende algoritmes worden als AI aangeduid.

Is AI neutraal en rationeel?

Computers hebben een rationeel en neutraal imago. Echter, AI is niet noodzakelijkerwijs neutraal en rationeel. Een bias in AI-applicaties kan ontstaan door een bias in het ontwerp (door de wereldvisie van de ontwerpers) en/of een bias in de trainingsdata. Ook is een AI-applicatie niet altijd rationeel, bijvoorbeeld doordat er pseudowetenschappelijke theorieën in het model verwerkt zijn of dat correlaties worden geïnterpreteerd als causaliteiten. De effecten hiervan kunnen leiden tot aantasting van mensenrechten. Een belangrijk knelpunt hierbij is dat met name de generatieve AI-systemen als een black box fungeren: omdat er miljarden parameters in het spel zijn bij het gegenereerde resultaat, is niet te achterhalen hoe zij tot dat resultaat gekomen zijn, zelfs niet voor de ontwerpers zelf. Kortom, naast bias zijn de uitlegbaarheid en het ontbreken van transparantie over de totstandkoming van de resultaten knelpunten.

Zorgen om energieverbruik door AI

Er zijn ook zorgen over de energieconsumptie door AI – met name door de generatieve AI-modellen.³ Zo is berekend dat één opdracht gelijk staat aan een lamp die een uur lang aan staat⁴ en bedraagt de uitstoot van het trainen van een foundation model (taalmodel) meer dan 300 ton CO₂.⁵

Is het nog te stoppen?

De ontwikkelingen in AI-applicaties gaan heel erg snel en spelen zich voor een belangrijk deel buiten de invloedssfeer van de overheid en publieke organisaties af. Er wordt bijvoorbeeld naar verwachting 200 miljard dollar in AI geïnvesteerd door het bedrijfsleven vanaf nu tot 2025.⁵ En omdat het een systeemtechnologie is, is de impact van AI op heel veel vlakken groot. Enkele uitspraken van geïnterviewden voor dit rapport zijn daarbij tekenend: 'een opt-out is niet mogelijk', 'de tijd van niks doen is voorbij' en 'we moeten de regie pakken'. Bij een vraag of een pauze van zes maanden waarin alle ontwikkelingen van AI worden stilgelegd mogelijk is, schieten minister Dijkgraaf en staatssecretaris van Huffelen tijdens een interview dan ook in de lach.⁶ Nee, AI is niet te stoppen.

Hoofdstuk 2

AI en digitale inclusie

Er zijn verschillende groepen die niet of niet voldoende kunnen meekomen met de digitale wereld. Daar gaat dit hoofdstuk over. AI biedt hen de kans op empowerment door andere vormen van interactie met de digitale wereld mogelijk te maken. Maar er zijn ook belangrijke kanttekeningen. De verschillende zienswijzen hebben een plek in dit hoofdstuk. Er wordt afgesloten met drie conclusies.



2. AI en digitale inclusie

2.1 Digitale inclusie

De basis op orde

Digitale inclusie is een essentieel element in de missie van de Alliantie Digitaal Samenleven: iedereen in Nederland mee kunnen laten doen met de digitale samenleving. De basis dient op orde te zijn: men moet beschikken over een computer, een internetverbinding en moet lokaal en laagdrempelig toegang hebben tot hulp en mogelijkheden om digitale basisvaardigheden op te doen. Als dat het geval is, kan iemand in de basis deelnemen aan de digitale samenleving en is digitaal burgerschap de volgende uitdaging. Dat onderwerp wordt behandeld in hoofdstuk 3.

Digitale inclusie – miljoenen mensen kunnen onvoldoende meedoen

De Alliantie Digitaal Samenleven hanteert de volgende definitie van digitale inclusie: 'Mensen zijn gemotiveerd en vaardig, én beschikken over de middelen en hulp om mee te kunnen doen in de digitale wereld'.

Dit geldt echter niet voor een belangrijk deel van de Nederlandse bevolking. In tabel 1 zijn een aantal gegevens over groepen weergegeven die moeite hebben om digitaal mee te komen. Deze tabel illustreert twee zaken:

1. Er zijn uiteenlopende redenen om digitaal niet mee te kunnen komen, variërend van geen toegang of onvoldoende vaardigheden tot beperkende sociale of fysieke omstandigheden.
2. Er is overlap tussen de verschillende groepen en daarom ontbreekt een precies getal over de totale groep van mensen die niet of onvoldoende mee kunnen komen met de digitale wereld. Maar het is duidelijk dat het om meerdere miljoenen mensen gaat.

Handelingsperspectieven	Aantal	Bronvermelding
Geen toegang of geen gebruik		
Geen toegang tot internet	900.000	7
Nog nooit internet gebruikt	763.000	8
Geen toegang tot laptop of tablet	17% Nederlanders	9
Vaardigheden beperkend		
Aantal burgers die onvoldoende digitaal vaardig zijn	4,5 miljoen	7
Haalt niet OESO-basisniveau voor digitale vaardigheden	20% Nederlanders van 16 tot 65 jaar	10
Kunnen niet geheel zelfstandig online bankzaken uitvoeren	2,6 miljoen	11
Mensen met moeite met lezen en schrijven	2,5 miljoen	7
Mensen met beperkte gezondheidsvaardigheden	25% Nederlanders	12
Omstandigheden beperkend		
Aantal burgers die leven in armoede	475.000	7
Fysieke beperkingen		
Slechtziend of blind	345.000	7
Matige of ernstige gehoorproblemen	575.000	13
Beperkte handfunctie	320.000	7
Licht verstandelijk beperkt	370.000	7

Tabel 1 Cijfers m.b.t. mensen die moeite hebben om digitaal mee te komen

Werken aan digitale inclusie

Er zijn meerdere partijen die werken aan het bevorderen van digitale inclusie. We noemen er hier enkele ter illustratie:

- De Stichting Allemaal Digitaal helpt mensen zonder laptop of tablet door middel van gedoneerde laptops die vervolgens opgeknapt worden.¹⁴ De Digihulplijn helpt mensen met vragen over digitale zaken via Whatsapp en een gratis telefoonnummer.¹⁵
- Cybersoek¹⁶ heeft meerdere fysieke plekken in Amsterdam waar per jaar 30.000 mensen met een digitale achterstand persoonlijk geholpen worden.
- Het programma 'Digitale Inclusie'¹⁷ van de bibliotheken focust op mensen in kwetsbare posities die moeite hebben om deel te nemen aan de digitale samenleving. In dit programma worden o.a. niet-digitaal vaardige burgers gestimuleerd cursussen te volgen bij de bibliotheken en worden zij ondersteund in hun contacten met de digitale overheid via de IDO's (Informatiepunt Digitale Overheid).

2.2 Potentieel positieve effecten van AI

Betere toegankelijkheid voor mensen met fysieke beperkingen

De ontwikkelingen van AI-applicaties voor mensen met fysieke beperkingen gaan momenteel erg snel. Een gesprekspartner: 'In mijn 25-jarige carrière zijn de ontwikkelingen nog nooit eerder zo vlot gegaan'. Er zijn inmiddels al een flink aantal voorbeelden van AI-applicaties, die mensen met beperkingen kansen geven om gelijkwaardig mee te doen. Een aantal voorbeelden:

Apps voor visueel gehandicapten:

- Het programma Midjourney heeft de optie 'describe' waarmee een afbeelding beschreven wordt.
- De applicatie Be my eyes is een app voor smartphones waarmee blinde of slechtziende gebruikers hulp kunnen vragen aan ziende vrijwilligers. Deze app heeft inmiddels als toevoeging een Be my AI functie. Je kan de app daarmee bijvoorbeeld vragen om een foto te beschrijven en zelfs doorvragen om bepaalde details van de foto nader te beschrijven. Voor velen is de drempel om een computer een beschrijving te vragen lager dan dezelfde vraag aan een mens te stellen.
- Envision Glasses zijn AI-powered "slimme" brillen met een geïntegreerde camera en ingebouwde speakers. De applicatie kan tekst herkennen en voorlezen, beschrijven wat de camera waarneemt, en biedt ook de mogelijkheid om iemand te bellen die dan via zijn/haar smartphone kan zien wat de camera ziet en de visueel gehandicapte vervolgens kan helpen.

Apps voor mensen met andere beperkingen:

- Mensen met auditieve beperkingen worden ondersteund door gehoorondersteuning, maar ook door automatische ondertiteling.
- Motorisch gehandicapten worden geholpen door een 'slim' huis met AI-applicaties die voorspellen wat je wilt gaan doen zodat bijvoorbeeld de deuren al geopend worden. Er zijn al veel elektronische toepassingen op dit gebied, maar ze worden steeds slimmer door AI-ontwikkelingen.

Er wordt ook gewerkt aan één applicatie waarin alle toepassingen worden gebundeld zodat de bediening gemakkelijker wordt.

- Daarnaast worden er diverse apps ontwikkeld voor mensen met andere beperkingen zoals autisme.

Sneller = beter:

- Mensen met beperkingen hebben vaak een energieprobleem: ook daarin kunnen AI-applicaties helpen. Als je bijvoorbeeld een vraag bij Google intikt, krijg je een lijst met verwijzingen naar documenten die je allemaal zou moeten raadplegen. ChatGPT geeft daarentegen direct een antwoord, wat natuurlijk veel tijd en energie scheelt.

Betere toegankelijkheid voor mensen met moeite met lezen en schrijven

- Taal-apps: Leessimpel¹⁸ is een app die mensen helpt om ingewikkelde brieven van bijvoorbeeld de belastingdienst of de zorgverzekering te lezen. Deze app is een succesvol, recent ontwikkeld voorbeeld van AI-applicaties die teksten samenvatten, in eenvoudige taal of in een andere taal omzetten of juist een in simpele taal opgestelde tekst kunnen omzetten naar bijvoorbeeld een juridisch onderbouwde tekst. Uit kwalitatief onderzoek blijkt overigens dat degenen die dat nodig hebben vaak al op AI gebaseerde applicaties gebruiken, zoals Google Translate en Google Lens.
- Spraaktechnologie: Spraaktechnologie maakt het mogelijk om in te spreken in plaats van typen/schrijven. Daarnaast is het mogelijk om teksten voor te laten lezen.
- Apps voor mensen met beperkte gezondheidsvaardigheden: Er wordt gewerkt aan een label voor gezondheidsapps¹⁹ die betrouwbaar zijn én begrijpelijk voor mensen met beperkte gezondheidsvaardigheden.

Beloften voor de toekomst

- 'Weg van het toetsenbord en weg van het Engels': De gesprekspartner van Meta stelt dat de ontwikkeling van de taalmodellen andere manieren van interactie met internet mogelijk maken: 'weg van het Engels' door middel van de taalmodellen en 'weg van het toetsenbord' door middel van spraaktechnologie. In een stap verder werkt Meta zelfs aan een koppeling van zintuiglijke waarnemingen met internet met ImageBind.²⁰ Dit is een AI-model dat zes modaliteiten aan elkaar verbindt zoals onder meer afbeeldingen en video, audio, en tekst. Het achterliggende idee is om zintuiglijke waarnemingen aan elkaar te koppelen. Dus bijvoorbeeld een plaatje van een koe, kan vertaald worden in een tekst, of in geluid.
- AI-apps als 'mentaal exoskelet': Een exoskelet is een omhulsel van het lichaam dat menselijke kracht ondersteunt en zo helpt met het verrichten van zware werkzaamheden. AI-applicaties hebben de belofte in zich om een mentaal exo-skelet te worden waarbij AI-technologieën worden gebruikt om menselijke cognitieve capaciteiten te versterken, uit te breiden, of aan te vullen. In het bijzonder betekent dat extra potentieel voor laagtaalvaardigen en mensen met beperkingen.

