

# Maatschappelijke input thematische focus NWA- ORC 2024

Dit rapport beschrijft de maatschappelijke input voor de  
thematische focus van de NWA-ORC call 2024 voor de route  
Waardecreatie door verantwoorde Artificial Intelligence en Big  
Data

Jikke van den Ende  
Tessel van Leeuwen  
Simone van der Burg

31 augustus 2023

waag  futurelab

# Inhoud

- Inleiding
  - Route Waardecreatie door verantwoorde AI & Big Data
  - Achtergrondinformatie (NWA)
  - Onderzoek 2023: aanleiding (NWA-ORC) & doel
- Onderzoek 2022: een verkenning van het veranderde speelveld sinds 2015
- Onderzoek 2023: methoden
  - Onderzoeksmethode: citizen panel
  - Onderzoeksmethode: online enquête
- Resultaten:
  - Citizen Panels & 12 geïdentificeerde thema's
  - Online enquête
- Conclusie, discussie en aanbevelingen

# Inleiding

Niemand kan het zijn ontgaan: de ontwikkeling van Artificial Intelligence (AI) is in een stroomversnelling gekomen. In allerlei sectoren wordt technologie ontwikkeld die patronen herkent in volumineuze en zeer diverse informatiestromen en ook in staat is zichzelf te verbeteren. Dit wordt ontwikkeld voor een veelheid aan toepassingen, variërend van het automatisch detecteren van (cyber-)criminaliteit, het selecteren van sollicitanten op , het voorspellen van erupties van vulkanen, tot het verbeteren van methodes voor preventie, diagnose en behandeling van ziekten of verduurzamen van land- en tuinbouw. AI maakt niet alleen gebruik van veel data, maar ook van nieuwe combinaties van data en stromen van gegevens. Ook kan AI 'leren' van eerdere (goede of juist foute) prestaties. Dit alles leidt tot nieuwe vragen over infrastructuur en technologie, maar ook over de samenleving die steeds vaker met AI te maken krijgt. Denk bijvoorbeeld aan recente debatten in de krant over studenten die hun werkstuk laten schrijven door ChatGPT of Siri die in een handomdraai al je vragen beantwoordt, maar tegelijk stiekem meeluistert op het moment dat je daar als gebruiker niet om vraagt.

Binnen de route *Waardecreatie door verantwoorde Artificial Intelligence en Big Data* van de Nationale Wetenschapsagenda (NWA) worden deze kwesties geadresseerd. De NWA is een maatschappelijk onderzoeksprogramma dat wordt uitgevoerd door de Nederlandse Organisatie voor Wetenschappelijk onderzoek (NWO). De agenda is in 2015 in opdracht van het ministerie van Onderwijs, Cultuur en Wetenschap opgesteld met als belangrijkste doel de wetenschap en de samenleving dichter bij elkaar te brengen. De overtuiging die hieraan ten grondslag ligt is dat complexe maatschappelijke vraagstukken om gedegen wetenschappelijke inzichten vragen en dat de kracht van wetenschappers, beleidsmakers, scholen, bedrijven én bewoners gebundeld zou moeten worden. Bij het opstellen van de agenda is er bij een breed publiek vragen aan de wetenschap opgehaald. Deze vragen zijn vervolgens ondergebracht in 140 clustervraagstukken verdeeld over 25 verschillende routes die samen het fundament vormen van het onderzoek en de innovatie in Nederland.

De route *Waardecreatie door verantwoorde Artificial Intelligence en Big Data* richt zich op de maatschappelijke betekenis en impact van AI en de grote diversiteit aan data in onze samenleving. In de toekomst neemt het vermogen van AI toe door wetenschappelijk onderzoek, maar er moet op worden toegezien dat die ontwikkelingen betrouwbaar en verantwoord blijven. De verbeterde analysetechnieken en algoritmes moeten in staat zijn om een grote hoeveelheid en diversiteit aan data op een betrouwbare manier te verwerken. Deze data moet uitwisselbaar zijn en moet kunnen worden geanalyseerd zonder dat gevoelige informatie naar buiten lekt. Daarnaast is het belangrijk dat waarden van burgers geadresseerd worden in het ontwerp van AI, want dat biedt de grootste kans dat de ontwikkelde technologie uiteindelijk ook wordt geaccepteerd en gebruikt. De betrokkenheid van de gebruikers van AI staat daarin centraal. Een beschrijving van de ontwikkelingen binnen de route en een toelichting op het ontstaan, de functie en de organisatiestructuur van alle routes wordt in een video weergegeven op de overzichtspagina van de route.

Het routemanagement is in handen van boegbeelden Tibor Bosse van de Radboud Universiteit en Freek Bomhof van TNO. Zij dragen zorg voor de ontwikkeling van de route, de coördinatie van het netwerk van onderzoekers daar omheen, de verbinding met de andere routes en het uitvoeren van activiteiten die bijdragen aan de doelstellingen van de agenda als geheel. Binnen de route bestaan verschillende programmalijnen waarvoor financiering vanuit NWO beschikbaar wordt gemaakt voor onderzoekers. De programmalijn Onderzoek op Routes door Consortia (ORC) is daar een van. Deze lijn is in 2018 opgezet en is bedoeld voor grootschalig en langdurig onderzoek door multidisciplinaire consortia.

In 2022 is er een evaluatierapport uitgebracht over het volledige NWA-programma door een onafhankelijke evaluatiecommissie. Deze evaluatie liet onder meer zien dat er ruimte is voor verdere versteviging van de verbinding van de agenda met de maatschappij. Daarnaast was een van de aanbevelingen om duidelijker op thema's te profileren, prioriteren en agenderen. Daarom heeft de NWO besloten dat de programmalijn ORC met ingang van de ronde 2023 meer focus krijgt. Elke route krijgt nu elke twee jaar de kans om één thematische keuze aan de NWO voor te leggen dat kan rekenen op zowel wetenschappelijk als maatschappelijk draagvlak. Vervolgens wordt er door een programmacommissie besloten welke van deze thema's geprogrammeerd zullen worden in de komende NWA-ORC ronde. Voor de route *Waardecreatie door verantwoorde AI en Big Data* is er gekozen om voor de ronde in 2024 een thema voor te dragen.

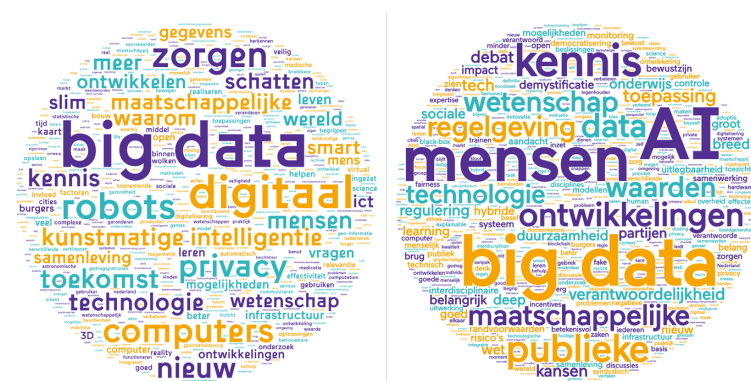
Voor het vaststellen van de thematische focus voor deze NWA-ORC ronde kreeg Waag Futurelab in 2023 de opdracht om een methode te ontwerpen en uit te voeren voor het ophalen van maatschappelijke thema's op het gebied van AI en Big Data. De vragen uit 2015 vanuit de samenleving waar de route in eerste instantie op gebaseerd is waren voornamelijk gericht op Big Data. De stormachtige ontwikkelingen op het gebied van AI maken dat het belangrijk is de maatschappelijke relevantie van deze route opnieuw uit te vragen. Daarnaast hebben de 16 clustervragen die als fundament voor de route dienen een breed bereik: van vraagstukken over klimaat, de ruimte, de gezondheidszorg tot vraagstukken over bestuur en de rechtstaat. Dat de vragen zo uiteenlopen is niet zo gek gezien de impact van AI op vele facetten van de wetenschap, economie en maatschappij. Maar het zorgt wel dat een thematische focus aanbrengen voor de ORC-programmalijn een complexe opgave is. Met als vertrekpunt de evaluatie van de route die Waag in 2022 al uitvoerde is er dit jaar voor gekozen om het brede publiek mee te laten denken over welke richting het toekomstig onderzoek binnen de route op zou moeten gaan. Door burgers mee te laten denken over zowel kansen als uitdagingen voor de ontwikkelingen in AI en Big Data heeft Waag in co-creatie met geïnteresseerde leden uit de samenleving onderzocht welke thema's er spelen en zo de maatschappelijke relevantie van de thematische focus van de NWA-ORC gedefinieerd.

# Onderzoek 2022: een verkenning van het veranderde speelveld sinds 2015

Sinds het ontstaan van de NWA en de route rond Big Data in 2015 hebben er stormachtige ontwikkelingen op het gebied van AI plaatsgevonden. Juist omdat die ontwikkelingen zo snel zijn gegaan is er in 2022 door Waag Futurelab gekeken naar de maatschappelijke vraagstukken waar de route in 2015 op was gebaseerd en in hoeverre er in onze huidige samenleving nu andere thema's spelen. De vraag van het onderzoek was: waarop zou de route zich in de toekomst op moeten gaan richten? Daarbij was het van belang dat betrokkenen invloed zouden krijgen op de thematiek van de route.

Om een beeld te krijgen van de ontwikkelingen sinds 2015 en het gesprek tussen wetenschappers en verschillende niet-wetenschappelijke partijen te faciliteren heeft Waag de oorspronkelijke vragen uit 2015 onderzocht en twee co-creatie sessies georganiseerd. Samen met wetenschappers, beleidsmakers en ontwikkelaars op het gebied van Big Data en AI zijn we in gesprek gegaan over zowel wetenschappelijke als maatschappelijke ontwikkelingen en hebben we een eerste blik geworpen op een mogelijke toekomst voor de route. Tijdens deze sessies kwamen er drie thema's naar voren: de ontwikkelingen op het gebied van *deep learning*, de regulering van AI en de interactie tussen menselijke en kunstmatige intelligentie. Deze drie thema's zijn meegenomen in de herformulering van de route in 2023. Daarnaast vormde de gesprekken een startpunt voor het aanpassen van de naam van de route, door het toevoegen van Artificial Intelligence aan Big Data.

