

Scenario's AI Agents

Gemaakt voor de conferentie:

Waarden gedreven Dialogen – Agentic AI: De volgende stap in AI

op 30-10-2025

Marleen van Leengoed – VBI Studio

Daniel Tijink - ECP

Inleiding

Deze scenario's zijn gemaakt voor de Conferentie 'Waarden gedreven Dialogen – Agentic AI: De volgende stap in AI' op 30-10-2025. De scenario's zijn bedoeld om bestuurders en beleidsmedewerkers van zowel publieke als private instellingen te inspireren en gesprekken te uit te lokken. Ze verkennen mogelijke toekomstige ontwikkelingen rond AI-agents, waarbij we anticiperen op zowel wenselijke als onwenselijke situaties. Bewust zijn de scenario's enigszins extreem en provocatief geformuleerd - niet per se realistisch, maar wel geloofwaardig genoeg om je te laten nadenken. Deze scenario's kunnen dan ook helpen om strategische opties te bedenken waarvan men nog niet bewust was.

De scenario's zijn gebaseerd op de uitkomsten van drie sessies met de [aanpak begeleidingsethiek](#) over AI-agents. Deze sessies, in de domeinen fintech, inkoop en recruitment, bieden een breed perspectief op de toekomst van AI-agents. De verslagen zijn te vinden op de site van begeleidingsethiek. Hoewel niet volledig, bevatten de uitkomsten uit de cases voldoende informatie om betekenisvolle, nuttige scenario's op te baseren. De scenario's staan los van de bedrijven en de cases en zeggen niets over de visie of plannen van deze bedrijven. Wel zijn we dankbaar dat we hun input mochten gebruiken.

Deze scenario's zijn geen voorspellingen, maar verkenningen die ons denken uitdagen. Ze onthullen mogelijke blinde vlekken en bereiden ons voor op de complexe ethische vraagstukken rond AI-agents. Door de scenario's kritisch te beschouwen, kunnen organisaties en professionals proactief nadenken over AI-implementatie. Uiteindelijk nodigen ze uit tot reflectie, dialoog en strategische voorbereiding op de toekomstige rol van AI-agents in onze samenleving. Wat gebeurt er als we niets doen, en waar is nu echt actie en besluitvorming nodig?

Inhoud

Inleiding	2
Inhoud.....	3
Introductie tot de vier scenario's.....	4
Ontwikkelp proces.....	4
Kenmerken van de scenario's.....	4
De vier scenario's.....	7
❶ Stevige stapjes.....	8
Omschrijving scenario.....	8
Kenmerken	9
Fictieve voorbeelden	10
Feedback uit break-outs van de conferentie.....	11
❷ Iedereen zijn eigen AI-agent.....	12
Omschrijving scenario.....	12
Kenmerken	13
Fictieve voorbeelden	14
Feedback uit break-outs van de conferentie.....	15
❸ Goede voorbereiding is het halve werk	16
Omschrijving scenario.....	16
Kenmerken	17
Fictieve voorbeelden	18
Feedback uit break-outs van de conferentie.....	19
❹ Continue maatschappelijke aanpassing	20
Omschrijving scenario.....	20
Kenmerken	21
Fictieve voorbeelden	22
Feedback uit break-outs van de conferentie.....	23
Conclusies	24
Nawoord.....	26
Colofon	27

Introductie tot de vier scenario's

Ontwikkelproces

Om tot scenario's te komen met de resultaten van de drie sessies begeleidingsethiek, zijn de positieve en negatieve effecten en de handelingsopties die voorkwamen uit de sessies eerst geclusterd. Deze clusters vormden de basis voor de scenario's.



Figuur 1 Illustratie clustering

Daarnaast is het belangrijk te vermelden dat in de sessies vanuit de volgende zes centrale waarden over handelingsopties is nagedacht. Deze waarden zijn zichtbaar in de scenario's:

- Betrouwbaarheid
- Efficiëntie
- Kwaliteit
- Rechtvaardigheid
- Uitlegbaarheid
- Werkgeluk

Maar dat betekent niet dat andere waarden niet relevant zijn. In de verslagen van de sessies worden ook andere relevante waarden genoemd als verbondenheid, kansgelijkheid, eenvoud, vrijheid, veiligheid, onafhankelijkheid en autonomie, menselijkheid, gebruiksgemak, eerlijkheid. Deze waarden komen naar voren in de scenario's maar waarschijnlijk in mindere mate.

Kenmerken van de scenario's

De clustering van handelingsopties en positieve en negatieve effecten is in eerste instantie gestart rondom twee assen. Met op de eerste as de reikwijdte van de AI, die hieronder uitgebreider wordt toegelicht, en op de tweede hoe er op problemen wordt gereageerd, namelijk proactief of reactief. Daarna kwamen twee andere kenmerken ook duidelijk naar voren, namelijk de machtsconcentratie en de innovatie drijfveer.

De reikwijdte van de AI-agents, samen met de correctie en toezicht zegt iets over de autonomie van de AI-agents. Een kleine reikwijdte en veel correctie en toezicht geeft een beperkte autonomie van de AI-agent

en een grote reikwijdte en weinig correctie en toezicht heel veel autonomie. Een andere vraag die hierbij hoort is wie het toezicht en de macht over de correctiemechanismen heeft.

De drijfveer van innovatie heeft hier invloed op, wanneer dat voornamelijk marktgedreven is, is er weinig motivatie om proactief problemen op te lossen die zich manifesteren in de samenleving. Een van de redenen is dat men denkt dat de 'first mover' een grotere kans heeft om marktleider te worden, en dat 'move fast and break things' inmiddels een motto is van veel commerciële partijen. Wanneer de overheid een grotere rol heeft om deze drijfveren in goede banen te leiden en/of wanneer meer vanuit behoeften uit de gehele samenleving wordt gedacht krijg je een heel ander verhaal, en zullen de reikwijdte en correctie en toezicht meer met elkaar in evenwicht zijn.

Hieronder lichten we ze 1 voor 1 kort toe.

Reikwijdte van AI en de arbeidsverdeling tussen mens en AI - agent. Er ontstaan nieuwe werkprocessen, uitdagende, complexere taken, en minder repetitieve taken. Verwachting is dat het betere analyses en andere invalshoeken oplevert en inspirerend kan zijn, maar er is ook kans op blinde autorisatie, 'de-skilling' en verlies aan intuïtie.

- *Doel* – Wie definieert het doel? Is dat de gebruiker, de ontwikkelaar van de AI of de AI-agent zelf?
- *Acties* - Wat voor acties mag de AI-agent doen?
- *Omgeving* - Tot welke data en systemen heeft de AI-agent toegang?
- *AI- AI interactie* - Met welke andere AI-agents mag de AI communiceren?
- *AI-netwerken* - Welke ketens van AI-agents laten we toe? En welke niet? Hoe houden we dit in de hand?
- *Mens- AI interactie* - Zitten er restricties op de input die je kan geven of de output die gegeven kan worden?

Correctie en toezicht – Over het voorkomen dat AI als zodanig problemen oplevert

Verwachting is dat beleid en besluitvorming de toenemende complexiteit niet kan bijbenen, en dat aanvullende maatregelen op het wettelijk kader nodig zijn om problemen te voorkomen.

- *Controlemechanismen* – Zijn er veel verschillende controlemechanismen? Monitoring door werknemers, juristen, management, externe audits, maatschappijonderzoekers, zelf-rapportages van bedrijven of service leverancier, ontvanger van de informatie/ beslissing, AI die AI controleert. Of gebeurt dit achter gesloten deuren bij een bedrijf?
- *Verantwoordelijkheid* - Er ontstaat het risico op onduidelijke verantwoordelijkheid en aansprakelijkheid als het fout gaat als hier geen goede afspraken over gemaakt worden. Wie is er verantwoordelijk als het fout gaat? En waar kan iemand aankloppen over wie wordt beslist, wat voor rechten heeft deze persoon/ entiteit?
- *AI Orkestratie* - Hoe vastomlijnd is de rol van de AI-agents in de flow en wie orkestreert, is dat de mens, een andere AI-agent?
- *Rol van de mens in flow* - Zit de mens 'in the loop' of controleert ze op afstand ('on the loop')?
- *Fallback* - Wat is het fallback scenario als het AI-systeem faalt? Is dat een mens, en hoe zorgen we er dan voor dat we de taken kunnen blijven overnemen?