2.3 Kanttekeningen

De mogelijkheden van AI om digitale inclusie te faciliteren zijn veelbelovend. Maar er zijn ook belangrijke kanttekeningen:

- Businessmodellen toegankelijkheids-apps: Bovengenoemde toegankelijkheids-apps maken gebruik van AI-technologie zoals ChatGPT. Dat is op dit moment nog gratis, maar het is niet duidelijk of dat zo blijft. Ontwikkelaars van de eerder genoemde app Leessimpel zijn daarom op zoek naar een niet-commercieel businessmodel om de app gratis en duurzaam beschikbaar te houden en deze verder te kunnen ontwikkelen. Mocht er toch betaald moet worden voor dergelijke apps, dan is er beleid nodig over de voorwaarden die daaraan gesteld worden. Een andere mogelijkheid is dat de overheid zelf datasets beschikbaar gaat stellen.
- AI-paradox – betere toegang tot apps die kunnen discrimineren: AI faciliteert de toegang voor mensen met beperkingen, maar de AI-applicaties waartoe toegang geboden wordt, zijn gebouwd op modellen die zelf weinig of geen gegevens over mensen met beperkingen bevatten en zijn als gevolg daarvan vaak biased in hun output. Dat zou je de AI-paradox kunnen noemen: AI vergemakkelijkt de toegang tot AI-applicaties die mensen met beperkingen tegelijkertijd kunnen discrimineren [zie voor meer details hoofdstuk 3].
- Mogelijk mindere motivatie om te leren lezen en schrijven: Applicaties die helpen bij het begrijpelijk maken van geschreven teksten zouden mogelijk de motivatie van mensen om hun vaardigheden op dit gebied bij te spijkeren kunnen verminderen.
- Onbegrip en angstgevoelens: Bij mensen met een digitale achterstand leidt AI tot veel vragen en overheerst een gevoel van angst en zelfs paranoïde gevoelens. Cyberzoek ziet bijvoorbeeld dat banken hun klanten via de bankapp berichten sturen met voorspellingen over hun financiële situatie. Sommige klanten kunnen deze berichten niet plaatsen waardoor ze het gevoel krijgen in de gaten gehouden te worden of gehackt te zijn. Dat baart hen zorgen, en veroorzaakt gevoelens van frustratie, boosheid en angst. Er is behoefte aan menselijk contact om gerustgesteld te worden. Cybersoek heeft ook ervaring opgedaan met een onderzoek onder bezoekers van Cybersoek over videobellen met een avatar. Een publieke organisatie wilde onderzoeken of deze manier van contact hun bereikbaarheid zou kunnen vergroten. Bij het gebruik van het prototype bleken de ondervraagden zich genept te voelen: ze waren boos dat ze niet met een echt iemand gesproken bleken te hebben.
- Belang van design for all: Voor mensen met fysieke beperkingen zijn er al applicaties die toegankelijkheid bevorderen, gebruikmakend van AI. Ook daar zijn er punten van zorg: de techniek wordt meestal ontwikkeld door een homogene groep ontwikkelaars waardoor diversiteit vaak nauwelijks een rol speelt. Daarom wordt vanuit het European Disability Forum²¹ en Leder (in)²² gepleit voor de toepassing van Design for All²³, een Europese standaard waarbij doelgroepen meegenomen worden bij de ontwikkeling van een product of dienst om toegankelijkheid te vergroten en bias te verminderen. De gesprekspartner van Leder (in) geeft een aantal voorbeelden die het belang hiervan onderstrepen. Zo is biometrische identificatie, bijvoorbeeld het met een vingerafdruk ontgrendelen van een smartphone, een probleem als je geen handen hebt. Ook de privacy-waarborging is extra belangrijk voor mensen met beperkingen omdat hun gebruik gevoelige gegevens kan opleveren.

2.4 Discussie en conclusies

Kunnen, willen en meedoen via anderen

Onderzoek over hoe mensen met moeite met lezen en schrijven in het dagelijks leven omgaan met de digitale wereld brengt twee inzichten naar voren: ten eerste hebben deze mensen vaak een afwerende houding t.a.v. digitale zaken zoals DigiD en sinds kort ook AI. Dit wordt veroorzaakt door een negatief zelfbeeld enerzijds en anderzijds negatieve ervaringen met instanties waardoor zij het vertrouwen in deze technologie en de organisaties die er gebruik van maken verliezen. Ten tweede doen veel mensen uit deze groep wél mee in de digitale samenleving, maar doen zij dat via vrienden, burens, sociaal werkers en/of de bibliotheek. Op basis van hun sociale kapitaal kunnen ze meedoen door hulp te vragen aan hun omgeving, waarbij ze via sleutelfiguren en -instanties toch participeren in de digitale samenleving.

Menselijk contact

Cybersoek heeft de ervaring dat hun doelgroepen – mensen met een digitale achterstand – eerst op hun gemak gesteld moeten worden. Hun aanpak gaat daarom uit van drie stappen: eerst ontmoeten en elkaar leren kennen door menselijk contact en niet via een scherm, vervolgens toegang bieden en tenslotte het opdoen van kennis en vaardigheden stimuleren. De gesprekspartner van Cybersoek stelt dat de mogelijkheid van AI om door spraak te interacteren met internet impliceert dat iemand al weerbaar is en kan duiden dat hij/zij met de computer praat. Dat is niet voor iedereen het geval. Verantwoord gebruik van technologie is daarom niet voor iedereen mogelijk. Daar komt het digitale burgerschap dus in de knel.

Verskillende zienswijzen

Dit alles betekent dat verschillende gesprekspartners verschillende zienswijzen hebben op AI en digitale inclusie: sommigen zijn positief en benadrukken de mogelijkheden die AI biedt om bepaalde groepen te betrekken bij de digitale wereld en hen een stem te geven. Anderen zijn meer sceptisch en waarschuwen ervoor dat organisaties de mogelijkheden van AI aangrijpen om hun dienstverlening verder te automatiseren en menselijke contactmogelijkheden te versoberen met als gevolg dat de digitale kloof juist groter wordt.

Conclusies

Dit alles leidt tot de volgende conclusies:

1. De groep mensen die niet of onvoldoende digitaal kan meekomen is groot (meerdere miljoenen) én zeer divers met verschillende redenen waarom deelname aan de digitale wereld moeilijk is.
2. AI-applicaties bieden kansen om een deel van deze groepen te 'empoweren': op AI gebaseerde spraaktechnologie en vertalingen in eenvoudigere taal geven hen de mogelijkheid om aan de digitale wereld deel te nemen, mits deze applicaties op een verantwoorde manier zijn ontwikkeld (zie voor verantwoorde AI paragraaf 3.4).
3. Een ander deel van deze mensen zal nooit zelf deelnemen aan de digitale wereld omdat zij dit niet kunnen of niet willen. Daarom moet er bij digitale dienstverlening altijd een alternatief zijn in de vorm van menselijk contact.

Hoofdstuk 3

AI en Digitaal Burgerschap

Na een definitie en korte bespreking van digitaal burgerschap, worden in dit hoofdstuk de positieve effecten en schaduwzijden van AI behandeld. Met verantwoorde AI wordt geprobeerd de schaduwkanten te minimaliseren om zo de kansen van AI te benutten. Verantwoorde AI houdt zich bezig met de ELSA-aspecten: de ethical, legal, societal aspecten van AI-applicaties. De op handen zijnde AI Act, evenals het ethische instrumentarium en de ELSA-labs komen aan de orde. Er wordt afgerond met een discussie over de verschillende zienswijzen en een aantal conclusies.



3. AI en Digitaal Burgerschap

3.1 Digitaal burgerschap

Digitaal burgerschap gedefinieerd

Digitale inclusie en digitaal burgerschap zijn twee zijden van dezelfde munt. Wanneer de basis op orde is, is er ruimte voor het uitoefenen van digitaal burgerschap: begrip van de kansen en risico's van digitalisering, zowel op het individu als op de samenleving. Met voldoende begrip kun je je daartoe verhouden en de regie nemen. Ook dat is van belang om mee te kunnen doen. Door de snelle ontwikkelingen komt de lat echter steeds hoger te liggen wat betreft digitale vaardigheden en digitale geletterdheid. 'Je moet steeds meer en er komt steeds meer bij'.

De Alliantie Digitaal Samenleven (ADS) hanteert de volgende definitie voor digitaal burgerschap: 'mensen kennen de effecten van digitalisering op de menselijke waarden en hebben het vertrouwen om zelf de regie te pakken in het (digitale) leven'. Het Rathenau Instituut onderscheidt drie niveaus van digitaal burgerschap (zij noemen het ook wel technologisch burgerschap²⁴):

- het individuele perspectief (empowerment door technologie)
- het gemeenschapspectief (hoe willen we samenleven en beschermen we gebruikers tegen risico's van de technologie?)
- het perspectief van participatie in het maatschappelijk debat (wat is de invloed die een burger kan uitoefenen in de politieke arena?).

Werken aan digitaal burgerschap

Er bestaan diverse initiatieven om digitaal burgerschap te stimuleren. In deze verkenning zijn de volgende gesignaleerd (niet uitputtend):

- ADS is actief op diverse thema's binnen digitaal burgerschap, zoals digitale nalatenschap en AI²⁵.
- De bibliotheken zijn actief met een programma Digitaal Burgerschap²⁶ dat gericht is op burgers die in de basis wél digitaal vaardig zijn. In dit programma wordt aandacht besteed aan de ontwikkeling van complexere vaardigheden en aan een kritische houding waardoor zij actief en weerbaar kunnen blijven participeren in de digitale samenleving.
- De Nederlandse AI Coalitie heeft een nationale AI-cursus²⁷ gelanceerd en er zijn verschillende AI-cursussen per sector ontwikkeld.²⁸
- De Patiëntenfederatie ontwikkelt een e-learning-programma voor patiëntenvertegenwoordigers.

AI-competenties blijven achter en vertrouwen daalt

Hoe staat het met de digitale vaardigheden met betrekking tot AI? Met het onderzoeksinstrument DigiQ worden in een breed opgezet periodiek onderzoek relevante digitale kennis en vaardigheden van Nederlanders gemeten. Uit de meting van 2022 blijkt dat kennis en vaardigheden voor wat betreft AI onder alle leeftijdsgroepen het minst ontwikkeld zijn.²⁹

Een ander onderzoek laat zien dat de bekendheid met algoritmes onder Nederlanders in 2022 is toegenomen tot 63%. Over het algemeen is het vertrouwen in algoritmes sinds 2019 gedaald (24% heeft minder vertrouwen gekregen, 6% meer). Technologische vooruitgang en groei van de economie worden door ruim 40% als toegevoegde waarde van AI gezien. Maar slechts 19% denkt dat algoritmes bijdragen aan een eerlijke en rechtvaardige wereld.³⁰

3.2. Potentieel positieve effecten van AI

Eerder in dit rapport is AI beschreven als een systeemtechnologie en dat de technologie als zodanig vele toepassingen heeft en zal krijgen. In deze verkenning lichten we enkele van deze toepassingen toe.

Arbeid

Een rapport van McKinsey berekende het potentieel om 2100 werkactiviteiten van ca. 850 beroepen door generatieve AI te automatiseren.³¹ Zij concluderen dat 60% tot 70% van deze werkzaamheden geautomatiseerd kan worden. Dit zal echter niet van de ene op de andere dag gebeuren: zij schatten dat ongeveer de helft daarvan rond 2045 inderdaad zal zijn geautomatiseerd. Generatieve AI-applicaties zullen vooral effect hebben op hoger-opgeleide kenniswerkers – in tegenstelling tot andere automatiseringgolven die juist uitvoerende, lager opgeleide werknemers raakten. Ter illustratie: kenniswerkers besteden gemiddeld 1 dag per werkweek aan het zoeken en verzamelen van informatie. Iets wat generatieve AI-applicaties in de toekomst zouden kunnen overnemen. McKinsey concludeert dat AI voor de economie een belangrijke productiviteitssprong kan betekenen mits er adequate maatregelen genomen worden om de transitie van de beroepsbevolking in goede banen te leiden.