Op basis van zowel de vragen uit 2015 waarop de route gebaseerd is als de gedocumenteerde gesprekken tijdens de co-creatiesessies zijn er twee *wordclouds* gemaakt die weergeven welke onderwerpen voornamelijk ter sprake kwamen. De verschillen tussen deze twee geven een kijk op het veranderde speelveld sinds 2015. Waar er in 2015 een focus lag op robots, digitalisering, computers en technologie zien we bij de huidige gesprekken meer aandacht voor maatschappij, publieke waarde en regelgeving. Het onderzoek laat zien dat er in de huidige tijd meer aandacht is voor de maatschappelijke implicaties, waar de route zich in de toekomst meer toe kan verhouden. Alle resultaten van dit onderzoek zijn terug te vinden in het eindverslag op de projectpagina: <https://waag.org/nl/project/evaluatie-nwa-route-big-data/>.



Figuur 1. Links: wordcloud 2015, rechts: wordcloud 2022

# Onderzoek 2023: methoden

Met als aanleiding de benodigde thematische focus voor de NWA-ORC ronde van 2024 kijkt Waag Futurelab in 2023 opnieuw samen met de samenleving naar de route die nu *Waardecreatie door verantwoorde AI en Big Data* heet. Voor het uitvragen van de maatschappelijk relevante thema's gerelateerd aan deze route heeft Waag een werkvorm ontwikkeld waarbij het brede publiek wordt gevraagd om mee te denken over toekomstig onderzoek binnen de route. Dit keer wilden we naast betrokken wetenschappers, beleidsmakers en ontwikkelaars meer leden van de samenleving betrekken die niet vanuit hun expertise al met het onderwerp te maken hebben. Het is de bedoeling dat de NWA aansluit bij de interesses en zorgen vanuit de samenleving. Daarom is er gekozen voor een twee-sporen methodiek: twee citizen panels gevolgd door een online enquête. De werkvorm van het citizen panel is ontwikkeld op basis van bestaande literatuur met een soortgelijk doel, namelijk het betrekken van burgers bij het vormen van een onderzoeksagenda. De resultaten van de georganiseerde panels, 12 geïdentificeerde thema's, zijn vervolgens middels een online enquête verspreid onder verschillende maatschappelijke organisaties en het bredere publiek. De resultaten van deze enquête hebben geleid tot een prioritering van maatschappelijke thema's waar de NWA-ORC call zich op zou moeten focussen en een overzicht van mogelijke vraagstukken voor toekomstig onderzoek.

## Onderzoekmethode: Citizen Panel

De vormgeving van het citizen panel is gebaseerd op de nationale toekomstvisie workshop van CIMULACT – een Europees project wat liep tussen 2015 en 2018 (European Commission, 2015). Het doel van CIMULACT was het vergroten van de relevantie en adequaatheid van de onderzoeks- en innovatieagenda's ontwikkeld door de Europese commissie binnen het Horizon 2020 programma. Het project was gebaseerd op het idee dat onderzoek moet aansluiten bij collectieve maatschappelijke intelligentie. Burgers, wetenschappers, maatschappelijk initiatief en politici voeren met elkaar dialoog om oplossingen aan te dragen voor een gewenste en duurzame toekomst. De betrokkenheid van burgers zorgde ervoor dat visies, verwachtingen en zorgen vanuit de samenleving opgehaald werden die vervolgens vertaald konden worden naar politieke of wetenschappelijke adviezen en voorstellen. Onderzoek dat hierop aansluit is dus responsief naar maatschappelijke waarden en zorgen en heeft een grotere kans om door de samenleving te worden geaccepteerd en gewaardeerd, Het draagt bij aan verantwoord onderzoek en innovatie binnen de EU.

Voor het vormgeven van de sessies heeft Waag Futurelab gekozen geen workshop van een hele dag te organiseren, zoals dat binnen CIMULACT het geval was. De gedachte was dat het voor mensen moeilijk zou zijn om een hele dag vrij te maken voor deelname aan een panel. Daarnaast hadden we te maken met een strakke deadline, vanwege het aanleveren van de thematische focus begin september. Daarom is de werkvorm omgeschreven naar een interactieve sessie van 2 uur. Daarnaast is ervoor gekozen om de sessie in te leiden met een introductie van het onderwerp – een toegankelijke presentatie over de huidige ontwikkelingen binnen het onderzoek naar AI, voorbeelden van toepassingen en een

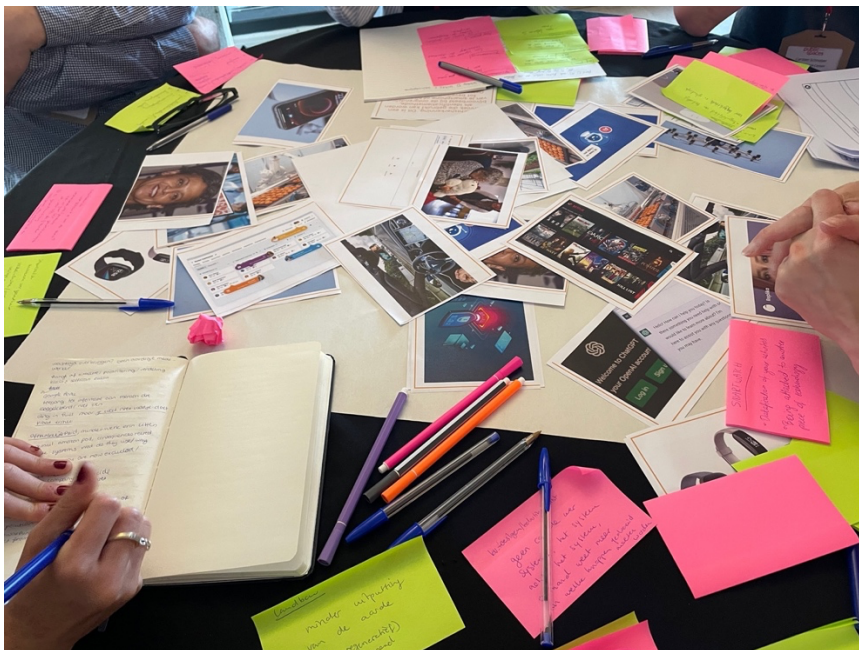
overzicht van enkele mogelijke (maatschappelijke) implicaties. De stappen van de werkvorm staan beschreven in Tabel 1.

Stap	Omschrijving	Resultaat
Inleiding en overzicht	Welkom & kennismaking Waag geeft introductie en een overzicht van het programma	Deelnemers doorzien belang onderzoek, intentie en impact
Introductie van het onderwerp	Introductie: huidige toepassingen, ontwikkelingen en (maatschappelijke) implicaties van AI aan de hand van enkele voorbeelden.	Common ground onder deelnemers
Betekenisgeving	Een onderdeel dat dient als ijsbreker: er liggen voorbeelden van toepassingen van AI op tafel en deelnemers geven welke voorbeelden hen aanspreekt en welke betekenis dat voorbeeld voor hen heeft.	Elke deelnemer heeft het woord gehad, men kent achtergrond anderen, kennismaking
Positieve en negatieve consequenties	Deelnemers wordt gevraagd drie positieve en drie negatieve consequenties van de voorbeelden van AI die op tafel liggen op te schrijven. Dit kunnen consequenties zijn voor jezelf, je naasten of de samenleving als geheel.	Persoonlijke positieve en negatieve consequenties van AI toepassingen
Inzichten bespreken	Gezamenlijk worden de positieve en negatieve consequenties besproken. Er wordt gekeken of er consequenties overeenkomen of juist verschillen. Clusters worden aangebracht: eerste thema's kunnen zo geïdentificeerd worden.	Overzicht van kansen en zorgen
Prioriteren	Het overzicht wordt bekeken en controleert op missende onderdelen: deelnemers kijken of ze er nog een onderwerp bij zouden willen. De clusters worden samengevat en er wordt gekeken welke onderwerpen het meest aan bod komen. Iedereen geeft zijn of haar prioritering van onderwerpen door en zo ontstaat er een algemene top 3.	Drie clusters van kansen en uitdagingen, geprioriteerd
Ombuigen naar vragen	Uitgaande van de top 3 die in de vorige ronde gedestilleerd is: als we nu rond deze thema's onderzoeksvragen moeten formuleren, welke vragen zouden dat dan zijn? Wat, hoe, waarom, waartoe?	Een aantal onderzoeksvragen
Inzichten oogsten	Facilitator haalt de resultaten op, vraagt door bij deelnemers om verduidelijking en ervaringen, en vraagt om reflectie. Vraagt kort aan mensen of ze willen toelichten wat er uit is gekomen.	Verduidelijking & reflecties

Tabel 1. Het algemene stramien van het citizen panel zoals door Waag ingericht

Wat betreft het werven van deelnemers voor de citizen panels was er gezien de nabije deadline geen tijd en capaciteit voor uitgebreide werving van een geïnformeerde, grote en diverse groep. Er is daarom gekozen voor een tweetal publiek toegankelijke citizen panels op twee verschillende locaties: één bij de Ru Paré Community (een buurthuis in Amsterdam Nieuw-West in een wijk met een diverse groep bewoners) en één tijdens de PublicSpaces Conferentie in Pakhuis de Zwijger (een culturele ontmoetingsplaats in Amsterdam Centrum). Ru Paré is een community waarbij tientallen organisaties en initiatieven uit de buurt zijn

aangesloten. Er is specifiek voor deze plek gekozen om een publiek aan te trekken dat gewoonlijk geen deel uitmaakt van het Waag Futurelab netwerk. De promotie verliep middels een flyer campagne op verschillende plekken in de buurt (en in de rest van Amsterdam) zoals bij bibliotheken, buurtcentra, hogescholen en universiteiten. Daarnaast is het panel via verschillende sociale mediakanalen van Waag aangekondigd (Twitter, Mastodon en LinkedIn), is het in verschillende nieuwsbrieven meegenomen en is het geplaatst op websites van wijken die hun platform onderhouden op GebiedOnline. Een dag voor aanvang waren er 25 deelnemers die zich hadden aangemeld. Op de dag zelf (donderdag 22 juni 2023) kwamen er slechts 5 deelnemers opdagen, waarmee we onder leiding van twee facilitators de werkvorm in twee groepen hebben kunnen doorlopen. Redenen waarom deelnemers niet kwamen opdagen zijn niet opgegeven, maar we denken dat het te maken had met a) het dreigende onweer die avond en b) de locatie in een buitenwijk van Amsterdam, waardoor de mensen die gewoonlijk op onze evenementen afkomen verder moesten reizen. De sessie begon met een presentatie van Jurriaan van Diggelen van TNO. Deze presentatie diende als introductie van het onderwerp – de tweede stap in Tabel 1. Aan de hand van enkele voorbeelden lichtte hij in grote lijnen de ontwikkelingen op het gebied van AI toe en welke effecten dit heeft of kan hebben op onze samenleving.