- *Observeerbaarheid* - In hoeverre wordt het systeem geobserveerd met dashboard en logs? In hoeverre is het proces nog observeerbaar en is het mogelijk om toezicht te houden en fouten op te sporen? Is het proces opgedeeld in stapjes zodat het iets duidelijker is waar het fout is gegaan? En wie kan en mag het systeem observeren?
- *Updates* - In hoeverre is het AI-systeem veranderlijk? Zijn er updates? Is het zelflerend?

Machtsverhoudingen en machtsconcentraties en digitale autonomie. Door het gebruik van AI- agents kunnen deze machtsconcentraties nog meer toenemen. Maar er is ook een tegenbeweging om AI-agents zelf te ontwikkelen omdat dit de digitale geletterdheid verhoogd en de afhankelijkheid verlaagd.

- *Innovatiecapaciteit* – In hoeverre zijn publieke en private instellingen zelf in staat om te innoveren en te experimenteren?
- *Toetreding barrières* - Hoeveel ruimte is er voor kleine spelers om te ontstaan?
- *Scale-ups en digitale autonomie* - Hoeveel ruimte is er voor kleine spelers om uit te groeien tot grote spelers? Of worden investeringen en overnames vooral door de grote spelers gedaan?
- *Controlemechanismen* – Zijn er veel verschillende controlemechanismen, zodat er niet veel macht bij 1 partij kan komen?

Innovatie - drijfveer

De groepen uiteten zorgen over het verlies aan menselijkheid en verbondenheid. De AI-agent heeft beperktere mogelijkheden dan de mens en kan iets als menselijkheid nooit vervangen. Deze effecten hangen sterk af van wat we precies gaan automatiseren met AI en wat niet.

Verwachting is dat de efficiëntie qua uitvoering van taken toeneemt en vooral sneller en goedkoper wordt. Maar de ontwikkeling en onderhoud van AI is niet gratis en het inrichten van monitoring en corrigerende mechanismen ook niet! Het energiegebruik dat nodig is voor het verrichten van de taak neemt waarschijnlijk toe.

- *Drijfveer* - Wordt innovatie met name gedreven door de markt en waar men aan kan verdienen, of kijken we meer naar behoeften vanuit de gehele samenleving? Wat voor rol speelt de overheid hierin?
- *Weloverwogen keuzes* - In hoeverre zijn mensen digitaal geletterd en kunnen ze bewuste keuzes maken voor producten die passen bij hun behoeften, ook op lange termijn? In hoeverre maken we als samenleving bewuste keuzes wanneer we AI-agents wel en niet inzetten? Wat voor rol speelt de overheid hierin?
- *Keuzevrijheid* – in hoeverre kunnen consumenten kiezen, of zijn ze min of meer gedwongen om een bepaald systeem te gebruiken?

De vier scenario's

Hieruit zijn de volgende vier scenario's ontstaan. In de volgende hoofdstukken zullen deze nader worden toegelicht.

- **Scenario 1: Stevige stapjes** – Gedreven door bedrijfsleven met focus op betrouwbare en hoogkwalitatieve AI, omdat het alleen dan echt meerwaarde blijkt op te leveren. Beweging gaat langzamer maar met zekere stappen.
- **Scenario 2: Iedereen zijn eigen AI-agent** – Enkele grote spelers domineren met veel laagrisico toepassingen, AI-agents zijn heel toegankelijk maar niet transparant. Er is veel (risico op) automation bias en onvoorziene systemische effecten.
- **Scenario 3: Goede voorbereiding is het halve werk** – Ketens van AI werken samen. Veel controlemechanismen en regelgeving remmen, maar zorgen voor effectieve inzet.
- **Scenario 4: Continue maatschappelijke aanpassing** – AI-agents vormen complex netwerk dat transparant moet zijn over de optimalisatie-doelen. Bedrijven zijn verplicht opening van zaken te geven.



① Stevige stapjes

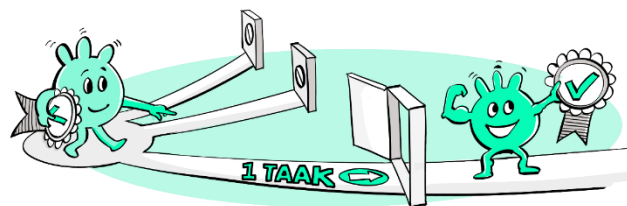
Omschrijving scenario

In 2025 was de opkomst van AI-agents en general purpose AI-modellen. Geavanceerde taalmodellen zoals ChatGPT of Claude en geïntegreerde assistenten zoals Microsoft Copilot werden dagelijks door miljoenen mensen gebruikt. En er komen steeds meer specifieke toepassingen op de markt. Maar naar verloop van tijd bleken de AI -agents vooral een dure gimmick. De AI-agent die het supply-chain managementproces moest stroomlijnen maakte onverwachts kostbare fouten. Sommige MKB-bedrijven stonden zelfs op de rand van faillissement omdat ze te veel overlieten aan onbetrouwbare AI, zonder zelf voldoende kennis in huis te hebben.

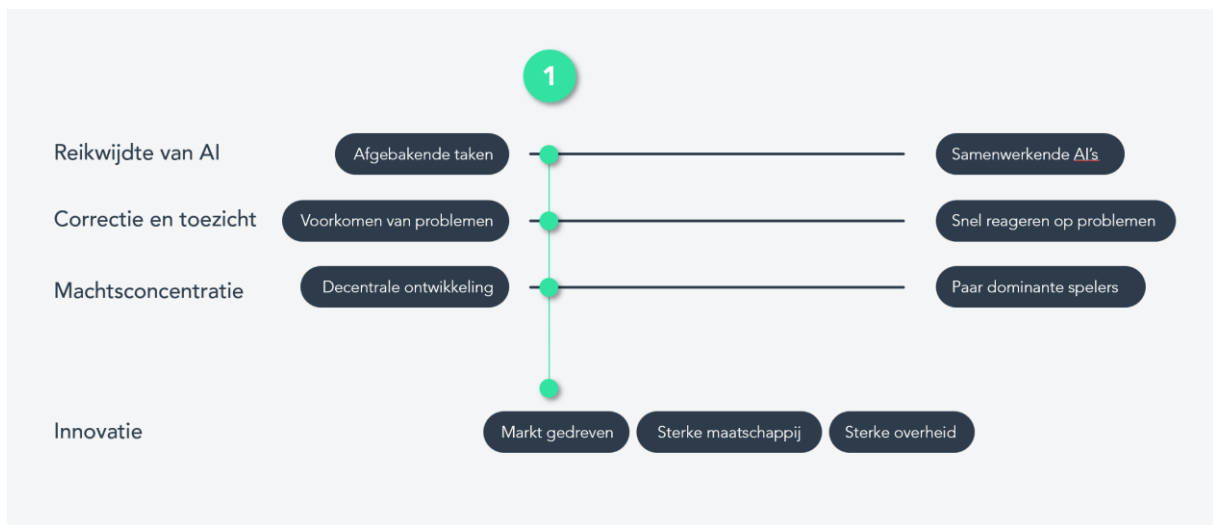
De les is helder: AI-agents moeten betrouwbaar en van hoge kwaliteit zijn om écht meerwaarde te bieden. Anders kost het controleren en corrigeren van fouten meer tijd en geld dan ze opleveren. Dit lijkt het beste te gaan door te focussen op heel specifieke taken en dit takenpakket in kleine stapjes uit te bereiden. Deze aanpak werkt het beste in sterk gereguleerde sectoren zoals zorg en financiën. Daar zijn vanaf het begin duidelijke afspraken gemaakt over wie verantwoordelijk is als het misgaat. Er zijn ‘codes of conduct’ opgesteld, maatregelen tegen ‘automation bias’, en heldere afspraken door de hele keten heen: van de AI ontwikkelaars tot de eindgebruikers.

De AI-agents hebben strikt omliggende taken en kunnen alleen bij bepaalde data. Er is veel aandacht gestoken in een technische goede basis: AI die getraind is op schone, representatieve data en systemen die op elkaar aansluiten. Dit is een grote investering. Maar deze agents zijn ontworpen met privacy en security by design en zijn ontzettend goed in de taak die ze doen. Een bijkomend voordeel: de gespecialiseerde AI-modellen zijn lichter en gebruiken veel minder energie dan de zware algemene modellen. De processen zijn opgedeeld in begrijpbare stappen, zodat mensen kunnen zien waar het eventueel fout gaat en kunnen ingrijpen. Sommige bedrijven trainen deze paraatheid door een paar keer per maand zonder AI te werken. Als een gebruiker met een mens wil praten, kan dat altijd. Het systeem leert van die interactie en wordt zo steeds beter.