Overheidsdienstverlening

Gesprekspartners voorzagen een persoonlijke digitale assistent die vragen van burgers kan beantwoorden, maar de burger ook proactief kan wijzen op bepaalde overheidsregelingen waar hij/zij mogelijk recht op heeft. AI biedt mogelijkheden om de huidige verticalisatie van de overheid te doorbreken: wanneer een burger bijvoorbeeld een bepaalde aanvraag doet kan hij/zij gewezen worden op regelingen van andere overheden. Daartoe zouden AI-applicaties ook afwegingen van eisen en voorwaarden van verschillende regelingen mogelijk moeten maken teneinde te voorkomen dat deze langs elkaar heen werken. Eén gesprekspartner pleit ervoor dat wanneer een overheid een tool ontwikkelt voor intern gebruik, deze ook aan de burgers beschikbaar wordt gesteld. Als een tool wordt ingezet om bijvoorbeeld een vergunningsaanvraag te beoordelen, kan de aanvrager de tool zelf gebruiken om vooraf bekijken of de vergunningsaanvraag een kans maakt en op welke gronden deze bijvoorbeeld wordt afgewezen, zodat hij/zij de mogelijkheid heeft om de aanvraag aan te passen. Hiermee wordt de informatiepositie van de burger opgebouwd en komt er meer balans in de machtsverhouding. Het is tevens een grote kans voor de overheid om vertrouwen terug te winnen.

E.e.a. is voornamelijk toekomstmuziek, maar er wordt wel degelijk aan dergelijke AI-applicaties gewerkt (zie kader op de volgende pagina).

Gebruikersinterface met natuurlijke taal

De SVB verkent momenteel de toepassing van een op generatieve AI gebaseerde chat-interface voor haar website. De uitvoeringsorganisatie past de werkwijze van hun eigen innovatielab Novum toe om op 'slimme' wijze effectiviteit na te streven met inachtneming van privacy en veiligheid. 'Slim' omdat er op voorhand aannames zijn vastgesteld die helpen om de verkenning stap voor stap vorm te geven. Zo zal het eerste experiment draaien om de vraag of de gebruiker een chat-interface als positief ervaart. Andere experimenten gaan over de juistheid van antwoorden en over hoe de AI-functionaliteit goed te beheren is. De veiligheid van het traject wordt onder andere bewaakt door alleen te werken met reeds publiek toegankelijke informatie van de website. Ook draait de chatbot software op een aparte computer en niet op de bestaande "live" productie-omgeving. Het experiment zet deelnemers van een onderzoekspanel in die de instructie krijgen om geen eigen gegevens in te voeren. De chatbot software maakt gebruik van een van de grote generatieve AI-aanbieders.

Onderwijs

De Onderwijsraad ziet twee belangrijke effecten van AI in het onderwijs: (1) de inzet van AI brengt veranderingen met zich mee in leergemeenschappen; sociale gemeenschappen gericht op leren zoals klassen, scholen en opleidingen.³² Dit heeft implicaties voor de rollen van onder andere leraren en docenten; (2) De inzet van AI verandert het onderwijzen van kennis, vaardigheden en houdingsaspecten. In deze verkenning zijn voor de laatst genoemde impactie een aantal toepassingen naar voren gekomen:

- Adaptieve leermiddelen worden al toegepast in het onderwijs en zijn veelal gebaseerd op regel-gebaseerde algoritmes. Er worden ook steeds meer zelflerende algoritmes gebruikt. AI kan ook door reacties van de leerlingen te analyseren bepalen wat de optimale aanpak is qua methode of presentatie. De technologie houdt de belofte in van gepersonaliseerd onderwijs waarbij leerlingen hun eigen leerpad volgen en waarin zij zich kunnen ontwikkelen op hun eigen talenten en interesses.
- Virtual Reality-brillen maken betekenisvoller simulatieonderwijs voor met name het MBO mogelijk. Een student kan bijvoorbeeld in Virtual Reality oefenen in het aanspreken van een foutparkeerder.
- Sociaalpsychologische monitoring van leerlingen en interpretatie daarvan met behulp van AI de onderwijsprestaties te verbeteren, maakt opgang in het onderwijs. Het is bijvoorbeeld opgenomen in de Leerversnellers van Microsoft³³ voor het primair en voortgezet onderwijs en in Perusall in het hoger onderwijs³⁴ (zie tekstkader). Monitoring wordt ingezet omdat sociaal emotioneel welbevinden van belang is om te kunnen leren. Het is de vraag of meer monitoren ook altijd beter is voor de leerlingen of studenten.

Enkele praktijkvoorbeelden van AI-applicaties in het onderwijs

Leerversnellers: De applicatie Leerversnellers van Microsoft is toegevoegd aan Teams. Deze applicatie biedt leerhulpmiddelen die bijdragen aan het maken, beoordelen en analyseren van oefenopdrachten terwijl leerlingen in real time worden geholpen een inhaalslag te maken, bij te blijven en/of verder te komen. Het geeft docenten inzicht in de voortgang van de leerlingen op leesen rekensvaardigheid, maar ook emotioneel welzijn.

Perusall: Perusall is een online platform waar de docent artikelen en andere materialen kan uploaden. Studenten loggen in en lezen de teksten, kunnen annotaties toevoegen en met elkaar in discussie gaan. De docent kan zien hoeveel tijd een student bezig is geweest en hoeveel annotaties deze heeft gemaakt. Perusall voegt ook automatisch scores toe aan annotaties van studenten. Studenten kunnen geholpen zijn met de geautomatiseerde feedback, maar worden ook geconfronteerd met een alomtegenwoordige observerende AI-applicatie terwijl ze aan het studeren zijn.

Deliberatieve democratie

AI kan burgers ondersteunen in de onderlinge dialoog én de dialoog met de overheid. Door het faciliteren van een grootschalige, gestructureerde dialoog en door het analyseren van grote hoeveelheden feedback helpt AI burgers om tot genuanceerde beleidsinzichten te komen en doordachte voorstellen te formuleren. Voorbeelden van dergelijke applicaties zijn Citizens Foundation³⁵ en Pol.is³⁶. Een gemeente kan met deze AI-hulpmiddelen bijvoorbeeld een kwestie in klare taal uitleggen aan groepen burgers en hen vragen om mee te helpen richting te geven aan de oplossing ervan. AI kan mensen die niet gewend zijn aan politieke beleidstaal een stem geven aan mensen en ervoor zorgen dat ze participeren.

Persoonlijke digitale assistent

AI eerder is een persoonlijke, op AI gebaseerde digitale assistent ter sprake gekomen. De ontwikkelingen lijken hierin razendsnel te gaan. Zo heeft ChatGPT recent de mogelijkheid gecreëerd om gebruikers hun eigen persoonlijke assistenten te laten creëren. Het vooruitzicht is dat AI-assistenten zaken gaan regelen zoals e-mailen, vergaderingen plannen of boekingen uitvoeren.³⁷

3.3. Vier schaduwkanten van AI

Data in AI

Bias

AI heeft het vermogen om grote hoeveelheden data, gestructureerd en ongestructureerd, van verschillende bronnen samen te voegen en te analyseren naar gelang de toepassing. Maar hoe zitten die datasets in elkaar? Hoe zit het met de datakwaliteit? Zijn de datasets representatief? Het antwoord is in veel gevallen negatief: de datasets zijn vaak niet representatief voor de totale bevolking omdat niet van alle groepen evenveel data verzameld is. Dit is onder andere het geval voor mensen met fysieke beperkingen. Ook bevestigen data uit het verleden vaak historische stereotypen waar we juist vanaf willen, denk bijvoorbeeld aan bepaalde man-vrouw verhoudingen.³⁸ En last but not least: de datasets kunnen überhaupt foutieve, incorrecte data omvatten. Kortom, datasets hebben veelal een bias, met als gevolg dat de AI-applicaties die erop gebaseerd worden ook biased kunnen zijn.

Data-honger en zeggenschap over eigen data

AI-applicaties verzamelen vaak data op een ongebreidelde manier. Het wordt gebruikers moeilijk gemaakt om hun data af te schermen. Zo passen Big Tech bedrijven gemakkelijk hun gebruiksvoorwaarden aan en als individu is de gemakkelijkste weg om deze te accepteren. Je eigen data beschermen is vaak wel mogelijk, maar kost veel inspanning, zelfs met een app daarvoor.³⁹ Het is een non-keuze: je maakt gebruik van de app en levert data, of je maakt er geen gebruik van. In de zorg ontstaat druk op patiënten om mee te doen aan onderzoek en/of hun data af te staan voor onderzoek met het moreel dwingende narratief dat zij hiermee andere patiënten helpen. Zeggenschap over de eigen data is dus een belangrijk punt van aandacht.

Privacy en de optie van synthetische data

De datasets van AI-applicaties zijn vaak bijeengeschrapt van internet. Daar zitten onmiskenbaar ook persoonlijke data tussen. Er zijn gevallen bekend dat prompts die bij ChatGPT zijn ingevoerd in het model van ChatGPT zijn opgenomen, hoewel ChatGPT zelf dit tegenspreekt.^{40 41 42} Het is overigens belangrijk hierbij een onderscheid te maken tussen het rechtstreeks gebruik van de ChatGPT-app en het gebruik van het taalmodel van ChatGPT via de API. In het laatste geval geldt er een ander privacy-regime: die data worden niet gebruikt door ChatGPT en blijven slechts voor een periode van 30 dagen bewaard, zodat moederbedrijf OpenAI kan nagaan of de toepassing niet voor illegale doeleinden wordt gebruikt.¹⁸

AI-applicaties in bijvoorbeeld de gezondheidszorg of de financiële sector hebben per definitie privacygevoelige data nodig. Technieken zoals anonimiseren of pseudonimiseren van datasets zijn vaak niet afdoende om de herleidbaarheid tot een individu te voorkomen. Een recent ontwikkelde AI-applicatie lijkt hier een oplossing voor te kunnen bieden door deze privacygevoelige data om te zetten naar zogenaamde synthetische data: deze data hebben dezelfde karakteristieken als de originele data waardoor ze bruikbaar zijn voor onderzoek en applicaties, maar zijn niet herleidbaar tot personen.

Ook werkt deze app op een decentrale wijze zodat de data bij de verschillende betrokken instellingen kunnen blijven (multi-party computation). Deze applicatie is ontwikkeld door een spin-off van de TU Delft en vormt tevens een goed voorbeeld van de snelle ontwikkelingen in AI.⁴³

Auteursrechten

Vele nieuwsorganisaties en wetenschappelijke uitgeverijen hebben hun data afgesloten voor ChatGPT en vergelijkbare AI-applicaties. Zij zien het ongeautoriseerde gebruik door deze apps van hun data als inbreuk op hun auteursrechten. Ook schrijvers en kunstenaars protesteren tegen opname van hun boeken in de datasets van deze apps en zien hun auteursrechten en identiteit aangetast als deze apps teksten of beelden creëren in hun stijl. Sowieso is onze samenleving niet goed voorbereid op grote hoeveelheden door AI gegenereerde content. In het laatste anderhalf jaar blijken er bijvoorbeeld al 15 miljard afbeeldingen met AI gemaakt, een aantal waar de fotografie 50 jaar over heeft gedaan.⁴⁴

AI als Black Box

Betrouwbaarheid en hallucineren

In hoeverre zijn de uitkomsten van AI-applicaties betrouwbaar? Zijn de algoritmes gebaseerd op wetenschappelijke kennis of schuilen er misschien pseudowetenschappelijke theorieën achter? Worden correlaties wel onderscheiden van causaliteiten? Kortom, wat zit er onder de motorkap van een AI-applicatie? Dat zou transparant moeten zijn, maar die transparantie wordt bemoeilijkt doordat met name de generatieve AI-applicaties zoveel data, rekenregels en parameters omvatten, dat zelfs voor de ontwikkelaars zelf het niet duidelijk is hoe antwoorden van deze systemen tot stand zijn gekomen en waarop deze zijn gebaseerd. Van ChatGPT-3 wordt bijvoorbeeld gezegd dat deze 175 miljard parameters heeft.⁴¹ In de praktijk blijkt zelfs dat deze systemen kunnen hallucineren of (een beter woord) confabuleren: feiten en referenties worden dan verzonnen om de hiaten in een onvolledig antwoord aan te vullen, maar worden vervolgens als waarheid gepresenteerd.⁴¹

Uitlegbaarheid, bewijslast en verantwoordelijkheid

De resultaten van AI-applicaties zijn soms dus niet betrouwbaar. Daarom wordt het belang van uitlegbare AI benadrukt, waarbij de onderliggende bewijslast dient te worden meegeleverd. Bij het gebruik van AI-applicaties in een organisatie komt ook de vraag naar voren wie er verantwoordelijk is voor de resultaten ervan. De gedachte is dat als dit voor specifieke toepassingen belegd is bij specifieke functionarissen er ook een effectievere toetsing op de AI-applicatie zal plaatsvinden.