Figuur 2. Enkele voorbeelden van AI-toepassingen ter introductie van de workshop

Voor het citizen panel tijdens de PublicSpaces Conferentie is er een iets andere aanpak gebruikt voor de werving. PublicSpaces is een coalitie van verschillende publieke organisaties die samen werken aan een alternatief digitaal ecosysteem op basis van publieke waarden. Tijdens de jaarlijkse conferentie komen professionals, experts, politici en beleidsmakers bijeen om te bedenken hoe een eerlijke markt met alternatieve software en ethische infrastructuur eruit zou kunnen zien en onderzoek of de democratie bestand is tegen Big Tech. Het citizen panel werd opgenomen in het programma van de gehele conferentie waardoor het automatisch meeging in de algemene promotie. Mensen konden zich niet specifiek voor de sessie inschrijven, alleen voor de gehele conferentie, dus het was



afwachten hoeveel mensen daadwerkelijk op de sessie zouden afkomen. Tijdens het citizen panel (dinsdag 27 juni) meldden zich 27 deelnemers die in groepjes van 5 onder begeleiding van een eigen facilitator de werkvorm doorlopen hebben. Voorafgaand aan de sessie vond er een in het programma van PublicSpaces een panelgesprek plaats over AI met publieke waarden als kern. De panelleden gingen dieper in op de risico's die gepaard gaan met de huidige dynamiek waarmee AI-systemen worden ontwikkeld. Deze sessie diende als introductie van het onderwerp, waarna we de deelnemers vroegen zich aan te sluiten bij het citizen panel wat daarna plaatsvond. Zo begon de sessie zelf met de tweede stap uit Tabel 1.



Figuur 3. Citizen panel tijdens de PublicSpaces Conferentie

In totaal zijn de panels met 32 deelnemers uitgevoerd. Facilitators gebruikten allen eenzelfde template voor het vastleggen van de gesprekken aan tafel. Deze zijn na afloop aangevuld met de informatie van de papieren waarop deelnemers aantekeningen konden maken. Op grond van deze gespreksverslagen zijn er 12 maatschappelijke thema's geïdentificeerd. Deze thema's zijn toegelicht op basis van de input vanuit de deelnemers van de panels.

### Onderzoeksmethode: online enquête

Aan de hand van de input die is opgehaald tijdens de twee citizen panels is er een online enquête ontwikkeld middels het programma Typeform. Hierin worden de 12 thema's toegelicht en wordt er gevraagd om mee te denken over toekomstig onderzoek binnen de route. De enquête bestaat uit de volgende onderdelen:

- Opening: Vraagstelling en doel van de enquête
- Introductie: Omschrijving van het totstandkomingsproces van de twaalf thema's plus uitleg over de structuur van de enquête
- Toelichting + aanvullen: omschrijving per thema plus de mogelijkheid om zelf iets toe te voegen - een korte omschrijving van iets wat volgens de deelnemer in de toelichting ontbreekt of een voorbeeld uit het eigen leven, die van zijn/haar naasten of uit de maatschappij die het belang van het thema onderstrepen.
- Prioriteren: deelnemers geven aan wat zij de drie belangrijkste thema's vinden

- Dankwoord en optie toevoegen van thema: Mogelijkheid om een thema (of thema's) toe te voegen die volgens de deelnemer ontbrak(en) in de lijst met een korte toelichting en eventueel een voorbeeld.
- Optioneel achterlaten persoonsgegevens: deelnemers krijgen de mogelijkheid om te kiezen of zij wel/niet persoonsgegevens willen achterlaten.
- Persoonsgegevens: geslacht, leeftijd, opleiding en niveau, huidige woonplaats.
- Afsluiting: mogelijkheid om emailadres achter te laten om in de toekomst updates te ontvangen over het onderzoek wat gedaan wordt binnen de route.



Figuur 4. Openingsblad van de online enquête.

De keuze om een online enquête te ontwikkelen is gemaakt zodat er bij een breed publiek input opgehaald kon worden op de twaalf thema's, een prioritering van de 12 thema's vast te stellen en om eventueel missende thema's in kaart te brengen. Het doel was om de representativiteit van de deelnemers die input leveren op het thema te vergoten en te toetsen of de thema's resoneren in de samenleving. De online enquête is op verschillende manieren verspreid:

- Diverse nieuwsbrieven: 5.637 abonnees.
- Sociale mediakanalen (X, LinkedIn, Mastodon, Instagram) van Waag: 51.846 volgers.
- Direct mailing naar het netwerk van Waag, o.a. een oproep aan masterstudenten van de opleiding Urban Management en burgers die vaker deelnemen aan citizen science activiteiten van Waag.
- Activatie van maatschappelijke organisaties om de oproep op eigen kanalen te publiceren, zo is de enquête geplaatst op het platform Amsterdam Smart City en verspreid door PublicSpaces en de Maatschappelijke Coalitie.

### Inzichten respondenten

De Typeform statistieken laten zien dat de online enquête uiteindelijk 514 keer is geopend. Daarvan is er 269 keer doorgelikt naar de introductie, maar slechts 22.5% heeft de enquête daadwerkelijk afgerond, wat neerkomt op 63 respondenten. Zij deden er gemiddeld 44 minuten over om de enquête in te vullen. Het percentage mensen dat nog tijdens de

introductie afhaakte is 25%. Ook bij de vraag voor toelichting of aanvullingen voor het eerste thema haakte er opnieuw 37% af. Voor het verdere verloop van de enquête neemt dit percentage steeds verder af.

Afleidend uit het hoge percentage afhakers en de tijd die het heeft gekost voor de respondenten om de volledige enquête in te vullen, kunnen we concluderen dat de enquête complexer is dan op voorhand was gedacht. De uitleg die bij de thema's gegeven wordt is aan de lange kant en mogelijk dat de uitleg voor non-experts lastig te volgen was. Echter betekent dit wel dat de respondenten de tijd hebben genomen de enquête naar behoefte in te vullen hebben. Dat betekent dat de reacties niet zonder na te denken zijn ingevuld en dat de respondenten de enquête serieus genomen hebben. Mogelijk komt dat de kwaliteit van de reacties ten goede.

Sommige respondenten hebben ook feedback gegeven op de opzet van de online enquête. De teksten werden als lang en complex gezien, wat ervoor gezorgd heeft dat sommige mensen zijn afgehaakt. Dat zagen we ook al terug in de statistieken van Typeform zelf. Mensen gaven aan dat zonder kennis de thema's lastig te begrijpen waren en dat het daarom onmogelijk was een prioritering door te geven. Van tevoren hadden wij daarover ook nagedacht. We kozen ervoor een uitgebreide toelichting te geven, juist om te zorgen dat iedereen de thema's begrijpt en de kwaliteit van de uiteindelijke prioritering te behouden. Maar veel mensen haakten ook af omdat de uitleg lang was. Voordeel was wel dat de thema's aanleiding gaven tot reflectie, waardoor de mensen die de hele enquête afmaakten er ook diep over nadachten en veel van hun reflecties opschreven. De thema's leverden veel verschillende reflecties, ideeën en vragen op. Dit is misschien ten koste gegaan van de kwantiteit (het aantal respondenten hadden we liever hoger gehad) maar de kwaliteit van de reacties was erg hoog. Een respondent stelde voor om per thema een schaal toe te voegen waardoor de belangrijkheid direct gekwantificeerd kon worden (1-10 bijvoorbeeld). Dit was mogelijk een goede toevoeging geweest, waarbij er sneller naar het belang van bepaalde thema's gevraagd kon worden, ook als mensen geen zin hadden om de hele tekst te lezen, en misschien meer mensen betrokken waren gebleven.

### **Spreiding respondenten**

Van 52 van de 63 respondenten zijn er enkele persoonsgegevens verzameld. De genderverspreiding is redelijk gelijk verdeeld: 58% procent van de respondenten is man, 40% vrouw en 2% geeft aan zijn geslacht liever niet door te geven. De gemiddelde leeftijd is 51 jaar. Het niveau waarop mensen een opleiding gevolgd hebben is voor het grootste deel wetenschappelijk onderwijs (71%). 25% gaf aan het hoger beroepsonderwijs te hebben afgerond en 2% het middelbaar onderwijs. Enkele respondenten hebben hun huidige woonplaats doorgegeven. Het overgrote deel is woonachtig in de randstad (19 respondenten) en daarnaast wonen er mensen in Almere, Lelystad, Midwolda, Groningen, Geldrop, Eindhoven, Enschede, Franeker, Middelharnis en Zundert.

# Resultaten

## Resultaten van de citizen panels

Tijdens beide citizen panels kwamen er al gauw interessante gesprekken op gang. Opvallend was dat deelnemers in het algemeen het meest reflecteerden op de ongewilde consequenties van AI-ontwikkelingen en het voorkomen daarvan. Er werden wel degelijk kansen gezien, zoals in de zorg, het onderwijs of de verduurzaming, maar het bleek makkelijker de verliezen of de risico's te bedenken die gepaard gaan met de komst van AI dan de positieve toekomstbeelden. Een onderbouwing die hiervoor werd gegeven is dat de winst die we zouden kunnen behalen niet zou opwegen tegen de mogelijke verliezen en/of schade die AI ons zou kunnen toebrengen.