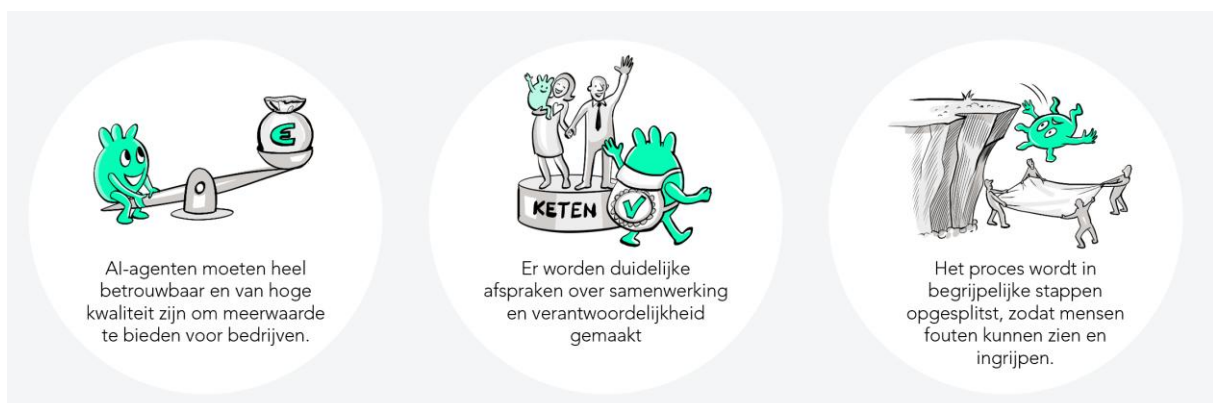
Kennisdeling tussen sectoren, en tussen publiek en private instellingen zorgt voor veel innovatie. Er ontstaat een keurmerk voor algoritmen die voldoen aan wet- en regelgeving. Waar arbeid duur of schaars is, wordt deze stevige aanpak snel overgenomen. Na een valse start waarin we allerlei kanten op zijn gaan rennen, hebben we nu een solide aanpak gevonden, met stevige maar kleine stapjes in de goede richting.



Kenmerken



- **Reikwijdte:** AI Agents mogen alleen specifieke acties doen en hebben beperkte toegang tot data en systemen. De gebruiker bepaald het optimalisatiedoel. AI-Agents zijn betrouwbaar en van hoge kwaliteit.
- **Correctie en toezicht:** De focus ligt op het voorkomen van problemen, want problemen veroorzaakt door AI kosten veel geld. Het takenpakket wordt in kleine stapjes uitgebreid, systemen zijn goed observeerbaar en er zijn duidelijke fallback scenario's en goede samenwerking en afspraken over verantwoordelijkheid in de keten
- **Machtsconcentratie:** Gedecentraliseerde ontwikkeling, veel ruimte voor kleine spelers
- **Innovatie:** Innovatie is met name markt gedreven, de kosten en baten van AI-agents worden geoptimaliseerd.



Fictieve voorbeelden

AI-agents die getraind zijn voor bepaalde taken, in een bepaald domein volgens specifieke regels van een bepaald bedrijf/ domein. Agents doen (complexe) analyses en patroonherkenning maar hebben beperkte gegevensverwerking. Potentieel zijn er meerdere agents die elkaar controleren zodat de foutmarge heel klein is. Denk aan:

- *Klantenservice*: Chatbots die alleen helpen bij het afhandelen van specifieke problemen (en daar goed in zijn).
- *Software schrijven*: Het schrijven van software in bepaalde taal, volgens bepaalde regels.
- *Consumentenmarkt*: Nieuw kapsel uitproberen in digitale spiegel bij kapper, zonder dat gezichtsgegevens en andere privacygevoelige informatie worden opgeslagen.
- *Financieel dataverkeer*: Bij een golf aan issues in de communicatie tussen banken, automatisch betrouwbaar prioriteren, zodat de verantwoordelijke organisatie zich kan concentreren op het meest urgente probleem.
- *HR en recruitment*: AI-agent beantwoordt basis vragen over bedrijf en sollicitatieprocedure via een chat aan de sollicitant. De AI-agent beperkt zich tot een goedgekeurde set aan informatie, zoals jaarverslagen en website.
- *Inkooptoepassingen*: Contracten checken tegen specifieke EU-regelgeving waarvoor ze heel goed getraind zijn.
- *Zorg*: Automatische verslaglegging tijdens een medische procedure, op basis van accurate data, AI-patroonherkenning en sensoren.

Feedback uit break-outs van de conferentie

Positief

- Taakvolwassen Net als andere taakuitvoerders moet AI bewijzen dat het de gevraagde taken kan doen.
- Mens floreert - Mensen vooruithelpen zonder mensen te vervangen
- Betrouwbaar - Systeem is betrouwbaar, specifieke toepassingen in tegenstelling tot generieke
- Wendbaar - Kleine stapjes maakt je wendbaar
- Optimalisatie - Iteratief proces, door te doen leert men
- Duidelijk controlemechanismen - Mens blijft verantwoordelijk. Fouten zijn makkelijk op te sporen.
- Efficiëntie - Minder energie
- Gedeelde waarden - Ontwikkelen samen met domein experts en diverse gebruikers/consumenten
- Geen machtsconcentratie - Intrinsieke motivatie bedrijven, big Tech bepaald niet

Negatief

- Beperkend, vertragend, missen we dan de 'boot', en benutten we de kansen dan wel voldoende?
- Het kost tijd en geld om kwaliteit te leveren
- Internationale druk - er zullen altijd plekken/landen zijn waar het anders gaat
- Haastige implementatie - Sommige partijen verliezen geduld en gaan door met minder betrouwbare systemen
- Digitale kloof - Beperkt aantal mensen zijn digitaal vaardig
- Externe controle nodig - Wie controleert op privacy en technische veiligheid?
- Onafhankelijke monitoring en controle nodig - Wie monitort invloed van mensen en hun taken en invloed op de maatschappij?
- Waar ligt grens van zelflerendheid - Hoe ver moeten we het zelflerende onderdeel laten gaan?
- Controle grote spelers - Hoe houd je de Tech giganten in de hand?

② Iedereen zijn eigen AI-agent

Omschrijving scenario

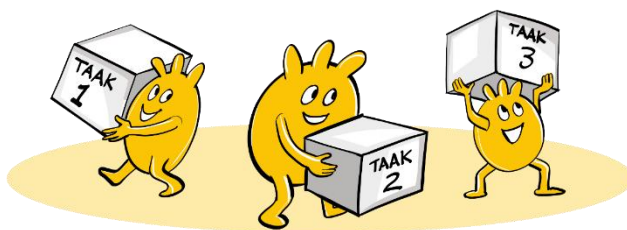
In 2026 zijn **AI-agents alomtegenwoordig in ons dagelijks leven. Ze helpen bij talloze taken** - van vakantie boeken, als persoonlijke assistent in ons werk, tot het helpen bij belastingaangifte - en zijn breed beschikbaar voor laag-risico toepassingen.

De markt wordt gedomineerd door grote bedrijven. Ondanks de diversiteit aan ontwikkelaars blijven we afhankelijk van enkele grote platforms. Succesvolle kleine initiatieven worden vaak overgenomen, wat leidt tot verdere marktconcentratie.

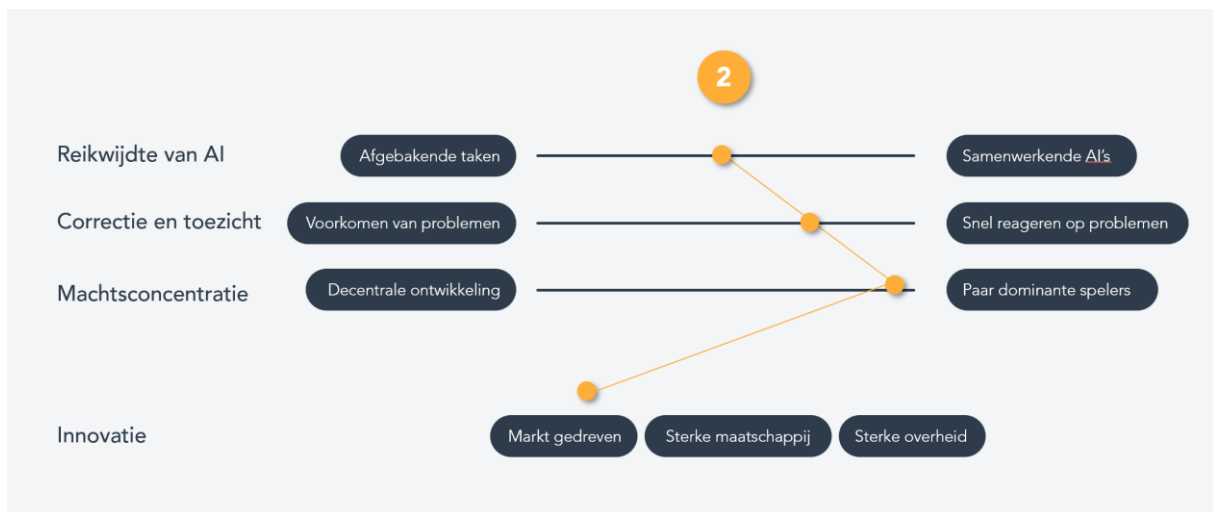
AI-agents zijn heel toegankelijk: ze communiceren vloeiend in natuurlijke taal, maken gebruik van geavanceerde patroonherkenning en worden continu verbeterd qua performance en gebruiksvriendelijkheid. Cybersecurity blijft een kat en muis spel, waarbij AI-agents worden ingezet voor zowel het inbreken in systemen, als het verbeteren van veiligheid van de systemen.