Mens-machine interactie en automation bias

Er worden drie vormen van mens-machine-interactie met AI-applicaties onderscheiden:

- De mens als onderdeel (Human in the loop). Bij beslissingsondersteunende AI-applicaties wordt de gebruiker geassisteerd bij het nemen van beslissingen op basis van onder andere data-analyse en voorspellende modellering. Het AI-systeem doet suggesties aan de gebruiker die deze al dan niet opvolgt. In principe kan dit efficiëntie en effectiviteit met zich meebrengen wanneer de voordelen van de mens en die van de computer worden gecombineerd (complementariteit van mens en computer). In de praktijk hebben gebruikers echter gauw last van 'automation bias' - het idee dat de computer objectief en neutraal is. Dit, in combinatie met menselijke gemakzucht, kan ertoe leiden dat de gebruiker het voorstel van de applicatie klakkeloos overneemt. Dit is te vatten in de metafoor van de Centaur: half-mens, half-paard, waarbij de mens de kop is met de reflectie en de computer de benen met de rekenkracht, wat zou leiden tot de gewenste complementariteit. Door automation bias wordt dit omgekeerd: de kop wordt door de computer gevormd en de benen door de mens. In dat geval wordt de bias van de AI-applicatie versterkt waardoor er minder ruimte overblijft voor de menselijke maat. Dat benadeelt van de norm afwijkende, vaak kwetsbare groepen onevenredig.
- De mens monitort (Human on the loop). In dit geval werkt het AI-systeem zelfstandig, maar wordt wel gemonitord door een mens die kan ingrijpen. Het gaat hier voornamelijk om autonoom handelende systemen bij vliegtuigen, zelfrijdende auto's e.d. die – als zij zich buiten bepaalde veiligheidsmarges bevinden – overgenomen kunnen worden door een mens.
- Autonome AI-apps (Humans out of the loop). Deze AI-systemen werken zonder tussenkomst van een mens. Een voorbeeld hiervan is Sophie, een op AI-gebaseerde avatar in de vorm van een menselijke gedaante.⁴⁵ Zij is een intelligente chatbot (die het gezicht kan krijgen van een medewerker die bij de organisatie werkt) waaraan de klant vragen kan stellen. De chatbot is gebaseerd op het LLM-model van OpenAI. Deze ontwikkeling kan gezien worden als een mogelijkheid voor een betere dienstverlening die tegelijkertijd arbeidsbesparend is.

Bedreiging van de digitale autonomie

Scheve marktmacht, afhankelijkheid van Big Tech en digitale autonomie

Het systeem achter met name generatieve AI wordt gekenmerkt door een scheve marktmacht. Voor generatieve AI zijn grote datasets en enorme rekenkracht nodig. Door de kosten die dat met zich meebrengt, zijn er maar een paar bedrijven op de wereld (m.n. in de VS) die ermee kunnen werken.³¹ In het systeem is daardoor het DNA van deze Big Tech-bedrijven terug te vinden met een focus op winstmaximalisatie. Het is overigens een bekend verschijnsel dat bij de implementatie van nieuwe systeemtechnologieën er aanvankelijk monopolies ontstaan. Het inperken van de macht van Big Tech wordt dan ook gezien als een zaak van lange adem. Ondertussen veroorzaakt deze scheve marktmacht een afhankelijkheid van Big Tech-bedrijven en bedreigt dit de digitale autonomie van overheid en publieke instellingen.

b Een recente AI ontwikkeling zou het misschien mogelijk maken om ook met kleinere modellen met minder data te trainen, daardoor minder energie te verbruiken en bereikbaarder te zijn voor kleinere bedrijven.⁴⁶

Waarvoor geoptimaliseerd?

Zoals hierboven vermeld, zijn veel AI-applicaties in handen van Big Tech-bedrijven. Door hen ontwikkelde applicaties worden vaak niet geoptimaliseerd met als doel maatschappelijke vooruitgang. Zo hebben sociale media door AI aangestuurde perverse mechanismen om de aandacht vast te houden, met als gevolg verslavingsgedrag bij gebruikers. Bij bijvoorbeeld jongeren wordt er bij 5,3% van leerlingen aan het voortgezet onderwijs problematisch gebruik van sociale media vastgesteld.⁴⁷ Sowieso zijn er zorgen over het digitaal opgroeien van de jeugd en wordt gepleit voor meer aandacht voor de veelal op AI gebaseerde mechanismen die schadelijk online gedrag in de hand werken.⁴⁸

Kwaadwillenden

Er zijn grote zorgen over AI-applicaties in handen van kwaadwillenden:

- Kwaadwillenden kunnen met AI-applicaties grote hoeveelheden desinformatie en deepfake-afbeeldingen of -video's produceren, die niet van echt te onderscheiden zijn. Daarom wordt gepleit voor een watermerk of een ander herkenningmechanisme waarmee met name door AI-gegenereerde audiovisuele materialen herkend kunnen worden.
- Kwaadwillenden kunnen met behulp van AI allerlei vormen van cybercriminaliteit bedrijven. Als voorbeeld geldt het crimineel gebruik van voice cloning, waarbij iemands stem m.b.v. AI wordt nagebootst en daarmee een familielid wordt gebeld met een dringende vraag om geld over te maken vanwege een vermeende noodsituatie.

3.4 Verantwoorde AI

Introductie

Met het streven naar verantwoorde AI wordt geprobeerd de bovengenoemde schaduwkanten van AI te minimaliseren om zo de kansen van AI te benutten. De zogenaamde ELSA-aanpak waarbij ethical, legal en societal aspecten van een AI-applicatie worden beoordeeld, is een manier om verantwoorde AI te ontwikkelen. Je zou kunnen zeggen dat onverantwoorde AI deze aspecten negeert en de mensenrechten en publieke waarden aantast. Daarom worden deze rechten en waarden eerst besproken om vervolgens in te gaan op verantwoorde AI en de ontwikkelingen daarin.

Publieke waarden, mensenrechten en de ‘dikke’ opvatting over democratie

mensenrechten

De volgende clusters van mensenrechten kunnen door het inzetten van algoritmen en/of AI in het geding komen⁴⁹:

- **Privacy:** Dit betreft het recht op respect voor het privéleven, familie- en gezinsleven. Daarmee samenhangend bestaat er een aantal wettelijke beperkingen op de verwerking van persoonsgegevens (recht op dataprotectie). In Nederland is dit vastgelegd in de AVG (Algemene Verordening Gegevensbescherming).
- **Gelijkheidsrechten/non-discriminatie:** In Nederland heeft iedereen het recht op gelijke behandeling in gelijke gevallen. Daarnaast is er een verbod op direct of indirect onderscheid op grond van godsdienst, levensovertuiging, politieke gezindheid, ras, geslacht, handicap en seksuele gerichtheid. Deze wetgeving bestaat om ervoor te zorgen dat iedereen gelijke en eerlijke toegang heeft tot de mogelijkheden in de samenleving.
- **Recht op informatie (onderdeel van de vrijheidsrechten).** Bij AI-applicaties gaat het vooral om het recht van een burger op informatie. Er worden hierin twee typen rechten onderscheiden:
 - **Passief recht om toegang tot informatie te verkrijgen:** bijvoorbeeld recht op toegang tot informatie over de totstandkoming van een overheidsbesluit. De term passief betekent dat de betreffende burger de overheid hierom moet vragen.
 - **Actief recht om informatie te ontvangen.** De term actief betekent hier dat bijvoorbeeld de overheid deze informatie actief dient te verstrekken. Het gaat dan bijvoorbeeld om de plicht om overheidsbesluiten te motiveren.
- **Procedurale rechten:** Deze rechten zorgen ervoor dat burgers bij de aantasting van hun individuele belangen bezwaar kunnen maken en dat geschillen op een effectieve en objectieve manier worden opgelost. Het gaat bijvoorbeeld om de mogelijkheid om vanuit een gelijkwaardige informatiepositie overheidsbeslissingen aan te kunnen vechten (equality of arms), het recht op hoor en wederhoor en het uitgangspunt dat iemand onschuldig is tot het tegendeel bewezen is.

Publieke waarden en de 'dikke' opvatting van democratie

De maatschappelijke vertaling van deze rechten heet publieke waarden. M.b.t. AI zijn de volgende relevant:

- Privacy en zeggenschap over persoonlijke data
- Non-discriminatie; kansengelijkheid en dus toegankelijkheid voor iedereen (inclusie)
- Transparantie en uitlegbaarheid; betrouwbaarheid
- Autonomie
- Vrijheid van meningsuiting
- Menselijke maat

Deze publieke waarden worden gezien als een set waarden die een essentieel onderdeel van de democratie vormen. Dit wordt ook wel de 'dikke' opvatting over democratie genoemd, in tegenstelling tot een 'dunne' opvatting die democratie beperkt tot periodieke verkiezingen waarbij degene met de meerderheid wint.⁵⁰

Wat is verantwoorde AI?

Verantwoorde AI (responsible AI) richt zich op de ontwikkeling en inzet van AI-applicaties die gebaseerd zijn op menselijke waarden. Op die manier kunnen de kansen van AI benut worden, terwijl voorkomen wordt dat op grote schaal mensen worden achtergesteld of de AI andere negatieve gevolgen heeft voor individu of samenleving. Dit betekent in de praktijk dat AI-applicaties voldoen aan bepaalde randvoorwaarden. E.e.a. is overigens nog in ontwikkeling en nog niet helemaal uitgekristalliseerd. Het gaat om randwoorden op de volgende gebieden:

- transparantie en betrouwbaarheid
- geen bias die mensen achterstelt
- waarborgen van privacy
- toegankelijkheid en inclusiviteit

Verantwoorde AI is ook in Nederland en Europa het uitgangspunt. Daarin onderscheidt Europa zich van de VS (met een kapitalistische benadering) en van China (met een focus op staatscontrole).⁵¹

Verantwoorde AI omvat zowel de ontwikkeling van AI-applicaties als de implementatie in de praktijk en de impact/effecten op de gebruikers c.q. de maatschappij. In een recent wetenschappelijk artikel wordt een overzicht gegeven van alle elementen van verantwoorde AI (zie tabel 2) en worden op basis daarvan ook wetenschappelijke onderzoeksvragen uitgewerkt.⁵²

c Vrijheidsrechten zijn rechten om bepaalde keuzes en handelingen onbelemmerd te kunnen uitvoeren. Voorbeelden zijn het demonstratierecht en het kiesrecht.