Op basis van de gespreksverslagen van beide panels zijn er 12 maatschappelijke thema's geïdentificeerd. De toelichting op deze thema's bestaat voor het grootste gedeelte uit de input gegeven door de deelnemers tijdens de sessies en is aangevuld met voorbeelden of uitleg op basis van bureauonderzoek. De 12 thema's zijn als volgt:

### Machtsverhoudingen en machtsconcentraties

Verschillende deelnemers vinden het een probleem dat de macht en controle over data en AI in handen is van een handjevol grote Tech bedrijven. Zij hebben de macht en het geld en zijn dus grotendeels verantwoordelijk voor de innovatie in deze technologie. Men geeft aan dat de gebruiker al gauw *high tech* producten wil hebben: deze zijn efficiënt, verleidelijk, goedkoop en snel. Dit maakt dat het moeilijk is om geen gebruik te maken van de mogelijkheden die de bedrijven bieden en zo worden we afhankelijk van hun aanbod. In een competitieve samenleving en economie lopen mensen die geen gebruik maken van AI al snel achter op degenen die dit wel gebruiken. Dit maakt hen vervolgens weer machtiger. De zorg is dat de bedrijven de technologie in eerste instantie ontwikkelen om meer geld te verdienen en niet om per se iets goeds te doen voor de wereld en haar bewoners. "Je moet hopen dat de makers van robots weldoeners zijn" zegt één van de deelnemers. De geldschieters hebben het voor het zeggen en ondermijnen de democratie. De maatschappelijke impact van de technologie zou tenminste in overweging genomen moeten worden door de ontwikkelaars. Hoewel technologie ontwikkelaars zich vaak onttrekt aan wetten en regels, zou het eigenlijk zo moeten zijn dat die regels worden aangescherpt en dat er controle komt op naleving. Eigenlijk zou je moeten bepalen wat AI betrouwbaar maakt en vervolgens al het onderzoek naar AI monitoren en beoordelen op betrouwbaarheid. Net zoals de rekenkamer een monitorende en beoordelende functie heeft, zouden we volgens sommige deelnemers ook een organisatie moeten hebben die AI monitort.

### Nepnieuws, filterbubbels en polarisatie

Er wordt veel gesproken over nepnieuws, filterbubbels en de gevolgen daarvan op de samenleving. Gebruikers van (sociale) media krijgen op basis van hun eigen internetgedrag nieuws en berichten aangeboden die passen bij hun eigen voorkeuren. De algoritmen die deze selectie maken zijn gebaseerd op AI en worden steeds slimmer. Hier kleeft volgens

sommigen een groot nadeel aan: mensen komen in een bubbel terecht waarin ze alleen nog maar berichten te zien krijgen die hun eigen opinies, standpunten en voorkeuren bevestigen. Alternatieve meningen en ideeën krijgen ze niet meer te zien. Dit heeft een polariserende werking op de samenleving: mensen worden minder geconfronteerd met de wereld buiten hun eigen bubbel en verliezen een kritische blik van buitenaf en worden niet meer uit hun comfort zone getrokken. Daarnaast kan AI zelf ook onjuiste informatie genereren. Voorbeelden als ChatGPT laten zien dat het voor AI heel makkelijk is om tekst, afbeeldingen of video's te maken. Voor mensen is het soms erg lastig te zien of een foto echt is of niet. Daarnaast weet je soms niet of een tekst gebaseerd is op feiten en bronnen of dat het verzonnen is. "Vroeger zei men: het staat gedrukt dus het is waar. Maar dat is helemaal niet meer zo" zegt een deelnemer. Dat geldt overigens niet alleen voor AI gegenereerde content: mensen kunnen ook nep nieuws verspreiden. Maar de polariserende werking van AI via het bevorderen van nepnieuws, filterbubbels en verslavende technologieën wordt gezien als een belangrijke negatieve consequentie van AI, die ondervangen moet worden.

### **Verantwoordelijkheid en aansprakelijkheid**

Deelnemers delen zorgen over het steeds zelfstandiger worden van systemen gebaseerd op AI. Technieken als *machine* en *deep learning* maken het mogelijk dat deze systemen zelf keuzes kunnen maken en zich kunnen aanpassen aan een steeds veranderende omgeving. Maar er heerst een belangrijke vraag: wie is er vervolgens verantwoordelijk (en aansprakelijk) als deze systemen schade aanrichten? Denk bijvoorbeeld aan een zelfrijdende auto die een ongeluk veroorzaakt. Een deelnemer noemde een militaire robot die zelfstandig zijn doelwit uitkiest op basis van gezichtsherkenning. Er zijn voorbeelden van AI-systemen die een selectie maakten uit een lijst met sollicitanten die gebaseerd bleek te zijn op (racistische) vooroordelen. Volgens sommigen ligt de verantwoordelijkheid bij de ontwikkelaar, anders noemen de gebruiker of het systeem zelf. Verantwoordelijkheid van AI is volgens hen tweedelig: AI zou moeten handelen volgens ethische principes en achteraf moet het kunnen uitleggen hoe bepaalde keuzes zijn gemaakt.

### **Vooroordelen en inclusie**

Het wordt al even genoemd in vorig thema, maar er wordt in beide sessies gereflecteerd op de vooroordelen die kunnen bestaan in AI-systemen. Wanneer je AI bijvoorbeeld inzet voor het beoordelen van sollicitaties is de data waarop in het verleden keuzes zijn gemaakt van invloed. Als in een organisatie vaak witte mannen zijn aangenomen terwijl er ook vrouwen en mensen van kleur gesolliciteerd hadden, dan zal het AI-systeem ook soortgelijke witte mannen met een vergelijkbaar CV selecteren. Mensen met alternatieve kwaliteiten, een beperking, een buitenlandse achternaam, vrouwen of heel jonge of juist oude mensen vallen dan buiten de selectie. Als een systeem zulke maatschappelijke vooroordelen meeneemt in beslissingen, zijn deze vaak lastig te herkennen. Laat staan te corrigeren. De vooroordelen kunnen zo onderdeel worden van de 'black box' in een AI. Andere deelnemers geven aan dat het omgekeerde ook mogelijk is. AI kan juist ook ingezet worden om vooroordelen in de samenleving te herkennen en te corrigeren. Dat betekent wel dat daar bij onderzoek expliciet op gelet moet worden. Daarnaast is AI ook in staat om inclusie te vergroten: mensen die eerst niet mee konden doen aan de samenleving kunnen profijt hebben van de technologie.

Tools kunnen helpen zelfstandiger te leven. Echter, de zorg is wel dat deze mensen in toenemende mate afhankelijk worden van deze tools en zo weinig zeggenschap hebben over de producten omdat ze de deze niet zelf begrijpen of kunnen repareren. De bedrijven achter de technologie blijven nodig om reparaties te leveren en bepalen de richting van de innovatie. Dat AI in zo'n geval inclusie bevordert heeft dus tegelijk het gevolg dat mensen meer afhankelijk worden van de Tech bedrijven.

### **Sturing van gedrag en beïnvloeding van keuzes**

Deelnemers noemen de mogelijkheden van positieve *nudging* door AI: bijvoorbeeld door mensen te stimuleren om gezonder, klimaat bewuster of veiliger te leven. Maar AI kan mensen ook aansporen tot ongewenst gedrag. Voorbeelden als reclame gebaseerd op persoonlijke voorkeuren dat buitensporig winkelgedrag kan veroorzaken of slimme algoritmes die zorgen voor eindeloos scrollen op sociale media komen naar voren. Deze beïnvloeding kan maatschappelijk en psychologisch nadelige gevolgen hebben, zoals verslaving. Ook kan de drempel voor het maken van keuzes lager worden bij gebruik van AI ten opzichte van een mens zoals bij het doden van mensen in een oorlog, spionage of bij losbandig seksueel gedrag met een robot. Een mens zou misschien langdurig over nadenken maar voor een AI is deze keuze een stuk makkelijker vanwege het ontbreken van een moreel kompas. Men vraagt zich af wanneer de sturing of beïnvloeding gewenst is en hoe je kunt zorgen dat AI het gedrag in de juiste richting stuurt.

### **Onderwijs, ontwikkeling en vaardigheden**

Het gebruik van AI in het onderwijs kan positieve gevolgen hebben. AI kan de leerprestaties monitoren en zo leerlingen herkennen die gebaat zijn bij alternatieve leermethoden. Door gemaakte fouten te analyseren kan AI verschillende leerstijlen definiëren en vervolgens ondersteunen in een persoonlijk lesprogramma. Als je het hebt over de inhoud van het onderwijs, kun je je afvragen of leerlingen nog wel dezelfde vaardigheden moeten leren als daar ook AI voor kan worden ingezet. Nu worden er allerlei vaardigheden gestuurd op scholen en universiteiten: het schrijven van werkstukken of essays, het zoeken van informatie, het formuleren van een goede onderzoeksvraag of het toepassen van bronvermelding. Een language model als ChatGPT kan dat sneller en efficiënter. Dat scheelt tijd, waardoor er door mensen op andere zaken gefocust kan worden. Daarnaast kan AI ook worden ingezet om te compenseren voor vaardigheden die sommige mensen niet kunnen leren. Taalbarrières, leerachterstanden, leerproblemen (zoals dyslexie of dyscalculie) kunnen zorgen dat mensen heel verschillende schoolloopbanen hebben. Mogelijk kan AI compenseren voor die verschillen en mensen gelijkere kansen geven. Echter, in al deze voorbeelden zit een risico. Het gevaar is dat mensen afhankelijk worden van de technologie en mogelijk ook geen moeite meer hoeven doen om vaardigheden aan te leren. Maar om te kunnen beoordelen of AI de taken goed uitvoert, moeten mensen het ook zelf goed kunnen. Wat is een goed moment om te beslissen dat het beter is om taken aan technologie te delegeren in plaats van te proberen het (alsnog) zelf te leren? Hoe zorgen we dat mensen voor dit soort taken niet afhankelijk worden van AI zonder de nodige vaardigheden te leren om te kunnen beoordelen of deze het wel goed doen? Als de technologie betrouwbaar is, dan is dit misschien minder een probleem. Maar als Big Tech erachter zit en dit vertrouwen

niet hebben, dan is dat wel het geval. Tot slot is het de vraag wat mensen dan moeten gaan doen met alle vrijgekomen tijd, die ze niet meer hoeven te besteden aan het aanleren van vaardigheden of het doen van tijdrovend werk. Zijn er dan nog taken die een mens nodig (blijft) hebben? En als die er niet zijn, wat is dan een zinvolle vulling van een mensenleven?