Er ontstaan echter fundamentele uitdagingen. Op papier zijn gebruikers en ontvangers van informatie verantwoordelijk voor controle en correctie van het eindresultaat. Zij hebben gebruiksvoorwaarden getekend. Maar in de praktijk gaat het mis. Gebruikers zijn vaak niet in staat de resultaten kritisch te controleren, en zijn zich vaak niet bewust van de optimalisatiedoelen van de AI-agents, of hoe hun input en data wordt gebruikt als verdienmodel. Door 'automation bias' glippen fouten ongemerkt door. De risico's stapelen zich op, en de systemische effecten zijn te complex voor individuele gebruikers om te doorgronden.

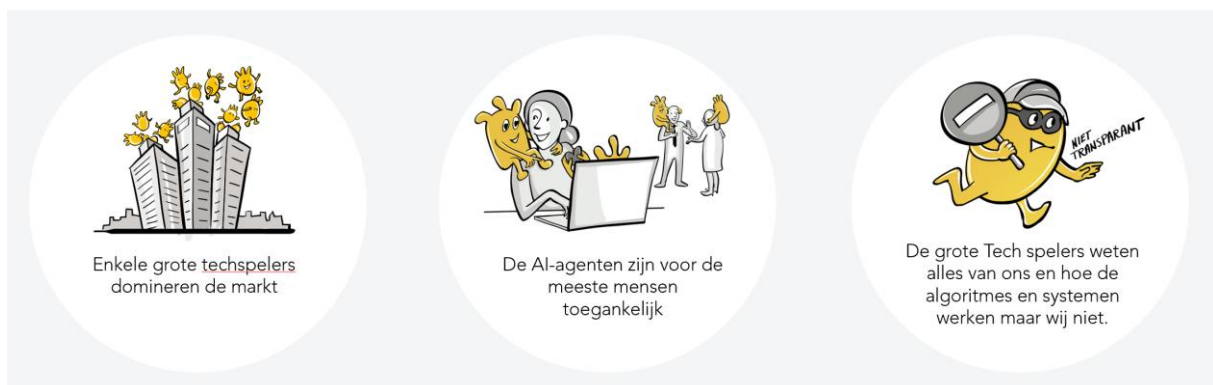
De informatieasymmetrie is zorgwekkend: Tech bedrijven beschikken over vergaande kennis van gebruikers en de systemische effecten, terwijl gebruikers hier nauwelijks inzicht in hebben. Als reactie hierop worden rechtszaken gestart om controle uit te oefenen op grote Tech spelers.



Kenmerken



- **Reikwijdte van AI:** AI Agents mogen alleen bepaalde acties doen en se hebben beperkte toegang tot data en systemen. De AI Act zorgt voor focus op 'laag-risico' toepassingen. Mens kan het proces stoppen maar niet begrijpen.
- **Correctie en toezicht:** Focus op het reageren op problemen, verantwoordelijkheid voor correctie en toezicht wordt hoofdzakelijk bij gebruikers gelegd. Maar de gebruiker kent de optimalisatiedoelen van de algoritmes niet, want systemen zijn niet transparant. De onderhandelingspositie van gebruikers is nihil en correctie vindt plaats door grote rechtszaken.
- **Machtsconcentratie:** Er zijn een paar dominante Tech spelers, succesvolle kleine initiatieven worden vaak overgenomen.
- **Innovatie:** Markt gedreven beslissingen domineren en winstmaximalisatie is belangrijk.



Fictieve voorbeelden

Agents die heel veel redelijk kunnen, generatieve AI gebruikt voor allerlei toepassingen waarbij het wordt geaccepteerd dat er soms fouten worden gemaakt. Denk aan:

- *Werk algemeen:* Werklijstprioritering (in allerlei domeinen) waarbij al het werk wel gedaan moet worden binnen een bepaalde tijd, en mocht het foutief een initiële lage prioriteit krijgen het geen probleem is.
- *Werk algemeen:* Maken van een PowerPoint, samenvattingen van een meeting etc.
- *Klantenservice:* Chatbots als klantenservice, geoptimaliseerd om zo min mogelijk vragen door te sturen naar echte mensen, desnoods met tegoedbonnen.
- *Consumententoepassingen:* Een eigen personal assistant, boodschappen bijhouden met slimme koelkast en agenda, kijken hoe een nieuw kapsel staat in de spiegel van de kapper, voor je aan je knipbeurt begint, nieuwe kleding uitproberen.
- *Financieel:* Hulp bij belastingaangifte (geen financieel advies, alleen formulier uitleggen)
- *HR en recruitment:* AI-agent helpt sollicitant met sollicitatie en basis vragen over bedrijf en sollicitatieprocedure beantwoorden via een chat.
- *Inkooptoeepassingen:* Risicoprofielen van leveranciers maken op basis van publieke data

Feedback uit break-outs van de conferentie

“dit lijkt het pad waar we nu op zitten”

Positief

- Gepersonaliseerde begeleiding - Op maat gemaakte begeleiding bijv. in het onderwijs, en persoonlijke ontwikkeling kan echt op eigen tempo, diepgang en thema's.
- Persoonlijke assistent - die de rottige klusjes uit handen neemt.
- Verhogen van productiviteit - en mensen hebben meer tijd over voor wat ze zelf willen doen
- Informatietoegang - Veel meer informatie binnen handbereik
- Snelle innovatie - De koplopers mogen gratis uitproberen en de volgers hebben veel profijt.
- Toegankelijkheid – ook voor de minder digitaal vaardigen, dit regelt de markt, dat gaat ‘vanzelf’.

Negatief

- Nieuwe verslaving - weer een nieuwe verslaving waar tijd in gestopt gaat worden
- Sociaal isolement -Afnemend menselijk contact, toenemende vereenzaming en maatschappelijke ongelijkheid, minder gemeenschapszin
- Mentale gezondheidsrisico's - Toename van burn-outs doordat repetitieve taken die rust geven vervallen en bore-outs omdat niemand meer zelf nadenkt.
- Onderscheid mens en AI - Waar trekken we de scheidslijn AI vs. mensenwerk. Hoe herkennen we dat?
- Groeiende afhankelijkheid van AI - Waar men niet meer zonder kan
- Andere waarden - Gebrek aan gedeelde waarden en authentieke menselijke dialoog
- Beïnvloeding – (Nog meer) misbruik van persoonlijke informatie voor winstdoeleinden, ongevraagde advertenties en beïnvloeding. Nog meer filterbubbel. AI gaat mens aansturen
- Privacy risico's - Persoonlijke data ligt opgeslagen bij regimes die autoritair worden.
- Grondrecht risico - Verleiding om je grondrechten op te geven voor meer gebruiksgemak (net als je persoonsgegevens bij sociale media)
- Bias en waarheid - Regie op bias en de controleerbaarheid van waarheid, in handen van een paar partijen, en vormt daarmee een dreiging. We kunnen het zelf niet meer controleren.
- Macht in handen van een paar commerciële partijen, “big brother” maatschappij.
- Feature creep, de technologie lost alles wel op. Maar dit is niet altijd het juiste middel.
- Systeemrisico's -potentieel desastreuze gevolgen door kleine fouten op grote schaal
- Wetgeving kan de razendsnelle technologische ontwikkelingen niet bijbenen
- De-skilling Verlagend van de autonomie en creativiteit, kennis verdwijnt

③ Goede voorbereiding is het halve werk

Omschrijving scenario

Persoonlijke AI-assistenten waren overal: ze hielden agenda's bij, controleerden koelkasten, coachten bij het sporten. Maar al snel werden de nadelen duidelijk. Mensen werden eenzamer door minder persoonlijk contact, AI maakte bizarre fouten (hallucinaties) en niemand wist meer wanneer ze met een echt mens praatte. Het MKB werd getroffen door cybersecurityproblemen door gekoppelde AI-systemen, en het energieverbruik schoot omhoog. Bedrijven, burgers en belangengroepen trokken aan de bel.