Dimension	Description
Fairness	AI systems should enable inclusion and diversity and not lead to discriminatory outcomes.
Transparency	AI systems should be open and transparent regarding processes and outcomes and facilitate traceability
Accountability	AI systems should be developed considering the responsibility and accountability of their outcomes with ethics and principles.
Robustness and safety	AI systems should be developed with a preventative approach to risks and in a manner that they behave as intended while minimising unintentional and unexpected harm.
Data governance	AI systems should ensure that adequate data governance covers the quality and integrity of the data used throughout the entire lifecycle.
Laws and regulations	AI systems should adhere to the respective laws and regulations that dictate their functioning.
Human Oversight	AI systems should generate tangible benefits for people and always stay under human control.
Societal and Environmental well-being	AI systems should promote sustainability

Tabel 2 Dimensies van verantwoorde AI⁵²

Groeiende ambitie om risico's te beheersen

De Autoriteit Persoonsgegevens – in Nederland coördinerend toezichthouder voor algoritmes – stelt dat er op nationaal niveau een duidelijke ambitie is om de risico's van algoritmes beter te beheersen, te beginnen bij de overheid.⁵³ Er is een toetsingskader ontwikkeld – het Impact Assessment Mensenrechten en Algoritmes.⁵⁴ Daarnaast is recentelijk een eerste versie van het implementatiekader 'Verantwoorde inzet van algoritmen' verschenen dat tot doel heeft overheden te ondersteunen met een overzicht van de belangrijkste normen en maatregelen bij de ontwikkeling en implementatie van algoritmen.⁵⁵ In dit verband is het ook belangrijk om de Handreiking non-discriminatie by design te noemen.⁵⁶ Andere maatregelen betreffen het Algoritmeregister waarin steeds meer door overheden ontwikkelde en gebruikte algoritmen die aanmerkelijke impact kunnen hebben op burgers zijn opgenomen (er waren in november 2023 209 algoritmes geregistreerd).⁵⁷

Een onderzoek naar AI-applicaties bij het Uitvoeringsinstituut Werknemersverzekeringen (UWV), de Sociale Verzekeringsbank (SVB) en het Centraal Justitieel Incassobureau (CJIB) laat zien dat overheidsinstanties bij het ontwikkelen en inzetten van algoritmes ook de risico's ervan toetsen en afwegen.⁴⁹ De algoritmes in dit onderzoek hadden verschillende doelen: werkzoekenden helpen bij gericht zoeken naar werk; gericht foutief gebruik van regelingen detecteren en een algoritme gericht op het innen van boetes en het voorkomen van oplopende schulden.

Deze algoritmes werden met name getoetst op risico's m.b.t. privacy en non-discriminatie door een ethisch adviseur of een ethische commissie aan de hand van normenkaders met ethische principes.

Toch gaat het niet altijd goed. De Autoriteit Persoonsgegevens constateert dat een verantwoorde inzet van algoritmes m.b.t. opsporen van fraude van sociale voorzieningen bij gemeentelijke organisaties nog niet mogelijk is vanwege de 'onvoldoende volwassenheid' van deze organisaties.⁵³

Ethisch instrumentarium en infrastructuur voor verantwoorde AI

Het ethisch instrumentarium en de ethische infrastructuur om de ontwikkeling en de inzet van verantwoorde AI-applicaties in overheidsorganisaties te begeleiden is sterk in ontwikkeling.⁵⁸ Qua menskracht zijn er externe audits en interne ethische commissies, maar steeds vaker nemen overheidsorganisaties ook ethici in dienst. Zo heeft bijvoorbeeld de Sociale Verzekeringsbank een Ethics Center ingericht dat zich o.a. met AI-applicaties bezighoudt. Voor de ontwikkeling van een AI-app wordt inclusive design bepleit: een aanpak waarbij vanaf het begin van de ontwikkeling álle stakeholders (dus ook de doelgroep zelf) worden betrokken en er veel aandacht is voor de datakwaliteit en het voorkomen van bias. Maar er is ook aandacht nodig voor de context waarin de AI-app zal functioneren. Daarvoor is o.m. de aanpak begeleidingsethiek ontwikkeld: een methode om AI in de praktijk op een ethisch verantwoorde manier in te zetten.⁵⁹ De gesprekspartners in dezen geven aan dat er meerdere van deze ethische frameworks ontwikkeld zijn en stellen dat sommige ervan (zoals IAMA⁵⁴) verplicht zouden moeten worden. Maar er moet wel gekeken blijven worden naar wat goede ethische instrumenten zijn voor een specifieke context. Er is een spanning tussen het willen standaardiseren van ethiek als onderdeel van reguliere processen en het voorkomen dat het een checklist wordt die zonder voldoende reflectie afgewerkt wordt. Uiteindelijk dient er een goed verhaal te liggen, waarin een aantal keuzes dat in de applicatie is gemaakt voldoende gedocumenteerd moet zijn om verantwoording voor die keuzes mogelijk te maken.

ELSA-labs ontwikkelen zich verder en breiden uit

De ELSA-labs zijn een initiatief van de Nederlandse AI Coalitie. Het doel is om in een co-creatie-proces met overheden, maatschappelijke organisaties, bedrijfsleven, onderzoeksinstituten én met betrokkenheid van de betreffende doelgroepen verantwoorde AI-applicaties te ontwikkelen die helpen een maatschappelijk probleem op te lossen. Er zijn inmiddels 22 ELSA-labs met dit label, waarvan er vijf gesubsidieerd zijn door NWO en een met subsidie van het ministerie van BZK. Andere ELSA-labs hebben óf andere financiering gevonden óf zijn nog zoekende naar financiering. In de komende jaren zal het aantal labs naar verwachting verder uitgebreid worden.

ELSA-labs richten zich op verantwoorde AI-toepassingen in verschillende domeinen. Uitgangspunt is daarbij een maatschappelijk probleem of onderwerp om vervolgens te onderzoeken wat de inzet van AI daarin kan betekenen.

Een voorbeeld van een ELSA-lab in een meer gevorderd stadium is het ELSA-lab Noord Nederland. Dit lab richt zich op de ontwikkeling en implementatie van verantwoorde AI in de zorg en neemt daarbij uitdrukkelijk het perspectief mee van mensen met een lage SES-score (lage sociaaleconomische status). Het gaat om de vraag hoe men, áls AI wordt ingezet in de zorg, ervoor kan zorgen dat inwoners met een lage SES-score daarbij betrokken worden. Hierbij spelen twee aspecten een rol: aan de ene kant inventariseert het lab wat hun vragen, behoeften en beleving is en aan de andere kant stimuleert het bewustwording van AI. Daartoe is een 'patiënt en burger-klankbordgroep' opgericht. Deze klankbordgroep is breed samengesteld en omvat mensen met een hogere en lagere SES-score en mensen met- en zonder ervaring als patiëntvertegenwoordiger. Het is de bedoeling dat het lab deze groep vijf jaar lang volgt. Naast de rol van klankbord om informatie op te halen, is het streven om deze groep ook in het project te laten participeren met een adviesrol en mogelijk ook in de rol van co-onderzoeker.

Overige ontwikkelingen rond verantwoorde AI

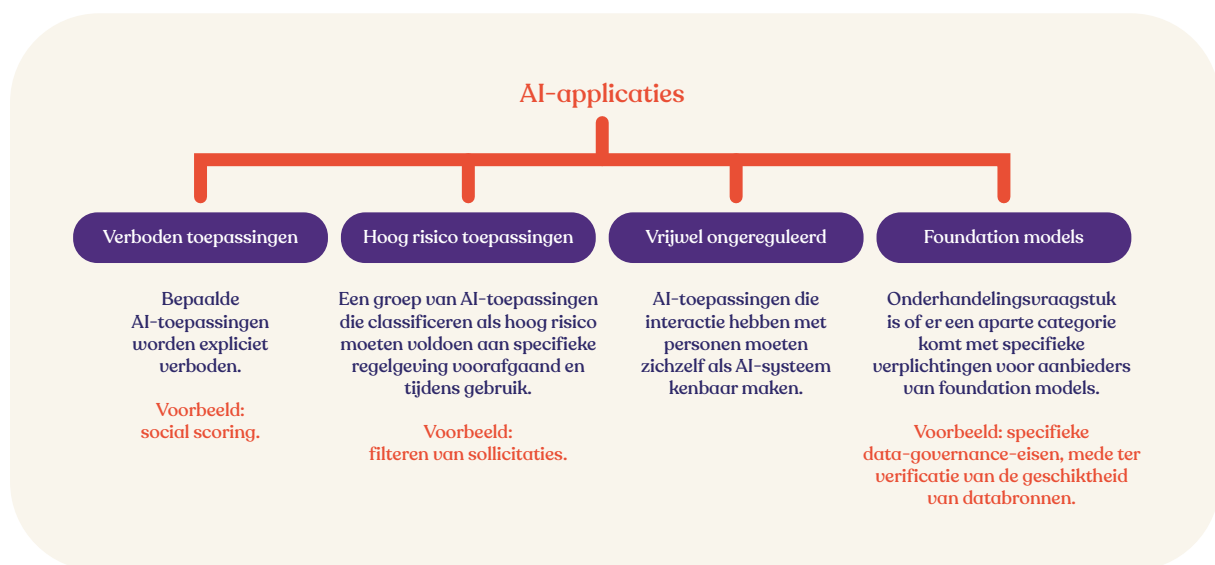
- Ontwikkeling betrouwbaarheidslabel voor gezondheidsapps: Gezondheidsapps op de consumentenmarkt kunnen de zorg, zelfzorg en preventie versterken. Zo blijkt dat apps o.a. kunnen leiden tot een gezondere levensstijl, meer ziekte-inzicht en minder last van symptomen. Maar wat is de kwaliteit en de betrouwbaarheid van deze apps? In een Europees project wordt momenteel een betrouwbaarheidslabel voor deze apps ontwikkeld met als doel vertrouwen te wekken en het gebruik ervan te stimuleren¹⁹.
- Watermerken voor AI-genererde AV-materiaal: Generatieve AI kan inmiddels foto's en video's maken die niet van echt te onderscheiden zijn. Meta is deze AI-genererde content aan het watermerken, maar dat heeft alleen zin wanneer er standaarden worden afgesproken zodat ook andere aanbieders deze praktijk overnemen. Zie het tekstkader voor een recente uitspraak van Nick Clegg hierover.

'We need standards now on common watermarking standards. We can do whatever we can do to watermark our AI-generated content, which we are doing now. If others don't, it is totally confusing to users, and we cannot police the content that comes off-platform on our platforms.'

Nick Clegg, Meta, tweet 6 oktober (transcript door MvdG van videofragment)

AI Act in de maak

In de figuur hieronder staat een kort overzicht van de verwachte regulering van AI door de AI Act van de Europese commissie. Er komt een aantal toepassingsgebieden waarin gebruik van AI verboden is, zoals in social scoring, waarbij een burger op basis van zijn/haar gedrag punten krijgt toebedeeld en op basis van het puntentotaal wel of niet bepaalde privileges krijgt. Daarnaast wordt een aantal toepassingsgebieden met hoog risico gedefinieerd waar AI-applicaties aan regels gebonden zullen worden voor wat betreft o.m. transparantie, menselijk toezicht, de juistheid van data en het monitoren van de risico's voor de veiligheid en grondrechten tijdens het gebruik. Voor overige toepassingsgebieden worden er weinig regels opgesteld, behalve dat het voor gebruikers duidelijk moet zijn of een systeem al dan niet een AI-systeem is. Wat betreft generatieve AI is het nog niet duidelijk óf en zo ja welke regels er worden opgesteld. De onderhandelingen over de AI Act worden naar verwachting eind 2023 afgerond, zodat de AI Act mogelijk in 2024 van kracht zal worden.



Figuur 1 Overzicht van de AI Act volgens de RAN⁵³

3.5 Discussie en conclusies

Kansen én risico's voor digitaal burgerschap

Uit het voorgaande is duidelijk geworden dat AI ingrijpt op alle drie te onderscheiden niveaus van digitaal burgerschap.²⁴

Op individueel niveau biedt AI kansen doordat AI-applicaties mensen helpen (persoonlijke assistent) of hen nieuwe mogelijkheden bieden om bijvoorbeeld computercode of beelden te generen. Tegelijkertijd brengt AI ook forse risico's met zich mee voor het individu, waarbij met name privacy-risico's en risico's voor de digitale autonomie opvallen. Ook op het niveau van organisaties, sectoren en de gemeenschap als geheel is de impact van AI groot: de kansen liggen vooral in 'beter', 'meer' en 'toegesneden op specifieke behoeften', terwijl de risico's zich vooral bevinden op het terrein van discriminatie en achterstelling en het verlies van digitale autonomie. Tenslotte biedt AI ook kansen op participatie in het maatschappelijk debat, maar zijn de risico's daar zo mogelijk nog groter door o.a. nepnieuws en misbruik van deepfake-technologie.

Naar verantwoorde AI

Hét antwoord op deze risico's wordt gevormd door verantwoorde AI. Eigenlijk waren alle gesprekspartners hierover eensgezind. Wel werd er verschillend gedacht over de manier waarop dit wordt vormgegeven en gereguleerd en over de vraag of het instrumentarium van verantwoorde AI al voldoende ontwikkeld is en voldoende breed gedeeld en geaccepteerd om de 'rijdende AI-trein' bij te sturen.

Effecten op de mens op langere termijn

Tenslotte wordt er ook aandacht gevraagd voor de sociaalpsychologische effecten van AI op de langere termijn. In de praktijk blijkt dat sommige mensen de computer als een maatje kunnen gaan beschouwen. Zo zijn er in Japan al mensen die liever leven met een avatar dan met een mens. E.e.a. kan leiden tot 'social de-skilling', een devaluatie van intermenselijke relaties en sociale vaardigheden. Het rapport van Rathenau hierover concludeert dan ook dat er meer kennis nodig is om de lange termijn-effecten van AI op de mens, ons mens-zijn en de mensheid te beoordelen.²

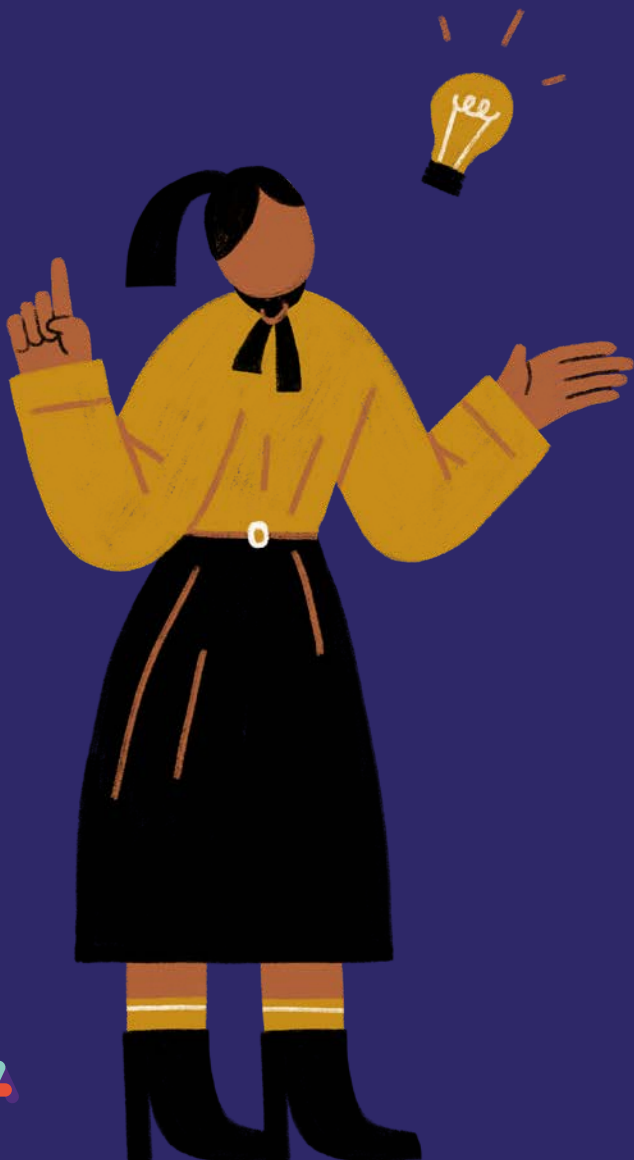
Conclusie

Dit alles leidt tot de volgende conclusies:

- De schaduwzijden van AI kunnen samengevat worden in vier punten: (1) bias/vooringenomenheid; (2) black box – gebrek aan transparantie en betrouwbaarheid van AI; (3) bedreiging van de digitale autonomie van burgers en organisaties door de macht van Big Tech-bedrijven; (4) misbruik van AI door kwaadwillenden.
- 'Onverantwoorde' AI – d.w.z. AI die niet aan de hieraan gekoppelde randvoorwaarden voldoet – brengt belangrijke risico's met zich mee voor digitaal burgerschap en kan voor meer uitsluiting zorgen.
- Verantwoorde AI legt de schaduwzijden van AI aan banden door middel van strikte randvoorwaarden. Deze randvoorwaarden zijn (1) geen bias die mensen achterstelt; (2) transparantie en betrouwbaarheid van de resultaten; (3) waarborging van de privacy; (4) toegankelijk en inclusief ontwerp.
- Aan wettelijke en ethische instrumenten wordt gewerkt. Naast de AI Act is een ethisch instrumentarium ontwikkeld en nog in ontwikkeling bestaande uit codes, instrumenten en/of benaderingen voor het stellen van ethische vragen. Daarnaast ruimen steeds meer organisaties capaciteit en menskracht in om dit ethisch instrumentarium toe te passen.
- Ondanks al deze ontwikkelingen moet worden vastgesteld dat er nog een lange weg te gaan is, waarin ADS mogelijk een rol kan spelen.

Hoofdstuk 4

Conclusies en een voorstel voor een handelingskader met vijf actierichtingen



We ronden af met vier hoofdconclusies die tegen een achtergrond van een vijftal trends en observaties moeten worden gezien. Hierop baseren we een handelingskader met vijf actierichtingen voor de Alliantie Digitaal Samenleven.

4. Conclusies en een voorstel voor een handelingskader met vijf actierichtingen

4.1 Conclusies en trends

Uit de resultaten van dit verkennende onderzoek over AI in relatie tot digitale inclusie en digitaal burgerschap kunnen de volgende vier hoofdconclusies getrokken worden:

- AI-mogelijkheden en AI-applicaties ontwikkelen zich zeer snel: De ontwikkelingen in de technologische mogelijkheden van AI en de AI-applicaties begeven zich momenteel in een stroomversnelling die naar verwachting nog enige tijd zal voortduren. Dit raakt alle lagen van de samenleving; vrijwel niemand kan zich eraan onttrekken. Dit betekent dat er een grote urgentie is om deze ontwikkelingen in goede banen te leiden.
- Verwevenheid van AI met de digitale wereld is groot: AI wordt steeds meer onderdeel van software die we dagelijks gebruiken. Het verschil tussen 'gewone' digitale applicaties en AI vervaagt hierdoor. AI wordt een – soms weinig zichtbaar – integraal onderdeel van de digitale wereld. AI is daarmee zeer relevant voor de Alliantie Digitaal Samenleven.
- AI-applicaties kunnen die digitale inclusie bevorderen: De groep mensen die moeite heeft om digitaal mee te komen is zeer divers. Denk aan mensen met lees- en schrijfproblemen, mensen met een lichamelijke of licht verstandelijke beperking, en mensen die door armoede, ziekte of andere omstandigheden geen toegang hebben tot het internet. AI heeft het potentieel om een deel van deze mensen beter toegang te geven tot de digitale wereld. Nu al zijn er op AI gebaseerde applicaties voor mensen met fysieke beperkingen en voor mensen met moeite en lezen en schrijven, die de toegang voor hen vergemakkelijken. In de nabije toekomst zijn er verdere snelle ontwikkelingen te verwachten doordat er naast lezen, typen en navigeren met de muis andere manieren van interactie mogelijk worden. Spraaktechnologie kan voor veel mensen uit deze groepen een belangrijk vorm van empowerment betekenen. Deze ontwikkelingen in de manier van interacteren met de digitale wereld geven hiermee een belangrijke kans voor méér digitale inclusie, waarop ADS kan inspelen. Tegelijkertijd moeten we erkennen dat AI-technologie niet voor iedereen in deze doelgroepen de oplossing is. Daarom moet er bij digitale dienstverlening altijd een alternatief zijn in de vorm van menselijk contact.
- Zonder de invulling van strikte randvoorwaarden voor verantwoorde AI kunnen de schaduwzijden van AI voor meer uitsluiting zorgen en negatieve impact op digitaal burgerschap hebben: De schaduwzijden van AI kunnen samengevat worden in vier punten: (1) bias/vooringenomenheid; (2) black box – gebrek aan transparantie en betrouwbaarheid van AI; (3) bedreiging van de digitale autonomie van burgers en organisaties door de macht van Big Tech bedrijven; (4) misbruik van AI door kwaadwillenden. Als gevolg hiervan kan AI voor meer uitsluiting zorgen en een negatieve impact hebben op digitaal burgerschap.

Maar de gevolgen van deze schaduwzijden kunnen aan banden gelegd worden door applicaties aan de strikte randvoorwaarden van verantwoorde AI te laten voldoen. Deze randvoorwaarden zijn (1) geen bias die mensen achterstelt; (2) transparantie en betrouwbaarheid van de resultaten; (3) waarborging van de privacy; (4) toegankelijk en inclusief ontwerp. Aan deze randvoorwaarden kan worden voldaan door bij de ontwikkeling, implementatie en het gebruik van de AI-applicaties de ethische, wettelijke en maatschappelijke aspecten te bezien. Daartoe is een ethisch instrumentarium ontwikkeld (en nog in ontwikkeling) bestaande uit codes, instrumenten en/of benaderingen voor het stellen van ethische vragen. Daarnaast ruimen steeds meer organisaties capaciteit en menskracht in om dit ethisch instrumentarium toe te passen in hun organisatie (ethische infrastructuur). Het toenemend aantal actieve ELSA-labs is een belangrijke illustratie van de ontwikkeling van verantwoorde AI. Tegelijkertijd moet worden vastgesteld dat hier nog een lange weg te gaan is, waarin ADS mogelijk een rol kan spelen.