### **Arbeidsverdeling tussen mens en AI**

Veel deelnemers kozen voorbeelden van mens-machine interacties en dachten na over een passende arbeidsverdeling tussen AI en de mens. AI zou relatief eenvoudig vervelend of routinematig werk kunnen overnemen, zoals in fabrieken aan de lopende band. Maar dat zorgt ook voor werkeloosheid. Anderen opperen ook de werkmogelijkheden die ontstaan voor onderhoud en de ontwikkeling van de technologie. Bovendien zouden mensen met een achterstand tot de arbeidsmarkt mogelijk gebaat zijn bij hulp van toepassingen van AI. Sommigen richtten hun aandacht specifiek op de zorg: hulpbehoevenden krijgen minder menselijk contact wanneer een zorgrobot het werk overneemt, maar het lost wel het personeelsprobleem in de zorg op. Daarnaast spreken deelnemers spreken zich nog uit over vervlakking: creatieve of complexe banen kunnen ook worden overgenomen door AI waardoor vaardigheden en kennis bij de mens verdwijnen. Echter, AI kan ook gebruikt worden voor het creëren van nieuwe kunstvormen.

### **Relaties en psychologische ontwikkeling**

Verlies aan contact tussen mensen wordt vaak genoemd als nadeel van de inzet van AI. Chatbots worden gebruikt voor de klantenservice en sociale media vermindert menselijk contact in de echte wereld. "Ik vraag altijd om een mens" zegt een deelnemer over de klantenservice. "Als diegene stottert, is het vervelend. Maar toch liever een mens." Daarnaast kunnen zorgtaken voor hulpbehoevenden door een robot worden uitgevoerd, terwijl veel deelnemers liever een menselijke zorgverlener hebben. De zorg is dat dit zou kunnen leiden tot eenzaamheid en uitsluiting van het sociale leven. Relaties spelen een grote rol in de ontwikkeling van de menselijke psychologie. Een relatie met een AI kan handig zijn voor mensen die verminderd sociaalvaardig zijn of erg verlegen om te oefenen. Maar het roept ook vragen en zorgen bij de deelnemers op. Als je beste vriend een AI is, in de vorm van een Chatbot, wat doet dat met je psychologische ontwikkeling? Stel dat een AI vriend alleen maar aardige dingen zegt en geen kritische reflectie biedt. En bovendien onbereikbaar en niet echt is. Hoe verhoudt de mens zich tot AI en wat doet het met de psychologische ontwikkeling? Er worden ook vragen gesteld als hoe ziet liefdesverdriet eruit als je beseft dat de relatie met jouw AI vriend nooit iets kan worden of hoe draagt intimiteit met AI bij aan (de ontwikkeling van) een relatie?

### **Vervuiling en milieu**

Sommige deelnemers observeren dat AI enorm veel stroom gebruikt en niet heel milieuvriendelijk is. Men vraagt zich daarom af of de vlugge ontwikkelingen en het grootschalig gebruik van AI impact zullen hebben op het milieu. Aan de andere kant kan er met AI ook juist een bijdrage worden geleverd aan verduurzaming. Denk aan monitoring van gewasgroei in de landbouw en adviseren op behandeling van de grond of door het voorspellen van weersomstandigheden. AI kan verschillende data combineren en leren van

het verleden, en op basis daarvan aangeven wat er moet gebeuren om bijvoorbeeld de kwaliteit van de bodem, het water of de lucht te verbeteren. Zo wordt verduurzaming mogelijk gemaakt. Ook zou voedsel sneller, goedkoper en met minder belasting voor het milieu kunnen worden geproduceerd. Deelnemers vragen zich af hoe we ervoor kunnen zorgen dat er een balans gevonden wordt in de vervuilende gevolgen van AI en de kansen die toepassingen hebben op het gebied van verduurzaming.

### **Gezondheid**

Over het algemeen werd door de deelnemers erkend dat AI veel mogelijkheden biedt voor het stimuleren en verbeteren van de gezondheid van mensen. Grootschalig onderzoek op basis van grote hoeveelheden data kan tot nieuwe inzichten leiden en er kan meer inzicht komen in gepersonaliseerde gezondheidszorg. Ziektes kunnen hierdoor beter, sneller en efficiënter worden gedetecteerd en bepaalde werkzaamheden van artsen kunnen worden overgenomen door AI wat hen meer tijd geeft voor overige taken. Ook kunnen mensen betere en gepersonaliseerde informatie krijgen over wat bijdraagt aan hun eigen gezondheid. Maar men ziet ook risico's. Een deelnemer vraagt zich af wat er gebeurt als er diagnoses gesteld worden door een AI die onjuist blijken te zijn? Als mensen meer leren over hun eigen gezondheid en dat continu monitoren, kan dat onrust of angst veroorzaken en zelfs een schuldgevoel creëren. Ook is veel gezondheidsdata noodzakelijk maar niet iedereen wilt zijn of haar data zomaar prijsgeven.

### **Surveillance en vrijheid**

Om de soevereiniteit van Nederland te bevorderen, zouden we volgens veel deelnemers technologie zelf kunnen gaan ontwikkelen. Dat geldt ook voor AI. Aan de andere kant is het ook nodig om goed in de gaten te houden wie het voor het zeggen heeft: wie gaat er over de technologie? Wie genereert de data? Wat wordt daarmee gedaan en welke beslissingen worden er genomen op basis van de data? De overheid kan mensen monitoren met behulp van AI en op basis van de data beoordelen wie of wat een risico vormt in de samenleving. Denk aan het risico op geweldpleging, inbraak, fraude of verslaving. Preventief ingrijpen helpt om deze risico's te voorkomen maar dat betekent ook dat mensen al geprofileerd worden voordat ze daadwerkelijk iets gedaan hebben. Iemand noemt het voorbeeld dat er op basis van data uit het verleden kunnen mensen verkeerd worden geprofileerd, waardoor ze ten onrechte gekenmerkt worden als 'crimineel'. Mensen vinden dat er daarbij onvoldoende rekening gehouden wordt gehouden met de capaciteit van mensen om zelf een vrije keuze te maken, en dus gedrag te ontwikkelen dat afwijkt van het profiel waarin ze (op grond van data) zijn ingedeeld. Het verzamelen van data over burgers zou kunnen leiden tot surveillance; een samenleving waarin iedereen voortdurend wordt gevolgd. Deelnemers delen de zorg dat dit de vrijheid van mensen serieus zou kunnen beperken.

### **Intelligenties: kunstmatig en biologisch**

Een deelnemer be vraagt de I in AI. Wat betekent het als we een digitaal systeem 'intelligent' noemen? En hoe verschilt die intelligentie van wat we intelligent noemen in mensen of dieren? Deelnemers suggereren dat hier onderzoek naar gedaan kan worden. Computers en systemen kunnen op basis van grote hoeveelheden data leren en kunnen vervolgens bijna

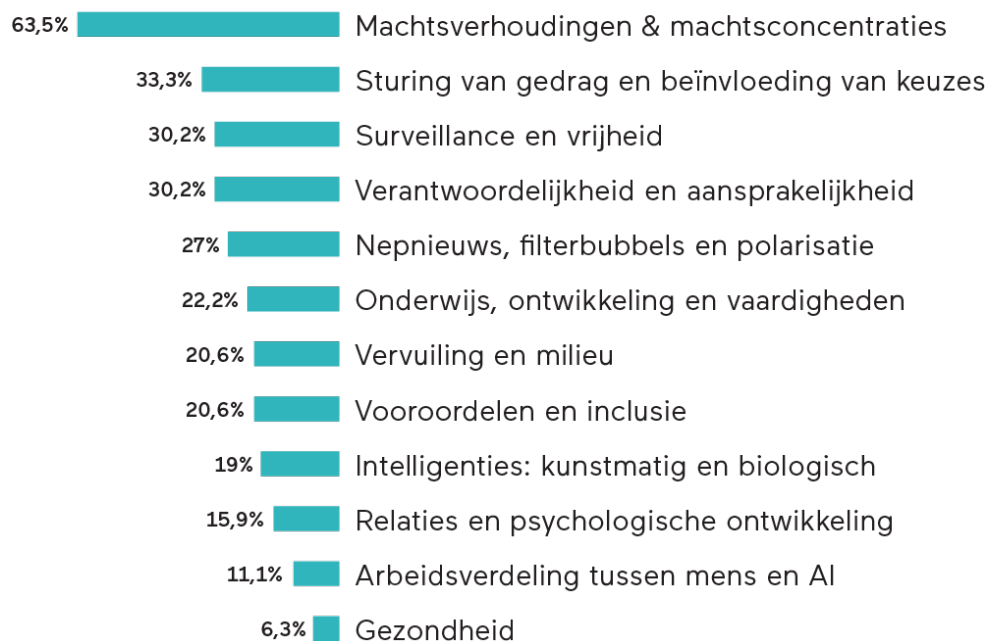


hetzelfde handelen en communiceren als mensen. Maar als computers onnavolgbaar en onbegrijpelijk worden en dus eigenlijk de menselijke intelligentie overstijgen, kunnen we ze dan nog wel 'intelligent' noemen? Is er een grens aan wat we intelligent kunnen noemen? Interessant is ook hoe intelligentie is verbonden met andere zintuigen. Een computer leert bijvoorbeeld op basis van data en visuele input. Maar de menselijke intelligentie is ook afhankelijk van andere zintuigen zoals gehoor, tastzin en geur. Wat betekent dit voor de menselijke intelligentie en hoe deze zich verhoudt tot de intelligentie van AI-systemen?

## Resultaten online enquête

De 12 geïdentificeerde thema's zijn vervolgens in de online enquête aan een breder publiek voorgelegd. De thema's werden in hoofdlijnen toegelicht. Op basis van de prioritering van de thema's en de aanvullingen die de respondenten op de verschillende thema's gegeven hebben, is er een analyse gedaan. Opvallend was dat veel van de respondenten zelf (onderzoeks-)vragen opperden, zonder dat daar expliciet om werd gevraagd. Een overzicht van de reacties op de thema's en de (afgeleide) vragen geeft een verhelderend beeld van waar zowel de deelnemers van de citizen panels als de respondenten hopen dat het onderzoek binnen de route zich op zal gaan richten.

Allereerst kijken we naar de prioritering van de thema's. Op de vraag welke thema's het belangrijkste zouden moeten zijn voor het toekomstig onderzoek binnen de route *Waardecreatie door verantwoorde AI en Big Data* konden de respondenten 3 thema's kiezen. De uiteindelijke prioritering is te vinden in Figuur 5. Overduidelijk is er door de respondenten het meest gekozen voor het thema 'Machtsverhoudingen en machtsconcentraties'. Daarna staat het thema 'Sturing van gedrag en beïnvloeding van keuzes' op nummer 2 en wordt plaats 3 gedeeld door 'Surveillance en vrijheid' en 'Verantwoordelijkheid en aansprakelijkheid'.