Er kwam strenge regelgeving: nieuwe AI-functies en koppelingen moesten eerst gevalideerd worden voordat ze vrijgegeven mochten worden. Er komen juridische kaders voor energieverbruik, effectiviteit en kwaliteit. De juridische kaders voor betrouwbaarheid, rechtvaardigheid en uitlegbaarheid worden herzien. Er komt een ethische code voor ontwikkelaars, en keurmerken voor algoritmes die aan de wet voldoen (AVG, AWB). Met als doel AI veilig en effectief maken in gebruik.

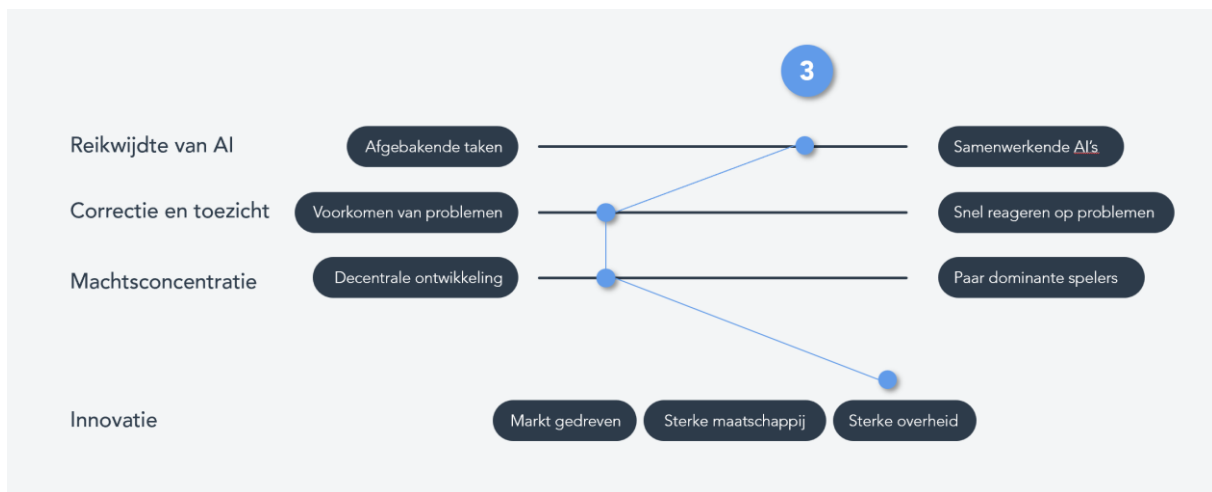
Door een combinatie van regelgeving voor validatie en vrijgave van AI-agents en goede technische oplossingen voor security en privacy by design, lukt het om AI-agents veilig en effectief te maken in gebruik. Een belangrijke doorbraak hierbij zijn AI-systemen die kunnen aangeven wanneer ze iets niet weten of ergens onzeker over zijn. De meeste AI-agents hebben een breed takenpakket en communiceren met elkaar zonder menselijke tussenkomst, bijvoorbeeld via mail. De interactie tussen systemen is complex en niet altijd uitlegbaar, maar er zijn wel verschillende stevige veiligheidsmaatregelen. De maatschappelijke kosten en baten worden periodiek gemonitord. Als het nodig is, wordt bijgestuurd. AI-systemen kunnen tijdelijk uit de markt gehaald worden bij het niet voldoen aan de kaders of teruggedraaid naar een eerdere versie. Alle meldingen worden door een onafhankelijke instantie geanalyseerd. Fouten worden geanalyseerd en leiden tot structurele verbeteringen. De mens-machine (AI) interactie wordt continu gemonitord en er wordt een bewuste afweging gemaakt tussen de taken van de AI-agent en verantwoordelijkheid van de mens. Limieten van automatische besluitvorming en benodigde training voor de mens worden aangepast zodat de risico's beheersbaar blijven.

Er is meer inzicht in waar menselijk contact nodig is in een bedrijfsproces. Persoonlijk contact wordt dan ook bewust ingezet. Gebruikers kunnen altijd een second opinion door een mens aanvragen. Er is meer inzicht in wat wij als mensen nodig hebben om tot bloei te komen, wat niet met AI-systemen ondersteund kan worden, zoals fysieke ervaringen en menselijk contact en creativiteit. Er komt een verschuiving in werk voor mensen. Aan de ene kant wordt mensenwerk meer handenwerk, in de bouw, zorg, fysieke winkels. En aan de andere kant komen er nieuwe intellectuele banen: maatschappijonderzoekers, data-anthropologen, AI-ontwerpers, systeemdenkers en AI-geletterdheidsonderwijzers.

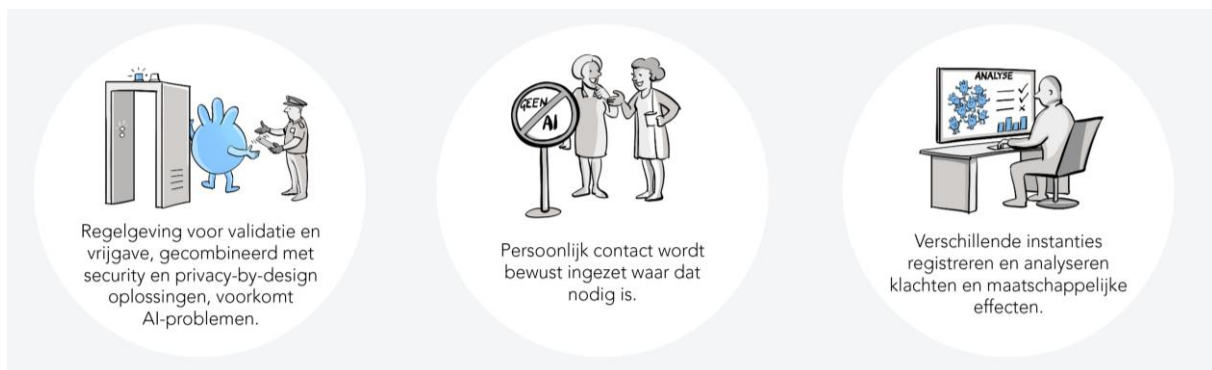
De strenge regelgeving remt innovatie af en bedrijven moeten schakelen om dit goed in te richten. Maar het leidt wel tot creatieve oplossingen die de samenleving versterken. Met als conclusie dat goede voorbereiding meer is dan het halve werk.



Kenmerken



- **Reikwijdte van AI:** AI-agents vormen een netwerk en communiceren met elkaar, daardoor is de interactie tussen AI-systemen complex en minder goed uitlegbaar. Maar de optimalisatiedoelen van de AI-agents dienen het algemeen belang. Bepaalde AI-ketens zijn niet toegestaan, en ook zijn er restricties in toegang tot data en systemen.
- **Correctie en toezicht:** Problemen worden voorkomen door AI-systemen vooraf in een sandbox omgeving, met echte gebruikers te testen. Regelgeving voor validatie verplicht dit te doen, voor een systeem de markt op mag. Ook moet er gebruik gemaakt worden van privacy en security by design. Er zijn veel verschillende controlemechanismen ook wanneer systemen eenmaal op de markt zijn.
- **Machtsconcentratie:** Er is geen grote machtsconcentratie, maar veel decentrale ontwikkeling met een grote onderhandelingspositie voor gebruikers, waarmee ook kwetsbaren worden beschermd.
- **Innovatie:** Een sterke overheid bepaald de regels waarbinnen ontwikkeld mag worden.



Fictieve voorbeelden

AI-agents die met elkaar samenwerken, maar van tevoren wel zijn getest onder realistische omstandigheden en waarvan de risico's worden gemitigeerd, doordat gebruik beperkt is, of door technologische aanpassingen. Denk aan:

- *Klantenservice*: Chatbots die met elkaar samenwerken als eerste aanspreekpunt, en kunnen aangeven als ze iets niet kunnen oplossen of weten. Een gebruiker kan altijd aangeven een mens te willen spreken.
- *Zorg*: Virtuele verpleegkundige begeleidt mensen tijdens MRI-scan en stelt hen gerust, terwijl een andere AI-agent de scan instelt en start.
- *Financieel dataverkeer*: Patronen herkennen in transacties om criminele activiteiten op te sporen.
- *HR en recruitment*: Zowel HR en recruitment als de sollicitant gebruiken AI-agents, maar deze ketens worden wel van tevoren getest en voortdurend gemonitord, wat betekent dat alleen bepaalde AI-agents mogen worden gebruikt.
- *Inkooptoeepassingen*: Ook hier zijn ketens van AI-agents aan het werk. Voor het schrijven van de aanbesteding, het schrijven van de inzendingen door leveranciers, het maken van een overzicht van de inzendingen, het doen van de legal checks (aan beide kanten). Om de betrouwbaarheid in de keten te waarborgen moeten de individuele AI-agents aan bepaalde eisen voldoen, zo moeten AI-agents bijvoorbeeld aan elkaar delen hoe zeker ze zijn van bepaalde inzichten.

Feedback uit break-outs van de conferentie

Positief

- Sterke overheid - Een sterkere overheid voor een sterkere maatschappij
- Visie-ontwikkeling op politiek niveau
- Grondrechtenborging, en maatschappelijke waarden centraal: veiligheid en privacy i.p.v. winstbejag, voorkomen van sociale media scenario
- Duidelijke kaders en richtlijnen die richting en duidelijkheid bieden voor organisaties. Minder cowboys
- Herwaardering van de mens en niet-digitaal contact, focus op winst/geld wordt minder
- AI bewust inzetten – AI-agents inzetten met een doel – de mens is de baas. AI tegenhouden die niet voldoet aan waarden: Tegenkracht big Tech en big money. Keuze voor AI als dat waarde levert
- Eerst denken, dan doen – Kritische blik, maar ook ruimte om te experimenteren. AI-management systeem ISO 42001, ISO 47001
- Nieuwe beroepen (data antropologen), jong beginnen, ook met nieuwe opleidingen is key.
- Stimuleren persoonlijk contact “stel een set met basisregels op”, meer ontwikkeling op de kwaliteit van menselijk contact.
- Hoe kom je hier? Met universeel basis inkomen en inzetten op egaliserend beleid?
- Hoe kom je hier? Principe gebruiken van de vervuiler betaald, dus AI- leveranciers moeten betalen voor toezicht en validatie?
- Decentrale ontwikkeling, zodat macht wordt verdeeld. Digitale soevereiniteit.

Negatief

- Systeem complexiteit – Dit scenario is gericht op vergaande digitalisering. Hoeveelheid AI & verknoping van al die agents. Dit is te complex om te controleren
- Bureaucratische barrières - Hoge kosten compliance, veel bureaucratie, niet wendbaar, niet werkbaar, afhankelijkheid van regulators, mogelijke bottleneck, ontmoedigd praktische implementatie en adoptie. Vertraging door te veel regelgeving. Kan het simpeler, wat is je doel en hoe bereik je dat?
- Toetredingsdrempels - Te hoge eisen om als bedrijf in te stappen, vrijheid burgers verminderd
- Sterke overheid - Overheid & goede besluitvorming is niet evident, kan mensenrechten ook verslechteren. Overheid heeft steeds meer behoefte aan centrale afspraken, en standaardisatie m.b.t. AI, decentrale ontwikkeling kan hier haaks op staan. Overheid kan AI-systemen van de markt halen, hoe bepalen we wanneer dit gebeurt?
- Concurrentie en defensiepositie: Als dit alleen in Nederland is, verzwakt dat onze positie
- Onafhankelijke instanties bestaan niet, iedereen heeft een belang.
- Omzeiling van regels: Mensen gaan geitenpaadjes zoeken
- Regelgeving vs. eigen verantwoordelijkheid: Meer regelgeving maakt de wereld niet per se beter, wat trekt de overheid naar zich toe? En wie gaat daardoor in afwachtende stand? Er is een betere balans nodig tussen regelgeving en menselijke maat en eigen verantwoordelijkheid.
- Vertrek van kennis
- Hotel California: wat als je bent aangemerkt als fraudeur, hoe kom je ooit van het stempel af?
- Onwettelijk dat het mogelijk zou zijn om dingen ‘voor’ te zijn.
- Nieuwe beroepen - Neem als uitgangspunt menselijke waardigheid niet (gereguleerde) efficiëntie. Mensen moeten kunnen blijven concurreren met de techniek. Zijn al die nieuwe beroepen echt nodig of moeten we het anders oplossen?
- Decentraal is risicovol, beter een iets stabiele ontwikkeling, dan te klein, wereld wordt toch overgenomen
- Energie impact

④ Continue maatschappelijke aanpassing

Omschrijving scenario

Stel je voor, we leven in 2030 in een wereld waar AI-agents autonoom met elkaar communiceren, en zich continue aanpassen omdat ze zelflerend zijn.

In 2026 besteden bedrijven en overheidsorganisaties nog veel tijd aan vaste afspraken over verantwoordelijkheid en aansprakelijkheid, maar naarmate AI-systemen steeds complexer en zelflerend werden, bleek dit praktisch onmogelijk. Vaak was er een keten van AI-agents aan het werk. Als er iets misgaat, kon men lastig begrijpen wat er precies was gebeurd. Je kon ook niet altijd goed verschil zien tussen echte mensen en AI-agents, wat extra problemen opleverde. Journalisten en onafhankelijke instanties onderzoeken de maatschappelijke impact, maar hun studies en rechtszaken kosten jaren, terwijl de systemen ondertussen alweer zijn veranderd. Een paar grote spelers zoeken voortdurend mazen in de wet.

Conclusie: papieren afspraken schieten tekort bij dynamische, complexe systemen. Daarom is de **focus verschoven van vooraf vastgelegde regels naar transparantie, digitale geletterdheid en snelle respons op incidenten en meer dynamische regelgeving.**

Nieuwe transparantieregels verplichten bedrijven om:

- de optimalisatiedoelen, leermethoden en databronnen van hun algoritmes te openbaren.
- 360-graden feedback op AI-agents te verzamelen, met daarin een duidelijk overzicht van de mogelijkheden en beperkingen en risico's van hun systemen.
- onafhankelijke instanties de verstrekte informatie te laten controleren.

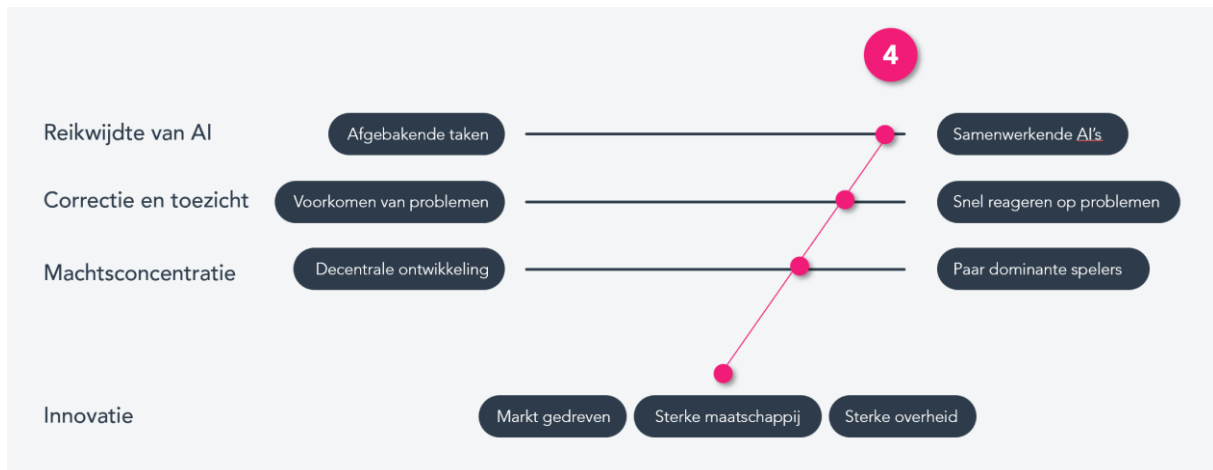
Daarnaast komt er een goede regeling om klokkenluiders te beschermen, en een meldpunt waar incidenten kunnen worden gerapporteerd. Er wordt intensief geïnvesteerd in digitale geletterdheid en maatschappelijke normontwikkeling. Er komt een soort rijbewijs voor AI, een AI-bewijs. En burgers vormen netwerken om kennis te delen. Sommige groepen kiezen voor privacy vriendelijke alternatieven. Tegelijkertijd gebruiken sommigen AI om de rechtsstaat te ondermijnen en nepnieuws te verspreiden. Verschillende mensen gebruiken de technologie voor verschillende doeleinden. Mensen verenigen zich om algoritmes rondom verschillende idealen. Er is voortdurend debat over de wenselijkheid van bepaalde GAI-toepassingen.

De markt wordt gedomineerd door grote bedrijven. Ondanks de diversiteit aan ontwikkelaars blijven we afhankelijk van enkele grote platforms. Succesvolle kleine initiatieven worden vaak overgenomen, wat leidt tot verdere marktconcentratie.

Menselijk werk verschuift naar werk en taken die AI niet kan vervangen, waarvoor persoonlijk contact belangrijk is. Opdrachtnemers worden bijvoorbeeld gevraagd om feedback zonder tussenkomst van AI. Mensen worden zich steeds meer bewust van waarvoor menselijk contact essentieel is en wat je niet met een AI-systeem kan oplossen. Er komt steeds meer aandacht voor niet-digitale kennis en het overdragen van die ervaring. Daarnaast is er meer werk gericht op emoties en contact maken: Dans, muziek, ontspanning in de natuur.

Regelgeving wordt meer dynamisch om snel op de effecten van 'wicked' systemen te kunnen reageren. Zo kunnen toepassingen die eerst werden toegestaan worden verboden. **Het vergt veel flexibiliteit van mensen, continue maatschappelijke aanpassing.**

Kenmerken



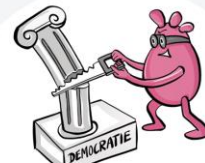
- **Reikwijdte van AI:** AI-agents vormen een netwerk en communiceren met elkaar, interactie tussen AI-systemen is complex en minder goed uitlegbaar. Bepaalde AI-ketens worden niet gebruikt, omdat de bevolking weet dat dat gevaarlijk kan zijn. AI-systemen zijn veelal zelflerend en passen zich continue aan. De reikwijdte van de AI-agents is dus heel hoog.
- **Correctie en toezicht:** Er is verplichte transparantie over systemen en optimalisatiedoelen. Er is decentraal toezicht door de maatschappij. De maatschappij reageert snel op problemen door goede informatie-uitwisseling tussen burgers. Afspraken vooraf werkten niet, continue aanpassing is nu de norm.
- **Machtsconcentratie:** Markt wordt gedomineerd door grote spelers, succesvolle kleine spelers worden vaak overgenomen.
- **Innovatie:** Markt gedreven, maar maatschappij beïnvloedt markt door kritische keuzes. Digitale geletterdheid onder bevolking is heel hoog, iedereen die AI gebruikt is verplicht een AI-bewijs te halen. Mensen verenigen zich rondom verschillende idealen.



De digitale geletterdheid is hoog. Mensen weten welke systemen voor welk doel te gebruiken, en wat de optimalisatiedoelen zijn van de AI-systemen.



Bedrijven moeten transparant zijn over de optimalisatiedoelen en beperkingen van de AI systemen.



AI agenten worden voor allerlei toepassingen gebruikt, ook om bijvoorbeeld fake-news te maken en de democratie te ondermijnen.

Fictieve voorbeelden

AI-agents die met elkaar samenwerken, die worden gebruikt door mensen die een AI-bewijs hebben gehaald en op de hoogte zijn van de optimalisatie doelen van de algoritmes. Het onderscheid tussen mens en algoritme is heel duidelijk. Denk aan:

- *Klantenservice:* Chatbots die met elkaar samenwerken in een klantenservice en waarvan klanten goed op de hoogte zijn wat ze kunnen verwachten van deze chatbots.
- *Zorg:* Grootschalig onderzoek naar verbanden tussen levensstijl/omgeving en ziekten met diverse databronnen (luchtfoto's, grondwatermetingen, gezondheidsdata)
- *Consumententoepassing:* Tussenpartijen die flexibiliteit bieden om verschillende algoritmes naast elkaar te gebruiken en uit te proberen wat voor jou het beste werkt (vergelijkbaar met DuckDuckGo die verschillende GPAI aanbiedt)
- *Financieel dataverkeer:* Automatische effectenhandel: we grijpen alleen in als het écht mis gaat maar begrijpen eigenlijk niet meer wat er gebeurt
- *HR en recruitment:* Zowel HR en recruitment als de sollicitant gebruiken AI-agents in het hele proces. Tijdens het sollicitatiegesprek worden heel bewust geen chatbots gebruikt ter ondersteuning. Een dataspecialist en securityspecialist monitoren de AI-agents met rapportages over doorlooptijden, conversieratio's, feedback en voorspellende waarde. Ze vergelijken kandidaten met en zonder AI-toepassing en monitoren automation bias. Deze informatie wordt gebruikt in de rapportage over het algoritme.
- *Inkooptoepassingen:* Ook hier zijn ketens van AI-agents aan het werk. Voor het schrijven van de aanbesteding, het schrijven van de inzendingen, het maken van een overzicht van de inzendingen, het doen van de legal checks (aan beide kanten), kwaliteitscontrole. Ook hier worden de algoritmes goed in de gaten gehouden en bijgestuurd waar nodig.

Feedback uit break-outs van de conferentie

Positief

- Enorm productievermogen, veel innovaties mogelijk
- Autonomie - Voor iedereen behoefte vervulling, iedereen staat sterk in hun waarden
- Burgers als watchdog - Sterk maatschappelijk middenveld.
- Zelf reinigend systeem - Het klinkt als de situatie die ook met de AI-verordening wordt beoogd, een zelf reinigend systeem.
- Verantwoordelijkheid nemen - Snel reageren op problemen is verantwoordelijkheid nemen, schade oplossen i.p.v. juridisering en kosten externaliseren.
- Kennisdeling, vertrouwen in kennis als oplossingsmiddel
- Saamhorigheid, eigenaarschap
- Waardenvolle besteding van tijd - Vervelende processen worden makkelijker, we kunnen focussen op andere zaken die echt verschil maken, meer mensen in cruciale beroepen.
- Transparantie: Loggen van welke input resulteert in welke actie en output.
- Mensen worden steeds meer digitaal geletterd, en herkennen nu ook al content die ze een half jaar geleden nog als echt zouden hebben bestempeld.
- Nodig: onderwijs en levenslang leren, anticorruptie, verenigingszin, sterke onafhankelijke overheid, experts en bedrijven met goede bedoelingen. We moeten open source gebruiken, mits niet anders kan. We moeten ook een infrastructuur bouwen voor AI. En er moeten veel centra zijn waar de niet-digitaal geletterden hulp kunnen krijgen. Er is een enorme cultuurverandering nodig voor dit scenario, een alternatief voor het liberaal kapitalisme.

Negatief

- Misbruik mogelijk - Door digitale geletterdheid kunnen mensen ook misbruik maken doordat ze kennis hebben, bijvoorbeeld voor het ondermijnen van de democratie, criminaliteit. Dit moeten we tegenhouden door wet-en regelgeving, en afdwingen door tooling opgelegd aan bedrijven.
- Digitaal geletterdheid zal beperkt blijven, dit zien we in de misinformatie die bepaalde politieke partijen tijdens de verkiezingen hebben verspreid en worden geloofd.
- Digitale kloof - AI vergroot de kenniskloof, gaat zorgen voor veel leed. Digibeten mogen niet worden uitgesloten.
- Technocratie - Burger wordt opgejaagd, vergroot onrust en wantrouwen, en dan krijg je een omgekeerde wereld, AI die overneemt, vervreemding in werk, verlies van controle.
- Geen gezamenlijke richting - Talenten, intenties en waarden van mensen zijn niet eenduidig, we gaan veel verschillende kanten op.
- Geconcentreerde macht - Kapitaal flow naar Silicon Valley en/of Shanghai. Veel macht bij een paar bedrijven ondermijnt de transparantie en dynamische respons van dit scenario.
- Afhankelijkheid - Ongeremde super AI die het goede wil doen, maar zo complex geworden is dat wanneer het stopt met werken geen mens het meer kan fixen.
- Klokkenluiders - We hebben alternatieven nodig voor klokkenluiders, want het loopt voor hen niet goed af.
- Wetgeving weg gelobbyd
- Flexibele regelgeving is slecht voor ondernemerschap. Er moeten stabiele juridische kaders zijn waar burgers en bedrijven op kunnen terugvallen.
- Waardenvolle besteding van tijd - Tijd krijgen voor andere zaken die zinvol zijn is een utopie zolang financiële prikkels de drijvende krachten zijn.
- Hoog energieverbruik, te kort aan grondstoffen
- Voor wie doen we dit?
- Accountability - Wie is er uiteindelijk verantwoordelijk als er iets misgaat?
- Transparantie – bedrijven zullen dit nooit écht doen, want dan verliezen ze hun positie.
- Fouten door AI-agents worden wel gemaakt, we gaan uit van een goede reactie en controlerende mechanismen achteraf, maar is dat voldoende of is het leed dan al geschied?

Conclusies

In de scenario's en de feedback op de scenario's komen een aantal hoofdthema's naar voren. We zien dat er een aantal acties zijn die in alle situaties relevant lijken. Dit zijn acties die in alle scenario's werden gewaardeerd of aangedragen om een negatief gevolg uit het scenario te voorkomen. Deze acties zijn geclusterd in drie hoofdthema's:

- 1) Mensenrechten beschermen
- 2) Soevereine technologie en controlemechanismen op technologie
- 3) AI-agents bewust inzetten en verantwoordelijkheid

Daarbinnen kunnen we drie soorten acties, oftewel handelingsopties, onderscheiden:

- de omgeving van de AI-agents, regelgeving en organisatie
- het menselijk gedrag en benodigde vaardigheden/training en de cultuur
- de technologie zelf, in dit geval de AI-agents, de kenmerken van het systeem en dienst

Daarbij zie je dat deze soorten acties ook een beetje in evenwicht moeten zijn, te veel van het een of andere is niet goed, en lost het probleem ook niet voldoende op.

1) Mensenrechten beschermen

Regelgeving is nodig om mensenrechten te beschermen en controlemechanismen in te richten. Dit moet zonder te veel bureaucratie, zodat kleine partijen en nieuwe toetreders tot de markt niet te veel barrières ondervinden. Er moet een balans zijn tussen eigen verantwoordelijkheid en regelgeving. Er moeten stabiele juridische kaders zijn waar burgers en bedrijven op kunnen terugvallen en soms is een snelle ingreep vereist om kwetsbaren te beschermen. Systeemrisico's moeten in kaart worden gebracht en de invloed op de maatschappij gemonitord.

Een van de belangrijke elementen hierin is om onderscheid te kunnen maken tussen mens en AI. Met de european digital identity wordt het mogelijk om dit op een veilige en privacy vriendelijke manier te doen. De ontwikkeling moet dan wel wat versneld worden.

Daarnaast moet er altijd ruimte zijn voor een menselijke benadering omdat er altijd bedrijven, mensen en situaties zijn die niet in de gedefinieerde hokjes passen, of zelfs niet meetbaar zijn, maar wel belangrijk. Mensen mogen niet worden uitgesloten en kwetsbaren moeten worden beschermd, maar dit is in bijna alle scenario's een grote uitdaging.

2) Soevereine technologie en controlemechanismen op technologie

Tijdens de sessie wordt een aantal keer de zorg geuit hoe we onze markt kunnen beschermen tegen ongewenste ontwikkelingen buiten de EU, want er zullen altijd landen zijn waar het anders gaat. Een pluspunt is dat de Europese markt voor veel bedrijven groot genoeg om op te focussen. Er is ook zorg over het gebrek aan controle op grote technologie spelers en onze afhankelijkheid hiervan. In alle scenario's komt naar voren dat een diverse robuuste set aan controlemechanismen nodig zijn, onder andere de regie over bias en waarheid, denk aan onafhankelijke journalistiek, en maatschappelijke organisaties bijvoorbeeld. En een derde punt van zorg is de mogelijkheid voor kleine bedrijven om zelf uit te groeien tot een grote speler. Kapitaalflow richting kleine ethisch opererende bedrijven is nodig. Maar er is ook bescherming nodig, andere businessmodellen en bedrijfsstructuren, zodat deze kleine bedrijven niet voor ze kunnen uitgroeien tot grote spelers, overgenomen worden door een van de grote technologie bedrijven.

Het is mogelijk AI-agenten minder te laten hallucineren en privacy en security by design in te bouwen. (scenario1 en 3). Dit is noodzakelijk om mensenrechten te beschermen en voor controlemechanismen om goed te functioneren in een democratie. Maar dit is niet vanzelfsprekend, ook hiervoor zijn andere businessmodellen, inkoop strategieën en regelgeving nodig.

Het uitgebreid valideren van een AI-agent voor het de markt op mag, net zoals we bij medische apparatuur doen zal te veel vertraging opleveren verwacht men, maar dat betekent niet dat de technologie dan maar zomaar kan worden uitgerold en dan maar kijken wat er gebeurt. Het verplichten van een sandbox omgeving voor het testen van de AI-agent en monitoring van effecten en gedrag van echte gebruikers, en het gecontroleerd opschalen van de technologie zijn allemaal tussenstappen die genomen kunnen worden.

Er is wel snelle actie vereist om dit allemaal goed in te richten.

3) AI-agents bewust inzetten en verantwoordelijkheid

Een kernuitdaging blijft hoe we AI-agents bewust kunnen inzetten waar dat meerwaarde biedt, en welke ketens van AI-agents we wel en niet toestaan. Soms botst efficiëntie met menselijke waardigheid, soms versterkt het die. Datzelfde geldt voor duurzaamheid. Hoe leiden we dit in goede banen? Er moet een balans zijn tussen eigen verantwoordelijkheid van bedrijven en burgers en regelgeving en toezicht vanuit de overheid. Daarnaast wordt genoemd dat bewust inzetten nodig is om de complexiteit in de hand te houden, omdat het systeem anders niet meer te controleren of te repareren is.

We moeten nu nadenken over welke ketens van AI-agents we toestaan, en aan welke randvoorwaarden deze moeten voldoen, zoals de reikwijdte van de AI-agent en optimalisatiedoelen die de AI-agent wel en niet mag hebben. In hoeverre laten we doelen toe die tot doel hebben mensen zo lang mogelijk op een platform te houden of zoveel mogelijk persoonlijke details te laten delen? In hoeverre laten we toe dat een AI-agent mag hallucineren? Moeten deze doelen zichtbaar zijn voor de gebruiker? In hoeverre kunnen mensen zelf verantwoordelijk zijn voor de uitkomsten van een AI-agent? Wat voor training en skills zijn nodig? Wanneer kunnen ze deze verantwoordelijkheid wel en niet dragen? Systemische effecten zullen voor het individu niet zichtbaar zijn. Waar ligt de verantwoordelijkheid van de technologie leverancier?

Ai-agents zullen in de toekomst wellicht ook 360 graden feedback ontvangen en een functioneringsgesprek hebben en een beoordeling. Goede afspraken over verantwoordelijkheid en aansprakelijkheid in de keten zijn nodig voor als er iets misgaat. Achteraf corrigeren met rechtszaken gaat te langzaam, dus moeten vooraf goede afspraken en terugvalmechanismen ingericht worden.

De analyse van de scenario's en de feedback daarop laten een aantal acties en randvoorwaarden zien die in alle situaties relevant zijn. Door deze acties te identificeren, kunnen we een gestructureerd kader ontwikkelen dat zowel de bescherming van mensenrechten als de soevereine technologie en de verantwoordelijke inzet van AI-agents mogelijk maakt. Zo kunnen we de risico's van AI-technologieën minimaliseren en de maatschappelijke impact ervan maximaliseren. Dit vereist een geïntegreerde aanpak gericht op de omgeving van de AI-agents, het menselijk gedrag en de technologie zelf, en die dus zowel technische, juridische als culturele aspecten omvat.

Nawoord

We kregen veel aanvragen om deze scenario's te delen. Tijdens de conferentie hebben de scenario's uitgenodigd tot reflectie en dialoog en we hopen dat ze zullen bijdragen aan strategische voorbereiding op AI-agents en snelle besluitvorming rondom dit thema. We nodigen jullie uit om met dit document verder aan de slag te gaan. Dank aan iedereen die bij de conferentie aanwezig was en input heeft gegeven op deze scenario's.

Colofon

Scenario's AI Agents © 2025 by Marleen van Leengoed, Daniel Tijink is licensed under CC BY-NC-SA 4.0.
Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International

To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-sa/4.0/>

This license requires that reusers give credit to the creator. It allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, for noncommercial purposes only. If others modify or adapt the material, they must license the modified material under identical terms.

Auteurs: Marleen van Leengoed ([VBI Studio](#)), Daniel Tijink ([ECP](#))

Illustratie tekeningen: Mary-Ann Pepers

Copy-editing: Myrthe Tiemessen