Deze vier conclusies moeten gezien worden tegen de achtergrond van de volgende constatering en trends:

4. Kennis en bewustzijn over AI in het licht van publieke waarden is nog onvoldoende ontwikkeld: Het WRR-rapport uit 2021 riep op tot het demystificeren van AI. Die oproep is twee jaar later nog steeds relevant. De discussie rondom AI kenmerkt zich door óf overdreven positieve verwachtingen óf doemscenario's. Er is behoefte aan een nuchter en realistisch verhaal over wat AI kan betekenen, welke impact het heeft op mensen en de samenleving en welke voorwaarden nodig zijn om het in goede banen te leiden. Dit betekent dat we moeten inzetten op het vergroten van kennis, bewustzijn van AI en weerbaarheid tegen misbruik van AI, zowel in voorlichting als in educatie.
5. Digitale autonomie van publieke organisaties staat onder druk: De digitale autonomie van burgers en publieke organisaties stond al onder druk door de afhankelijkheid van grote tech-bedrijven. De razendsnelle ontwikkeling van AI-technologieën kan leiden tot een nog grotere afhankelijkheid en digitale autonomie verder aantasten.
6. Regulering loopt achter ontwikkelingen aan: Er is wet- en regelgeving in de maak om de AI-ontwikkelingen zoveel mogelijk in goede banen te leiden: de Europese commissie werkt aan de AI Act die naar verwachting in 2024 geïmplementeerd wordt. De Nederlandse overheid werkt daarop aansluitende nationale regelingen en handreikingen uit. Gelet op de snelle ontwikkelingen van AI is de verwachting echter dat de kloof tussen wat er bij wet geregeld is en hoe AI kan worden ingezet de komende jaren zal blijven groeien. Dit betekent dat er in voortdurend een maatschappelijke discussie zal zijn over wat gereguleerd dient te worden.
7. Wetenschappelijke kennis moet sneller gebruikt worden: Wetenschappelijke onderzoeksprojecten duren vaak enkele jaren, terwijl AI-technologie zich razendsnel ontwikkelt. Dit betekent dat wetenschappelijke inzichten vaak te laat komen. We moeten dus op zoek naar manieren om sneller gebruik te maken van wetenschappelijke kennis, ofwel een soort 'wetenschappelijke snelweg' ontwikkelen.
8. Kwaadwillenden en AI: In het kielzog van de ontwikkelingen van AI nemen ook nepnieuws, misbruik van deepfake-technologie en cybercriminaliteit toe. Dit maakt kritische vaardigheden belangrijker dan ooit. Daarnaast hebben we dringend behoefte aan een herkenningsmechanisme waarmee met name door AI gegenereerde audiovisuele content herkend kan worden; een soort watermerk.

4.2 Vijf actierichtingen voor een handelingskader rond AI, digitale inclusie en digitaal burgerschap

Vijf actierichtingen voorgesteld

Gebaseerd op de bevindingen van dit rapport en de inbreng uit twee groepsdialogen, stellen we de volgende met elkaar verweven actierichtingen voor. Deze kunnen dienen als leidraad voor de inspanningen van Alliantie Digitaal Samenleven (ADS):

1. Voorlichting en educatie over AI in relatie tot publieke waarden organiseren;
2. Stimuleren van verantwoorde ontwikkeling van AI-applicaties die toegang faciliteren voor mensen met beperkingen en het gebruik ervan stimuleren;
3. Bewustzijn van het belang van ieders digitale autonomie, van burgers en van organisaties, bevorderen;
4. Digitale inclusiviteit organiseren in de ontwikkeling, inzet en regulering van AI middels participatie van ervaringsdeskundigen;
5. Rol van wetenschappelijk onderzoek naar de impact en effecten van AI op digitale inclusie en digitaal burgerschap versterken.

Uitgangspunten bij het handelingskader

Inclusive design

Dit rapport heeft als doel de mogelijke effecten van AI-technologie op digitale inclusie en digitaal burgerschap te begrijpen. Bij het vervolg - het oppakken en uitwerken van activiteiten rond de genoemde actierichtingen - hanteert ADS de principes van inclusive design. Hierbij is het vanzelfsprekend dat de mensen om wie het gaat vanaf de start actief betrokken zijn en meedenken.

Een evenwichtig verhaal

Veel voor dit rapport geraadpleegde gesprekspartners pleitten voor een evenwichtig verhaal over AI, digitale inclusie en digitaal burgerschap. Dit houdt in: een focus op initiatieven rondom ontwikkeling en inzet van AI-applicaties die digitale inclusie kunnen bevorderen en die digitaal burgerschap stimuleren. Denk aan het verbeteren van AI-vaardigheden en AI-bewustzijn tot het stimuleren van actieve deelname aan de digitale samenleving. Dit alles mét verantwoorde AI die voldoet aan strikte randvoorwaarden:

- Transparant en betrouwbaar;
- Geen bias die mensen achterstelt;
- Waarborgen van privacy;
- Toegankelijk en inclusief.

Actierichting A: Organiseren van Voorlichting en educatie over AI in relatie tot publieke waarden

De Nederlandse AI Coalitie heeft reeds een nationale AI-cursus²⁷ gelanceerd, met diverse verdiepingmodules per sector.²⁸ De bibliotheken zijn actief met het programma digitaal burgerschap, de Patiëntenfederatie ontwikkelt een e-learning-programma voor patiëntenvertegenwoordigers. En er zijn ongetwijfeld nog vele andere initiatieven die niet in deze verkenning meegenomen zijn. Kortom, er is veel actie op dit gebied.

Toch is het aanbod versnipperd en blijkt uit onderzoek dat het vertrouwen in AI onder een breed publiek is gedaald en dat kennis en vaardigheden over AI achterblijven. Er is dus meer nodig, vooral op het gebied van het koppelen van kennis en inzicht over AI aan inzicht in de door AI bedreigde publieke waarden. Het evenwichtige verhaal over AI dient daarbij centraal te staan. Een aspect hiervan is digitale weerbaarheid bevorderen en het stimuleren van kritische vaardigheden met betrekking tot nepnieuws, misbruik van deepfake-technologie en cybercriminaliteit. Het is een uitdaging om dit te doen zonder mensen bang te maken en hen daarmee juist af te schrikken van deelname aan de digitale samenleving.

ADS kan helpen door de verschillende initiatieven op dit gebied met elkaar verbinden, kennisuitwisseling te bevorderen en zo deskundigheid te verhogen. Ook kan ADS helpen bij het identificeren van eventuele lacunes. Een overkoepelende landelijke voorlichtingscampagne kan daarbij een optie zijn.

Actierichting B: Stimuleren van verantwoorde ontwikkeling en gebruik van AI-applicaties die op een verantwoorde manier toegang faciliteren voor mensen met beperkingen

We hebben gezien dat er een aantal apps is ontwikkeld (en dat er nog veel meer te verwachten zijn) die andere manieren van interactie met internet mogelijk maken. Daardoor zal de toegang tot de digitale wereld voor veel mensen met moeite met lezen en schrijven of met fysieke beperkingen vergemakkelijkt worden.

Niettemin zijn hieraan een aantal cruciale vragen verbonden: Welke van deze toegankelijkheid bevorderende apps voldoen aan de randvoorwaarden van verantwoorde AI? Wat is het business model van deze apps – moeten gebruikers betalen na een gratis proefperiode of betalen ze door hun gegevens af te staan? En zijn deze apps daadwerkelijk nuttig in het dagelijks leven van mensen met beperkte digitale vaardigheden of leiden ze tot andere problemen?

ADS zou hierin een rol kunnen oppakken. Bijvoorbeeld door wetenschappelijk onderzoek naar de impact van deze apps op de gebruikers te stimuleren/initiëren. Of door zich hard te maken voor certificering van apps die voldoen aan de randvoorwaarden van verantwoorde AI. Ook kan ADS de ontwikkeling van niet-commerciële businessmodellen voor deze apps ondersteunen.

Actierichting C: bevorderen van Bewustzijn van het belang van ieders Digitale autonomie, van zowel burgers als organisaties

Uit het rapport blijkt dat zowel burgers als publieke organisaties moeite hebben om hun digitale autonomie te behouden. Zo worden patiënten onder morele druk gezet om hun medische gegevens af te staan bij het gebruik van medische AI-applicaties en wordt AI al dan niet zichtbaar toegevoegd aan software zonder dat burgers en publieke instellingen daarover de regie hebben.

Voor het waarborgen van de digitale autonomie van publieke instellingen is in de interviews een aantal opties naar voren gekomen. Het gaat om het ontwikkelen van inkoopvoorwaarden voor AI-applicaties met oog voor publieke waarden. De inkoop kan per sector collectief georganiseerd worden om zo meer onderhandelingskracht te hebben t.o.v. de technologieleveranciers (zie ook digitale spoelkeuken⁶⁰). Een meer vergaande stap daarin is om de afhankelijkheid van de Large Language Modellen van Big Tech te vermijden door de ontwikkeling van een publiek gefinancierd taalmodel voor Nederlandse publieke instellingen: dit is inmiddels opgepakt door de start van het project GPT-NL, dat een eigen open taalmodel gaat ontwikkelen met het oog op het behouden van digitale autonomie.⁶¹ De digitale autonomie van de burger waarborgen kan bijvoorbeeld door inzet van het door de Patiëntenfederatie gewenste zeggenschapsregister voor medische gegevens en de ontwikkeling van de Europese ID-wallet.⁶²

ADS kan de digitale autonomie van publieke organisaties stimuleren en een sector-overkoepelende rol spelen door middel van kennisuitwisseling en deskundigheidsbevordering. Dit geldt ook voor de digitale autonomie van burgers, waarbij ADS aanvullend de inbreng van ervaringsdeskundigen kan verzorgen.

Actierichting D: organiseren van Digitaal inclusieve regulering van AI middels Participatie van ervaringsdeskundigen

De randvoorwaarden van verantwoorde AI zijn in dit rapport gekarakteriseerd als essentieel en noodzakelijk. Indien deze randvoorwaarden niet adequaat ingevuld zijn, kunnen AI-applicaties juist digitale exclusie tot gevolg hebben. Het invullen van deze randvoorwaarden voor verantwoorde AI moet gebeuren door een voortdurende aandacht voor de ethische, wettelijke en maatschappelijke aspecten, ook wel bekend als 'ELSA'. De ethische infrastructuur hiervoor – instrumenten om deze randvoorwaarden in te vullen én capaciteit om dit in de praktijk te implementeren – is nog sterk in ontwikkeling en niet verplichtend. Zo is recent het instrument Impact Assessment voor Mensenrechten (IAMA) ontwikkeld, maar nog niet verplicht gesteld.

Wat betreft wet- en regelgeving, is de verwachting dat de Europese AI Act in de loop van 2024 van kracht zal worden (ook in de VS zijn regelingen in de maak⁶³). Dit zal betekenen dat de AI Act zijn doorwerking zal hebben in standaarden, van de wet afgeleide regelgeving, etc. terwijl de ontwikkelingen in AI onverminderd doorgaan.

Dit alles – samen met inzichten uit de zich ontwikkelende ethische infrastructuur – betekent een grote dynamiek in wet- en regelgeving en ethische standaarden rondom AI in de komende jaren. ADS kan hierin het belang van digitale inclusie en digitaal burgerschap vertegenwoordigen door partners en ervaringsdeskundigen in te brengen en hen mee te laten denken binnen de verschillende verbanden.

Actierichting E: Versterken van de Rol van wetenschappelijk onderzoek naar de impact en effecten van AI op digitale inclusie en digitaal burgerschap

Meerdere gesprekspartners hebben gepleit voor meer wetenschappelijk onderzoek naar het gebruiken van AI-applicaties door de verschillende groepen mensen met beperkte digitale vaardigheden en de impact ervan. Daarnaast is het van belang om AI-datasets representatiever te maken door het toevoegen van data over deze groepen. Dit betekent dat er wetenschappelijk verantwoorde datasets over bijvoorbeeld mensen met moeite met lezen en schrijven of met fysieke beperkingen gecreëerd dienen te worden ten behoeve van verantwoorde AI-applicaties.

Daarnaast wordt er gepleit om wetenschappers intensief bij beleidsontwikkeling over AI te betrekken. Traditionele vormen van wetenschappelijk onderzoek kunnen de ontwikkelingen rondom AI momenteel niet bijhouden. Daarom werd gesuggereerd om een 'wetenschappelijke snelweg' te ontwikkelen. Hoe dit zou moeten worden vormgegeven is niet uitgewerkt. Ook de Rijksoverheid zoekt naar manieren om wetenschappelijke kennis en inzichten te betrekken bij het duiden van en reageren op ontwikkelingen in AI. Zo wil de minister van OCW voor AI een rapid response team van wetenschappers inrichten om hun wetenschappelijke inzichten direct te gebruiken bij de ontwikkeling van AI-beleid (zie tekstkader).⁶⁴

ADS zou enerzijds een rol kunnen spelen in het verbinden van wetenschappers met partijen die AI-applicaties ontwikkelen of willen gebruiken en anderzijds in het initiëren of stimuleren van wetenschappelijk onderzoek (al dan niet via een 'snelweg'-methode) naar de impact van AI-applicaties op het leven van mensen die niet of onvoldoende digitaal mee kunnen komen.

“Een ander veelkoppig vraagstuk is AI. Daarbij volgen de ontwikkelingen zich in een razendsnel tempo op. Samen met collega's binnen het kabinet kijk ik hoe we de laatste wetenschappelijke inzichten het beste structureel kunnen benutten in het AI-beleid. Zo zijn we aan het bezien of we een rapid response team kunnen oprichten waarbij overheid, wetenschap, bedrijfsleven en maatschappelijke organisaties op een innovatieve en flexibele wijze in gesprek kunnen komen, dit in lijn met de razendsnelle ontwikkelingen op dit terrein.”

[uit de ROB lezing 2023 van de Minister van OCW].

Totstandkoming van dit rapport

Ten behoeve van dit Trendrapport werden 23 experts geïnterviewd (zie tabel hieronder). Van alle interviews werd een samenvatting gemaakt die voor feedback werd voorgelegd aan de geïnterviewden, m.u.v. een respondent, die aangaf hieraan geen behoefte te hebben. Daarnaast werd aanvullende deskresearch gedaan. Dit alles werd uitgevoerd door Maurits van der Graaf. Voor de twee dialoogsessies in oktober werd een tussenrapport opgesteld en toegezonden aan de deelnemers ervan (zie tabel hieronder). Femke Hammer heeft de dialogen georganiseerd en samen met John Leek begeleid. De resultaten van de dialogen werden als input gebruikt bij paragraaf 6.2 over de actierichtingen. Het eindrapport is gereviewd door Valerie Frissen [Bijzonder Hoogleraar Digitale Technologie en Sociale Verandering en Directeur SIDN fonds], Marcel Broersma [hoogleraar Media and Journalism Studies, en voorzitter Digital Literacy Coalition Groningen], John Leek [voorzitter Stuurgroep Alliantie Digitaal Samenleven] en Femke Hammer [aanjager digitaal bewustzijn Alliantie Digitaal Samenleven].



Graag citeren als: Van der Graaf, M.; Effecten van AI op Digitale Inclusie en Digitaal Burgerschap; Trendrapport Alliantie Digitaal Samenleven, 2023.

Geïnterviewden

Naam gesprekspartner	Organisatie
Alex Corra	Sociale Verzekeringsbank
Alexander Smit	Rijksuniversiteit Groningen
Anke van Dam	Pharos
Dirk van Roode	NLdigital
Duuk Baaten	SURF
Edo Haveman	Meta
Eric Velleman	HAN University of Applied Sciences
Erik Schrijvers	Wetenschappelijke Raad voor het Regeringsbeleid
Hester Benedictus	Leessimpel
Ildiko Vajda	Patiëntenfederatie
Jeroen Vonk	Novum - innovatielab Sociale Verzekeringsbank
Karien Sondervan	Cybersoek
Linda Kool	Rathenau Instituut
Lotte Dondorp; Harriet Leget; Eva Leurink	Kennisnet
Marjolein Oomes	Koninklijke Bibliotheek
Martijn Bekking	Nederlandse AI Coalitie
Mellouki Cadat-Lampe	Movisie
Mirjam Plantinga	ELSA-lab Noord Nederland
Nitesh Bharosa	Digicampus - TU Delft
Tim Groenenboom	Stichting Lezen en Schrijven
Victor Zuydweg	ICTU / Gebruiker Centraal
Wouter Bolier	Ieder(in)

Deelnemers AI-dialogen

Naam gesprekspartner	Organisatie
Alex Corra	Sociale Verzekeringsbank
Alex van der Zeeuw	Universiteit Twente
Anke van Dam	Pharos
Barbara Bufkens	Alliantie Digitaal Samenleven
Eric Velleman	HAN
Erik Schrijvers	Wetenschappelijke Raad voor het Regeringsbeleid
Femke Hammer	Alliantie Digitaal Samenleven
Hester Benedictus	Leessimpel
Ildikó Vajda	Patiëntenfederatie
Joeri Hazelaar	Filosofie in actie
John Leek	Alliantie Digitaal Samenleven
Josanne Ganzevles	Alliantie Digitaal Samenleven
Marianne Schaaphok	TNO
Martijn Bekking	NLAIC
Olaf Simonse	ELSA-lab Heerlen, Brightlands
Réka Deuten	Alliantie Digitaal Samenleven
Sanne Giessen	ICTU / Gebruiker Centraal
Sophie van Eck	NLAIC
Tim Groenenboom	Stichting Lezen en Schrijven
Victor Zuydweg	ICTU / Gebruiker Centraal
Wietske Kamsma	Alliantie Digitaal Samenleven

Bronnen

1. Prins, C. et al. Opgave AI. De nieuwe systeemtechnologie. <https://www.wrr.nl/publicaties/rapporten/2021/11/11/opgave-ai-de-nieuwe-systeemtechnologie> (2021).
2. Bronzwaer, S. 'De ontwikkelingen in AI gaan sneller dan het tempo waarin de overheid werkt'. NRC (2023).
3. Klerk, M. Mensen met lichamelijke of verstandelijke beperkingen (factsheet). (2012).
4. Groot, A., Wesseling, W., van Drie, E. & van Uden, D. Literatuuronderzoek. (2019).
5. de Vries, D. A., Piotrowski, J. T. & de Vreese, C. H. Resultaten Onderzoek Digitale Competenties (DIGCOM). <https://www.uva.nl/shared-content/faculteiten/nl/faculteit-der-maatschappij-en-gedragswetenschappen/nieuws/2022/05/hoe-digitaal-vaardig-zijn-wij-nederlanders.html> (2022).
6. Van Praat, F. Algoritme Vertrouwensmonitor 2022. <https://kpmg.com/nl/nl/home/topics/digital-transformation/artificial-intelligence/vertrouwen-in-algoritmes/algoritme-vertrouwensmonitor.html> (2022).
7. Hamer, J., Lemmens, A. & Kool, L. Algoritmes afwegen. <https://www.rathenau.nl/nl/digitalisering/algoritmes-afwegen> (2022).
8. Biller-Andorno, N. et al. Opinion on democracy in the digital age. (2023) doi:10.2777/0787819.
9. Fowler, G. Delete your digital history from dozens of companies with this app. Washington Post (2023).
10. AP vraagt om opheldering over ChatGPT | Autoriteit Persoonsgegevens. <https://www.autoriteitpersoonsgegevens.nl/actueel/ap-vraagt-om-opheldering-over-chatgpt>.
11. Generative AI: Implications for Trust and Governance. https://aiverifyfoundation.sg/downloads/Discussion_Paper.pdf (2023).
12. How your data is used to improve model performance | OpenAI Help Center. <https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model-performance>.
13. Leessimpel. <https://www.leessimpel.nl/>.
14. BlueGen.ai | AI-based synthetic data generation platform. BlueGen AI <https://bluegen.ai/>.
15. AI Image Statistics: How Much Content Was Created by AI. <https://journal.everyapixel.com/ai-image-statistics> (2023).
16. Manifest Mensgerichte Artificiele Intelligentie. (2023).
17. I AM Sophie. <https://d3mdyry31n8x9i.cloudfront.net/>.
18. Chui, M. et al. Economic potential of generative AI. <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier> (2023).
19. Cijfers over mediagebruik | Nederlands Jeugdinstituut. <https://www.nji.nl/cijfers/mediagebruik>.
20. Autoriteit Persoonsgegevens; Directie Coördinatie Algoritmes (DCA). Rapportage Algoritmerisico's Nederland | Autoriteit Persoonsgegevens. <https://www.autoriteitpersoonsgegevens.nl/documenten/rapportage-algoritmerisicos-nederland> (2023).
21. Zaken, M. van A. Impact Assessment Mensenrechten en Algoritmes - Rapport - Rijksoverheid.nl. <https://www.rijksoverheid.nl/documenten/rapporten/2021/02/25/impact-assessment-mensenrechten-en-algoritmes> (2021).

22. Koninkrijksrelaties, M. van B. Z. en. Implementatiekader 'Verantwoorde inzet van algoritmen' - Rapport - Rijksoverheid.nl. <https://www.rijksoverheid.nl/documenten/rapporten/2023/06/30/implementatiekader-verantwoorde-inzet-van-algoritmen> (2023).
23. Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. Handreiking non-discriminatie by design - Rapport - Rijksoverheid.nl. <https://www.rijksoverheid.nl/documenten/rapporten/2021/06/10/handreiking-non-discriminatie-by-design> (2021) doi:10/handreiking-non-discriminatie-by-design.
24. Ministerie van Binnenlandse Zaken en Koninkrijksrelaties. Kamerbrief algoritmen en AI - Kamerstuk - Rijksoverheid.nl. (2023) doi:10/kamerbrief-over-algoritmen-reguleren.
25. Werkgroep Mensgerichte AI. Ethiek en AI - Zeven methoden in theorie en praktijk. <https://nlaic.com/nieuws/publicatie-ethiek-en-ai-zeven-methoden-in-theorie-en-praktijk/> (2022).
26. Home - Begeleidingsethiek. <https://begeleidingsethiek.nl/>.
27. Lohr, S. What Ever Happened to IBM's Watson? The New York Times (2021).
28. Patiëntenfederatie Nederland. Zeggenschap over gezondheidsgegevens bij secundari gebruik van data. (2023).
29. Label2Enable. NeLL <https://nell.eu/projecten/labelenabel>.
30. Onderwijsraad. Inzet van intelligente technologie - Advies. <https://www.onderwijsraad.nl/publicaties/adviezen/2022/09/28/inzet-van-intelligente-technologie> (2022).
31. Bok, C., 't Veld, I., Bomas, E., Dondorp, L. & Pijpers, R. WaardenWijzer voor digitalisernig in het onderwijs.
32. Microsoft. Leerversnellers: hulpmiddelen voor studenten. Microsoft Education <https://eduv2.msftedu.com/home/index?page=ROW/learning-tools/learning-accelerators&lang=nl-nl&var=3292a07e-a0b5-4b23-bb52-64ce063d1cb7>.
33. Walker, J. & Baten, D. Promises of AI in Education. <https://www.surf.nl/en/promises-of-ai-in-education> (2022).
34. Das, D., de Jong, R. & Kool, L. Werken op waarde geschat. (2020).
35. Ethische Code Artificial Intelligence: gedragscode NLdigital. NLdigital <https://www.nldigital.nl/ethischecodeai/>.
36. Meta and Microsoft Introduce the Next Generation of Llama. Meta <https://about.fb.com/news/2023/07/llama-2/> (2023).
37. ImageBind: Holistic AI learning across six modalities. <https://ai.meta.com/blog/imagebind-six-modalities-binding-ai/>.
38. Ministerie van Onderwijs, C. en W. De ROB-lezing 2023 door minister Robbert Dijkgraaf - Toespraak - Rijksoverheid.nl. <https://www.rijksoverheid.nl/documenten/toespraken/2023/09/13/rob-lezing-door-robbert-dijkgraaf> (2023).
39. De Nationale AI-Cursus. De Nationale AI-Cursus <https://www.ai-cursus.nl/>.
40. Nationale AI-Cursussen. Nederlandse AI Coalitie <https://nlaic.com/bouwstenen/human-capital/nationale-ai-cursussen/>.
41. De Digitale Spoelkeuken. PublicSpaces <https://publicspaces.net/de-digitale-spoelkeuken/>.
42. Bakker, C. De Europese id-wallet: hoe, wat en waarom? Digitale Overheid <https://www.digitaleoverheid.nl/achtergrondartikelen/de-europese-id-wallet-hoe-wat-en-waarom/> (2022).



Alliantie
Digitaal Samenleven