Figuur 5: Prioritering van maatschappelijke thema's naar aanleiding van de online enquête

Naar aanleiding van het thema 'Machtsverhoudingen en machtsconcentraties' komen de volgende reflecties vanuit de respondenten terug: de verantwoordelijkheid/zeggenschap van grote bedrijven wat de verantwoordelijkheid van staten overstijgt, (gebrekkige) controle die de overheid over die bedrijven kan uitoefenen, marktwerking en hoe dit zorgt dat sommige bedrijven AI sneller op de markt kunnen brengen, afhankelijkheid van grote bedrijven die AI maken, soevereiniteit (mogelijkheden en beperkingen van monitoring en controle vanuit overheid), databeheer en eigenaarschap van de data waarop AI is gebaseerd, of er moet worden betaald voor data, open bronnen, inclusie, mogelijkheden om als gebruiker zelf AI te sturen of er inspraak op te hebben, en betrouwbaarheid en transparantie van technologie. Een volgende lijst met onderzoeksvragen kunnen we afleiden op basis van de input op dit thema:

Onderzoeksvraag
Hoe kunnen gebruikers van AI meer transparant inzicht krijgen in hoe de technologie werkt en hoe dit is gebaseerd op hun data?
Wie mag beslissen over AI? Wie mag er geld mee verdienen? Gaan de grote Tech bedrijven mensen ook betalen voor het gebruik van de data waarop AI functioneert?
Wat zijn gevolgen van AI voor de wijze waarop individuele mensen en organisaties (bedrijven/overheden) afhankelijk worden van AI?
Wie bepalen de regels voor AI en wie controleert die? Hoe kun je dat op grote schaal (en automatisch) doen?
Wat moet er veranderen in de governance van een bedrijf, als je wil zorgen dat er niet alleen een ethiek afdeling is die tegenspraak biedt, maar dat die ook macht heeft zodat ernaar wordt geluisterd?
Hoe organiseren we tegenmacht? En wie moeten er in die tegenmacht vertegenwoordigd zijn? Overheden? Kunnen collectieven van burgers en onderzoekers iets betekenen?
Hoe kunnen we de technologie transparant maken en inspraak organiseren op algoritmes door mensen die geraakt worden door algoritmes?
Is het mogelijk om met minder data ook betrouwbare AI te ontwikkelen, zodat kleine startups meer kans krijgen op de markt (en de grote bedrijven niet de hele markt controleren)?
Zijn er andere publieke instanties die de controlerende rol op AI op zich kunnen nemen? Of moeten burgers dit zelf gaan doen?
Hoe ziet een goede samenwerking tussen mensen en technologie eruit, zonder dat we afhankelijk worden van de tech?
Kunnen we het onderzoek dat gedaan wordt naar AI monitoren en beoordelen op betrouwbaarheid?
Is het mogelijk om AI te ontwikkelen zonder dat dit tot machtsconcentraties leidt of de democratie ondermijnt?

Op basis van het thema 'Sturing van gedrag en beïnvloeding van keuzes' wordt door meerdere respondenten de vraag gesteld of je je als individu mag onttrekken aan de sturing door AI en onder welke voorwaarden sturing of *nudging* een goed idee is. De verantwoordelijkheid die men zelf heeft (voor keuzes en gedrag) moet daarbij niet vergeten worden, of worden ondermijnd door AI. Daarnaast zouden we zelf invloed moeten kunnen uitoefenen op AI en het gebruik ervan, wat ook meer inzicht geeft in AI-sturing van keuzes en gedrag. Er wordt ook een link gelegd met het thema over machtsverhouding: hoe gaan we om met de sturing en beïnvloeding van gedrag door grote Tech bedrijven en de overheid? Dit onderwerp maakt het belang van onderzoek naar AI alignment duidelijk: de sturing van de systemen richting de beoogde doelen, voorkeuren en ethische principes van mensen. De volgende onderzoeksvragen zijn leidend binnen dit thema:

#### Onderzoeksvraag

Mag AI het gedrag van mensen sturen of keuzes beïnvloeden? En – zo ja – onder welke voorwaarden en voor welke doelen mag dat dan? Wie beslissen die doelen?

Kunnen gebruikers ook macht krijgen over het gedrag van AI? En waar gaan die mensen AI dan voor inzetten?

Op welke manieren kun je AI inzetten om gedrag in de goede richting te sturen? Op wat voor manier past AI-beïnvloeding toe? Zou je *nudges* kunnen personaliseren op verschillende type motivaties om zo de meegaandheid voor goed gedrag te vergroten?

Zijn er manieren om sturing/*nudges* van gedrag door AI acceptabeler te maken voor gebruikers? Wat zijn goede manieren om de ontwikkeling van AI die stuurt acceptabel te maken: moet je AI transparant maken zodat mensen de sturing/*nudge* zien (en kunnen weigeren)?

Welke rol kan wet- en regelgeving spelen in het omgaan met sturing van gedrag en beïnvloeding van keuzes door AI?

Wat betreft het thema 'Surveillance en vrijheid' geven respondenten aan dat de sleutel tot het voorkomen van misstanden ligt bij de omgang en regelgeving omtrent de gegevens van mensen. Anderen stellen zich voor hoe een surveillance maatschappij eruit zou zien, en suggereren dat mensen ze zich anders gaan gedragen als ze worden bekeken. Mensen zouden zo veel mogelijk eigenaar moeten blijven van hun eigen gegevens. Er moeten duidelijke regels worden opgesteld voor het monitoren van mensen zonder expliciete toestemming en er moet uitgelegd kunnen waarom het nodig is. Daarnaast moeten mensen ook de vrijheid krijgen om fouten te maken. En ze moeten inspraak hebben in bestuur, en dus ook in of, hoe en wanneer ze worden gesurveilleerd. De volgende onderzoeksvragen kunnen we hieruit afleiden:

#### Onderzoeksvraag

Kunnen er ook ruimtes komen waar mensen niet worden gesurveilleerd? Hoe herken je die gebieden? Hoe zouden sensor-vrije gebieden eruitzien?

Wat is de meerwaarde van gebruik van AI bij surveillance? Kunnen we ook op kleinere/ decentrale/menselijke schaal surveillance regelen, waarbij er ruimte is voor persoonlijke interpretatie?

In hoeverre kan de privacy beschermd worden met AI? Hoe verhoudt de aandacht voor privacy kwesties in Nederland zich tot die in andere landen?

Een algoritme identificeert verdachten op basis van eerdere data. Kun je voorkomen dat sommige groepen mensen bij voorbaat gecriminaliseerd worden?

Hoe voorkomen we dat het uitgangspunt 'onschuldig, tenzij anders bewezen' omslaat in 'verdacht bij voorbaat'?

Voor het thema 'Verantwoordelijkheid en aansprakelijkheid' klinken er vanuit de respondenten instemmende geluiden met de beschrijving uit de enquête. Iemand noemt nog dat het juridisch systeem achterloopt bij de technische ontwikkelingen en dus niet is toegerust om aansprakelijkheidsvraagstukken te beantwoorden. Er is herziening nodig van wet- en regelgeving. Mensen vragen om transparantie van AI gebruik (een disclaimer) en ook om uitleg over hoe het werkt. Dan pas kan er verantwoordelijkheid genomen worden (los van de vraag door wie). Daarnaast wordt er gewezen op het feit dat bias en error in een systeem altijd voorkomt, ook bij de mens. Hieruit kunnen we de volgende onderzoeksvragen afleiden:

#### Onderzoeksvraag

Hoe werkt verantwoordelijkheid bij AI? Moet AI handelen volgens ethische principes? Moet AI achteraf kunnen uitleggen welke keuzes er zijn gemaakt en waarom?

Wie is verantwoordelijk voor AI? De maker? De gebruiker? AI zelf? Moeten we van de inzet van AI een gedeelde verantwoordelijkheid maken?

Wie heeft verantwoordelijkheid om te verduidelijken hoe AI werkt? Moeten ontwikkelaars die verantwoordelijkheid krijgen?

Wat is de waarde van een AI die zich verantwoordt voor de samenleving en binnen een juridisch systeem? Krijgt het rechten en plichten en kan het gestraft worden voor zijn keuzes?

Wie is verantwoordelijk voor de *legacy* van AI: wie checkt of AI na verloop van tijd nog steeds doet waar het voor ontworpen is?

Als we normen en waarden integreren in het ontwerp van AI, en normen en waarden veranderen over tijd, kunnen systemen zich dan ook aanpassen? Hoe passen systemen zich aan? Hoe kunnen we zien dat een systeem nog niet is aangepast?

Hoeveel beter moet AI zijn dan mensen om het gebruikt (en de fouten) acceptabel te maken? Mensen maken ook fouten.

De Algemene Rekenkamer onderzoekt of de rijksoverheid goed omgaat met publiek geld. Zou er ook een organisatie opgericht kunnen worden die het onderzoek en de ontwikkeling naar AI monitort en beoordeelt op betrouwbaarheid?

## Overige thema's

Voor de overige thema's focussen we ons hier op de onderzoeksvragen die we hebben kunnen afleiden naar aanleiding vanuit zowel de input vanuit de citizen panels als de online enquête. Een samenvatting van de reacties van de respondenten is te vinden in de bijlage.

Nepnieuws, filterbubbels en polarisatie

### Onderzoeksvraag

Hoe werkt vertrouwen in informatie? Aan welke voorwaarden moet betrouwbare informatie voldoen?

Is het mogelijk om AI te ontwikkelen die onderscheid kan maken tussen betrouwbare en onbetrouwbare informatie? Kunnen we daar een keurmerk voor ontwikkelen?

Hoe leren we onszelf en onze kinderen kritisch om te gaan met AI gegenereerde informatie?

Onderwijs, ontwikkeling en vaardigheden

### Onderzoeksvraag

In hoeverre kan AI leerachterstanden opsporen en helpen adresseren? En wat zien we dan als een goede manier om die te adresseren?

Hoe moeten we omgaan met AI in het onderwijs? Wat zijn de gevolgen op de lange termijn voor de ontwikkeling van kinderen, als ze vaardigheden niet meer ontwikkelen omdat AI de taken overneemt?

Wat hebben we nodig om te functioneren in deze maatschappij (samenwerken, creativiteit, flexibiliteit, analytisch vermogen) en welke rol heeft AI in de ontwikkeling daarvan?

Welke criteria gebruiken we om te bepalen of we de resultaten die AI biedt (bv een tekst, een analyse van data) kunnen vertrouwen?

Vervuiling en milieu

### Onderzoeksvraag

Hoe zorgen we ervoor dat er een balans gevonden wordt in de vervuilende gevolgen van AI en de kansen die toepassingen hebben op het gebied van verduurzaming

Moeten er milieueisen gesteld worden aan de ontwikkeling van AI? Hoe kun je AI zo klein en zuinig mogelijk maken zonder dat dit ten koste gaat van de kwaliteit?

Wie is er verantwoordelijk voor de vervuiling die AI veroorzaakt? De ontwikkelaar, de verkoper, de gebruiker of de overheid die de wet- en regelgeving maakt?

Welke rol kan AI spelen in de verduurzaming van verschillende sectoren (zoals bijvoorbeeld de landbouw)?

## Vooroordelen en inclusie

### Onderzoeksvraag

Hoe zorgen we ervoor dat AI inclusief functioneert, op grond van algoritmes die niet de voordelen van mensen herhalen?

Kunnen we vooroordelen herkennen en ter discussie stellen in een AI-algoritme? Aan wat voor voorwaarden moet AI voldoen om dit mogelijk te maken?

AI zal alleen inclusie bevorderen als er geld aan te verdienen is voor bedrijven die AI maken. Hoe zorgen we voor een ingebouwde 'leave-no-one-behind' mindset?

## Intelligenties: kunstmatig en biologisch

### Onderzoeksvraag

Wat voor verschillende soorten intelligentie bestaan er bij mensen? Wat is kenmerkend voor wat we intelligentie noemen in een machine? In hoeverre lijken die op elkaar?

Wat is bewustzijn? Is het mogelijk voor een machine om bewustzijn te hebben?

Als we niet meer kunnen volgen wat een machine doet, is de machine dan nog wel intelligent? Wat zijn de gevaren als we niet meer kunnen begrijpen waarom een machine tot bepaalde conclusies komt?

Wat is de invloed van AI op de ontwikkeling van de menselijke intelligentie?

## Relaties en psychologische ontwikkeling

### Onderzoeksvraag

Wat is het verschil tussen menselijke relaties en relaties tussen mens en machine?

Wanneer is een relatie tussen mens en AI problematisch? Wanneer maakt AI juist betere menselijke relaties mogelijk?

Heeft AI innerlijk leven en emoties? Wat is de waarde van het hebben van een innerlijk leven in relaties?

Onder welke voorwaarden kan AI in menselijke relaties worden ingezet? Hoe kunnen we de voordelen benutten en de nadelen mitigeren/voorkomen?

## Arbeidsverdeling tussen mens en AI

### Onderzoeksvragen

Wat is de waarde van arbeid in een mensenleven? Iets kunnen waarop je lang hebt moeten oefenen geeft zin aan een mensenleven. Wat betekent het voor een mensenleven als AI deze vaardigheden overneemt?

Wat is een inclusieve arbeidsmarkt? Kan AI hieraan ook bijdragen?

AI wordt vaak ingezet om dure mensen te vervangen of om een productieproces efficiënter te maken. Zijn er ook andere waarden, behalve kosteneffectiviteit en efficiëntie die AI kan dienen op de arbeidsmarkt?

Wat is de juiste arbeidsverdeling tussen mens en AI? Welke taken kan AI beter, en welke kun je beter overlaten aan mensen?

Wat zijn de gevolgen van de inzet van AI in arbeidsprocessen voor de kennis en vaardigheden van mensen?

## Gezondheid

### Onderzoeksvraag

Hoe kunnen AI *decision support* tools het beste in de workflow van artsen worden geïntegreerd? Hoe betrouwbaar moet AI zijn in het stellen van diagnoses t.o.v. artsen? Is een arts altijd eindverantwoordelijk?

Wie is er eigenaar van gezondheidsdata? Wat als je overlijdt, wat gebeurt er dan met de data? Hoe bewaken we de privacy als het gaat om patiëntgegevens? Hoe zit dat voor verzekeraars?

Wat is het effect van gepersonaliseerde medicatie op basis van AI voor verwachtingen ten aanzien van kwaliteit van leven en gezondheid? En ten aanzien van de verantwoordelijkheid van artsen?

# Conclusie, discussie en aanbevelingen

Vanuit de citizen panels zijn er 12 maatschappelijke thema's geïdentificeerd. Deze luiden als volgt:

1. Machtsverhoudingen en machtsconcentraties
2. Nepnieuws, filterbubbels en polarisatie
3. Verantwoordelijkheid en aansprakelijkheid
4. Vooroordelen en inclusie
5. Sturing van gedrag en beïnvloeding van keuzes
6. Onderwijs, ontwikkeling en vaardigheden
7. Arbeidsverdeling tussen mens en AI
8. Relaties en psychologische ontwikkeling
9. Vervuiling en milieu
10. Gezondheid
11. Surveillance en vrijheid
12. Intelligenties: kunstmatig en biologisch

Van deze thema's wordt de hoogste prioriteit gegeven aan 'Machtsverhoudingen en machtsconcentraties' (64%), 'Sturing van gedrag en beïnvloeding van keuzes' (33%), 'Surveillance en vrijheid' (30%) en 'Verantwoordelijkheid en aansprakelijkheid' (30%). We hebben ons over de prioritering van deze thema's verschillende dingen afgevraagd. Zo was het thema 'machtsverhoudingen en machtsconcentraties' het eerste thema van de lijst, en vroegen we ons af of dit een reden kon zijn waarom dit hoger scoorde in de prioriteitenlijst. Maar als we naar het tweede, derde en vierde thema in de oorspronkelijke lijst kijken, dan zet deze trend zich niet door. Bijvoorbeeld het derde thema 'surveillance en vrijheid' eindigde vrij hoog in de prioritering (op plek 3), maar staat in de oorspronkelijke lijst op nummer 11. Ook werd er pas gevraagd om thema's te prioriteren aan het einde van de enquête, dus als de deelnemer door alle thema's had heen gewerkt. Gemiddeld duurde dat 45 minuten en schreven deelnemers bij elk onderwerp veel reflecties op. Dat betekent dat mensen helemaal aan het einde van de enquête alsnog 'Machtsverhoudingen en machtsconcentraties'. Aanmerkten als belangrijkste thema.

Hieruit concluderen wij dat de volgorde waarin de thema's werden gepresenteerd geen invloed had op de uiteindelijke prioritering. Voorts valt op dat de reflecties van deelnemers op de thema's die het vaakst zijn geprioriteerd deels overlappen met wat ze opschrijven bij enkele van de andere thema's die later genoemd worden. Het wordt bijvoorbeeld problematisch gevonden dat AI het gedrag en de keuzes van mensen stuurt, juist omdat er soms grote machtige partijen zijn die AI vormen en die dus bepalen hoe dat gedrag en die keuzes worden gestuurd. Ook wordt er veel belang gehecht aan uitlegbaarheid of transparantie van de technologie, of de producten die AI oplevert (zoals bijvoorbeeld



nepfoto's, nepnieuws of een door AI gegenereerd voorkeurslijstje van sollicitanten), juist omdat het vaak onduidelijk is hoe dergelijke producten tot stand komen en mensen willen weten wie erachter zit. Grote machtige bedrijven (of overheden) zullen die producten met andere doelen en belangen in het achterhoofd ontwikkelen dan een kleinere non-profit partij die streeft naar een maatschappelijk doel. Hieruit blijkt dat deelnemers de thema's niet helemaal los van elkaar beschouwen, maar ze in hun reflectie vaak met elkaar verbinden, en dat machtsverhoudingen in veel van die reflecties een belangrijke rol spelen.

De invloed van marktwerking en marktbelangen in de ontwikkeling van AI, afhankelijkheid, gebrek aan transparantie en sturing van het individuele en sociale leven in richtingen die men niet zelf kiest, behoren tot de belangrijkste zorgen die mensen uitspreken. Andersom spelen kernbegrippen zoals transparantie, controleerbaarheid, uitlegbaarheid en verantwoordelijkheid een grote rol als het gaat om de ontwikkeling van meer acceptabele technologie. Deze aspecten zouden zeker een rol moeten hebben binnen de thematische focus voor de NWA-ORC ronde 2024.

Het onderzoek laat vooral zien dat mensen bezorgd zijn over de ontwikkelingen op het gebied van AI, en dat macht in die zorgen een grote rol speelt. Een verklaring voor die zorgen zou kunnen liggen in zich verder ontwikkelende ongelijkheid tussen mensen met veel kennis en expertise rond AI (de kennis have's) die de kans hebben om AI te ontwikkelen, en mensen die deze kennis en expertise ontberen (de kennis have nots). De have nots kunnen niet beoordelen of datgene wat AI biedt ook goed, betrouwbaar of waar is, en worden afhankelijk van de have's voor informatie, verklaringen en verantwoordingen. Er is kortom een groeiende disbalans tussen mensen, wat gemakkelijk kan leiden tot machtsmisbruik. Dat mensen in hun antwoorden bij de verschillende thema's dan zoeken naar manieren om greep te krijgen op AI, en tal van suggesties doen om dat voor elkaar te krijgen (zoals, maak disclaimers voor zaken geproduceerd door AI, maak wetten, zorg dat de werking van machines inzichtelijk is, laat zien hoe 'beslissingen' van AI tot stand komen, leer kinderen hoe ze met AI moeten omgaan etc.) is dan niet zo vreemd.

Al deze zaken geven wel aan dat het belangrijk is om aandacht te geven aan de zich ontwikkelende machtsongelijkheid die in de toekomst kan verergeren. Net zoals we tal van manieren hebben om nadelige effecten van machtsongelijkheid te verkleinen in andere relaties, zoals arbeidsrelaties, huwelijksrelaties, arts-patiënt relaties of onderzoeker-onderzoekssubject relaties, moeten we ook verstandig nadenken over hoe we machtsongelijkheid kunnen tegengaan in relaties met AI. Activiteiten rond educatie en wetenschapscommunicatie kunnen mensen 'empoweren' door bij te dragen aan betere kennis over AI in de samenleving. Daarnaast kan de ontwikkeling van ethische en juridische instrumenten (wetten, keurmerken, garanties etc...), helpen om de nadelige effecten van machtsongelijkheid te mitigeren tussen de mensen met en zonder kennis en expertise op het gebied van AI. Tot slot is het verstandig om AI te vormen in cocreatie met mensen, die daarmee in staat worden gesteld om de technologie van binnenuit te leren kennen en ermee te experimenteren, en om hun eigen waarden en behoeften mee te geven als ontwerp-eisen aan de ontwikkelaars van de technologie.

Door op allerlei manieren te kunnen meekijken, meeleren en mee-creëren en door controle-mechanismen te ontwikkelen voor AI, worden mensen in staat gesteld om vertrouwen te hebben in hun relatie met AI. Vertrouwen ontwikkelen is een persoonlijk en sociaal proces, dat in de tijd evolueert en dat continu zorg en aandacht nodig heeft. Net zoals mensen na een incident vertrouwen in elkaar kunnen verliezen, kan dat ook tussen een mens en AI. Het is zaak uit te vinden wat dat vertrouwen helpt vormen, wat er nodig is om het in stand te houden, en wat er moet gebeuren als dat vertrouwen is gebroken. In de relatie tussen mensen en AI gaat het namelijk om een relatie die onderhoud vergt, want noch mens noch AI is statisch. Mensen en AI evolueren in interactie met elkaar waarbij mensen zich aanpassen aan AI en AI (middels machine learning) zich aanpast aan mensen. Het gaat erom een evenwicht te vinden in die adaptieve relatie over en weer, zodat mensen de interactie met AI in de samenleving gaan waarderen en vertrouwen, maar ook tot de orde weten te roepen wanneer dat nodig is.

# Bijlage

Hier onder staan de uitgebreidere beschrijvingen van de andere thema's die mensen niet de hoogste prioriteit hadden gegeven, maar waarop zij toch hebben gereflecteerd.

## Nepnieuws, filterbubbels en polarisatie

Een democratische samenleving kan alleen goed functioneren als er persvrijheid is en mensen dus beschikken over goede informatie. Controle vanuit de overheid op het verbieden van door AI gegenereerde content lost niets op, want dan denken mensen dat de overheid censuur oplegt. Misschien moeten distributeurs van informatie de verantwoordelijkheid krijgen om informatie te checken, alvorens het te verspreiden? De accuraatheid van AI kan ook verstoord worden door een filterbubbel effect, en hoe voorkomen we dat? Mensen horen graag wat ze willen horen, maar ze zouden moeten leren om informatie te beoordelen en te onderscheiden tussen echte informatie en informatie gevormd door AI. Mensen hebben agency en moeten kritisch blijven nadenken. Daarbij kan het helpen als er een keurmerk komt of een disclaimer op AI (en zijn gegenereerde content), want dat kan helpen om de vraag te beantwoorden: wanneer is iets écht en wanneer nep?

## Onderwijs, ontwikkeling en vaardigheden

Technologie en AI kan inclusie in het onderwijs bevorderen: mensen die nu benadeeld worden door een leerprobleem zijn hier mogelijk bij gebaat. Maar het kan ook leiden tot stigmatisering: mensen worden sneller in een hokje geplaatst op grond van een karakterisering door AI. Is er met AI nog wel erkenning voor de verschillende menselijke intelligenties (erfelijk, aangeboren, cultureel)? Daarnaast noemt men dat ook docenten moeten worden getraind (*train the trainers*): mensen moeten kritisch leren reflecteren op technologie & ethiek van AI juist leren op scholen net zoals de *immersive skills* en digitale vaardigheden. AI zou een tool moeten zijn, geen vervanger van docenten in het onderwijs. Zo blijft er in het ideale geval meer tijd over voor het zorgen voor en het vooruithelpen van mensen met meer leerproblematiek. Ook is het de vraag of inzet van AI ook vaardigheden en kennis overbodig maken terwijl het juist veel voldoening geeft om die te leren. Respondenten zijn bang voor de gevolgen voor cognitieve vermogens van de mens: creatieve intelligentie, originaliteit, keuzes maken, analytisch denken, ordenen en structureren. De leerprocessen die nodig zijn om resultaten te behalen die AI nu overneemt, zullen dan vervallen. Het is de vraag of we dan nog de vaardigheden hebben om te beoordelen of AI die taken goed doet.

## Vervuiling en milieu

Als AI ondersteuning kan bieden om keuzes te maken die beter zijn voor het milieu en gevolgen productieprocessen op klimaatverandering te monitoren en beheersen, uitstoot van broeikasgassen te verminderen en het gebruik van hernieuwbare energiebronnen te verhogen, dan is dat groene innovatie. Daarnaast zou tijdens de ontwikkeling van AI altijd de vraag gesteld moeten worden: hoe houden we het functioneren van deze technologie zelf ook zo duurzaam mogelijk? Denk aan duurzame productiematerialen en grondstoffen, lokaal en minder complex (zonder kwaliteitsvermindering). Dit thema is misschien niet geschikt als

thematische focus voor de route, maar mensen brengen dit wel naar voren als een aspect dat verstandig is om altijd mee te nemen in het onderzoek en de ontwikkeling.

#### Vooroordelen en inclusie

Reacties die mensen hebben geformuleerd bij dit thema, laten zien dat dit thema voor hen nauw samenhangt met het thema over machtsverhoudingen en machtsmisbruik. Er ligt een verantwoordelijkheid bij de ontwikkelaar om te laten zien hoe AI werkt en eventuele vooroordelen bloot te leggen: transparantie is daarin een kernbegrip. Voorts zou AI getoetst moeten worden op ethische principes en mensenrechten. Dit zijn manieren om vooroordelen te herkennen en exclusie van (groepen) mensen te voorkomen.

Toegankelijkheid (openheid) van de AI-systemen is belangrijk. Daarnaast zou duidelijk moeten zijn dat de input niet garant kan staan voor de kwaliteit van de output van AI: deze disclaimer moet centraal worden gezet. Er moet bewustwording gecreëerd worden dat dit in systemen zit (en bij wie het ontwikkelen). Bij open-source alternatieven is AI minder afhankelijk van grote bedrijven. Misschien zou prioriteit gegeven moeten worden aan ontwikkeling van onderzoeksprojecten die werken met open source technologie.

#### Intelligenties: kunstmatig en biologisch

Respondenten benadrukken het onderscheid tussen verschillende soorten intelligentie bij mensen en hoe dat overeenkomt en verschilt van wat we intelligent noemen bij machines zoals AI. Daarbij noemen ze ook de grenzen van intelligentie: zijn systemen wel intelligent als we ze niet meer kunnen begrijpen? Het is belangrijk om betere en meer gedifferentieerde definities te maken van intelligentie, die ruimte laten voor variatie, en ons af te vragen of de term 'intelligentie' bij machines wel de lading dekt. 'We noemen een rekenmachine ook geen 'artificiële intelligentie' omdat die sneller kan rekenen dan een mens, we noemen het een 'machine die kan rekenen'. Dat AI ook menselijke intelligentie kan overstijgen, hoeft geen probleem te zijn, als het begrijpelijk blijft en uitlegbaar en als het menselijke doelen blijft dienen. AI alignment (met menselijke doelen) is van belang: de doelen van AI moeten in het teken staan van het menselijk welzijn, zeker als het de menselijke intelligentie overstijgt.

#### Relaties en psychologische ontwikkeling

Er zou onderzoek moeten worden gedaan naar verschil tussen menselijke relaties en relaties tussen mensen en AI: wat is de aard en waarde van beide typen relaties? Wanneer is een relatie met AI problematisch? Sommige emotionele relaties (in de zorg, vriendschappen en liefdesrelaties) lijken moeilijk te vervangen door AI, terwijl andere relaties (bijvoorbeeld met een klantenservice, een informatiepunt in een bibliotheek) gemakkelijker door AI kunnen worden vervuld. Dat zou betekenen dat mensen meer tijd overhouden voor de complexere aspecten van een relatie en contact. De inzet van AI kan dus tot voordelen leiden (co-pilot, personeelstekort, kosten, versterking offline relaties) maar ook tot vereenzaming, vervreemding en achterdocht.

#### Arbeidsverdeling tussen mens en AI

Vaardigheden en kennis zijn vaak waar mensen het meest trots op zijn. Mensen verwerven ze door tijd en moeite erin te investeren. Uiteindelijk geeft het voldoening als je ze hebt. Het is

volgens sommigen zelfs 'een deel van onze menselijkheid'. Arbeid overhevelen naar AI is dan ook een zaak die pijnlijk kan zijn. Het kan ervaren worden als een verlies, geven sommigen aan. Anderen leggen een link met thema's als inclusie van mensen met verschillende capaciteiten op de arbeidsmarkt, of stellen de vraag voor welke taken AI meer geschikt is en welke taken beter door mensen kunnen worden uitgevoerd. Arbeidsverdeling en goed samenwerken tussen mensen en AI wordt als thematiek naar voren gebracht. Anderen geven aan dat bepaalde menselijke vaardigheden niet verloren mogen gaan, zoals creatieve vaardigheden. Ook moet het duidelijk zijn wie bepaald werk heeft geleverd: een mens of AI. Authenticiteit van een door mens gemaakt kunstwerk is belangrijk (keurmerk voor kunst?). AI zou moeten dienen als hulpmiddel, niet als vervangmiddel: kan problemen oplossing (arbeidstekort, efficiëntie, kosten). AI zelf kan ook eentonig werk creëren.

### Gezondheid

Weinig mensen hebben deze thematiek prioriteit gegeven, hoewel het thema gezondheid wel in verschillende reacties bij andere thema's terugkomt. Verschillende respondenten geven aan dat er al veel onderzoek wordt gedaan naar dit thema, dus dat het de vraag is of dat voor deze route prioriteit zou moeten zijn. Respondenten noemen verschillende onderzoeksrichtingen van AI in de gezondheidszorg, maar delen ook de zorg dat het onderzoek op basis van grote hoeveelheden data ander type onderzoek gebaseerd op meer specifieke observaties en reflecties kan overvleugelen. Of dat gemeenschappen waarover weinig data beschikbaar zijn door de toepassingen worden benadeeld. Ook bestaat de zorg omtrent preventief medisch onderzoek, die het verloop van iemands gezondheid kan voorspellen, en hoe dit verzekeringen kan veranderen. De volgende vragen kunnen worden afgeleid uit de reacties op dit thema: