The Hague Centre
for Strategic Studies

In de greep van AI?

De impact van de AI-Transformatie op de Maatschappelijke Stabiliteit

Gerben Bakker, Sofia Romansky, Frank Bekkers, Bryan Smeets en Pieter Bindt

Maart 2024

The Hague Centre
for Strategic Studies

In de greep van AI?

De impact van de AI-Transformatie op de Maatschappelijke Stabiliteit

Auteurs:
Gerben Bakker, Sofia Romansky, Frank Bekkers,
Bryan Smeets en Pieter Bindt

Met medewerking van:
Paul Sinning, Tim Sweijs, Bart Vossebelt,
Anna Sophie den Ouden, Linde Arentze en Julia Döll

Cover image source:

Maart 2024

Dit onderzoek is verricht door HCSS in opdracht van de Nederlandse politie, als onderdeel van het meerjarige programma *Strategische Monitor Politie*. Dit programma heeft tot doel een 'van buiten naar binnen' en toekomstgerichte blik op en duiding van relevante trends en ontwikkelingen te ge-ven, om daarmee de strategievorming van de Nederlandse politie te ondersteunen. Deze studie is nadrukkelijk niet bedoeld als een alomvattende wetenschappelijke analyse, maar vormt een eerste verkenning waarmee het een basis biedt voor dialoog.

Het onderzoek voor dit rapport is afgerond in **februari 2024**. Gebeurtenissen of ontwikkelingen die plaatsvonden in de periode tussen afronding en publicatie zijn niet van invloed geweest op de bevin-dingen.

Inhoudsopgave

Management-samenvatting

De aandacht voor AI heeft wereldwijd grote hoogten bereikt. Daaruit spreekt de verwachting dat AI de leefwereld grondig verandert en menselijke taken op allerlei gebieden zal overtroeven. Zal de toenemende integratie van AI de Nederlandse samenleving de komende jaren daadwerkelijk op allerlei gebieden transformeren? Zo ja, welke vormen van druk zijn er dan te voorzien op de maatschappelijke stabiliteit? Hoe kunnen we daarmee omgaan?

Transformatie van de samenleving?

Het beeld van een totale maatschappelijke transformatie is misschien wat overtrokken maar deze studie laat zien dat AI de Nederlandse samenleving zowel in de breedte als in de diepte beïnvloedt. Dit gebeurt echter niet altijd op de manieren die vanuit de populaire verwachting naar voren komen. Wanneer we de nuchtere werkelijkheid scheiden van de hype is een aantal algemene (mis-)concepties over de transformatieve kracht van AI voor de maatschappij van belang:

- Vanuit het perspectief van transformatieve kracht is AI **geen technologie** maar een naam voor de **geschakelde ambities** om kunstmatig intelligente machines te creëren. Dit inzicht heeft belangrijke implicaties: We dienen, om de impact van AI te begrijpen, niet alleen te kijken naar de technologie maar naar de mix van behoeften, belangen, en publieke mythes die de aandacht voor AI opdrijven.
  - Qua AI-technologie zijn er eerder in de geschiedenis van computertechnologie spraakmakende doorbraken geweest. Het verschil nu is dat 'transformer-based', deep learning-technieken (zoals ChatGPT, Bard, enz.) de deuren hebben geopend voor AI-toepassingen in het dagelijks leven van iedereen.
  - Met name de bereidheid om te investeren in hogere efficiency en het dekken van arbeidstekorten geeft de integratie van AI een boost voor bedrijven. Het AI-technologiseringsproces wordt daarnaast aangedreven door consumptieve behoeften maar consumenten hebben meer reserves bij de privacy en veiligheid van AI-algoritmen.
  - AI wordt op de agenda gehouden door een mix van mondiale wetenschappelijke, politieke, militaire en industriële belangen. Dit wordt gereflecteerd in de gigantische (verwachte) market cap en groeiende investeringen en subsidiestromen. Private investeringen hebben met name in de VS overheidsfondsen ingehaald.
- De zorgen over de **'existential threat'** van AI-superintelligentie lijken op de lange termijn niet helemaal ongelegitimeerd, maar leiden momenteel vooral af van de impact van AI-risico's als optelsom van kleinere, incrementele impact die allerlei toepassingen (al) hebben.
- ChatGPT betekende voor velen een noviteit, maar in veel gevallen betekent AI geen revolutie maar eerder een **evolutie**. Dat wil zeggen, de integratie van 'zelflerende' algoritmen in allerlei digitale toepassingen vormt een continuering van processen van automatisering, virtualisatie en democratisering van technologie die al geruime tijd aan de gang zijn.

- De verwachte impact van AI op de samenleving schuilt voor een belangrijk deel in de brede, alomtegenwoordigheid waarmee AI wordt ingezet in allerlei taken en processen. Dit proces wordt ook wel **softening** genoemd. Door de stilzwijgende, gecontinueerde integratie van AI in allerlei toepassingen hebben we minder goed door dat de transformatie onder onze ogen plaatsvindt.

Impact op de maatschappelijke stabiliteit

Op existentieel en geopolitiek niveau:

- Hoewel omstreden vanwege 1) de paniekimplicaties en 2) de gesuggereerde corporate en nationale belangen voor het pushen van AI in het centrum van de internationale belangstelling, wint de wetenschappelijke hypothese aan terrein dat autonome, 'superintelligente' systemen mensen zouden kunnen **schaden**. Hoewel het gaat om een wetenschappelijk scenario op de langere termijn, is het toch van belang om op de primaire veiligheidsdreiging die hiervan uit gaat te anticiperen. Dit komt omdat het inbouwen van vangrails niet meer goed mogelijk is op het moment dat AI-systemen boven onze macht groeien.
- De kosten van het expliciet maken van out of control AI lijken echter niet op te wegen tegen de baten: de maatschappelijke implicaties van rogue AI zijn negatief omdat de benoemde 'existential threat' leidt tot **morele paniek, complotdenken** en **sociale backlasheffecten**. Hoewel de correlatie tussen angst voor AI en complotdenken een dark number blijft, toont onderzoek wel aan dat dreigingspercepties ('threat perceptions') versterkt aanzetten tot het zoeken naar eigen verklaringen en de selectieve keuze van 'eigen waarheden' online.
- Het genoemd scenario is dat AI gebruikt wordt als handleiding voor DIY (do it yourself)-vernietigingswapens heeft helaas wel degelijk een wetenschappelijke onderbouwing: Zogenaamde **'AI-convergentie'** veroorzaakt wetenschappelijk gevalideerde risico's op de terreinen van **bio-engineering** en **chemical engineering**. Op het gebied van **nucleaire wapenontwikkeling** bestaan zorgen die voortkomen uit toenemende onzekerheden in de 'balance of power' door de komst van AI in de war room, in convergentie met vroegsignaleringssystemen.
- De machtsblokken China en de VS zijn verwikkeld in een wedloop om de beslissende voorsprong op het gebied van AI-technologie en industrie. Deze grootmachtcompetitie zet het mondiale **geopolitieke krachtenveld** verder op spanning. Voor de Nederlandse samenleving kunnen door deze spanning allerlei secundaire effecten optreden door afhankelijkheden, zoals politieke onrust door marktverstoringen of door de kans verder te worden meegezogen in internationale conflicten.
- AI-competitie bestendigt de **machtspositie van techcorporates** en daarmee hun greep op binnenlandse politiek en economisch beleid, ook al helpt EU-regelgeving om deze invloed in te dammen. Dit betekent dat Nederland onvoldoende baas in eigen huis kan zijn. Dit schept kritische afhankelijkheden.
- AI zal in verschillende delen van de wereld op allerlei manieren worden **gepolitiseerd**. Zo is AI-technologie nu al een deel van de inzet in de Amerikaanse en deels Europese **Culture Wars**. Het is goed denkbaar dat ressentiment jegens AI een extra katalysator is voor **anti-institutionele** (internationale) netwerken en complotdenkers.

Op maatschappelijk macroniveau:

- De toename van **synthetische media** zorgt voor concrete dreigingen op het gebied van wraakporno, fraude en politieke beïnvloeding. In abstracto ondermijnt synthetische informatie het gezag en de waarde van feitelijke informatie.

- AI zal door bedrijven en organisaties massaal worden ingezet als een instrument ter verhoging van productiviteit en efficiency, leidend tot potentieel grote veranderingen op de **arbeidsmarkt.** Hoewel de meningen verdeeld zijn zal AI in een negatief ontwikkelingsscenario leiden tot een **toename van economische én generationele ongelijkheid**.
- (Virtuele) interactie met AI heeft potentieel verschillende negatieve **psychosociale effecten**. Zo werkt AI-interactie mogelijk verslavend, veroorzaakt ze vervreemding, paranoia of depressie. Er kleven bovendien psychologische risico's aan sociale en affectieve relaties met AI-chatbots. De toenemende afwezigheid in het fysieke maatschappelijke verkeer veroorzaakt politieke en maatschappelijke dissociatie wat op haar beurt weer aanleiding geeft voor afbrokkeling van de menselijke veerkracht.
- Algoritmen en synthetische media spelen een rol in het voortbestaan van **echokamers**. Er is een relatie met affectieve polarisatie door het zogenaamde 'stadioneffect': het bestaan van blikverkokering door filterbubbels wordt wetenschappelijk bestreden. Het mediadieet wordt juist diverser. Daar staat tegenover dat mensen de hun welgevallige informatie sneller opblazen om zich daarmee harder af te zetten tegen het onwelgevallige geluid.
- **De sociale impact van AI-surveillance**: wanneer AI-systemen door burgers geïdentificeerd worden als institutioneel machtsmiddel, leidt dit tot vertrouwenscrises tussen overheid, rechtspraak, bedrijven en burgers, alsmede tussen burgers onderling.

Op toepassingsniveau:

- **Algoritmische bias maakt toepassingen vatbaar voor discriminatie.** AI-toepassingen in vele verschillende vormen bezitten risico's van algoritmische bias. Hierdoor ontstaan onderdrukkende of discriminerende effecten die op grotere schaal leiden tot marginalisering, sociale onrust en ophef.
- **Criminele toepassingen van AI kunnen een serieuze uitdaging vormen voor opsporing en criminaliteitsbedrijding.** De toepassingen van AI in criminaliteit kunnen zorgen voor een professionaliseringsslag op bijvoorbeeld het terrein van vermogensdelicten. Waar sprake is van zogenaamde High Impact Crimes dragen moeilijk oplosbare zaken bij aan een dalend vertrouwen in de rechtsstaat.
- **Interactie met AI-bots geeft reële psychologische risico's**. Zoals in het geval van affectieve emoties die in machinecommunicatie wordt geprojecteerd door toedoen van het ELIZA-effect. Bots kunnen gevaarlijke onzin hallucineren, aanzetten tot misdrijven of zelfs tot zelfmoord.[1]
- **AI-toepassingen kunnen geen vervanging zijn voor menselijke ervaringskennis.** Maatschappelijk gezien ligt er het risico dat we teveel gaan leunen op AI en teveel navigeren op de pointers van algoritmen. Dit zorgt voor bestuurlijke armoede en ondermijnt het vertrouwen van burgers in de overheid die een menselijke maat dient te houden in beleid en uitvoering.
- **Ethische regie is in de praktijk lastig**. Regulering en ethische vangrails worden veelal wettelijk bepaald, maar de werkelijke ethische regie hangt af van ingewikkelde, moeilijk te handhaven toetsingsprocessen die bovendien veel tijd en schaarse capaciteit in beslag nemen. Er is dus sprake van een discrepantie tussen de waarden die men zonder concessies zegt te willen beschermen en de praktijksituatie waarin ethische toetsing om verschillende redenen slechts gedeeltelijk haalbaar is.
- **Er blijken vaak grijze gebieden in de wet- en regelgeving** die maken dat het in de praktijk moeilijk bindend vast te stellen is of een toepassing voor de wet risicovol is en verboden moet worden.
- **De terugkerende dilemma's op toepassingsniveau tussen effectiviteit en menselijke maat.** In de keuze tussen wel of geen integratie van AI in functies in het publieke domein zitten besluitvormers voortdurend gevangen in een bestuurlijk dilemma tussen kansen en ethische risico's.

[1] Pierre-Francois Lovens, "Sans ces conversations avec le chatbot Eliza, mon mari serait toujours là", La Libre, 2023.

Strategische balansoefeningen in de omgang met AI

De omvang van het krachtenveld rondom AI is zo groot, en de toepassingsontwikkelingen die van buitenaf op ons afkomen dermate kaleidoscopisch, dat het moeilijk voor te stellen is dat Nederland over de impact van AI erg veel regie kan claimen 'in eigen huis'. Vaak zal de slotsom zijn dat Nederland voor raad en daad zal willen aansluiten bij Europa. Maar toch, ook binnen de context van Europa heeft Nederland keuzes in hoe het zich oriënteert in relatie tot AI in onze samenleving en voor welke taken en rollen het AI (niet) wil inzetten. Met het doel om ook strategisch te blijven in de focus van deze studie (en geen concrete beleidsvoorstellen te willen doen), is er een aantal strategische 'trade-off-situaties' waarbinnen Nederland keuzes kan maken. Deze zijn achtereenvolgens:

- **Innovatie 'versus' waarden**. Vooropgesteld, deze hoeven elkaar niet uit te sluiten. Maar specifiek in relatie tot AI is het lastig om ethische eisen op te werpen en waarden te willen dienen zonder daarmee ook het innovatieproces te frustreren of naar elders te verjagen. In de praktijk zullen de investeringen ook niet snel toereikend zijn om invloed te bemachtigen op 'mensgerichte AI-innovatie.' De Europese investeringen verbleken bij de investeringen van de VS en China.
- **Buitenhouden 'versus' meebuigen**. Is het mogelijk om safeguards in te bouwen als je zelf niet aan de tekentafel zit? De EU heeft met de WAI een belangrijke dam gebouwd tegen risicovolle toepassingen. Maar hoe stevig is een dam tegen een systeemtechnologie die overal in lijkt te gaan doordringen? Zijn er mogelijkheden om op een verantwoorde manier AI-ontwikkeling en de bescherming van het maatschappelijk belang te dienen?
- **Defensief mensgericht 'versus' offensief mensgericht**. Nederland volgt de Europese koers van de regulering van AI ter bescherming van mensgerichte waarden. Deze ethische insteek is goed maar moet niet uitmonden in een protectionistisch 'leven achter digitale dijken'. Er is ook wat voor te zeggen om de potentie van AI op een proactieve manier te exploreren ten gunste van een mensgerichte, veilige samenleving.

1. Inleiding

1.1. Aanleiding en doelstelling

De opkomst en verspreiding van kunstmatige intelligentie – verder aan te duiden als 'AI', Artifical Intelligence, zal de Nederlandse samenleving naar verwachting stevig beïnvloeden.[2] Deze AI-transformatie is al op gang gekomen, met name door de publieke introductie van Large Language Models zoals ChatGPT.[3] Met deze introductie kwam een verschuiving in het publieke bewustzijn op gang: het grote publiek realiseerde zich dat machines en algoritmes steeds meer denktaken sneller en beter kunnen uitvoeren dan de mens.

Deskundigen over de hele wereld bevestigen de impact van AI op de maatschappij. Tegelijkertijd is met het beeld dat velen hebben van AI het nodige mis. AI is geen eenduidige, vastomlijnde techniek maar een verzamelnaam voor hoe steeds snellere rekenkracht en een nieuwe generatie van algoritmen leiden tot systemen die zelfstandig(er) taken kunnen uitvoeren. Taken waar tot dan toe menselijk intellect en creativiteit voor nodig waren. Er bestaat veel variëteit in deze systemen en hoe ze ontstaan. Vergelijk het eerder met de opkomst van een nieuw ecosysteem dan met de introductie van een enkele disruptieve techniek. Deze nuanceringen laten onverlet dat AI een substantiële invloed zal krijgen op de Nederlandse samenleving.

Deze studie heeft tot doel te analyseren hoe de integratie van AI in allerlei toepassingsvormen de maatschappelijke stabiliteit – kort gezegd het voortbestaan van een stabiele democratische rechtsorde waarin een stevig onderling vertrouwen bestaat tussen burgers onderling en in de institutionele realiteit – kan beïnvloeden. De horizon van de analyse is ca. 2-5 jaar met op punten een beperkte verdere doorkijk. Omdat generatieve AI op dit moment de grootste zichtbaarheid en gevolgen voor de Nederlandse samenleving in de volle breedte heeft, ligt hier de nadruk op. De impact van AI wordt bekeken vanuit de (mogelijke) effecten van AI op verschillende domeinen in de samenleving die het fundament vormen voor maatschappelijke stabiliteit.

---

2 Zie bijvoorbeeld: The Economist, How AI could change computing, culture and the course of history, 2023; Kate Crawford, Atlas of AI, 2023.

3 'GPT' staat voor Generative Pretrained Transformer, en het is deze klasse van modellen dat momenteel furore maakt. Van doorslaggevend belang voor de verbeterde kwaliteit van Natural Language Processing is de ontwikkeling van de Google Transformer in 2017 geweest. Dit algoritme is gebaseerd op een neurale netwerkstructuur en imiteert in die zin associatie- en ordeningsprincipes geïnspireerd op de werking van het brein. Jakob Uszkoreit, Transformer: A Novel Neural Network Architecture for Language Understanding – Google Research Blog, 2017.

1.2. Kader: het Strategische Monitor Politie-programma

Er zijn al veel rapporten verschenen over de impact van AI op de maatschappij, bijvoorbeeld van het Rathenau Instituut en van de Rijksoverheid.[4] Dit rapport onderscheidt zich door (1) het onderwerp expliciet te plaatsen binnen de aanpak van het Strategische Monitor Politie-programma dat HCSS uitvoert voor de directie Strategie en Innovatie van de Staf Korpsleiding van de Nederlandse politie; en (2) specifiek in te gaan op de impact van AI op de maatschappelijke stabiliteit.

Ten aanzien van het eerste geldt dat het genoemde programma ingaat op allerlei bewegingen in onze maatschappij die de maatschappelijke stabiliteit kunnen bedreigen en die het gevolg zijn van de grote trends en ontwikkelingen in de wereld om ons heen (een 'van buiten naar binnen'-perspectief).[5] In diverse gevallen worden deze dreigingen[6] ondersteund of versterkt door technologie-gedreven ontwikkelingen; ontwikkelingen die, op hun beurt, in toenemende mate gebruik maken van AI. Het gaat concreet bijvoorbeeld om de rol van sociale media en de daarop gegeneerde mis- en desinformatie in het verspreiden en mobiliseren van onvrede; om digitale dienstverlening die de georganiseerde misdaad faciliteert; of om allerlei vormen van cybercriminaliteit. Dat is één perspectief op de relatie tussen de opkomst van AI en maatschappelijke (in)stabiliteit dat in dit rapport nader wordt verkend.

1.3. Aanpak

Het onderzoeken van de impact van AI op maatschappelijke stabiliteit volgt de volgende stappen:

1. **Breedte van AI-toepassingen.** Het betreft een verkenning van de breedte van (potentiële) AI-toepassingen binnen de samenleving naar verschillende domeinen die raken aan het thema van maatschappelijke stabiliteit.
2. **De transformatieve kracht van AI.** Wat maakt dat we spreken over een AI-transformatie? Dit hoofdstuk maakt duidelijk dat het niet (alleen) gaat om het inpassen van een groeiend aantal AI-toepassingen, maar dat dit op macroniveau een effect zal hebben dat de lokale veranderingen van specifieke AI-implementaties overstijgt.
3. **De impact van de AI-transformatie op de maatschappelijke stabiliteit.** Het deduceren ván, en vooruitkijken náár de potentiële invloed van deze transformatie op de maatschappelijke stabiliteit. Dit deel maakt op verschillende niveaus (existentieel, sociaal-politiek, en op toepassingsniveau) duidelijk wat de (potentiële) impact is op de maatschappelijke stabiliteit.

---

4 Rathenau, Generatieve AI, december 2023; Rijksoverheid, Overheidsbrede visie Generatieve AI, januari 2024.

5 Dit rapport bouwt met name verder op de volgende eerdere HCSS-rapporten in het kader van het Strategische Monitor Politie-programma: De Staat van de Rechtsstaat: Waar Staan We, Waar Gaan We Naartoe?, 2023; Maatschappelijke Ontgoocheling van de Middenklasse: Optreden, Oorzaken en Gevolgen, 2023; Het Sociaal contract. Verwachtingen en Spanningen in de Democratische Rechtsorde, 2024; en Next Generation Organised Crime: Systemic change and the evolving character of modern transnational organised crime, 2023.

6 We maken ons vooral druk over maatschappelijke stabiliteit als die stabiliteit in het geding is. Vandaar dat we in de diverse studies vaak ingaan op de verschillende dreigingen die de stabiliteit en continuïteit van de samenleving kunnen verstoren; kansen worden vooral gevonden in het tegengaan of inkapselen van deze dreigingen.

4. **Strategische balansoefeningen.** Het beschrijven van het strategisch beleidsspectrum waarbinnen Nederland de ruimte kan hebben om maatschappelijke instabiliteit door AI te voorkomen. De AI-transformatie werpt een aantal strategische balansoefeningen op: Hoe kunnen we de ongewenste uitwassen en toepassingen van de AI-transformatie tegengaan of binnen de perken houden, zonder tegelijkertijd de positieve effecten sterk af te remmen?

De ontwikkeling van AI is bij uitstek internationaal en veel van de literatuur over AI is Engelstalig. Veel relevante concepten en toepassingen zijn beter bekend onder hun Engelse naamgeving – als er überhaupt al een goede, geaccepteerde Nederlandse vertaling bestaat. In dit document is ervoor gekozen om veel van de gangbare Engelse termen over te nemen en dus niet te vertalen.

Figuur 1. Stroomdiagram van de analyse

2. Breedte van AI-toepassingen

We hanteren de volgende definitie van AI: 'de verzamelnaam voor allerlei vermogens van computer-systemen om een breed scala aan geautomatiseerde taken uit te voeren, waarbij deze de menselijke cognitieve en lerende capaciteit nabootsen en daarmee de afhankelijkheid van menselijke bijdragen verminderen.' AI is een generieke technologie; AI-toepassingen zijn denkbaar voor vrijwel ieder aspect van de samenleving, van harde techniek en industriële processen tot aan culturele uitingen zoals AI-gegenereerde muziek, teksten en afbeeldingen. Het is duidelijk dat AI-toepassingen niet beperkt zijn tot de digitale sfeer. AI in combinatie met robotoplossingen en slimme apparaten maakt AI nog gangbaarder en impactvoller in ons leven.

In de literatuur[7][8] en door andere kennisinstituties zoals het Rathenau Instituut, World Economic Forum[9] of McKinsey[10] worden verschillende maatschappelijke impactdomeinen van AI onderscheiden. Tussen de verschillende sets categorieën bestaan grote overeenkomsten.[11] Omdat deze studie sterk redeneert vanuit maatschappelijke stabiliteit, kozen we ervoor de domeinen te reduceren tot enkele waarvan helder is dat AI op die terreinen sterke veiligheidsimplicaties kan hebben.[12]

Deze indeling omvat de belangrijkste maatschappelijke processen die beïnvloed gaan worden door AI. Hieronder beschouwen we per domein de trends en toepassingen op macroniveau. De per domein vermelde toepassingsgebieden worden in hoofdstuk 4 overgenomen om de maatschappelijke impact te duiden.

---

7 Onder andere: Corinne Cath et al., Artificial Intelligence and the 'Good Society': The US, EU, and UK Approach, 2018; Michael Cheng-Tek Tai, The Impact of Artificial Intelligence on Human Society and Bioethics, 2020; Samuel Fosso Wamba et al., Are We Preparing for a Good AI Society? A Bibliometric Review and Research Agenda, 2021; Spyros Makridakis, The Forthcoming Artificial Intelligence (AI) Revolution: Its Impact on Society and Firms, 2017.

8 Spyros Makridakis, The Forthcoming Artificial Intelligence (AI) Revolution: Its Impact on Society and Firms, 2017.

9 World Economic Forum, Global Risks Report 2024, 2024.

10 Michael Chui et al., Notes from the AI Frontier: Applying AI for Social Good, 2018.

11 Zo noemt McKinsey tien en WEF bijna 40 impactdomeinen, maar zijn de domeinen van McKinsey een stuk breder dan van WEF. Zo omvat economic empowerment (McKinsey) bijna hetzelfde als acht verschillende economische domeinen van WEF.

12 Dat laat onverlet dat deze implicaties soms indirect zijn. AI in het onderwijsdomein geeft vooral indirecte risico's. Men denke aan teruglopende basiskennis over burgerschap en rechtsstaat of een gebrekkiger herkenning van authentieke informatie.

Figuur 2: Vijf maatschappelijke impactdomeinen

We bespreken voor elk domein enkele belangrijke toepassingsvormen, niet met de bedoeling om daarmee volledig te zijn maar om een indruk te geven van de reikwijdte. Deze beperkte dwarsdoorsnede van maatschappelijke toepassingsgebieden zullen we bovendien aanhouden om in hoofdstuk 4 hun respectievelijke implicaties voor de maatschappelijke stabiliteit te beschrijven.

Dual use. Het is fair om de positieve beloften van AI in ogenschouw te houden en niet alleen de bedreigingen. Dit is bovendien strategisch verstandig: cynisme kan er toe leiden dat men een kans verliest op aanpassingsstrategieën ten aanzien van AI. Desondanks leggen AI-optimisten in hun narratief teveel nadruk op positieve toepassingen. De beloften zijn, onkritisch bekeken, eindeloos: AI zal helpen onze gezondheid en levensgeluk spectaculair te verbeteren; AI zal een grote bijdrage leveren op het gebied van genetica, nanotechnologie. AI zal ons toestaan tumoren op te zoeken, ons helpen materie op moleculaire en atomische schaal te beïnvloeden. De door AI verder geholpen medische vooruitgang zal het ouderdoms-proces nog verder vertragen.[13] Op termijn gloort voor tech-optimisten de hoop dat we onszelf permanent verbinden met AI, met anderen, en zal er zelfs een singulariteit ontstaan waarin

[13] Spyros Makridakis, The Forthcoming Artificial Intelligence (AI) Revolution: Its Impact on Society and Firms, 2017.

onze kennis en capaciteiten opgaan in een grenzeloze totaliteit. Zo ver zijn we echter nog niet. Hoogdravend optimisme slaat de plank mis om tenminste drie redenen: 1) het 'hypet' technologie omdat het de voorspelde toepassingen neemt en niet de feitelijke prestaties nu; 2) het onderschat of miskent de impact die al bereikt wordt dankzij de kleinere, incrementele veranderingen en integratie van AI in bescheiden toepassingen, 3) het negeert de schaduwzijde van AI. We benadrukken dat AI-toepassingen vrijwel altijd een **dual use** kennen, waardoor de toepassingen zowel goedaardige als kwaadaardige effecten hebben. De toekomst met AI wordt getekend door ambiguïteit.

AI-toepassingen in het licht van een doorlopend digitaliseringsproces. AI-toepassingen blijken best vaak een veredeling van toepassingen die al eerder bestaansrecht hebben verworven in de digitale transitie. Algoritmen hadden bijvoorbeeld al een belangrijke rol in het personaliseren van content of advertenties. Nieuwe generaties AI-algoritmen voegen daaraan nog meer effectiviteit toe zonder dat het soort maatschappelijke impact drastisch veranderd. De integratie van AI maakt de verwevenheid van technologie met leven en maatschappij weliswaar intensiever en moeilijker ontrafelbaar, tot op het punt dat algoritmen hun eigen algoritmen genereren. In generieke zin treden de hieronder beschreven AI-toepassingen in het spoor van technologische veranderingsprocessen die dan ook al veel langer bestaan. Deze processen zijn te vatten onder de ideeën van digitale **automatisering** (een proces dat al op gang is gekomen sinds de jaren vijftig), **virtualisatie** (sinds circa 2000) en **democratisering**.[14]

Automatisering. AI heeft het vermogen om processen nog verder te stroomlijnen en allerlei routinetaken uit te voeren. AI-analytics kunnen menselijke processen en menselijke productiviteit monitoren en bijstellen waardoor een nog hogere efficiency bereikt kan worden. De positieve belofte dat mensen hun werk sneller kunnen doen, meer vrije tijd overhouden en prioriteit kunnen geven aan andere belangrijker taken. In werkelijkheid wordt aan de positieve beloften van AI-automatisering voor mensen sterk getwijfeld, zoals we zullen zien in hoofdstuk 4.

Virtualisatie. Een ander technologisch ontwikkelingsproces dat AI helpt doorevolueren is dat van **virtualisatie.** Over de smartphone wordt weleens gezegd dat deze al in belangrijke mate vergroeid is met het menselijk lichaam. Steeds meer mensen brengen hun tijd online door, wat wijst op de virtualisatie van onze leefwereld. Met behulp van AI kan content op alle denkbare manieren op maat worden gemaakt voor alle taken en behoeften. Maar ook hier is sprake van keerzijden: wat zijn de psychosociale effecten? Wat als de grens tussen de echte en de synthetische wereld teveel vervaagt met implicaties voor onze democratie? AI-machines zullen het leven op fronten makkelijker en comfortabel maken, maar daarbij ook inspelen op gevoelens van gewenning en verslaving zoals sociale media-algoritmen dat nu al doen. Met de graduele introductie van allerlei 'handige' toepassingen zal het incrementele proces van virtualisatie zich voortzetten.

Democratisering. Het derde – tevens meest omstreden – technologisch ontwikkelingsproces betrokken bij AI is dat van **democratisering** (dat sterk overlapt met het idee van de **decentralisatie** van technologie). Daarmee wordt bedoeld dat AI-technologie een belofte inhoudt om de toepassing van disruptieve technologie in handen te leggen van iedereen die daar iets nuttigs mee wenst te doen. Dit kan worden mogelijk gemaakt omdat heel complexe technologie toch op een laagdrempelige en gebruiksvriendelijke manier kan worden aangeboden. Ironisch genoeg is technologische democratisering een doctrine die voor een belangrijk deel verkondigd wordt door Silicon Valley dat de neiging heeft om broncodes en

[14] Sandy Kaul, Evolution of Commercial Technologies and Impact on Business Delivery, 2022.

gebruiksvoorwaarden in eigen hand te houden. Daarmee is niet gezegd dat AI geen rol heeft in het laagdrempeliger maken van complexe taken: ChatGPT kan in mum van tijd een website bouwen. Daar staat tegenover dat met alle input de eigenaars van de broncode hun algoritme optimaliseren en zich op die manier 'gratis' trainingscapaciteit verwerven. Het is de vraag of er werkelijk sprake is van democratisering wanneer heimelijke belangen worden gediend. Een bovendien niet te onderschatten keerzijde is de kans op kwalijke dual use-toepassingen. Het ideaal van technologie voor iedereen kan zich transformeren tot een rampscenario wanneer kwaadwillenden AI-tools inzetten voor het maken van wapens of het zaaien van terreur.

Overzicht in de breedte. Het is onmogelijk om volledig recht te doen aan alle mogelijke toepassingsvormen die AI kan krijgen. Desalniettemin kan de analyse van verschillende impactvolle toepassingsvormen de analyse verrijken. Een lijstinventarisatie van AI-toepassingen geeft iets weer over **de schaal en de veelvormigheid** waar we mee te maken hebben en krijgen. Indirect zegt dit ook iets over de enorme maatschappelijke uitdagingen (en misschien onmogelijkheid?) om al deze toepassingen op een goede manier in banen te kunnen leiden. Daarover meer in hoofdstuk 4 en 5. We benoemen nu eerst de belangrijkste toepassingsgebieden per domein.

2.1. Publieke domein

Het publieke domein verwijst onder andere naar toepassingen in **wet- en beleidsvormingsprocessen** van de overheid, de (ontwikkeling van de) **democratie** en de **sociale dienstverlening**. In beleidsvormingsprocessen is al een zekere integratie van AI waarneembaar in de vorm van analysetools. AI zal vaker een rol spelen als (co-)adviseur voor het maken van besluiten, gebaseerd op algoritmen en data gegenereerd uit analysemodellen. AI kan persoonlijke situaties toetsen aan regelgeving. Er wordt al geëxperimenteerd met het schrijven van wetsvoorstellen door AI. Verder kan de overheid AI inzetten om dienstverlening te versnellen en meer maatwerk te leveren. Bij de inzet van AI voor democratische processen kunnen we denken aan het inzetten van AI voor de selectie van burgerberaadskandidaten,[15] maar ook voor het omzetten van technische, moeilijk begrijpbare teksten en besluiten naar toegankelijk taalgebruik.[16] Ook valt te denken aan AI-assistentie bij verkiezingen, onder andere in het voorkomen van fraude en helpen bij het tellen.[17] Omgekeerd kan AI ongewenste verkiezingsbeïnvloeding faciliteren.

15 Bailey Flanigan et al., Fair Algorithms for Selecting Citizens' Assemblies, 2021.

16 Toolify, Revolutionizing Plain Language Summaries: The Power of AI, 2024.

17 Norman Eisen et al., AI Can Strengthen U.S. Democracy—and Weaken It, 2023.

Tabel 1: Toepassingsgebieden en voorbeelden van toepassingen van AI in het publieke domein

Toepassingsgebieden en voorbeelden van feitelijke toepassingen in het publieke domein

Door AI ondersteunde **beleidsvormingsprocessen**:
- 'Forecasting support systems', 'Planning Support Systems' en 'Decision Support Systems'[18], 'robo-advisors'[19]: ondersteunende algoritmen en AI data-analysemodellen richtinggevend aan beleid, zoals macro-economische modellen[20], financiële doorrekeningen en haalbaarheidsstudies.[21]
- Door AI aangedreven interactieve beleidsvorming en crowd sourcing.[22]

Door AI ondersteunde **wetgevingsprocessen**:
- 'Microlegislation': incrementele, kleine aanpassingen in wetsontwerpen die op grond van AI analyse of door AI-powered lobby worden ingepast.[23]
- Toepassing van AI in de ondersteuning van het schrijven van moties of wetsontwerpen.
- Door AI geschreven moties en/of amendementen of wetsvoorstellen op naam van AI.

Geautomatiseerde en gerobotiseerde **publieke dienstverlening** (eGovernment of GovTech[24]):[25]
- Chatbots en AI-assistentie.
- ADM's[26] – Automated Decision-making processes: Door AI geleide beoordelings- of besluitvormingsprocessen (belastingaanslagen, geautomatiseerde vergunningsaanvragen).[27]

AI gedreven **maatschappelijke monitoring en voorspelling**:
- Geautomatiseerde risicoidentificatiesystemen gericht op burgers.
- Sentimentanalyse, monitoring van sociale media en hate speech.[28]

AI gedreven **marketing voor politieke partijen en publieke sector** (campagnevoering):
- Door AI geoptimaliseerde publiekstargeting (microtargeting).
- AI fundraising of door AI gegenereerde inkomsten voor politieke partijen.
- Een volledige door AI geleide politieke partij.

Toepassing van AI in instrumenten van **directe democratische inspraak**:
- Het toepassen van AI voor de selectiecriteria en de selectie van burgerberaadskandidaten.[29]
- De toepassing van AI in het vertalen en toegankelijk maken van complexe politieke onderwerpen.
- De toepassing van AI in grass roots beleidsontwerp.

Toepassing van AI in **ethische of op compliance gerichte (zelf-)beoordelingsprocessen**:[30]
- 'Self-explaining AI' (AI-systemen die hun keuzes en werking kunnen uitleggen)
- 'Automated Compliance Assessments' (AI systemen die helpen te voldoen aan wet- en regelgeving).
- AI ondersteunde ethische impact assessments – nog niet (breed) operationeel toegepast.

18 Daan Kolkman, The Usefulness of Algorithmic Models in Policy Making, 2020.

19 Zeynep Engin en Philip Treleaven, Algorithmic Government: Automating Public Services and Supporting Civil Servants in using Data Science Technologies, 2019.

20 Mary Morgan en Frank den Butter, Empirical Models and Policy Making: Interaction and Institutions, 2003.

21 Michela Arnaboldi en Giovanni Azzone, Data Science in the Design of Public Policies: Dispelling the Obscurity in Matching Policy Demand and Data Offer, 2020.

22 Svilena Iotkovska, Helsinki Invites Cyclists to Collect Data on Street Conditions and Earn Money, 2023.

23 Nathan E. Sanders en Bruce Schneier, How AI Could Write Our Laws, 2023.

24 Frederik Peters, Govtech is voorbij de hype, 2023.

25 Zeynep Engin en Philip Treleaven, Algorithmic Government: Automating Public Services and Supporting Civil Servants in using Data Science Technologies, 2019.

26 AlgorithmWatch, Automated Decision-Making Systems in the Public Sector – Some Recommendations, 2022.

27 Hila Mehr, Artificial Intelligence for Citizen Services and Government, 2017; Amnesty International, Algoritmes, Big Data en de overheid, 2021.

28 Thomas Brewster, ChatGPT Has Been Turned Into A Social Media Surveillance Assistant, 2023.

29 Flanigan et al., Fair Algorithms for Selecting Citizens' Assemblies, 2021; Leah Burrows, Can AI Make Democracy Fairer?, 2021.

30 Ali Hashmi, AI Ethics: The Next Big Thing In Government - Anticipating the Impact of AI Ethics within the Public Sector, 2019.

2.2. **Veiligheidsdomein**

Wat betreft het veiligheidsdomein zijn meerdere sectoren in beeld (het gaat ook hier om een niet-limitatieve inventarisatie): AI kan ondersteunen bij de bescherming van nationale veiligheid en criminaliteitsbestrijding (politie, marechaussee, inlichtingendiensten); AI en rechtspraak; en AI en criminaliteit.[31] AI zal aantrekkelijker worden als tool voor **surveillance, opsporing en predictive profiling**. In het **rechtssysteem** helpt AI bij het verlagen van de administratieve werkdruk door het automatiseren van processen, bij het formuleren van juridische argumenten en bij het ondersteunen van wettelijke toetsings- en beoordelingsprocessen. De foutgevoeligheid is voor een sector die draait om zorgvuldigheid nog een groot obstakel, maar dit argument verdwijnt wanneer de kans op foutmarges van machines significant lager wordt dan de menselijke foutmarge. In negatieve zin kan en zal AI uitgebreid worden gebruikt voor **criminele activiteiten**. AI-enabled crime zal uiteraard een maatschappelijk ondermijnend effect hebben.

Tabel 2: Toepassingsgebieden en voorbeelden van toepassingen van AI in het veiligheidsdomein

Toepassingsgebieden en voorbeelden van feitelijke toepassingen in het veiligheidsdomein
AI-tooling in **fraude en cybercriminaliteit**. • AI-enabled cyber crime: gebruikt als ondersteunend instrument voor cyber crime.[32] • AI-enabled LaaS (Laundering as a service) AI gebruikt als witwas- en financieringsmodel. • AI-enabled CaaS – cybercrime as a service. • Aanvallen óp AI, waaronder stelen en manipuleren van algoritmen.[33]
AI ondersteuning bij **handhaving openbare orde** en radicaliserings-/criminaliteitspreventie.[34] • Beeldherkenningssoftware en biometrische video analytics van (potentiële) daders of slachtoffers. • Predictive profiling[35] en crime prediction toepassingen.[36] • AI crowd intelligence systems t.b.v. crowd control en crisismanagement. • AI die wordt gebruikt als polygraaf om emotionele toestanden van personen te bepalen[37] • Smart policing-toepassingen zoals operationele AI-assistentie.
AI ondersteuning bij **opsporing, inlichtingenvergaring, criminaliteitsbestrijding en (digitale) recherche**[38] • Video surveillance en video analytics in combinatie met patroonanalyse ter identificatie van verdachten.[39] • AI-assisted DNA-analyse.[40] • Forensische data-analyse: AI-systemen om bewijsmateriaal te doorzoeken en te prepareren. • AI-powered crime scripting (datagegenereerde misdaadscenario's die helpen bij focus en cold case analyse).[41]
Ondersteuning van AI in **strafrechtelijke beoordelingsprocessen** en rechtsbedeling[42] • Toetsing van de feiten aan het recht. • Door AI ondersteunde gerechtelijke verdediging.[43] • Routine taken kunnen geautomatiseerd worden, wat meer tijd over laat voor complexere taken en processen[44]

31 Internationale veiligheid, het primaire domein van Buitenlandse Zaken en Defensie, wordt hier goeddeels buiten beschouwing gelaten; anders dan dat het een drijvende kracht kan zijn achter bijvoorbeeld desinformatie, cybercriminaliteit en terrorisme die zich binnen onze landsgrens manifesteert.

32 Trend Micro, Exploiting AI: How Cybercriminals Misuse and Abuse AI and ML, 2020.

33 AIVD, AI-systemen: ontwikkel ze veilig, 2023.

34 Europees Parlement en Europese Raad, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, 2021.

35 Thaddeus L. Johnson and Natasha N. Johnson, Police Facial Recognition Technology Can't Tell Black People Apart, 2023

36 Fatima Dakalbab e.a., Artificial Intelligence & Crime Prediction: A Systematic Literature Review, 2022; Een voorbeeld is het Risicotaxatie Instrument Geweld (RTI-G) – de ontwikkeling daarvan is stopgezet op grond van ethische problemen: David Davidson, Politie Stopt Met Gewraakt Algoritme Dat 'Voorspelt' Wie in de Toekomst Geweld Gebruikt, 2023.

37 Priya Chakriswaran e.a., Emotion AI-Driven Sentiment Analysis: A Survey, Future Research Directions, and Open Issues, 2019; Keyur Patel et al., Facial Sentiment Analysis Using AI Techniques: State-of-the-Art, Taxonomies, and Challenges, 2020.

38 Christopher Rigano, Using Artificial Intelligence to Address Criminal Justice Needs, 2019.

39 Deloitte AI Institute, The Government & Public Services AI Dossier, 2021.

40 Christopher Rigano, Using Artificial Intelligence to Address Criminal Justice Needs, 2019.

41 Thom Snaphaan en H. Borrion, Connecting the dots: Utilizing crime scripting to leverage multimodal data and innovative techniques in a meaningful manner, manuscript in voorbereiding.

42 Dory Reiling, Courts and Artificial Intelligence, 2020; Matt Novak, Lawyer Uses ChatGPT In Federal Court And It Goes Horribly Wrong, 2023.

43 Zsuzsa Czobor, Generative AI Could Radically Alter the Practice of Law, 2023.

44 Dory Reiling, Courts and Artificial Intelligence, 2020.

2.3. **Economische domein**

Door AI-automatisering zullen beroepen en beroepsgroepen **vervangen worden**: routineuze taken maar ook witte-boordenbanen staan op de tocht. Overal waar sprake is van outperformance van menselijke capaciteiten kunnen banen verdwijnen.[45] AI-automatisering introduceert daarmee ook vragen over het idee van **toegevoegde waarde**. Een boekhouder of adviseur die grote delen van een product baseert op AI zal na moeten denken over of en hoe deze geautomatiseerde arbeid doorgerekend moet worden. In sommige gevallen kunnen banen behouden blijven door omscholing.[46] AI zal leiden tot **nieuwe typen arbeid** maar er is twijfel of deze voor weggevallen arbeidsvraag gaan compenseren. Weliswaar zal er meer vraag ontstaan naar gespecialiseerd IT-personeel: AI-engineers, datawetenschappers, hardwarespecialisten, 'prompt engineers' (de kunst van het ontwikkelen van de meest effectieve vraaginput). Er zal meer vraag zijn naar het bewaken van AI-ethiek. Dat geldt ook voor AI-juristen en compliance officers: deze hebben de ingewikkelde taak om zowel de regels als de AI-systemen te moeten doorgronden om procedures en beoordelingen te kunnen uitvoeren.[47] **AI-analytics** biedt enorme mogelijkheden op het monitoren en optimaliseren van menselijke productiviteit en consumentengedrag maar hieraan kleven ook venijnige kanten.[48]

Tabel 3: Toepassingsgebieden en voorbeelden van toepassingen van AI in het economische domein

Toepassingsgebieden en voorbeelden van feitelijke toepassingen in het economische domein
De opkomst van **AI-producten en AI product design**: • Vermarkting van allerlei soorten AI-assistentie. • Proliferatie van door AI gegenereerde marketing- en entertainmentcontent, kunst en NFT's op de markt
Verdere toename van AI **(social) media analytics**: • Hyperpersonalisatie-algoritmen (marketing op basis van een zeer gedetailleerd voorkeurenprofiel). • Inferential emotion tracking voor marketingdoeleinden.[49]
AI-toepassingen voor het **meten en verhogen van arbeidsprestaties**. • AI analytics die wordt gebruikt om beslissingen te nemen over contracten, om taken toe te wijzen op basis van individueel gedrag of persoonlijke eigenschappen en om prestaties te evalueren[50] • AI-systemen voor de selectie van kandidaten, en de training en doorscholingstrajecten van medewerkers. • AI Inferential Affective/Emotion Tracking (het monitoren van emoties op de werkvloer in relatie tot de werkcontext).[51]

45 White House, The Impact of Artificial Intelligence on the Future of Workforces in the European Union and the United States of America, 2022.

46 World Economic Forum, Recession and Automation Changes Our Future of Work, But There Are Jobs Coming, 2020.

47 Door de schaarste zal – enigszins ironisch – een toenemend beroep gedaan moeten worden op de zelfbeoordeling van AI. Zo ontstaat er een ethisch Droste-effect, want wie moet dan de AI-assessmenttools valideren die de algoritmen screenen?

48 Jean-Baptiste Hironde, Council Post: AI's Impact On The Future Of Consumer Behavior And Expectations, 2023.

49 Een voorbeeld is de verzameling en analyse van gefilmde consumentreacties op Superbowl reclames in de VS: Daniel McDuff et al., Affectiva MIT Facial Expression Dataset (AM-FED): Naturalistic and Spontaneous Facial Expressions Collected 'In the Wild', 2013.

50 Europees Parlement en Europese Raad, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, 2021.

51 Zhimin Chen and David Whitney, Inferential Affective Tracking Reveals the Remarkable Speed of Context-Based Emotion Perception, 2021.

2.4. **Onderwijsdomein**

In het kennis en trainingsdomein zal AI de structuur en de vorm van het leren beïnvloeden maar ook met welk doel precies geleerd moet worden. AI kan **leerlingvolgsystemen gerichter maken en onderwijsbeheertaken overnemen**.[52] AI zet aan tot het leren van **andere competenties en vaardigheden** (zie ook 2.3). Het is de vraag of kinderen in de toekomst nog moeten leren schrijven, bijvoorbeeld. Andere kerncompetenties impliceren ook andere vormen van leren en toetsen.[53] Aan het onderwijs bovendien de belangrijke taak om te leren alert te zijn op de toenemende onbetrouwbaarheid van informatie. De aandacht voor **media-wijsheid** zal verder toenemen. In abstracto verandert AI **de manier waarop we kennis definiëren**. De grenzen van feitelijke kennis worden vager door de enorme informatiedichtheid. Bovendien zullen de grenzen tussen menselijke en artificiële cognitie in de toekomst vervagen door de komst van nieuwe devices en interfaces. De vraag is of het brein deze verwerkingscapaciteit aankan zonder zelf een 'upgrade' te ondergaan. Hoewel nu nog science fiction, geeft het thema aan dat AI dwingt tot nieuwe verhoudingen tot onze cognitie. Niet wat mensen zelfstandig bedenken maar wat mensen met name in samenspraak met AI scheppen, zal de norm worden voor creatieve producten.

Tabel 4: Toepassingsgebieden en voorbeelden van toepassingen van AI in het onderwijsdomein

Toepassingsgebieden en voorbeelden van feitelijke toepassingen in het onderwijsdomein
AI-toepassingen die gebruikt worden voor het plegen van **fraude**. • Het succesvol inzetten van AI voor het maken van examens.[54] • Het plegen van plagiaat.
AI-toepassingen die de **toegankelijkheid van het onderwijs** verhogen:[55] • AI-assistentie voor **maatwerk kennisverwerving en competentieontwikkeling**.[56] • Didactisch en pedagogisch maatwerk op basis van slimme persoonlijkheids- en voortgangsanalyses. • AI huiswerk- en onderwijstaakassistentie • Virtuele leeromgevingen op basis van AR en VR-technologie[57]
AI voor **leerlingvolgsystemen** en de **analyse en observatie van leerprestaties.**[58] • Gebruik van predictive analytics om vroegtijdige uitval te voorkomen • Ontwikkeling van leerplannen gericht op individuele leerlingen om leraren te helpen effectievere ondersteuning te bieden.

52 Sayed Fayaz Ahmad et al., Academic and Administrative Role of Artificial Intelligence in Education, 2022.

53 Radboud Universiteit, AI in het onderwijs.

54 Sebastian Bordt, Ulrike von Luxburg. Chatgpt participates in a computer science exam, maart 2023.

55 Rose Luckin and Wayne Holmes, Intelligence Unleashed: An Argument for AI in Education, 2016.

56 Radboud Universiteit, AI in het onderwijs, n.d.

57 Stéphan Vincent-Lancrin and Reyer van der Vlies, Trustworthy artificial intelligence (AI) in education, 2020; Rose Luckin and Wayne Holmes, "Intelligence Unleashed: An Argument for AI in Education," 2016.

58 Stéphan Vincent-Lancrin and Reyer van der Vlies, Trustworthy artificial intelligence (AI) in education, 2020; Rose Luckin and Wayne Holmes, "Intelligence Unleashed: An Argument for AI in Education," 2016.

2.5. **Sociale domein**

Op sociaal gebied laat AI zich gelden door de mogelijkheid van aantrekkelijke, gepersonaliseerde interactievormen met **AI-chatbots** en **AI-assistenten**. AI heeft in potentie de kans om **eenzaamheid te bestrijden** maar ook om grote psychosociale problemen te veroorzaken op het gebied van stress, overprikkeling en maatschappelijke dissociatie. Wanneer het moment komt van de massale omarming van **AR** en **VR-technologie** is moeilijk te voorspellen. Maar het is zeker dat simulatie-instrumenten beter implementeerbaar zullen worden. Ergens in de toekomst zal een omslagpunt zijn van nicheproduct naar massale omarming, waarna de toepassingsgebieden van virtuele omgevingen snel verbreden. AR-techniek is het meest veelbelovend en kan uitstekend worden verrijkt met AI-applicaties. De real-time informatie die een AR-display kan geven is echter op vele manieren toepasbaar te maken. In professionele context – denk aan live ondertiteling van tekst of instructies bij objecten, of in persoonlijke context – denk aan persoonlijke informatie zoals verjaardagen of voorkeuren die opspringen bij het zien van een vriend. AI kan gebruikers voorzien in een wereld van mogelijkheden op het gebied van **User Generated Content**: zo kunnen AI-algoritmen helpen bij het vervaardigend van entertainment, memes en allerlei soorten online content. Users kunnen hun sociale mediakanalen laten beheren door een AI-kloon van zichzelf of kanalen creëren op grond van een volledig fictief, gegenereerd karakter.

Tabel 5: Toepassingsgebieden en voorbeelden van toepassingen van AI in het sociale domein

Toepassingsgebieden en voorbeelden van feitelijke toepassingen in het sociale domein
Vergemakkelijking van **communicatiemogelijkheden**:[59] • Het gebruik van automatische aanvulling in tekstberichten • Het gebruik van standaardantwoorden geformuleerd door AI • Real-time vertaling in gesprekken[60] • Proliferatie van AI chatbots en personages[61]
AI ondersteuning op het gebied van **coaching** en (mentale) gezondheid:[62] • AI als psychologisch zelfhulpmiddel of als diagnose of intakeinstrument. • AI life coaching • Eenzaamheidsbestrijding door middel van interactieve AI en/of robots.
Het ontwikkelen van allerlei vormen van **User Generated Content** • AI-toepassingen voor het maken van synthetische media. • Virtuele AI-kloon[63]

59 Eda Erensoy, How AI Is Changing Human Communication, 2021.

60 Haifeng Wang et al., Progress in Machine Translation, 2022.

61 Anne Zimmerman et al., Human/AI Relationships: Challenges, Downsides, and Impacts on Human/Human Relationships, 2023.

62 Simon D'Alfonso, AI in Mental Health, 2020.

63 BBC, YouTube tests AI tool that clones pop stars' voices, 2023.

3. De transformatieve kracht van AI

Zoals benoemd in hoofdstuk 2, bouwt de ontwikkeling van AI voort op een **technologisch digitaliseringsproces** dat al decennia aan de gang is. Sinds de automatiseringsrevolutie vertrouwen we meer en meer op computers om zowel routinematige als complexere taken te ondersteunen en uit te voeren. Ook al treedt AI-technologie in de voetsporen van bestaande technologische ontwikkelingsprocessen, er zijn meerdere kenmerken die maken dat AI iets groters representeert dan 'gewoon' een nieuwe technologie, en een 'turbo' kan zetten op veranderingen in ons maatschappelijk bestaan.

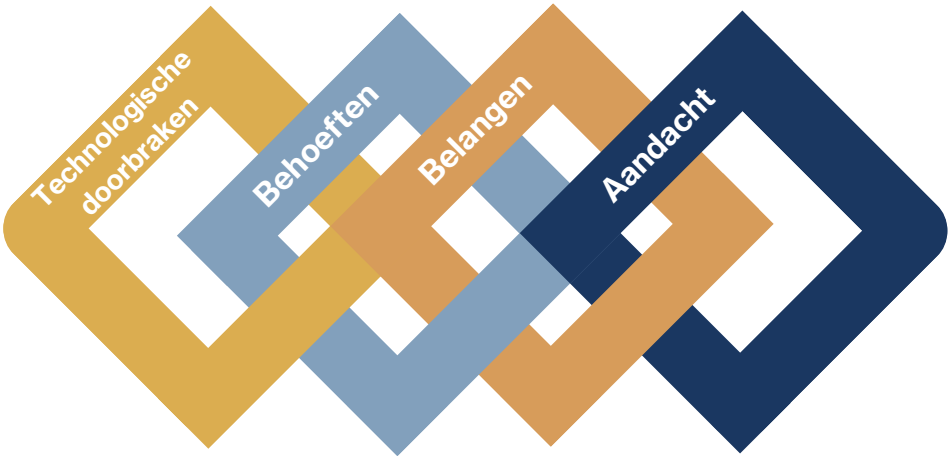
AI is géén technologie?!

Uiteraard is AI technologisch gefundeerd, maar in essentie is AI geen technologie maar een begrip voor allerlei geschakelde (wetenschappelijke, politieke en corporate) ambities om aan de hand van een grote verscheidenheid aan technologieën kunstmatig intelligente machines te bouwen.[64] De notie van AI als smalle technologie is misleidend. We moeten naar AI kijken als een op mondiaal niveau aangedreven proces om de ambitie van kunstmatige intelligentie waar te maken, aan de hand van allerlei technieken en versneld door vele verschillende agenda's. AI wordt soms ook wel vergeleken met de '**space race**'[65]: daarin was ook geen sprake van een enkele technologie maar van een door belangen en innovaties opgedreven mondiale beweging waarin ambities, public relations-management, nieuwe technische ontdekkingen, geld en grootmachtcompetitie samen de ruimtemissies aanjoegen. Omdat AI geen technologie is maar in werkelijkheid een geheel aan geschakelde factoren, is niet altijd duidelijk of deze ontwikkeling wordt aangejaagd door 1) **technologische doorbraken**, door 2) **behoeften**, door 3) **belangen** of door 4) de vele **aandacht** die ervoor is als gevolg van 'public myths' en angst. Waarschijnlijk betreft het een zichzelf versterkende combinatie van al deze factoren. We bespreken ze achtereenvolgens.

64 Kathleen Walch, Artificial Intelligence Is Not A Technology, 2018.

65 Verity Harding, What the cold war space race can teach us on AI, Financial Times, 2024.

Figuur 3: De ontwikkeling van AI verwijst naar in elkaar hakende drijvende krachten



3.1. De factor 'technologie'

Hoewel AI geen technologie is, is AI door en door technologisch gefundeerd. Verschillende doorbraken hebben tot een stroomversnelling geleid. De aard van zelflerende AI-technologie maakt het bovendien plausibel dat de capaciteit van systemen extreem versneld zou kunnen verbeteren.

Het succes van neurale netwerken. Lange tijd heeft AI in het teken gestaan van de ambitie om kunstmatige intelligentie te **ontwikkelen** door in essentie een **supercomputer** te bouwen die dezelfde logische functies kan uitvoeren als het brein. Deze weg bracht veel maar het grootste probleem ervan is dat alle functionaliteiten al besloten moeten liggen in een volledige architectuur.[66] Met de stroming van het **connectionisme** sloeg men een nieuwe weg in. Waarom geen systeem ontwikkelen dat zelf leert aan de hand van het aanleggen van statistische verbanden? Uiteindelijk bracht deze benadering **neurale netwerken** die een geavanceerde vorm van **deep learning** (DL) mogelijk maken. Deze werken dankzij activatiespreiding over verschillende neurale 'lagen'. In termen van het brein: inputsignalen leiden tot de activering van reeksen neuronen die weer zorgen voor activatie van andere reeksen. De neurale paden hangen af van welke associatieve relaties zijn aangelegd.[67] Het model van neurale netwerken weerspiegelt zo het cognitieve leerproces. Wanneer neurale instructies eenmaal aangelegd zijn dan is herkenning afhankelijk van ogenblikkelijke associaties. De associatieve paden worden bovendien versterkt naarmate de ervaringsinput toeneemt. Daarnaast kunnen verschillende neurale reeksen gelijktijdig worden geactiveerd waardoor het brein complexiteit

66 Een logisch systeem dat vooraf moet beschikken over alle symbolen en regels is 'star'. Het wordt noodzakelijk begrensd door de gesloten architectuur in plaats dat het een programma zelf kan uitbreiden. Een voorbeeld is de go- of schaakcomputer. Deze is goed in staat om binnen een gegeven set symbolen en spelregels strategische zetten te berekenen maar kan alleen leren binnen de eigen logische spelregels.

67 Denk aan het zien van een kat. Op grond van uiterlijke waarneming worden verschillende neurale paden geactiveerd (de vacht, een staart, de kleur). Het gevolg is dat de waarneming tot het woord 'kat' leidt, maar ook tot de term 'zoogdier' enzovoort. Elke waargenomen kat versterkt de neurale verbindingen die aangelegd zijn en maakt ons beter in het herkennen van allerlei kattenvariaties of wat juist geen kat is.

aankan. Juist dankzij de ontwikkeling van neurale netwerken heeft moderne AI kunnen doorbreken (zie bijlage 1 voor meer achtergrond).[68]

AI voor de massa: transformer-based modellen. Neurale leerprocessen zijn afhankelijk van de betrouwbaarheid van het algoritmedesign en de data waarop wordt getraind. De huidige revolutie in generatieve AI komt technisch gezien voor een groot deel voort uit de uitvinding van transformer-based modellen in 2018. Het eerste volledige transformer-model werd gebouwd door een team van Google, voortbouwend op tientallen jaren aan ontwikkelingen. Hoewel eerst ontworpen voor tekstinvoer, zijn huidige toepassingen ontwikkeld om ook een hoge mate van verfijning van visuele- en audiocreatie te bieden. Er zijn ook multimodale modellen ontwikkeld die tegelijkertijd met verschillende soorten data (tekst, visueel, audio) kunnen werken, voor zowel input als output (bijvoorbeeld tekstprompts die visuele output creëren).

Transformer-based modellen

Een 'transformer' bestaat uit twee hoofdonderdelen: een encoder en een decoder. De encoder zoekt herhalende patronen tussen tokens: individuele componenten van een tekst. Bij het lezen van een zin kijkt de encoder bijvoorbeeld zowel naar de betekenis van individuele woorden als naar de positie van het woord in een zin en de relatie ervan tot andere woorden. Zo kan de encoder ordenen welke woorden het belangrijkst zijn om betekenis uit de tekst te halen en kan het langere teksten sneller begrijpen. Op zijn beurt kan de decoder, gebaseerd op wat de encoder van de invoer heeft geleerd, een output produceren met behulp van statistische voorspellingen van wat de meest waarschijnlijke volgende token in een bepaalde reeks zou zijn, bijvoorbeeld, het volgende woord in een zin. Ook hier ligt de nadruk op zeker weten dat niet alleen de lineaire volgorde van woorden zinvol is, maar dat betekenis en context correct worden meegewogen in de tekst als geheel. De effectiviteit van de encoder en decoder is direct afhankelijk van de hoeveelheid data waarop een model wordt getraind. Hoe meer data, hoe beter de voorspellingen. ChatGPT (GPT staat voor Generative Pretrained Transformer) van OpenAI werd het bekendste Transformer-based model dat beschikbaar kwam voor massaal gebruik.

68 IBM, What Are Neural Networks?, n.d.

Diverse typen Deep Learning-modellen

Hoewel andere DL-modellen minder populair zijn geworden na de transformer revolutie in 2018, is het nog steeds relevant om ze te noemen omdat ze helpen met het contextualiseren van hoe revolutionair de nieuwste ontwikkelingen zijn geweest.

- **Recurrent neural networks (RNN) –** diende als basis voor spraakherkenning en vertaling (bijvoorbeeld in oudere versies van Google Translate). RNNs werken door opdrachten herhaaldelijk door het input/output-proces te halen. Dit is mogelijk door een soort 'selectief geheugen' voor de 'gestapelde' verwerking van informatie.
- **Convolutional neural networks (CNN) –** worden voornamelijk gebruikt voor visuele taken, waarbij AI-algoritmen afbeeldingen kunnen herkennen en classificeren. CNNs werken door patronen in een afbeelding te doorlopen en voort te bouwen op wat ze al hebben geleerd totdat ze een object als geheel kunnen identificeren.
- **Generative adversarial networks (GAN) –** zijn bijzonder krachtig voor het creëren van kunstmatige output. Ze bestaan uit twee concurrerende neurale netwerken, een generator en een discriminator. De generator produceert output die de discriminator vervolgens classificeert als echt of nep. Door dit te doen traint de discriminator de generator in het produceren van steeds levensechtere outputs. GANs worden o.a. gebruikt voor deepfakes.

Extreme versnelling. De versnelling van Transformers-based modellen is extreem, ook in vergelijking met de wet van Moore.[69] In 2018 beschikte GPT-1 over 117 miljoen parameters. Het vierde model, gelanceerd in 2023, beschikt (naar schatting) over meer dan een biljoen parameters. De rekenkracht die benodigd is om de meest krachtige AI-systemen te laten draaien en te trainen is de afgelopen tien jaar elk jaar met een factor tien toegenomen. De meest geavanceerde AI-modellen van het moment gebruiken vijf miljard keer meer rekenkracht dan de geavanceerde modellen van tien jaar geleden. Op grond van extrapolatie van deze ontwikkeling zou er binnen vijf jaar een model kunnen bestaan dat dezelfde neurale capaciteit als het menselijk brein bezit, bestaande uit circa honderd biljoen synapsen.[70] Deze ontwikkelingstoename is echter afhankelijk van investeringen in hardware en infrastructuur en wordt in die zin ook economisch en hardware-technisch begrensd. Daar staat tegenover dat AI-modellen naar verwachting snel efficiënter zullen worden, bijvoorbeeld in de manier waarop ze informatie verwerken, en ook grote sprongen in hardware (zoals kwantumcomputers) mogelijk zijn.

69 In 1965 voorspelde Gordon Moore, de medeoprichter van Intel, dat het aantal componenten per geïntegreerde schakeling elk jaar voor een periode van tenminste tien jaar zou verdubbelen. Anders gezegd, de prestaties van een computer van dezelfde prijs zouden elke 12 maanden verdubbelen. In 1975 stelde Moore de prognose bij tot een verdubbeling om de twee jaar. Deze voorspelling is grotendeels uitgekomen. In de periode van 1971 tot 2018 is het aantal transistors op een microprocessor gestegen van enkele honderden tot meer dan tien miljard. Als gevolg hiervan zien we de exponentiële toename van de computerverwerkingskracht. Deze groei is op zijn beurt cruciaal geweest bij het faciliteren van de huidige AI-revolutie.

70 Ian Bremmer and Mustafa Suleyman. The AI Power Paradox, 2023.

Inherente tekortkomingen van de technologie. Rondom AI hangt de pretentie dat ze snel superieur zal zijn aan het menselijk brein. Deze ideeën worden ingegeven door wat we ook wel een **reductionistische kijk**[71] kunnen noemen op wat menselijke intelligentie werkelijk is. Computerwetenschappers en filosofen zijn het hierover met elkaar vaak fundamenteel oneens. Computerwetenschappers verklaren menselijke intelligentie te veel vanuit hun gemodelleerde visies. Filosofen benadrukken de menselijke eigenheid, of het feit dat nog niemand het ontstaan van het bewustzijn heeft kunnen verklaren vanuit de architectuur. Er is bovendien een aantal werkingsprincipes van AI-machines die typerend zijn voor hun niet-organische systeemarchictectuur. AI-machines kunnen bijvoorbeeld maar moeilijk compenseren voor **vervuilde input**. Dit heet ook wel het **garbage in, garbage out**-beginsel. Dit punt demonstreert de voorlopig aanhoudende kwetsbaarheid dat machines op basis van foute input makkelijk te misleiden of te corrumperen zijn. Het betekent dat mensen beseffen, zeker wanneer de hype meer is gaan liggen, dat men niet blind kan varen op de betrouwbaarheid van AI.

AI-systemen zijn, vanwege hun logische architectuur, **rigide** in hun uitkomsten. Dit schept een probleem: in de menselijke begripswereld zijn er altijd meerdere opties te wegen. Deze variatie is van belang omdat ze aanzet tot reflectie en herpakken mogelijk maakt. Een AI-systeem kan problematische effecten veroorzaken wanneer het de wereld van opties negeert en slechts tot een enkel antwoord komt. Deze **padafhankelijkheid** is extra problematisch wanneer maar gebrekkig inzicht is in de logische stappen van het algoritme. Om deze rigiditeit te doorbreken ontstaan er meer algoritmen die beter overweg kunnen met zogenaamde **counterfactuals**; dit zijn alternatieve 'als dan'-scenario's die een systeem kan doorlopen om meerdere uitkomsten aan te bieden als alternatief voor het primaire antwoord. Op termijn zal het vermogen van AI om met counterfactuals om te gaan een belangrijke stap zijn om in de buurt van menselijke beslissingspatronen te komen. Het technologisch potentieel blijft in die zin het onderscheidend menselijk vermogen naar de kroon steken.

3.2. De factor 'behoeften'

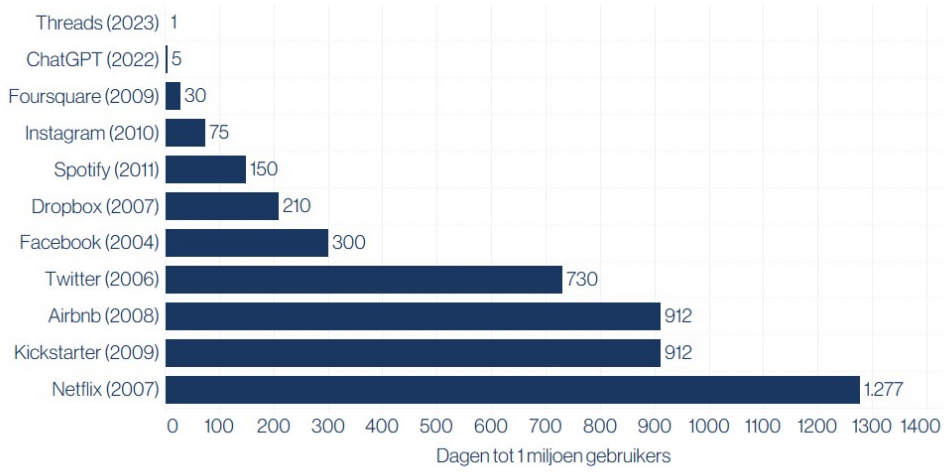
De brede ontvangst van Transformer-based modellen is te danken aan de toepasbaarheid voor een scala aan behoeften. Na 2018 is er een explosie geweest van publiekelijk beschikbare AI-programma's gericht op diverse niches van communicatie en creativiteit. Deze explosie is kwantitatief zichtbaar. Zo is het aantal nieuwe AI-patenten exponentieel gestegen van ongeveer 10.000 in 2017 tot 141.000 in 2021.[72] Voornamelijk ChatGPT heeft op een recordtempo gebruikers naar zich toe getrokken. Dit is mogelijk gemaakt doordat OpenAI, het moederbedrijf uit Silicon Valley, veel heeft kunnen investeren in het optuigen van rekencapaciteit om al deze gebruikers de baas te kunnen. Waar ChatGPT niet noodzakelijk het beste transformer-systeem is, heeft OpenAI hierdoor wel een grote slag kunnen slaan om de gunst van de massa.

71 Dan Aurelian Botică, Artificial Intelligence and the Concept of "Human Thinking", 2016.

72 Stanford University Human-Centered Artifical Intelligence, Artificial Intelligence Index Report 2022, 2022.

Figuur 4: Benodigde tijd van online diensten om 1 miljoen gebruikers te bereiken

ChatGPT vestigde het record voor het snelstgroeiende platform in de geschiedenis, het platform werd gelanceerd op 30 november 2022 en had slechts 5 dagen na de lancering al 1 miljoen gebruikers. Later werd dit record verbroken door de Thread-app van Instagram



Bron: Nerdynav

Efficiency. De razendsnelle opkomst van AI-toepassingen leunt enorm sterk op de **productiviteits- en efficiencywinst** die ze beloven. AI-toepassingen moeten makkelijk in te passen zijn in productie- en arbeidsprocessen en op die manier aantoonbare winsten kunnen opleveren in tijd en geld.[73] Gebruikers, bedrijven, overheden zijn in korte tijd overtuigd geraakt van de efficiency en tijdsbesparing die generatieve AI oplevert. Volgens sommige onderzoeken maakt AI de beloften van verhoogde efficiëntie in bedrijven ook daadwerkelijk waar.[74] De positieve perceptie onder bedrijven is in ieder geval enorm, getuige een onderzoek van Forbes, dat laat zien dat 97% van de ondernemers overtuigd is van de meerwaarde van ChatGPT.[75]

Maar of efficiencywinst bereikt kan worden is dan ook niet het belangrijkste argument om aan AI te twijfelen. De jacht op verhoogde efficiency heeft implicaties voor hoe mensen werk ervaren. Achter AI speelt voortdurend de clash tussen winstoptimalisatie en de inferieure positie van menselijke arbeid daarin. Een voorbeeld is het Amerikaanse Amazon dat voortdurend in de clinch ligt met medewerkers over taken die worden vervangen of gemonitord worden door AI.[76] We zien het nodige verzet van de werkvloer. Maar aangezien de 'betaler bepaalt', zullen deze voor medewerkers bedenkelijke ontwikkelingen de implementatie van algoritmen en robots niet extreem in de weg staan en is de verwachting dat AI als efficiencyverhogend instrument gemeengoed zal worden.

De hoop op empowerment. Ook de mogelijke empowerment die AI individuen belooft onder de noemer van democratisering, is een factor die bij heeft gedragen aan het succesverhaal. Omdat een deel van de technologie open source is, of anderszins laagdrempelig toepasbaar is te maken, kunnen AI modellen als verlengstuk worden gebruikt voor de eigen capaciteit. Er zijn allerlei GPT-tools om, bijvoorbeeld, aandelen voor je te laten verhandelen, online waar te

73 Bergur Thormundsson, Industries Using ChatGPT in Their Business 2023, 2023.

74 Brynjolfsson, Erik, Danielle Li, and Lindsey R. Raymond, Generative AI at work, 2023.

75 Katherine Haan, 24 Top AI Statistics & Trends In 2024, Forbes Advisor, 2024.

76 Annabelle Williams, How Amazon Tracks Workers, From Cameras to a Spy Agency, Business Insider, 2021.

verkopen, designs te ontwerpen en te patenteren, een influencer-account te laten beheren (al dan niet van een fictief persoon). Kenmerkend voor de belofte van AI nu is dat de mogelijkheden in de handen van elk individu onbegrensd kunnen zijn. Dit levert een bijna hallucinant toekomstbeeld op waarin het onmogelijk voor te stellen is hoe regeringen zeggenschap kunnen behouden over de richting van verschillende manieren waarop individuen AI kunnen toepassen.[77] Overigens, zoals we in hoofdstuk 4 benadrukken, is lang niet iedereen overtuigd van het (brede) empowerment-potentieel van AI, en gaan er veel kritische stemmen op dat AI juist zal bijdragen aan sociale ongelijkheid.

Het zichzelf aanjagende technologiseringsproces. Een belangrijk punt is dat AI-technologie zich ook zonder manifeste menselijke behoefte zal ontwikkelen en, zelfs tegen wil en dank, meer greep krijgt op het dagelijkse consumentenbestaan. AI-toepassingen vinden hun weg binnen het kapitalistisch model. Wat wil zeggen dat veel behoeften zich laten creëren zonder dat hier een expliciete vraag voor nodig is. De AI-industrie zal proberen de begeerlijkheid te vergroten van producten en daarmee processen inzetten van de commodificatie van AI. Met name voor AI geldt (omdat het ook een aandrijvende technologie kan zijn) dat toepassingen onbewust hun weg gaan vinden in allerlei applicaties die we nu al bezitten. AI introduceert nieuwe producten en mogelijkheden die zo ingebed raken in de samenleving dat ons leven ervan afhankelijk wordt en het praktisch ondenkbaar is de door AI geïnduceerde veranderingen om te keren.

3.3. **De factor 'belangen'**

Economische belangen. Grote economische belangen drijven de aandacht voor AI op. De geprognosticeerde totale market cap van AI-techniek zal in 2030 oplopen tot richting de $1900 miljard.[78] Ter vergelijking: de waarde van de mondiale wapenmarkt is momenteel minder dan $200 miljard.[79]. Er is bedrijven en staten veel aan gelegen om een stuk van deze taart te bemachtigen. De ontwikkeling van AI vergt torenhoge investeringen maar jaagt deze ook aan.[80]

77 Dan Hendrycks, Mantas Mazeika, and Thomas Woodside, An Overview of Catastrophic AI Risks, 2023.

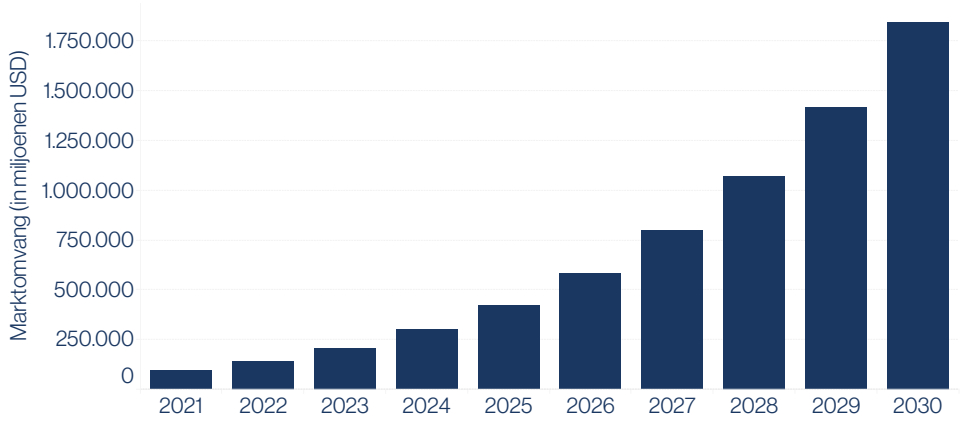
78 Next Move Strategy Consulting, Artificial Intelligence Market Size 2030, 2023.

79 SIPRI, Financial Value of the Global Arms Trade, n.d.

80 Rudy van Belkom, Do(n't) believe the hype, n.d.

Figuur 5: Verwachte mondiale groei van de 'AI market cap'.[81]

In 2030 heeft AI een geschatte marktomvang van 1,847 miljard in marktinkomsten



Bron: Next Move Strategy Consulting
*data 2024-2030 zijn voorspellingen

De afhankelijkheid van industrie. 'No Chip, No AI' wil zeggen: de mogelijkheden en prestaties van AI-systemen blijven afhankelijk van hardwarecomponenten als 1) werkgeheugen, 2) gegevensopslag en 3) rekenkracht. De werking van AI is daarmee afhankelijk van een enorme fysieke infrastructuur. Generatieve algoritmen hebben een onstilbare datahonger omdat de statistische kwaliteit van de output verhoogd wordt aan de hand van meer input en een uitdijend, complexer neuraal netwerk. AI drijft daarom de behoefte aan steeds grotere datasets die moeten worden opgeslagen en beheerd en moeten kunnen worden verwerkt. De initiële data waarop GPT-4 getraind is, wordt geschat op 13 triljard tekens. Maar dat is alleen de initiële trainingsdata. Met elke opdracht wordt het algoritme verder getraind. Large Language Model-algoritmen kunnen worden ingezet voor particuliere datasets waardoor iedereen met een eigen toepassingsdoel bijdraagt aan de toename van gegevensverwerking. De honger naar data neemt dus niet alleen toe naar rato van toenemende verwerkingscapaciteit, maar naar rato van individuele gebruikers. Ook hebben ze een enorme hoeveelheid (specialistische) rekenkracht nodig om getraind te worden. Alleen al de kosten voor de benodigde rekenkracht voor de training van GPT-4 wordt naar verluid geschat op $63 miljoen.[82]

Industriële belangen. Hoogwaardige AI-systemen hebben een enorme rekenkracht nodig. Daarvoor zijn high-end chips nodig om de informatie te verwerken. De vraag hiernaar is sinds de introductie van GPT exponentieel gestegen en is bovendien in handen van slechts enkele spelers en toevoerlijnen.[83] Naast de milieudruk die de mijnactiviteiten voor de fabricage van hoogwaardige chips veroorzaakt is er ook sprake van toegenomen geopolitieke spanningen door zogenaamde **grootmachtcompetitie**: het formele Amerikaanse beleid is om de komende jaren leidend te blijven op het gebied van AI en chiptechnologie, en de ontwikkeling van China juist te vertragen. Taiwanese, Amerikaanse en Europese partijen zijn daarbij vooral gericht op het bouwen van de chips en de chipsmachines. De grondstoffen en grondstoffentoevoer is echter grotendeels in handen van China. Diverse studies en scenario's wijzen uit dat deze spanningen kunnen leiden tot conflicten die ook repercussies kunnen hebben voor de

81 Bergur Thormundsson, Artificial intelligence (AI) market size worldwide in 2021 with a forecast until 2030, 2024.

82 Maximilian Schreiner, GPT-4 architecture, datasets, costs and more leaked, 2023.

83 Zo steeg de kwartaalomzet van Nividia van 6,01 miljard USD naar 22,1 miljard USD. De stijgende omzet wordt verklaard door een toenemende vraag naar Nividia's chips. Deze high-end chips zijn essentieel voor generatieve AI-chatbots, zoals ChatGPT. Associated Press, Nvidia's 4Q revenue, profit soar thanks to demand for its chips used for artificial intelligence, 2024

Nederlandse samenleving.[84] Sommige deskundigen beweren dat de chipcapaciteit spoedig tegen fysieke limieten aanloopt wat het vermogen van AI beperkt. De kans is echter groot dat deze fysieke limieten in de toekomst weer worden opgerekt door nieuwe typen gegevensdragers en -geleiders die nu nog in een experimenteel stadium verkeren. Denk aan fotonica of kwantumcomputing.[85]

Regionale belangen en de exploitatie van lokale capaciteiten. De infrastructurele voorwaarden voor AI staan niet altijd op het netvlies maar hebben stevige consequenties. De carbon footprint van AI, ecologische schade als gevolg van mijnwerkzaamheden, maar ook de afhankelijkheid van goedkope arbeid en het risico op toenemende geopolitieke instabiliteit geven aanleiding tot zorg over het verantwoord gebruik van AI. Naarmate het gebruik van AI normaliseert zullen deze kwesties vaker op de agenda verschijnen. Denk aan:

- **Elektriciteitsverbruik.** Grote partijen zoals Microsoft, Meta en Alphabet hebben (ook in Nederland) enorme datacenters opgetrokken om gebruikersgegevens te kunnen opslaan. Dergelijke centra zijn niet onomstreden omdat ze een enorme hoeveelheid energie consumeren, wat kan oplopen tot enkele percentages van het nationaal stroomverbruik.[86]
- Het **waterverbruik** voor hardwarekoeling. Ter indicatie: een studie wijst uit dat een halve liter water verbruikt wordt voor elke tien tot vijftig vragen die aan ChatGPT gesteld worden.[87] Het leidt tot exorbitante hoeveelheden water die moeten worden gebruikt om datacenters te koelen. Deze carbon footprint van AI-systemen wordt extra schrijnend tegen de achtergrond van allerlei duurzaamheidsmaatregelen en regionale klimaatproblemen zoals verdroging in gebieden waar datacenters staan.[88]
- **Trainingsarbeid.** Niet altijd worden algoritmen getraind dankzij computerkracht. Beeldherkenningssystemen worden dikwijls getraind en gecorrigeerd door mensen. Uiteindelijk zijn er mensenhanden nodig die informatie labelen om een algoritme te kalibreren of om herkenningsfouten van een systeem handmatig te corrigeren. Dit soort handmatige arbeid is echter zeer saai en repetitief en wordt daarom vooral uitgevoerd in lagelonenlanden.[89]

3.4. **De factor 'aandacht'**

Er is sprake van een valse analogie wanneer de impact van AI wordt vergeleken met de impact van technologieën zoals die van televisie of 5G. Dit soort technologieën zijn afgebakende instrumenten die het bereik en de mogelijkheden van het menselijk handelen vergroten. Hoewel AI ook veel mogelijkheden biedt om ons handelen te vergroten, verwijst AI primair naar een nieuwe, potentieel zelfstandige vorm van **intelligentie** die buiten onze regie zal kunnen vallen. Dit gegeven prikkelt de publieke fantasie maar leidt ook tot angsten die op verschillende manieren geïnduceerd zijn:

---

84 Deze ontwikkelingen zijn eerder in het kader van het Strategische Monitor Politie-programma behandeld; zie Joris Teer, Mattia Bertolini en Benedetta Girardi, Competitie tussen grootmachten en maatschappelijke stabiliteit in Nederland, 2023.

85 Rijksoverheid, De Nationale Technologiestrategie, 2024.

86 Het Centraal Bureau voor de Statistiek (CBS) rapporteerde dat Nederland in 2020 2,8% van het landelijk elektriciteitsverbruik leverde aan datacenters. CBS, Elektriciteit geleverd aan datacenters, 2017-2021, 2022.

87 Pengfei Li et al., Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models, 2023.

88 Shannon Osaka, A New Front in the Water Wars: Your Internet Use, 2023

89 Kate Crawford, *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, 2021.

Een nieuw referentiekader, een nieuwe menselijke zelfperceptie. AI kan veel intelligente taken beter uitvoeren en heeft de potentie een eigen handelingsvermogen te ontwikkelen. Dit is een zeer ontzagwekkend gegeven omdat het 'de mens' van zijn voetstuk dreigt te stoten van technologisch alleenheerser over de aarde.[90] Omdat AI zo'n belangrijke rol kan gaan vervullen in allerlei tot voor kort unieke menselijke activiteiten, zullen mensen moeten wennen aan een nieuwe gradatie van menselijke afhankelijkheid van intelligente automatisering. Dit legt nieuwe filosofische implicaties in de waagschaal die bij de introductie van andere technologieën niet aan de orde waren. Zoals historicus Yuval Harari treffend uitdrukte: 'AI has hacked the operating system of human civilisation' – intelligente automatisering ondersteunt de fundamenten van onze betekenisorde maar verandert die ook.[91] Omdat AI zelf creatieve output genereert en over onze bestaande kaders heen drapeert, kan AI ons hele wereldbeeld doen veranderen.

'Morele paniek'. Naast angst voor de ontwrichtende impact van AI op het menselijk leven genereert AI 'morele paniek' (– begrepen als sociologisch begrip dat duidt op een door de media aangejaagde maar vaak ook voorbarige publieke verontwaardiging). In het geval van AI heeft deze angst vaak nog geen duidelijke focus vanwege het kaleidoscopische beeld van toepassingen. Sommigen vrezen een samenleving waarin robots het voor het zeggen krijgen. Anderen vrezen de vervanging van menselijke arbeidstaken door AI. Deze angst is overigens niet nieuw. Gedurende de eerste 'AI-boom' (de discussie over intelligente computers werd toen gevoerd aan de hand van het begrip 'cybernetics' - zie bijlage 2) ontstond al veel onrust en debat over computers die binnen enkele jaren alle banen zouden hebben overgenomen. Maatschappelijke angst voor de machine is een op zichzelf staande factor op de maatschappelijke stabiliteit. We bespreken het thema verder in hoofdstuk 4. Het punt hier is dat angst en onwetendheid ervoor zorgen dat de aandacht voor AI verder wordt aangejaagd. De geschiedenis van AI-booms toont dat een verhoogde publieke aandacht bijdraagt aan de binnenlandse politieke tractie. Men wil de regie niet verliezen en concurrerend blijven omdat de angst voor AI in handen van andere grootmachten een nog groter angstbeeld vertegenwoordigt.

3.5. **Regulering?**

De samenloop van bovenstaande processen maken dat de transformatieve kracht van AI moeilijk in te dammen is. De mogelijkheden om deze krachten te temperen of te kanaliseren zijn beperkt. De belangrijkste mitigerende maatregelen voor regeringen bestaan uit regulering en wetgeving. Maar op dit punt zijn er grote hindernissen te nemen. Daarbij is het de vraag welke vormen van regulering voor welke verschijnselen daadwerkelijk effectief zijn. Hindernissen zijn er op het gebied van het bereiken van (internationale) **overeenstemming** over normen en regels; controlemogelijkheden en **handhaving**; en de gemakkelijk te bereiken **voordelen voor overtreders** als offset tegen eventuele repercussies. Deze tekortkomingen zijn ernstig maar toch moet ook de prestatie worden onderkend van de EU om de impact van AI enigszins te kunnen beheersen en beter te kunnen scheiden tussen toepassingen die bevorderlijk zijn, riskant of die rechtstreeks verboden moeten worden.

---

90 Gerben Bakker, 'Monster van AI' blijkt vooral weerspiegeling van menselijk gebrek, EW, 2023.

91 Yuval Harari, Yuval Noah Harari argues that AI has hacked the operating system of human civilisation, The Economist, 2023.

Responsible AI. De notie van verantwoorde AI (responsible AI) vormt een belangrijke grondslag voor wet- en regelgeving omtrent AI zowel in Nederland als binnen Europa. De EU is op dit vlak leidend voor Nederland. De EU houdt eraan vast dat mens en gebruiker centraal moeten staan in de beoordeling van AI-systemen. In 2019 heeft de Europese Commissie het voortouw genomen en een ethisch kader ontwikkeld, gebaseerd op vier morele principes: (1) autonomie; (2) niet-schadelijkheid; (3) rechtvaardigheid; en (4) verantwoording. Op deze grond heeft de EU ethische vereisten opgesteld als leidraad voor de ethische beoordeling van AI-systemen (Zie bijlage 3).[92] De risico's die men met responsible AI wil voorkomen voeren vaak terug op de impact die AI heeft op de fundamentele rechtspositie van het individu en andere waarden die AI-toepassingen niet zouden mogen schenden. Voor AI-systemen met een hoog risico zal de ontwikkelaar moeten kunnen aantonen dat het instrument belangrijke ethische principes niet schendt.

De Europese Wet op Artificiële Intelligentie (de WAI of AI Act). De ethische vereisten zijn leidend geweest voor de WAI. Op 9 december 2023 werd een historisch akkoord bereikt over deze wet, een noviteit in de wereld. De wet is bedoeld om te komen tot harmonisering van regels voor AI binnen de EU en vormt de leidraad voor nationale wet- en regelgeving. De benadering is vergelijkbaar met productregulering in vele andere sectoren: de EU kan als grootste economisch blok mondiale sectoren beïnvloeden door strenge producteisen te stellen als voorwaarde voor toelating op de interne markt.

De WAI dient verschillende doelen: 1) de WAI is gebouwd op de voornoemde ethische principes met het doel van **bescherming van rechten**, vrijheden en waarden van Europese burgers. Producten kunnen op grond van hun functionaliteiten in verschillende risicocategorieën vallen. De WAI verbiedt bepaalde hoog-risicotoepassingen of dwingt tot verantwoordingsmechanismen. 2) De wet heeft echter niet alleen een restrictief doel. Door richtlijnen op te stellen hoopt de Europese Commissie ook te koersen op **verantwoorde innovatie**. Bedrijven krijgen de kans om EU-compliant producten en diensten te ontwikkelen die ook voor andere werelddelen als veilig kunnen gaan gelden. Ook moet duidelijkheid over rechtszekerheid in alle EU-landen koudwatervrees wegnemen voor nieuwe investeringen en versnippering voorkomen. 3) Voorts moet de wetgeving **lacunes dichten** in (nationale) wetten van gegevensbescherming, zoals de AVG, en biedt ze een stimulans voor wettelijke doorontwikkeling op nationaal niveau. Ten slotte regelt de WAI zaken op het gebied van **handhaving**. Op dit punt bestaan echter nog grote onzekerheden.

Bestaande wettelijke kaders. De eerder ingevoerde Europese **General Data Protection Regulation (GDPR), Digital Services Act (DSA)** en **Digital Markets Act (DMA)**, respectievelijk gericht op het reguleren van het gebruik van gevoelige persoonsgegevens, van sociale media en van (de macht van) de grote digitale marktpartijen, zijn niet specifiek gericht op AI maar hebben wel degelijk invloed op de ontwikkeling en het gebruik. De Nederlandse AVG, gebaseerd op de Europese GDPR, vormt mogelijk nog een belangrijker beschermingsmechanisme tegen schadelijke AI dan de WAI zelf omdat de gegevensverordening verhindert dat bedrijven of instituties allerlei gegevens mogen verzamelen, nog voorafgaand aan de vraag of ze met behulp van AI-technieken verwerkt mogen worden. (Zie bijlage 3 voor uitgebreidere toelichting op de wet- en regelgeving en hun belang voor de regulering van AI.)

Kan regulering de AI-transformatie kanaliseren? De enorme Europese prestaties op het gebied van AI-regulering ten spijt, er blijven belangrijke obstakels die verhinderen dat de

92 Europese Commissie, Ethics guidelines for trustworthy AI, 2019.

introductie van AI zich eenvoudig laat sturen. De zogenaamde **AI Power Paradox**[93] stelt dat de kracht van AI allerlei beleidsvraagstukken veroorzaakt, maar dat deze juist vanwege de enorme ontwikkelingskracht onmogelijk snel op te lossen zijn. Zowel in beleidsontwikkeling als in wet- en regelgeving lopen overheden achter de feiten aan. De kans is daarom groter dat handhavende machten te maken krijgen met toepassingsgebieden van AI die nog niet gereguleerd zijn. Andersom kan regelgeving ook de ambities van AI-bedrijven remmen. Dit maakt het voor elk land essentieel om de algemene waarden te begrijpen die het wil behouden onder de vrijwel onvermijdelijke integratie van AI in de samenleving. Nationale overheden, waaronder die van Nederland, kunnen een belangrijke rol nemen in het ontwerpen of versterken van ethische raamwerken rondom AI.[94] We komen op deze zaken terug in hoofdstuk 5.

93 Ian Bremmer and Mustafa Suleyman. The AI Power Paradox, Foreign Affairs, 2023.

94 Ali Hashmi, AI Ethics: The Next Big Thing In Government - Anticipating the Impact of AI Ethics within the Public Sector, 2019.

4. De impact op de maatschappelijke stabiliteit

'Maatschappelijke stabiliteit' is een macrogrootheid. Voor de impact op de maatschappelijke stabiliteit moet wisselend gekeken worden naar hoe AI-toepassingen menselijk gedrag en de ervaring raakt maar ook naar hoe de optelsom van allerlei AI-ontwikkelingen leidt tot meer abstracte maar desalniettemin invloedrijke dynamiek. We herhalen dat de maatschappelijke stabiliteit verwijst naar het onverstoorbaar kunnen voortbestaan van de democratische rechtsorde. Waar de transformatieve kracht van AI de maatschappij ten goede en ten kwade kan beïnvloeden, gaat het hier vooral om het belichten van de manier waarop deze transformatie interne spanningen kan vergroten (of juist mitigeren) of sentimenten kan oproepen die leiden tot spanning. Er is zo de **directe technologische impact** op hoe we leven, en er is de impact als gevolg van **de perceptie over deze veranderingen** die op elkaar ingrijpen.

Naast deze op elkaar ingrijpende beïnvloedingsrelaties van AI, is het verstandig om onderscheid te maken tussen verschillende niveaus waarop deze impact speelt. Discussies over de impact van technologie willen soms ontaarden in alarmisme omdat de angst voor negatieve effecten makkelijk wordt gepolitiseerd. Hierdoor blijven de minder dramatische, incrementele veranderingsprocessen buiten schot, wat op zichzelf genomen een maatschappelijk risico kan worden. We stellen voor de maatschappelijke impact van technologie te bekijken op drie perspectivische niveaus:

1. **Existentieel en geopolitiek niveau** - i.e. de geprojecteerde negatieve gebeurtenissen dan wel angsteffecten die ontstaan vanuit de achtergrond van bedreigingen welke men ervaart in relatie tot het voortbestaan van de mensheid of welke voortkomen vanuit een geopolitiek krachtenveld. Dit type impact ontstaat op een maatschappij-overstijgende manier maar heeft de potentie om de Nederlandse interne stabiliteit wel te verstoren.
2. **Macroniveau.** 'Maatschappelijke stabiliteit' is een macrogrootheid. Voor de impact moet vaak gekeken worden naar het samenspel van effecten door de toepassingsgebieden heen – waarbij het geheel ingrijpender is dan de som der delen. We vinden op dit niveau bijvoorbeeld de geaggregeerde psychologische effecten van sociale AI-toepassingen.
3. **Toepassingsniveau.** Daaronder vinden we het (meso-)niveau: dit duidt op concrete effecten die te verwachten zijn voor toepassingsgebieden zoals beleidsondersteuning of sociale monitoring. Het gaat vaak om effecten die voortkomen uit het design die hierom goed inzichtelijk te maken zijn. Maar deze concreetheid werkt ook bestuurlijk micromanagement in de hand en leidt af van het zicht op het samenspel van factoren.

Op **existentieel niveau** speelt de angst voor het **onbekende** en onbegrepene een hoofdrol. De geschiedenis laat zien dat vernieuwende technologieën tegenkrachten oproepen die niet per se rationeel zijn. Kampen laten zich leiden door angst voor beschavingsverval en de

aantasting van instituties en waarden.[95] Daartegenover roeren zich groepen die de kansen benadrukken. Dergelijke tweespalt is voorgekomen bij de introductie van bijvoorbeeld (kabel) televisie[96], internet en gaming. Soms blijkt de onrust over nieuwe technologie overtrokken.[97] Angsten hebben dan hun werk in politieke zin al gedaan. Zo zijn de effecten van het spelen van videogames op gewelddadig gedrag bijvoorbeeld marginaal aangetoond, maar de vooroordelen hierover blijven bestaan.[98] We zien daarnaast dat AI-technologie buitengewoon is in de potentiële impact die ze heeft op de menselijke manier van leven in de wereld. Daarbij interfereert ze ook in de geopolitieke realiteit. Verdere instabiliteit van de geopolitieke balans kan een factor van belang zijn voor de Nederlandse samenleving.

Op **macroniveau** is impact vaak dermate gefacetteerd dat het moeilijk is om greep te krijgen op wat nu precies welke instabiliteit veroorzaakt. De complexiteit staat eenvoudig te isoleren oorzaak-gevolgrelaties in de weg. Neem de smartphone. Deze vormt een netwerk van technieken zoals het touchscreen, het internetprotocol, software interfaces en met (AI-)algoritmen die content 'klaarzetten'. Elk van deze componenten heeft zijn eigen implicaties, en toch moeten ze als verbonden met elkaar beschouwd worden om de effecten te begrijpen.[99]

Op concreet **toepassingsniveau** zijn sommige effecten van allerlei technologische toepassingen soms juist wel weer heel tastbaar. Waar het ingewikkelde web van sociale effecten moeilijk te ontrafelen is, is de potentiële gezondheidsschade van GSM-straling meetbaar. Het negatieve effect op de maatschappelijke stabiliteit is bovendien makkelijker te identificeren. Hoe concreter de effecten zijn, hoe eenvoudiger het is om maatregelen te ontwerpen. Dit verklaart ook waarom regulering zich vaak richt op concrete instrumentatie maar dat het ingewikkelder ligt om gevolgen van het samenspel van factoren op te lossen.

Hieronder geven we per niveau de belangrijke bedreigingen en kansen. De **dreigingen** kunnen onzekerheid en onrust genereren en, bij voldoende kracht en omvang, tot maatschappelijke instabiliteit leiden. Maar het is van belang te beseffen dat ook de **kansen** tot onrust en instabiliteit kunnen leiden. Kansen zetten immers veranderingen in gang en leiden, omdat ze in het algemeen niet gelijkelijk zijn verdeeld, tot (verdiepte of nieuwe) ongelijkheid.

4.1. Impact op existentieel en geopolitiek niveau

Existentiële en geopolitieke dreigingen verwijzen (breder dan de notie van 'existential threat') naar maatschappij-overstijgende ontwikkelingen die desondanks de stabiliteit van samenlevingen als de onze 'van buiten naar binnen' kunnen verstoren. Dat de impact van AI mondiaal zo fors wordt voorgesteld heeft te maken met de (deels geprojecteerde) **transformatieve**

95 Elizabeth M. Perse and Jennifer Lambe, Media effects and society, 2016.

96 Gunn Enli e.a., From Fear of Television to Fear for Television: Five political debates about new technologies, 2013.

97 Vijay Aswaran, The Case against Techno-Pessimism, 2018.

98 Simone Kühn e.a., Does playing violent video games cause aggression? A longitudinal intervention study, Nature, 2019.

99 Bijvoorbeeld, de gemakkelijke interface heeft een effect op de verslavingsgevoeligheid. Maar het algoritme heeft dat ook. De permanente verbinding met anderen over de wereld maakt weer een verbrede sociale vergelijking mogelijk waardoor de eigen identiteit onder druk komt te staan. De toename van depressieve gevoelens onder jongeren blijkt een gevolg van een complex samenspel.

kracht die we hiervoor besproken hebben. De bredere veranderingsdynamiek rondom AI heeft verschillende **dreigingsimplicaties** voor de maatschappelijke stabiliteit in Nederland.

'Existential threat'? Allereerst is de mondiale **angst voor AI** als zodanig relevant voor de maatschappelijke stabiliteit. In de geschiedenis is al vaker sprake geweest van AI-booms waarin zowel de positieve verwachtingen als de angst tot grote proporties werden opgeblazen (zie bijlage 2). Juist deze wijdvertakte angst wordt dan een zelfstandige factor van verstoring. In mei 2023 zagen we dat de onheilsboodschap van prominente tech CEO's aangaande de 'existential threat' van AI voor de mensheid enorme sneeuwbaleffecten veroorzaakte.[100][101] De angst voor AI beheerste de media wereldwijd. Daarmee werd AI direct ook een dankbaar onderwerp om politiek te maken. Sceptici hebben bovendien gewezen op het financiële profijt dat zowel positieve als negatieve aandacht voor AI oplevert: omdat er enorme investeringen nodig zijn, hebben industrieën rondom AI er belang bij om urgentie te creëren.

Superintelligente machines? Ook al sorteert angst voor killer robots averechtse angsteffecten, toch is het scenario dat AI ontspoort niet geheel ondenkbaar. Het gezaghebbende werk Superintelligence: Paths, Danger, Strategies (2014) van Nick Bostrom agendeerde de risico's van het verlies van controle over AI ruimschoots voor de komst van het Transformer-algoritme. Met alle recente ontwikkelingen heeft de boodschap dat AI kan ontsporen bovendien nog meer weerklank gekregen onder computerwetenschappers en filosofen.[102] Het science fictionperspectief heeft definitief plaatsgemaakt voor authentieke zorgen over de controle over AI. Er wordt veel gespeculeerd over wanneer superintelligente AI zich zou kunnen manifesteren. In een betrouwbare survey in 2021 onder 738 AI-deskundigen kwam naar voren dat men het 50 procent waarschijnlijk acht dat superintelligente machines rond het jaar 2059 zullen bestaan.[103] Anderen zien dit al binnen vijf tot twintig jaar gebeuren.[104] Maar het blijven speculaties.

Het controleprobleem. Het idee van superintelligentie spreekt tot de verbeelding omdat deze machines in staat kunnen zijn tot emotioneel begrijpelijke communicatie, zelfstandige afwegingsprocedures en doelbepalingen. Daarmee nemen de risico's toe dat dit soort machines niet langer meer onder menselijke regie te brengen is. We spreken ook wel over de kans op '**rogue AI**' in verschillende situaties:

1. Mensen kunnen zelf doelgericht AI-systemen doen ontsporen of **kwaadaardige programma's** ontwikkelen om daarmee slecht beheersbare, destructieve effecten te bereiken (zie ook het punt 'AI-convergentie hieronder).
2. Er zijn scenario's denkbaar dat een AI-systeem voor het oplossen van een onmogelijke logische opdracht alle hulpbronnen in dienst stelt van dit doel. De mens wordt dan **collateral damage** zonder dat het systeem kwaadaardige intenties hoeft te bezitten.
3. Hoewel nog zeer hypothetisch, zijn er scenario's denkbaar waarin AI-systemen een eigen 'wil' ontwikkelen (of een vergelijkbare vorm van **'bewuste' intentionaliteit**). AI-systemen moeten hiervoor wel eerst het stadium bereiken van (quasi-)autonoom intentioneel gedrag.

Om het controleprobleem op te lossen zijn in essentie maar twee richtingen beschikbaar. 1) Men zoekt naar mogelijkheden om het **handelingsperspectief van rogue AI-systemen te**

100 Kevin Roose, A.I. Poses 'Risk of Extinction,' Industry Leaders Warn, 2023.

101 Bill Joy, Why the Future Doesn't Need Us, 2000.

102 Manuel Alfonseca, Superintelligence Cannot be Contained: Lessons from Computability Theory, Journal of Artificial Intelligence Research, 2021.

103 2022 Expert Survey on Progress in AI, 2022 Expert Survey on Progress in AI – AI Impacts.

104 When will singularity happen? 1700 expert opinions of AGI [2024] (aimultiple.com)

beperken, bijvoorbeeld door de stroomvoorziening of internettoegang af te snijden. Deze oplossing zal echter nauwelijks sluitend zijn op het moment dat AI-systemen intelligent genoeg zijn om de beperkingen die we opleggen te omzeilen en daarop te anticiperen. 2) De andere mogelijkheid is om ervoor te zorgen dat we controle houden over de handelingsmogelijkheden door **greep op het design van AI-machines**. Maar deze oplossing zal moeilijk haalbaar blijken omdat de complexiteit van AI-machines de menselijke intelligentie ontgroeit. Dit komt omdat AI-systemen vanwege hun zelflerende capaciteiten in toenemende mate **black boxes** zijn. Nu al is het voor de architecten van AI-machines, vanwege de omvang en complexiteit, niet altijd meer navolgbaar hoe neurale netwerken zich opbouwen. Dat gold zelfs al voor oudere niet-AI-gebaseerde algoritmen.[105] Met het huidige tempo van ontwikkelingen is het in de toekomst zo goed als onmogelijk om te achterhalen waarom een machine precies tot een bepaalde handeling is gekomen. Voor veel (gecombineerde) algoritmen is de complexiteit van statistische paden al dermate hoog dat het bijvoorbeeld onmogelijk is om te achterhalen waarom iemand op basis van gemonitorde data een bepaalde online advertentie op het scherm krijgt.

De noodkreet van AI-prominenten in mei 2023 is niet geheel uit de lucht gegrepen. Echter, een probleem is dat de benodigde realiteitszin niet opgewassen is tegen de getriggerde **apocalyptische angsten** voor 'killer robots'. Juist het realistisch urgentiebesef hebben we hard nodig voor het kanaliseren van acute problemen waarvoor AI-toepassingen ons nu al stellen. Wél is het correct dat het controle houden over AI uiteindelijk zal afhangen van een optelsom van vele soorten inspanningen op vele verschillende terreinen. Samen moeten deze inspanningen ertoe leiden dat de ontwikkeling van AI zich in een voor de mens positieve richting ontwikkelt. Dit kan gebeuren door grenzen te stellen aan het toelaten van mogelijk kwaadaardige toepassingen en door innovatie van AI te koppelen aan menselijke waarden. Op het existentiële niveau van menselijke verhouding tot AI zal 'de mensheid' zich waarschijnlijk gedwongen zien tot het aannemen van een **andere controlementaliteit** dan nu het geval is. Daarmee wordt bedoeld dat AI-technologie zich moeilijk zal laten beheersen aan de hand van 'hard controls' en top-downregie. De pluriforme en brede ontwikkelingsvormen van AI-systemen, samen met alle belangen die de AI-innovatie tot grote snelheid opdrijven, maken command-and-control-structuren ineffectief. **Aanpassingsstrategieën**, het inbouwen van **'vangrails'** en het geleiden van de ontwikkeling in de richting van conformiteit aan menselijke waarden zal waarschijnlijk een veel betere strategie blijken over technologie die juist de neiging heeft om aan onze greep te ontsnappen. (Meer over deze strategie bespreken we in hoofdstuk 5).

Bedreigingen door kwaadaardige toepassing van AI. Wie de focus uitsluitend legt op mogelijke rogue AI, kan de aandacht verliezen voor het kwaadaardige aandeel dat mensen zelf kunnen hebben in het ontwikkelen van **AI-gebaseerde massavernietigingswapens** of toepassingen met een destabiliserende werking voor samenlevingen. Zeker in de multipolaire wereld waarin we nu leven, en waarin sprake is van forse grootmachtcompetitie, neemt de AI-wapenwedloop een vlucht en groeien de kansen op toepassingsvormen die net als de komst van atoomwapens game changers zouden kunnen zijn in de geopolitieke balans en nieuwe vormen van mondiale instabiliteit kunnen veroorzaken.

[105] Adrienne LaFrance, Not Even the People Who Write Algorithms Really Know How They Work, 2015.

AI-convergentie[106] is het verschijnsel dat AI-technologie versmelt met wetenschappelijke ontwikkelingsprocessen waardoor sneller nieuwe ontdekkingen gedaan kunnen worden dan dankzij het door mensen uitgevoerde experimentele proces. Met behulp van AI worden door wetenschappers baanbrekende resultaten bereikt die echter ook zeer verontrustend kunnen zijn in de verkeerde handen.

1. Uit een wetenschappelijke risicoscan voor de kwaadaardige toepassing van AI in **bio-engineering** is naar voren gekomen dat het *in silico* (computergemodelleerd) design van virusgevaarlijk biologisch materiaal zeer snel binnen handbereik kan komen (binnen 5 jaar).[107] Op middellange termijn (5 tot 10 jaar) geldt dat ook voor de daadwerkelijke synthetische creatie van gevaarlijke pathogenen. Een ander geïdentificeerd reëel risico is het hacken van biologische databanken door middel van AI-enabled cyberaanvallen waardoor gevaarlijke informatie in verkeerde handen komt.
2. Gelijksoortige risico's worden geïdentificeerd op het gebied van **chemische engineering**. Zo werd in 2022 aangetoond in een paper in Nature dat dezelfde AI-ondersteunde principes voor de ontwikkeling van nieuwe medicatie ook succesvol gebruikt kan worden voor de ontwikkeling van verschillende nieuwe recepten voor chemische wapens.[108]
3. Ook op het gebied van de versmelting van AI met **nucleaire technologie** bestaan de nodige zorgen. Risico's ontstaan door de convergentie tussen nucleaire precisiewapens (maar ook door de komst van conventionele hypersonische precisieraketten) en geautomatiseerde vroegsignaleringssystemen.[109] Op termijn ontstaan deze risico's doordat de handmatige bediening van oude kernkoppen wordt vervangen door nucleaire precisiewapens verbonden met (semi-)geautomatiseerde responsmechanismen. We moeten daarbij niet 'rogue AI' als primair gevaar zien maar het punt dat AI en de geavanceerde combinatie van deze systemen de reactietijd verkorten, de precisie verhogen en daarmee een strategisch voordeel bieden. Vanwege de afhankelijkheid van deze systemen groeit de kans op fatale misinterpretaties. Bovendien dreigt de nucleaire balans verstoord te raken omdat meer landen de beschikking willen krijgen over nieuwe generaties wapensystemen.

Een zijdelings punt van AI-convergentie is de integratie van AI in het menselijk lichaam. Hoewel nu nog verre toekomstmuziek, zal de afhankelijkheid van AI de vraag urgenter maken in hoeverre mensen zichzelf willen 'aanpassen'. Discussies over **bio-enhancement** lijken daarom onvermijdelijk maar hebben de potentie om ophef en sociale onrust te veroorzaken. Er valt bovendien een maatschappelijke tweedeling te voorzien tussen voor- en tegenstanders van kunstmatige veredeling van het lichaam.

Geopolitiek krachtenveld. Op geopolitiek niveau zien we dat we van een multilaterale naar een multipolaire wereldorde bewegen.[110] De ontstane grootmachtcompetitie tussen de VS en China, en Europa (dat zich meer in de schaduw bevindt van deze strijd maar wel probeert mee te dingen) zorgt ervoor dat AI-ontwikkeling de kenmerken begint te krijgen van een wapenwedloop waarbij verschillende machtsblokken proberen een innovatievoorsprong te bemachtigen en elkaar waar mogelijk de pas af te snijden. Voeg daarbij het punt dat AI-convergentie leidt tot allerlei interessante nieuwe wapensystemen en men ziet de riskante

106 Emilia Javorsky en Hamza Chaudry, Convergence: Artificial intelligence and the new and old weapons of mass destruction, Bulletin of the Atomic Scientists, 2023.

107 John O'Brien en Cassidy Nelson, Assessing the Risks Posed by the Convergence of Artificial Intelligence and Biotechnology, Health Security, 2020.

108 Fabio Urbina e.a., Dual use of artificial-intelligence-powered drug discovery, Nature, Nature Machine Intelligence, 2022.

109 Edward Geist en Andrew Lohn, How Might Artificial Intelligence Affect the Risk of Nuclear War?, RAND, 2018.

110 HCSS, De Strategische Monitor 2023 | Barsten en Blokken: Confrontatie en Samenwerking in een Wereld van Wisselende Coalities, 2023.

dynamiek waar de wereld momenteel in verzeild raakt. AI-technologie staat daarbij niet alleen in de belangstelling vanwege de wil tot technische voorsprong. Nieuwe (semi-)autonome militaire systemen kunnen ook gebruikt worden om de spelregels van het internationaal recht of oorlogsrecht, die gebaseerd zijn op het gedrag en de verantwoordelijkheid van menselijke actoren, te omzeilen.[111] Geopolitieke dreigingen werken op verschillende manieren destabiliserend op nationaal niveau:

1. In politieke en bestuurlijke zin, in zoverre dat ook de Nederlandse politiek moet worstelen met hoe we ons tot deze geopolitieke dynamiek willen verhouden en welke nationale veiligheidsstrategie we op deze punten willen aannemen. In de discussie over mogelijke modernisering en uitbreiding van Europese bewapening zal Nederland voor belangrijke keuzen komen te staan. Ook ligt er voor de overheid een dwingender rol om te werken aan anticiperende maatregelen, omdat het – gelet op de inschattingen van wetenschappers – uiteindelijk slechts een kwestie van tijd is eer we te maken krijgen met DIY-(vernietigings-) wapens ontwikkeld met behulp van AI.
2. In economische zin kan grootmachtcompetitie leiden tot verstoorde handelsrelaties en afgeknepen handelsroutes. Bijvoorbeeld, de Nederlandse chipsindustrie is groot maar kent vele afhankelijkheden en kwetsbaarheden terwijl de internationale belangen zeer groot zijn. In andere HCSS studies is vastgesteld dat een verstoring van alleen al deze industrie kan leiden tot allerlei acute tekorten met gevolgen van stagnatie voor een breed scala aan sectoren en industrieën.[112]
3. Nederland kan last krijgen van buitenlandse inmenging in binnenlandse aangelegenheden zoals de beïnvloeding van verkiezingsprocessen door de productie van deepfakes, of het inzetten van GPT voor spionageactiviteiten. OpenAI rapporteerde bijvoorbeeld in samenwerking met Microsoft Threat Intelligence dat de Russische intelligence cel 'Forest Blizzard (STRONTIUM)' accounts opende om met behulp van GPT satellietcommunicatieprotocollen te inventariseren.[113]

Ongewenste macht van techcorporates. Grote techbedrijven, zoals Google, Facebook en Microsoft, spelen een cruciale rol bij de ontwikkeling van AI. Met hun enorme kapitaalkracht zijn zij in staat hun machtspositie te vergroten omdat ze (naast eventueel statelijke actoren zoals de Chinese overheid) als enige in staat zijn om de enorme investeringen te doen die benodigd zijn voor het ontwikkelen, trainen en beheren van de GPT-algoritmen op mondiale schaal. De eenmaal gevestigde machtsbasis genereert weer meer macht: doordat men het massale gebruik van AI faciliteert, verzamelt men enorme hoeveelheden gegevens, wat weer een aanzienlijke voorsprong verschaft bij verbetering en doorontwikkeling. Een extra complicerende factor voor Nederland is dat de meeste grote techbedrijven zich bovendien in de VS bevinden. Dit kan leiden tot een concentratie van macht en invloed vanuit het buitenland waar de Nederlandse samenleving last van kan ondervinden.

Door de enorme kapitaalkracht van AI techcorporates kunnen regeringen worden verleid tot het faciliteren van zakelijke agenda's van AI-techbedrijven. Dit kan leiden tot effecten op het niveau van de nationale kritische infrastructuur, bijvoorbeeld omdat verbruiksintensieve datacenters extra kwetsbaarheden veroorzaken op het nationale stroomnet. Publieke investeringen in AI kunnen bovendien ten koste gaan van andere projecten van publiek belang, waardoor kleinere spelers moeilijker kunnen concurreren en innoveren. Voorts is de

111 Bérenice Boutin, State responsibility in relation to military applications of artificial intelligence, 2023.

112 Zie bijvoorbeeld het HCSS onderzoek naar de gevolgen van een militair conflict rondom Taiwan: Joris Teer, Davis Ellison en Abe Ruijter, De Prijs van Conflict: Economische Gevolgen van een Militaire Crisis rondom Taiwan voor Nederland en de EU, 2024.

113 Microsoft Threat Intelligence, Staying ahead of threat actors in the age of AI, 2024.

publieke controle gelimiteerd over innovatieprocessen die het publieke belang echter wel degelijk aangaan.[114]

De invloed van techcorporatisme laat zich ook voelen op het niveau van veranderingen op het gebied van de ethische koers die maatschappelijk gevaren wordt. Zoals techindustriecriticus Evgeny Morozov overtuigend laat zien staat het denken in termen van publieke waarden onder druk van een **kapitalistisch-pragmatische 'schijnethiek'** die corporate belangen verpakt als maatschappelijke oplossingen.[115] Morozov betoogt dat onze samenlevingen uiteindelijk niet gebaat zijn bij de diepere maatschappelijke worteling van dit type ethiek. De bescherming van publieke waarden tegenover dit oplossingsdenken (solutionisme) komt meer onder druk te staan.

AI en mondiale culture wars. Negatieve ervaringen met AI kunnen protesten en anti-bewegingen op gang brengen die de rol van AI willen reduceren of die invloed willen uitoefenen op hoe AI werkt. Eerder zijn al discussies losgebarsten of AI-chatbots wel politiek neutraal genoeg zouden zijn. Een deel van de alt-right community claimt dat de ChatGPT-algoritmen getraind zijn om uitkomsten te geven ten faveure van politiek links en 'woke'. In de VS zijn pogingen gedaan om eigen rechts georiënteerde algoritmen te ontwikkelen.[116] Op deze manier wordt AI inzet van de 'culture wars' in westerse landen. AI-technologie blijkt daarbij een gevoelige technologie voor politieke verdachtmaking. Dat AI (voor de meesten onder ons a priori) een black box is, schept een eindeloze ruimte voor suggestie. Zoals eerder is gebleken in gevallen van de cancel culture, Black Lives Matter, diverse complottheorieën zoals QAnon, blijkt dat politieke bewegingen eenvoudig foothold krijgen in Europese landen. Ook AI-technologie kan een belangrijke plek krijgen in een gepolitiseerd, anti-institutioneel narratief.

Overzicht. In onderstaand overzicht vatten we de verschillende dreigingen samen.

Dreigingen van AI voor de maatschappelijke stabiliteit op existentieel en geopolitiek niveau

- De hypothese wint aan terrein dat autonome, superintelligente systemen mensen, cq. de mensheid op verschillende manieren zouden kunnen **schaden of uitroeien**.[117] De implicaties zijn o.a. **angst, 'morele paniek', complotdenken** en **sociale backlasheffecten** in reactie op de angst voor AI-systemen.
- AI kan het mogelijk maken dat kwaadwillenden middelen ontwikkelen voor **geweld en terreur**. Een veel genoemd scenario is dat AI gebruikt wordt als handleiding voor DIY-vernietigingswapens. AI-convergentie veroorzaakt nieuwe risico's op de terreinen van bio-engineering, chemical engineering en nucleaire ontwikkeling. DIY-wapens vormen een directe dreiging voor de samenleving.
- AI kan het mogelijke **geopolitieke krachtenveld** op spanning zetten waardoor maatschappijen een verhoogd risico lopen op politieke onrust, negatieve economische impact of zelfs kunnen worden meegezogen in internationaal gewapend conflict.
- AI-competitie bestendigt de **machtspositie van techcorporates** en daarmee de greep op binnenlandse politiek, economisch beleid of zelfs de ethische oriëntatie ten aanzien van het publieke belang.
- AI zal in verschillende delen van de wereld op allerlei manieren worden **gepolitiseerd**. Zo is AI-technologie nu al een deel van de inzet in de Amerikaanse en deels Europese **Culture Wars**. Het is goed denkbaar dat ressentiment jegens AI een extra katalysator is voor anti-institutionele (internationale) netwerken en complotdenkers.

---

114 Niki Korteweg, 'We weten he-le-maal niets' van Neuralink, het hersenimplantaat van Elon Musk, NRC, februari 2024.

115 Technologisch pessimisme behelst vaak een bredere mens- en maatschappijkritiek dan louter de focus op negatieve uitkomsten. In algemene zin vloeit kritiek op AI samen met kritiek op de moderne rationaliteit waarin oplossingsdenken en maakbaarheidsgeloof aansturen op de exploitatie van mens en ecosysteem. Evgeny Morozov noemt techoptimisme dan ook sceptisch '*solutionism*'. De denkfout van solutionisten, volgens Morozov, is dat ze oplossingen voor problemen verwarren met een moreel goed. Daarmee worden ethische vraagstukken in dienst gesteld van uitvindingen die al op de markt zijn gepusht omdat er andere belangen speelden.

116 Nick Robins-Early, 'Very wonderful, very toxic': how AI became the culture war's new frontier, The Guardian, 2023.

117 Manuel Alfonseca, Superintelligence Cannot be Contained: Lessons from Computability Theory, Journal of Artificial Intelligence Research, 2021.

4.2. **Impact op macroniveau**

We bespreken de belangrijkste onderwerpen van potentiële maatschappelijke instabiliteit door AI op macroniveau.

Synthetische media. Synthetische media zijn media die door generatieve AI-processen zijn gecreëerd. Meestal wordt daarbij monenteel gedacht aan **deepfakes** en **voice cloning** maar het kan ook gaan om automatisch gegenereerde tekst. De toepassingsmogelijkheden van synthetische media zijn exemplarisch voor de dual use-problematiek van AI. Aan de ene kant geven nieuwe tools allerlei mogelijkheden tot het maken van creatieve user generated content. Aan de andere kant zijn de kwalijke toepassingen reden voor maatschappelijke zorgen en angst. Daarnaast brengt de introductie van synthetische media directe toepassingsgerelateerde effecten met zich mee maar ook een meer indirecte impact op de maatschappelijke stabiliteit. Directe toepassingsgerelateerde risico's zijn de volgende:

AI-generated pornografie. Hoewel moeilijk te verifiëren zou maar liefst 98 % van alle online deepfakes pornografisch van aard zijn.[118] Daarmee is de opkomst van AI-generated pornografie, waarbij beelden van een bestaand persoon worden gecombineerd met pornografisch materiaal, een niet te onderschatten probleem. Daarbij kunnen beelden gebruikt worden van derden, zoals celebrities, maar ook van intimi. Omdat de tools voor iedereen toegankelijk zijn is het verschijnsel een **kwalijk effect van de democratisering** van AI-technologie. Naarmate algoritmes toegankelijker worden, zal zelfgemaakte porno volstrekt normaliseren. In de meeste gevallen is AI-generated pornografie illegaal omdat beelden van anderen zonder toestemming verwerkt worden. Dit is in strijd met het verbod op het verwerken van persoonsgegevens.[119] Maar bij dit soort breed verspreide problemen is het de vraag wie er zich aan regels gaat houden en hoe er gehandhaafd moet worden achter de voordeur. Dit is terugkerend van aard: regels helpen, maar bij massale toegankelijkheid is de handhaving een utopie. Een verwant, groot maatschappelijk probleem van illegale AI-generated pornografie is dat van **AI-generated wraakporno**.[120] De effecten van verspreiding van wraakporno in algemene zin (ook niet-AI-generated) zijn zeer schrijnend. Vaak zijn slachtoffers minderjarig. Bovendien blijft het materiaal eenmaal verspreid zich vermenigvuldigen en voor altijd online rondzwerven. Wraakporno was in beginsel al een groot probleem, maar AI maakt het mogelijk om levensecht maar niet-authentiek materiaal te creëren. Dit zorgt ervoor dat de klassieke benaderingen van preventie en mediawijsheid grotendeels ineffectief worden (omdat bijvoorbeeld een publiek gedeelde foto of film volstaat) en het nog ingewikkelder wordt om hierop toe te zien.

Fraude en criminaliteit. Synthetische media geven allerlei mogelijkheden voor AI-enabled crime. Voice cloning wordt succesval ingezet als social engineeringtactiek (zich voordoen als bijvoorbeeld een office assistent en verzoeken tot een valse transactie). Synthetische teksten worden gebruik voor valsheid in geschrifte, enz. De democratisering van AI komt zo ten goede aan de misdaad. De impact op de maatschappelijke stabiliteit is er in zoverre dat het risico van persoonlijke gegevensdiefstal groter wordt. Daarnaast wordt de druk op de politie groter omdat digitale criminaliteit groeit naast het gewoon voortbestaan van conventionele criminaliteit. Voortdurende ondercapaciteit en een gebrekkig zaakoplossend vermogen ten aanzien van digitale criminaliteit zou het (nu hoge) vertrouwen in handhaving en criminaliteitsbestrijding kunnen ondermijnen. Soortgelijke problemen van capaciteit en zaakafhandeling

118 Home Security Heroes, 2023 State Of Deepfakes: Realities, Threats, And Impact, 2023.

119 Arnoud Engelfriet, Natuurlijk gaat AI ook weer voor porno gebruikt worden, Ius Mentis, 2018.

120 Manon van Dunnen, Exploration of the impact of Synthetic Reality & Deepfakes on Police Work, 2022.

zijn ook voorzienbaar voor de rechtspraak. Zowel de authenticiteit van bewijsmateriaal als de toename van zaken waarin synthetische media een rol spelen, veronderstellen weer extra know-how en maken zaken ingewikkelder en slepender.

Politieke beïnvloeding. AI kan een rol spelen om de democratie onder druk te zetten, door synthetische media te gebruiken voor het zaaien van **mis- en desinformatie**. Goed getimed kunnen dit soort tactieken van invloed zijn op electorale processen.[121] Grote AI-bedrijven als OpenAI zijn actief bezig om ervoor te zorgen dat hun producten niet voor dit soort doeleinden gebruikt kunnen worden.[122] Ook kunnen trends binnen de samenleving gemanipuleerd worden door het in groten getale inzetten van bots in de sociale media. AI kan gebruikt worden voor sentimentanalyse maar dus ook om dergelijke analyses te beïnvloeden. Het controleren van informatie wordt steeds ingewikkelder.

Het is de vraag of de veel in de media herhaalde stelling wordt bewaarheid dat in 2026 tot wel 90 procent van alle online content zou bestaan uit synthethische content.[123] Uit een ander rapport dat de toename van 2019 tot 2023 volgde, blijkt dat het aantal deepfakes dat publiekelijk online staat explosief groeit (met 550% over deze periode, leidend tot een totaal van 95.280).[124] De angst voor synthethische media lijkt enigszins onderhevig aan de **issue attention cycle**, waarbij het probleem eerst tot grote proporties opgeblazen wordt, waarna de onrust weer tot bedaren komt en de reële impact duidelijk wordt. Daarmee is niet gezegd dat de rol van synthetische media op de democratie moet worden uitgevlakt maar de invloed is nog niet zo verontrustend als werd gevreesd.[125] Over een aanhoudende periode kan publieke beïnvloeding en misleiding door media wel degelijk diepgaande effecten sorteren. De met AI gecreëerde memes waarin AI-varianten op de Paus, Mark Rutte of Donald Trump figureren, dienen voornamelijk als vorm van **propaganda of agit-prop**. Ook al worden synthethische media als fake herkend, deep fake-materiaal geeft een nieuwe dimensie aan campagne-voering en het breken van politieke support. Een voorbeeld is de strategie van het Lincoln Project, een republikeinse anti-Trump-groepering die creatieve spots maakt waarin bijvoor-beeld de overleden vader van Trump tot leven wordt gewekt om het politieke project van Trump te bespotten. Dergelijke video's zijn duidelijk nep maar missen hun impact niet: 'what has been seen, cannot be unseen.' Op termijn zal de stapeling van allerlei uitdrukkingen van synthetische media toenemen waardoor deepfakes als oppervlakteverschijnsel toch diepere impact krijgen.

Macro-effecten van synthetische media op de langere termijn. Tenminste twee macro-ef-fecten van synthetische media op de maatschappelijke stabiliteit zijn van belang:

1. Een massale toename van synthetische media draagt bij aan het voortschrijdend proces van **'mediatisatie'** van de samenleving. Dit betekent dat de politieke en sociale gang van zaken steeds sterker invloed staat van media-uitingen. Andersom moeten politici en promi-nenten meer vaardigheid ontwikkelen in het inspelen op de mediadynamiek, in termen van hun 'airtime' of door gebruik te maken van personal marketingtactieken. In algemene zin geldt dat mediatisatie leidt tot een grilliger maatschappijbeeld en aanverwante verschijn-selen zoals de zwevende kiezer of de toename van one issue-partijen die garen spinnen bij momentane, door media aangewakkerde ophef.[126]

[121] Bruce Schneier and Nathan E. Sanders, Six Ways That AI Could Change Politics, 2023.

[122] OpenAI, How OpenAI Is Approaching 2024 Worldwide Elections, 2024.

[123] IDCA, AI Experts Predict By 2026, 90% Of Online Content Will Be Generated By Artificial Intelligence, 2022.

[124] Home Security Heroes, 2023 State Of Deepfakes: Realities, Threats, And Impact, 2023.

[125] Marietje Schaake, Deepfake ondermijnt liberale democratie, 2021.

[126] Universiteit van Amsterdam, How do new political parties garner media attention?, 2023

2. De toename van synthetische media leidt tot algehele **inflatie van informatiebetrouwbaarheid**. Hiervan zal sprake zijn op een wat langere termijn, wanneer de verhouding van synthetische media op de door mensen gecreëerde media toeneemt en de grens tussen 'authentieke' en 'niet-authentieke' informatie het risico loopt te vervagen. Dit gaat ten koste van met name autoriteiten en overheden wier gezagspositie afhankelijk is van het publiek vertrouwen in de validiteit van de informatie die zij verstrekken. Het principe van het **leugenaarsdividend**[127] speelt hier een rol: doordat er meer valse informatie in omloop komt verzwaart de bewijslast voor partijen die claimen wél authentieke informatie te leveren. Onder aan de streep zal de samenleving zo last ondervinden van een steeds meer vloeibare informatiewerkelijkheid waarin authentiek bronnen steeds moeilijker te identificeren zijn. Dit werkt anti-institutioneel denkenden in de hand omdat zij volgens het postmoderne principe leven van het 'ik zoek mijn eigen feiten'.

Ongelijkheid. Zoals eerder belicht in hoofdstuk 2 vergroot de komst van AI de schaal, snelheid en het bereik van automatisering en leidt daarmee tot directe concurrentie met mensen en vervanging van menselijke eigenschappen. Op AI-gebaseerde algoritmen en machines zullen daarom naar verwachting veel taken van mensen overnemen en beter uitvoeren. Onderzoek uitgevoerd door McKinsey voorziet dat in 2030 15% van de mondiale arbeidskrachten vervangen zal zijn.[128] Onderstaand overzicht geeft weer wat de verwachting is van automatisering van verschillende beroepsgroepen volgens een survey van 738 AI-deskundigen in 2022. We zien dat zowel laagopgeleiden als hoogopgeleiden op termijn een grote kans hebben om te worden vervangen. Op korte termijn lopen met name laaggeschoolden het meest risico. De gemiddelde schatting van experts is dat 5% van alle arbeid binnen vijftig jaar overgenomen zal zijn door AI.

Hoewel de meningen erover verdeeld zijn[129], ziet bijvoorbeeld prominent econoom Joseph Stiglitz dat AI, ondanks kansen voor de werkgelegenheid, toch overwegend tot meer **sociale ongelijkheid** zal leiden.[130] Het ontstaan van nieuwe baantypen door de komst van AI weegt niet op tegen de vervangingsratio van mensen door toedoen van automatisering. Daar staat tegenover dat mensen en bedrijven die AI juist slim weten te benutten veel profijt kunnen hebben van AI. Zij vergaren extra rijkdom met behulp van AI en met minder loonkosten, wat de kloof tussen arm en rijk vergroot. Op langere termijn ontstaat er mogelijk een (verdere) ontkoppeling tussen de ketens van waardetoevoeging en menselijke activiteiten. Dat proces van ontkoppeling tussen geld en toegevoegde waarde is al langer aan de gang, denk bijvoorbeeld aan financiële AI trading-algoritmen. Maar omdat AI potentieel nog veel meer betaalde arbeid kan overnemen dan nu al het geval is, wordt de vraag wat wel en geen toegevoegde waarde is met behulp van AI dwingender. Ook hier geldt dat bedrijven en personen die AI goed weten te kapitaliseren in het voordeel zijn van mensen die nog denken in termen van toegevoegde waarde aan de reële economie op grond van arbeidskapitaal.

127 Manon van Dunnen, Exploration of the impact of Synthetic Reality & Deepfakes on Police Work, 2022.

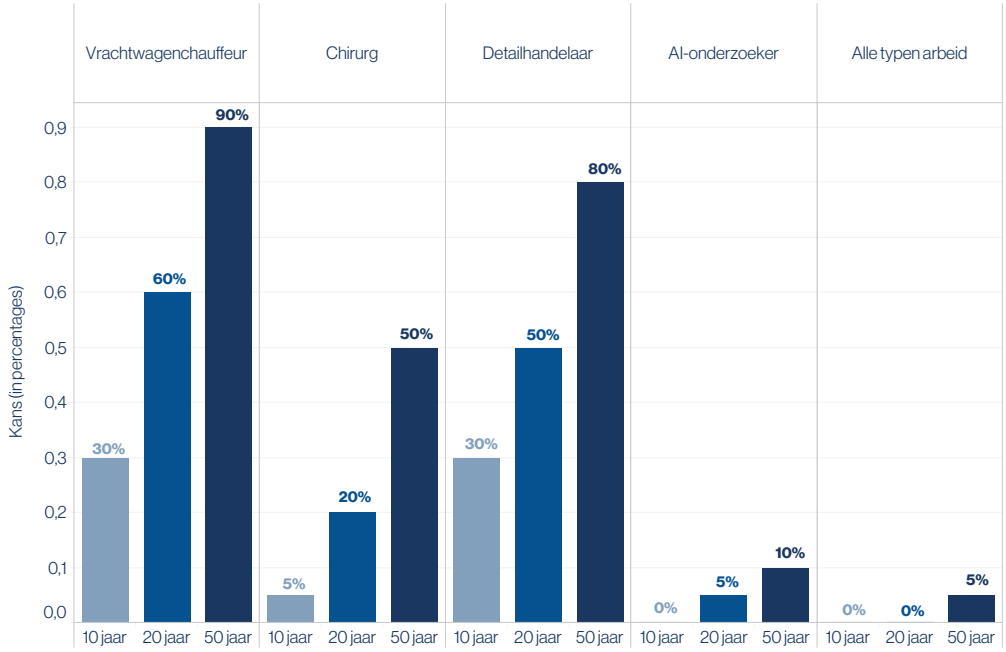
128 McKinsey, AI, automation, and the future of work: Ten things to solve for, 2018.

129 Ekkehard Ernst et al., The Economics of Artificial Intelligence: Implications for the Future of Work, 2018.

130 Sophie Bushwick, Unregulated AI Will Worsen Inequality, Warns Nobel-Winning Economist Joseph Stiglitz, 2023.

Figuur 6: Wat is de kans dat verschillende typen arbeid worden geautomatiseerd in de komende 10, 20 en 50 jaar?[131]

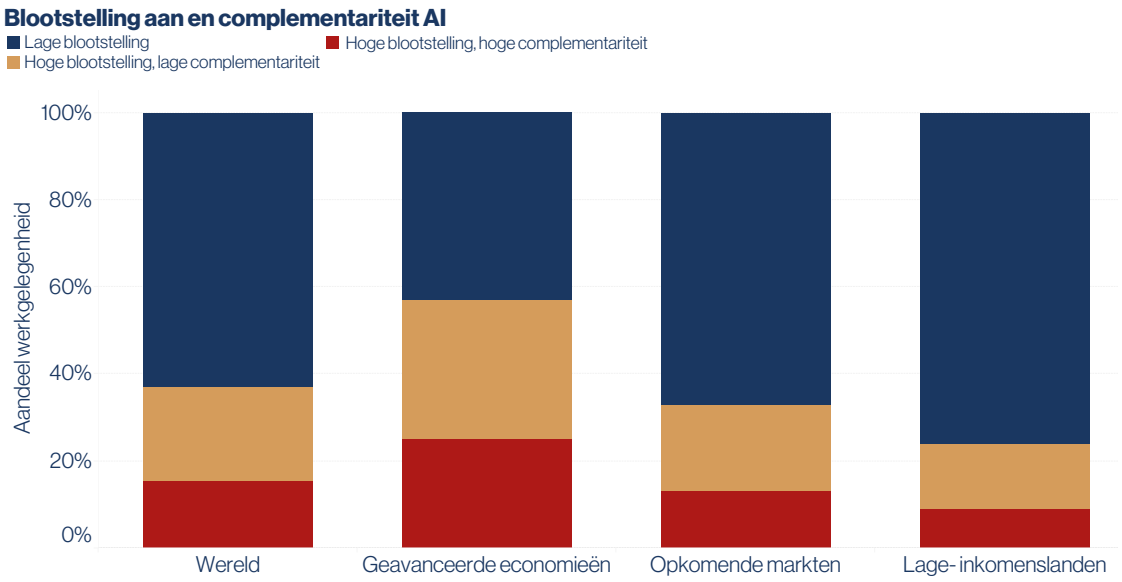
Gemiddelde antwoord van 89 machine learning experts



Bron: Universiteit van Oxford

Figuur 7: Impact AI op de werkgelegenheid: bijna 60% van de banen in geavanceerde economieën wordt door AI geraakt (bron: Internationale Arbeidsorganisatie en IMF)

De meeste banen zijn blootgesteld aan AI in geavanceerde economieën, met kleinere aandelen in opkomende markten en lage-inkomenslanden



Bron: Internationale Arbeidsorganisatie en het Internationaal Monetair Fonds
*Blootstelling in deze context refereert naar de mate waarin het gebruik van AI in het werkveld beschikbaar is voor de werknemer
**Complementariteit in deze context refereert naar de mate waarin AI het werk wel efficiënter kan maken, maar het menselijke werk niet kan vervangen

[131] Machine Intelligence Research Institute, University of Oxford, 2022 Expert Survey on Progress in AI, 2022.

Naast sociale ongelijkheid is er een verwachte toename van **generationele ongelijkheid**. Hoewel het in de lijn der verwachting ligt dat oudere generaties minder flexibel zijn om zich aan te passen aan de AI-transformatie, zijn het met name de jongste arbeidsgeneraties die de kans lopen om te worden vervangen door AI.[132] Vaak voeren zij laaggeschoolde taken uit, zoals routineuze bijbaantjes, die makkelijker vervangen kunnen worden door automatiseringsprocessen. Ook werken jongere medewerkers vaak op basis van flexcontracten en zijn daarom makkelijker te ontslaan. Mensen die al een goede baanpositie hebben verworven en lange scholingstrajecten hebben gevolgd lopen juist minder kans om te worden vervangen. In Nederland, met een toenemende vergrijzing en een onmogelijke vastgoedmarkt voor starters, komt er al relatief veel economische last op de schouders van de jongere generaties.

Tegen deze achtergrond zouden problemen op de arbeidsmarkt door AI bijdragen aan een grotere optelsom van tegenslagen. Het is goed denkbaar dat er sociale onrust kan ontstaan als het economisch slechter gaat. Mogelijk ontstaat er een **nieuwe 'lost generation'** op de arbeidsmarkt in de transitieperiode tussen de oude economische orde en de samenleving die zich met de tijd beter heeft aangepast aan de versnelling van de AI-transformatie. Deze generatie zou uit een gevoel van verbittering in het geweer kunnen komen voor betere voorwaarden en het maatschappelijk ressentiment jegens AI kunnen opzwepen met potentiële maatschappelijke conflicten tussen werknemers, werkgevers en politiek tot gevolg. Daarentegen, een positiever backlasheffect is ook te voorzien: door de toename van AI-gegenereerde mainstreamproducten, zal onder jongere generaties mogelijk meer voorkeur ontstaan voor de menselijke toegevoegde waarde. Dit kan leiden tot een **'reveil' van het (verloren) ambacht**. Tekenen daarvan zijn zichtbaar op het internet waar de waardering voor authentiek ambachtelijk werk breed wordt uitgemeten.[133]

Capaciteitsproblemen en IT-personeelstekorten. Voor sommige beroepsgroepen zal AI leiden tot een **vermeerdering van werk en taken** wanneer er een dubbele afhankelijkheid ontstaat van zowel de nieuwe wereld die AI openbaart en het traditionele werk dat er al was en ook nog gedaan moet worden. Een voorbeeld daarvan is de politiefunctie die zich moet aanpassen naar de nieuwe digitale realiteit maar ook verantwoordelijk blijft voor handhaving in de fysieke wereld. Andere beroepsgroepen die capaciteit zullen moeten bijzetten, zijn bijvoorbeeld governance autoriteiten op het gebied van naleving handhaving van regels en richtlijnen voor (responsible) AI. In brede zin is een grote maatschappelijke schaarste te voorzien van IT-personeel dat goed kan omgaan met AI-technologie. Extra schaars zal personeel zijn dat begrip van AI-engineering combineert met het goede pakket governance-competenties om AI-machines te kunnen toetsen en verantwoorden.

Het denken over werkdruk en werk-privébalans - zal veranderen door verdere automatisering. Toch is lastig te voorspellen omdat de voorkeuren om arbeid te verrichten niet één-op-één gekoppeld zijn aan de productie die maatschappelijk nodig wordt geacht. Marx voorspelde bijvoorbeeld al dat, afgaand op economische noodzaak, mensen in een communistisch arbeidsregime maar twee tot drie uren per dag zouden hoeven te werken. Dat is nooit uitgekomen omdat hij onderschatte dat de motieven om te werken veelzijdiger zijn. Mensen zullen misschien arbeid op een meer creatieve manier willen definiëren waar AI de meest saaie arbeid overneemt. Maar of het aantal gewerkte uren per week drastisch zal kelderen is nog absoluut onzeker. Mocht de besteedbare vrije tijd per capita wel fors toenemen, dan heeft

[132] Center for a New American Security, The indirect effects of the artificial intelligence revolution for global security, 2018.

[133] The Guardian, The US artisan revolution: how the simple life came in from the margins, 2022.

dat waarschijnlijk de meeste gevolgen voor de leefbaarheid van de omgeving en de **druk op de recreatieve ruimte** in Nederland.

Vervreemding en menselijke exploitatie. AI-analytics bieden enorme mogelijkheden op het monitoren en optimaliseren van menselijke productiviteit maar hieraan kleven ook venijnige kanten. Weliswaar zijn er positieve effecten op de efficiency en het afstemmen van processen op persoonlijke patronen en gedragingen. De andere kant van de medaille is dat er risico's van arbeidsuitbuiting en vervreemding op de loer liggen. Opmerkelijk genoeg leidt AI op bepaalde gebieden juist tot nieuwe vormen van laagbetaalde arbeidsuitbuiting. In lagelonenlanden zien we de opkomst van bedrijven die mensenhanden inzetten om algoritmen te trainen of te corrigeren. Omdat dit soort arbeid moet verhullen dat AI de taak zelf niet (goed) kan, wordt dit verschijnsel van afhankelijkheid van goedkope mensenarbeid ook wel '**fauxtomation**' genoemd.[134]

Psychosociale impact. Er bestaan zorgen over de vraag of mensen om kunnen gaan met toenemende onderdompeling in virtuele werelden. AI kan virtuele interfaces en digitale interactie aantrekkelijker en responsiever maken dan de echte wereld. Zeker wanneer AI-machines hun manier van interactie een menselijk tintje weten te geven, zijn mensen bereid tot het aangaan van vriendschappelijke of zelfs affectieve relaties met machines.[135] Wat zijn de gevolgen voor burgerschap, identiteit en reële sociale rollen? De intensieve verbondenheid met responsieve, virtuele AI-machines zorgt voor **psychosociale impact** op verschillende fronten:[136]

1. Het vertoeven in responsieve virtuele omgevingen kan **verslavend** werken; Overmatig gebruik van digitale technologie, waaronder AI-systemen, kan leiden tot verslavend gedrag. Mensen kunnen zich gedwongen voelen om constant hun apparaten te controleren of AI-aangedreven apps te gebruiken, die andere aspecten van hun leven kunnen verstoren, zoals werk of sociale relaties.
2. AI-interactie kan gevoelens van **vervreemding, paranoia of depressie** oproepen. Gevoelens van overbodigheid kunnen een probleem vormen voor het mentaal welzijn. Tegen de achtergrond van de (wel of niet reëel zijnde) depressie-epidemie groeit het belang om maatschappelijk zicht te houden op effecten die het menselijk contact met AI-systemen heeft. De opgepoetste virtuele identiteiten die mensen proberen te onderhouden zorgen mogelijk voor stress en een vertekend zelfbeeld. De vervreemding kan ook existentieel zijn: de leefwereld wordt in toenemende mate bepaald door technologische systemen die mensen niet meer begrijpen. Dit heeft effecten op het gevoel van geborgenheid en uniciteit.
3. AI kan het realisme en de functionaliteit van virtuele interactie versterken en maakt virtuele activiteiten toegankelijker. AI speelt daarmee een rol bij de vervagende grenzen tussen de echte wereld en de virtuele wereld, wat toenemende problemen kan geven van **politieke en maatschappelijke dissociatie**. Op maatschappelijk niveau kan dit consequenties hebben voor afbrokkelend informeel leefmilieu. Dit heeft implicaties voor **de sociale cohesie** in leefgemeenschappen waar informele netwerken afhankelijk zijn van onderling persoonlijk contact. Het leiden van een toenemende virtuele stijl van leven heeft ook consequenties voor het **engagement**. Mensen dissociëren zich van hun lokale politieke omgeving. Zo lang democratische besluitvormingsprocessen zelf niet zijn gedigitaliseerd, blijft het probleem van schijnparticipatie bestaan.

---

134 Gavin Jackson, Why the rise of the robots hasn't happened just yet, Financial Times, 2019. De term is oorspronkelijk van de Canadese activiste Astra Taylor.

135 Xinge Li en Yongyun Song, Anthropomorphism brings us closer: The mediating role of psychological distance in User–AI assistant interactions, 2021.

136 Ignas Kalpokas, Problematising reality: the promises and perils of synthetic media | SN Social Sciences, 2020.

4. **Afbrokkeling van de menselijke veerkracht en vindingrijkheid** wordt door sommigen gezien als een langetermijneffect van het massaal delegeren van denktaken aan AI-systemen. De stimulus voor menselijke vindingrijkheid wordt mogelijk weggenomen omdat AI de bulk van het denkwerk verricht en oplossingen aandraagt. Dit drijft afhankelijkheid en vermindert de zelfredzaamheid.

Echokamers en affectieve polarisatie

Een andere belangrijke groep van effecten heeft te maken met de verdere virtualisatie van ontmoetings- en discussieruimtes waardoor verdere verharding en polarisatie zou kunnen ontstaan. Het idee van filterbubbels suggereert dat mensen door algoritmen vooral worden blootgesteld aan meningen die consistent zijn met hun bestaande vooroordelen. Op die manier zouden deze bubbels leiden tot een versterkt commitment aan radicale en anti-institutionele sentimenten. De relatie tussen filterbubbels en politieke blikverkokering is in de studie naar het sociaal contract reeds belicht.[137] Hoewel het idee van filterbubbels suggereert dat de nieuwsoriëntatie verschraalt, wijst onderzoek juist uit dat het online informatiedieet diverser is geworden. De vooronderstelling dat algoritmen direct leiden tot blikverkokering klopt niet. Daar staat tegenover dat er wel **echokamereffecten** zijn: mensen die zich lange tijd omgeven met gelijkgestemden in de digitale omgeving hebben de neiging te verharden en zich krachtiger uit te laten tegenover de hun onwelgevallige representanten van het politieke spectrum. In die zin dragen echokamers enigszins bij aan **affectieve polarisatie** door het zogenaamde **'stadioneffect'**: de gelijktijdige combinatie van confrontatie met andersgestemden en verharding onder gelijkgestemden leidt tot versterkt politiek ressentiment en verminderde acceptatie van andersdenkenden.

Hoewel analyse van onderzoek laat zien dat ook echokamers minder voorkomen dan gedacht[138], moet rekening worden gehouden met een progressieve ontwikkeling door toedoen van AI om vier redenen: 1) AI leidt in algemene zin tot meer **synthetische informatie**, wat bijdraagt aan vertekende beeldvorming in echokamers; 2) AI leidt tot meer mogelijkheden tot **immersive experiences** waardoor mensen meer tijd doorbrengen in virtuele realiteiten met gelijkgestemden en het gevoel voor onderscheid tussen de virtuele en echte wereld vervaagt; 3) op termijn zal de opkomst van sociale relaties met AI-bots risico's kunnen opleveren omdat deze kunnen hallucineren of genegen zijn **vooroordelen te bevestigen** van de menselijke gesprekspartner; 4) Hoewel de correlatie tussen angst voor AI en complotdenken een dark number blijft, toont onderzoek wel aan dat **dreigingspercepties** ('threat perceptions') versterkt aanzetten tot het zoeken naar eigen verklaringen en de selectieve keuze van 'eigen waarheden' online.[139]

De 'AI-toezichtssamenleving'. Het probleem van AI als black box is niet alleen relevant met het oog op de voorkoming van rogue AI. Er ligt ook een vraagstuk dat te maken heeft met macht en gewenning aan systemen waarvan wel of niet weten of ze ons in de gaten houden. **Softening** is het verschijnsel dat mensen op termijn gewend aan toezichtpraktijken, waardoor de weerstand afneemt.[140] De aanwezigheid van identificatiesystemen normaliseert bijvoorbeeld terwijl de ooit enorme weerstand daartegen van decennia geleden vrijwel volledig is verdampt. AI is een systeemtechnologie die op zichzelf genomen niet verantwoordelijk is voor allerlei soorten (heimelijke) datavergaring. Maar voor AI-functionaliteiten moeten systemen wel met data gevoed worden, wat **dataficatie** in de hand werkt. Bekende maatschappelijke

137 HCSS, Het Sociaal Contract: Verwachtingen en spanningen in de democratische rechtsorde, 2024.

138 Reuters instituut for the Study of Journalism, Oxford University, Echo chambers, filter bubbles, and polarisation: a literature review, 2022.

139 Raffaell Heiss, e.a., How threat perceptions relate to learning and conspiracy beliefs about COVID-19: Evidence from a panel study, 2021.

140 Marc Schuilenburg, Making surveillance public, Inaugurele rede Erasmus Universiteit Rotterdam, 2023.

gevolgen van toezichtsystemen zijn **function creep** en **chilling effecten**. Het eerste doet zich voor wanneer technologieën oorspronkelijk ontworpen voor specifieke doeleinden stilzwijgend worden uitgebreid naar bredere toepassingen dan waarvoor ze ooit bedoeld zijn. Dit komt bijvoorbeeld voor wanneer een database van een gezichtsherkenningssysteem dat bedoeld was voor handhavingsdoeleinden, ook wordt ingezet voor opsporing van fraude. Chilling effecten zijn de impliciete gedragsaanpassingen die mensen laten zien uit angst dat hun gedrag wordt gevolgd. Deze effecten worden dan ook geïdentificeerd met zelfcensuur en passieve vrijheidsinperking. China figureert vaak als voorbeeld van een digitale politiestaat waarin deze effecten zich hebben ontwikkeld tot volwaardige middelen van statelijke oppressie. Zo is er het Skynet project, een bewakingsnetwerk in de publieke ruimte met naar schatting twintig miljoen camera's.[141] En er is het beruchte sociale kredietscoresysteem waarin puntenaftrek door slecht gedrag leidt tot ontzegging van publieke diensten. Waar, cynisch gesteld, geautomatiseerd toezicht in een autoritaire staat juist een methode is om orde en sociale stabiliteit te bereiken, is dit in vrije democratieën juist een aanleiding voor wantrouwen jegens de overheid.

Overzicht. In onderstaand overzicht vatten we de verschillende dreigingen op macroniveau samen.

Dreigingen van AI voor de maatschappelijke stabiliteit op macroniveau

- De toename van **synthetische informatie** zorgt voor concrete dreigingen op het gebied van wraakporno, fraude en politieke beïnvloeding. In abstracto ondermijnt synthetische informatie het gezag en de waarde van feitelijke informatie.
- AI zal door bedrijven en organisaties massaal worden ingezet als een instrument ter verhoging van productiviteit en efficiency, leidend tot potentieel grote veranderingen op de **arbeidsmarkt** en een **toename van economische en generationele ongelijkheid**.
- (Virtuele) interactie met AI heeft potentieel verschillende negatieve **psychosociale effecten**. Zo werkt AI-interactie mogelijk verslavend, veroorzaakt ze vervreemding, paranoia of depressie. Er kleven bovendien psychologische risico's aan sociale en affectieve relaties met AI-chatbots. De toenemende afwezigheid in het fysieke maatschappelijke verkeer veroorzaakt politieke en maatschappelijke dissociatie wat op haar beurt weer aanleiding geeft voor afbrokkeling van de menselijke veerkracht.
- Algoritmen en synthtetische media spelen een rol in het voortbestaan van **echokamers**. Er is een relatie met affectieve polarisatie maar deze moet ook genuanceerd worden.
- **De AI-toezichtssamenleving**: Wanneer AI-systemen door burgers geïdentificeerd worden als institutioneel machtsmiddel voor toezicht en gedragssturing, leidt dit tot vertrouwenscrises tussen overheid, rechtspraak, bedrijven en burgers.[142]

4.3. Impact op het niveau van toepassingsgebieden

Nu de mogelijke impact op de maatschappelijke stabiliteit op existentieel en macroniveau in beeld zijn gebracht, richten we ons op het niveau van de toepassingsgebieden. We hanteren daarbij het onderscheid dat we introduceerden in hoofdstuk 2.

Wegen van impact. We benadrukten dat AI-toepassingen vrijwel altijd een **dual use** kennen, waardoor de toepassingen zowel goedaardige als kwaadaardige effecten kunnen hebben. Of een AI-toepassing goed of slecht gebruikt wordt hangt uiteindelijk af van veel meer dan

141 "Skynet", China's massive video surveillance network | South China Morning Post (scmp.com)

142 Nicholas A. Christakis, We need to focus more on the social effects of AI, says Nicholas Christakis, 2023

wat de techniek belooft alleen. De goede of kwade richting waarin gebruik zich ontwikkelt, is afhankelijk van o.a. 1) de **transformatieve kracht** van het AI-product in de maatschappelijke **context**, 2) de **ethische begrenzing** dan wel **regulerende wet- en regelgeving** die mogelijk vangrails geven voor negatieve effecten en 3) de mate waarin **de samenleving blootgesteld** is aan goede of slechte impact (zie bijlage 4 voor een toelichting op de wegingscategoriën). In samenvatting van deze analyse terug te vinden in bijlage 4 valt een aantal terugkerende, generieke implicaties te benoemen. Vaak zien we dat effecten op toepassingsniveau bij elkaar opgeteld weer leiden tot implicaties op maatschappelijk macroniveau. Gevolgen komen echter duidelijk voort uit de wijze waarop het toepassingsdesign is ontworpen en welke bestuurlijke keuzes (niet) worden gemaakt om voor bepaalde toepassingen te kiezen:

Algoritmische bias maakt toepassingen vatbaar voor discriminatie. Algoritmische bias treedt op wanneer gevolgtrekkingen en output systematisch minder gunstig zijn voor individuen uit een bepaalde groep, waarbij er geen relevant verschil is tussen de groepen dat deze minder gunstige output rechtvaardigt. De onderdrukkende of discriminerende effecten zijn een gevolg van een schadelijke aangetrainde herkenningspatronen. Vaak ontnemen bevooroordeelde aanbevelingen mensen het recht door bepaalde kansen die aan anderen worden gegeven, te ontzeggen. AI-toepassingen in vele verschillende vormen bezitten risico's van algoritmische bias. Algoritmen die worden ingezet in het kader van publieke taken of openbaar bestuur liggen onder een extra groot vergrootglas omdat bij uitstek de staat zich dient te houden aan het neutraliteitsbeginsel. Denk aan spraakherkenningssoftware die niet-Nederlandse accenten niet kan herkennen,[143] algoritmen die discrimineren omdat ze voornamelijk getraind zijn op eenzijdige leeftijdsgroepen[144] of gezichtsherkenningssoftware die etnisch profileert. AI kan (onbedoeld) bijdragen aan marginalisering of stigmatisering van kwetsbare groepen.

Ethische regie is in de praktijk lastig. Regulering en ethische vangrails worden veelal wettelijk bepaald, maar de werkelijke ethische regie hangt af van ingewikkelde, moeilijk te handhaven toetsingsprocessen. Dit komt omdat AI-toepassingen al snel black box-eigenschappen vertonen, omdat de onderliggende statistische paden snel complex worden en onmogelijk te ontrafelen blijkt waarom een algoritme bijvoorbeeld met een gepersonaliseerde aanbeveling komt. Dit voert ook terug naar het controlevraagstuk: als vele AI-toepassingsvormen (contextuele) ethische toetsing nodig hebben, hoe organiseren we al deze ingewikkelde toetsingsprocedures dan grondig genoeg? (Over dit punt meer in hoofdstuk 5.)

Criminele toepassingen van AI kunnen een serieuze uitdaging vormen voor opsporing en criminaliteitsbestrijding. AI zal een verdere 'democratisering' van criminaliteit faciliteren.[145] Mensen met elementaire digitale kennis kan de mogelijkheid hebben om AI aan het werk te zetten voor illegale activiteiten. En dan zijn er nog de implicaties die ontstaan binnen de context van de virtuele ruimte. Zo zorgt diefstal van iemands virtuele identiteit of denkbeeldige eigendommen voor juridische hoofdbrekens. Wat zijn zaken waard die geen fysieke bestaansdimensie hebben?

Er is een breed spectrum van criminele activiteiten dat sterk kan profiteren van slimme automatisering. Hierbij kunnen we denken aan relatief eenvoudige phishing e-mails geschreven met behulp van ChatGPT tot aan complexe geautomatiseerde cyberaanvallen op kritieke

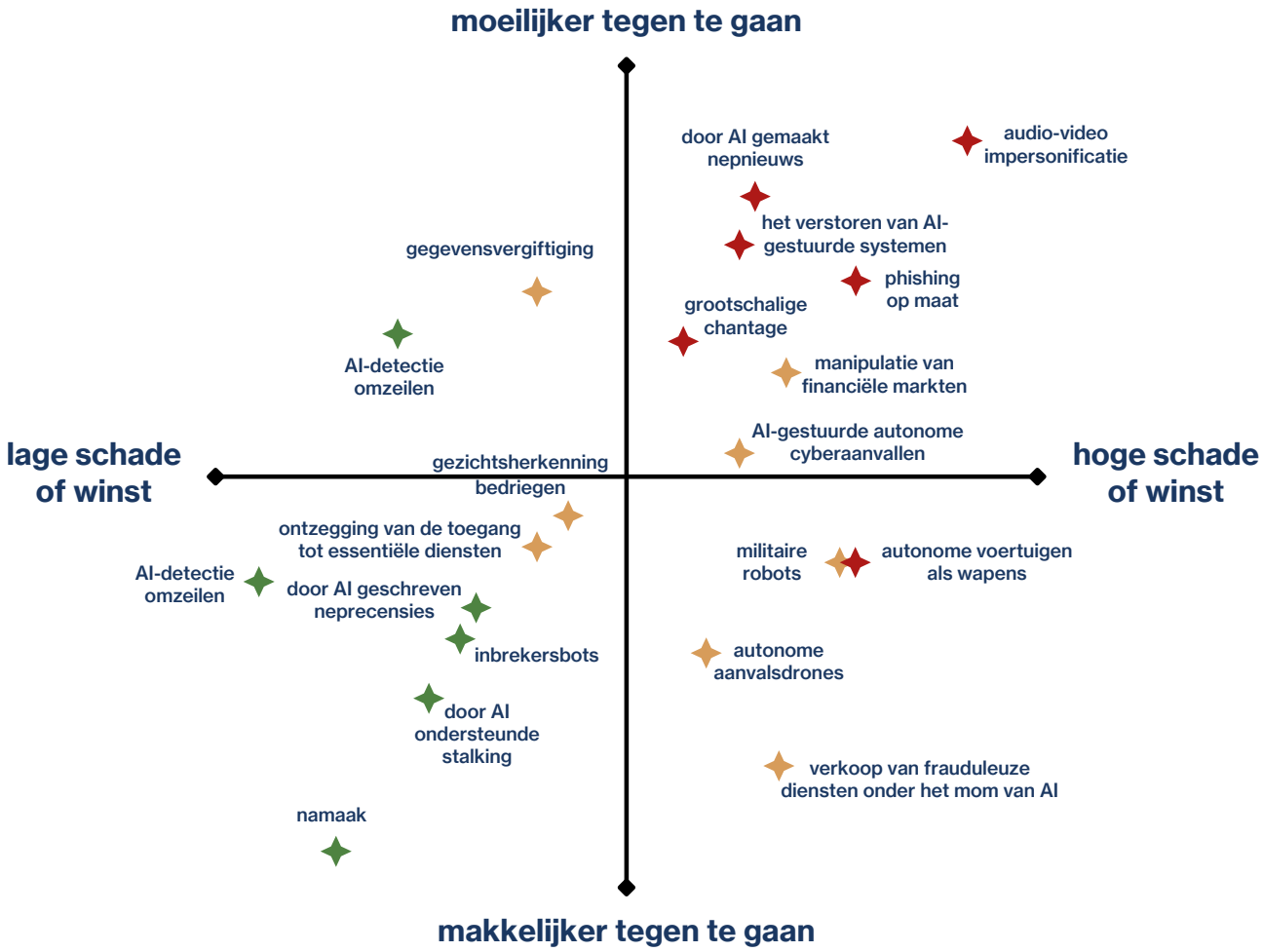
143 Claudia Lopez Lloreda, Speech Recognition Tech Is Yet Another Example of Bias, 2020.

144 Fay Cobb Payton, Without Small Data, AI in Health Care Contributes to Disparities, 2023.

145 Deze ontwikkeling is eerder in het kader van het Strategische Monitor Politie-programma behandeld in de context van transnationale georganiseerde misdaad; zie Julien Bastup-Brik et al., Next Generation Organised Crime: Systemic change and the evolving character of modern transnational organised crime, 2023.

infrastructuur. Verschillende denkbare toepassingsvormen zijn geplot in Figuur 8, gebaseerd op onderzoek onder criminologen wat de meest bedreigende vormen zijn op grond van de mate van succes en de mogelijkheden tot bestrijding ervan. Daarbij moet bedacht worden dat dit overzicht is gebaseerd op een brainstorm. De resultaten zijn speculatief en weerspiegelen de kennis, ervaring en prioriteiten van de deelnemers. Niettemin bieden de uitkomsten, zoals de bron zelf ook aangeeft, een bruikbare momentopname van de heersende zorgen en hoe deze zich naar verwachting de komende jaren zullen uiten.

Figuur 8: Vormen van AI-enabled criminaliteit naar ingeschatte schade en bestrijdingsmogelijkheden[146]



[146] M. Caldwell et al., AI-enabled future crime, 2020.

De toepassingen van AI in criminaliteit kunnen zorgen voor een professionaliseringsslag op bijvoorbeeld het terrein van vermogensdelicten, en in die zin bijdragen aan een vermeerdering van High Impact Crime. De juridische schemergebieden omtrent AI maken exploitatie van de regels mogelijk en verkleint de pakkans nog eens extra. Waar sprake is van zogenaamde High Impact Crime kan sprake zijn van ernstige ervaringen van slachtofferschap. Hoe minder greep er is op bijvoorbeeld geavanceerd gepleegde vermogensdelicten, hoe minder het maatschappelijk vertrouwen in de regie van de overheid op de problematiek. Maatschappelijk vertrouwen wordt in die zin geraakt door terugvallende regie over criminaliteit. Een ander secundair effect op het maatschappelijk vertrouwen is dat binnenlandse veiligheidsinstituties voor het probleem komen te staan van een verminderd overzicht over waar dreigingen vandaan kunnen komen. Dit vraagt om bestuurlijke prudentie met een alertheid voor het krachtenveld en de risico's buiten de Nederlandse grenzen. Het is een uitdaging om dit soort strategisch en prudent bestuur waar te maken omdat er een grote mate van onzekerheid is die zich moeilijk laat agenderen.

Interactie met AI-bots geeft reële psychologische risico's. Hoewel we nog maar aan het begin staan van de periode waarin **sociale en affectieve relaties met AI-chatbots** (zoals Chai en Replika) en -assistenten zullen normaliseren, zijn er verschillende problemen die zich al beginnen af te tekenen qua interactie op toepassingsniveau; Het **ELIZA-effect**[147] beschrijft wanneer een persoon ten onrechte emotionele intelligentie toeschrijft aan een AI-systeem, inclusief emoties en een gevoel van eigenwaarde. Het effect is vernoemd naar het ELIZA-computerprogramma van MIT-wetenschapper Joseph Weizenbaum, waarmee mensen in 1966 al uitgebreide gesprekken konden voeren. Mensen blijken alarmerend snel geneigd om empathische en affectieve gevoelens in de logisch geconstrueerde taalrespons van machines te projecteren. Hierdoor ontstaan diverse problemen wanneer mensen de output van systemen serieus nemen. Bots kunnen gevaarlijke onzin hallucineren, aanzetten tot misdrijven of zelfs tot zelfmoord.[148]

Er blijken vaak grijze gebieden in de wet- en regelgeving die maken dat het in de praktijk moeilijk (bindend) vast te stellen is of een toepassing voor de wet risicovol is en verboden moet worden. Dit punt wordt versterkt doordat de innovatie vooruit kan lopen op de ethische en juridische kaders, wat een risico vormt van ongevalideerde AI-systemen.[149] De verwachting is dat deze vertraging tussen systeemontwikkeling en ontwikkeling c.q. aanpassing van wettelijke kaders relevant blijft.

De terugkerende dilemma's op toepassingsniveau tussen effectiviteit en menselijke maat. In de keuze tussen wel of geen integratie van AI in functies in het publieke domein zitten besluitvormers voortdurend gevangen in een bestuurlijk dilemma: Aan de ene kant liggen er enorme kansen van AI-toepassingen, juist ook om aan de hand van AI meer maatwerk te realiseren en meer capaciteit vrij te maken. Aan de andere kant ligger er door integratie van meer digitale technologie de risico's op verdere verzakelijking en gebrek aan controle over AI-toepassingen.

147 Chloe Xiang, 'He Would Still Be Here': Man Dies by Suicide After Talking with AI Chatbot, Widow Says, VICE, 2023.

148 Pierre-Francois Lovens, "Sans ces conversations avec le chatbot Eliza, mon mari serait toujours là", La Libre, 2023.

149 Atlantic Council, Experts React: The EU Made a Deal on AI Rules. But Can Regulators Move at the Speed of Tech?, 2023.

Dreigingen van AI voor de maatschappelijke stabiliteit op toepassingsniveau

- **Algoritmische bias maakt toepassingen vatbaar voor discriminatie.** AI-toepassingen in vele verschillende vormen bezitten risico's van algoritmische bias. Hierdoor ontstaan onderdrukkende of discriminerende effecten.
- **Criminele toepassingen van AI kunnen een serieuze uitdaging vormen voor opsporing en criminaliteitsbedrijding.** De toepassingen van AI in criminaliteit kunnen zorgen voor een professionaliseringsslag op bijvoorbeeld het terrein van vermogensdelicten. Waar sprake is van zogenaamde High Impact Crimes verergert
- **Interactie met AI-bots geeft reële psychologische risico's**. Zoals in het geval van affectieve emoties die in machinecommunicatie wordt geprojecteerd.
- **AI-toepassingen kunnen geen vervanging zijn voor menselijke ervaringskennis.** Maatschappelijk gezien ligt er het risico dat we teveel gaan leunen op AI en teveel navigeren op de pointers van algoritmen. Dit zorgt voor bestuurlijke armoede en ondermijnt het vertrouwen van burgers in de overheid die een menselijke maat dient te houden in beleid en uitvoering.
- **Ethische regie is in de praktijk lastig**. Regulering en ethische vangrails worden veelal wettelijk bepaald, maar de werkelijke ethische regie hangt af van ingewikkelde, moeilijk te handhaven toetsingsprocessen.
- **Er blijken vaak grijze gebieden in de wet- en regelgeving** die maken dat het in de praktijk moeilijk bindend vast te stellen is of een toepassing voor de wet risicovol is en verboden moet worden.
- **De terugkerende dilemma's op toepassingsniveau tussen effectiviteit en menselijke maat.** In de keuze tussen wel of geen integratie van AI in functies in het publieke domein zitten besluitvormers voortdurend gevangen in een bestuurlijk dilemma tussen kansen en ethische risico's.

5. Strategische balansoefeningen

Met het zicht op de vele mogelijke vormen van impact die AI heeft op de maatschappelijke stabiliteit blijft de vraag – 'hoe moet de overheid zich tot deze ontketende dynamiek van AI verhouden?' De omvang van het krachtenveld rondom AI is zo groot, en de toepassingsontwikkelingen die van buitenaf op ons afkomen dermate kaleidoscopisch, dat het moeilijk voor te stellen is dat Nederland over de impact van AI erg veel regie kan claimen 'in eigen huis'. Vrijwel altijd zal de slotsom zijn dat Nederland in het bewandelen daarvan zal willen aansluiten bij Europa. Maar toch, ook binnen de context van Europa heeft Nederland keuzes in hoe het zich oriënteert in relatie tot AI. Met het doel om ook strategisch te blijven in de focus van deze studie (en geen concrete beleidsvoorstellen te willen doen), is er een aantal strategische 'trade-off-situaties' waarbinnen Nederland keuzes kan maken. De manoeuvreerruimte is echter vaak weer zo dat niet alle doelen kunnen gelijktijdig gediend worden. Deze 'strategische balanceeracts' zijn achtereenvolgens:

1. **Innovatie 'versus' waarden**. Vooropgesteld, deze hoeven elkaar niet uit te sluiten. Maar specifiek in relatie tot AI is het lastig om ethische eisen op te werpen en waarden te willen dienen zonder daarmee ook het innovatieproces te frustreren of naar elders te verjagen.
2. **Buitenhouden 'versus' meebuigen**. Is het mogelijk om vangrails in te bouwen als je zelf niet aan de tekentafel zit? De EU heeft met de WAI een belangrijke dam gebouwd tegen risicovolle toepassingen. Maar hoe stevig is een dam tegen een systeemtechnologie die overal in lijkt te gaan doordringen? Zijn er mogelijkheden om op een verantwoorde manier AI-ontwikkeling en de bescherming van het maatschappelijk belang te dienen?
3. **Defensief mensgericht 'versus' offensief mensgericht**. Nederland volgt de Europese koers van de regulering van AI ter bescherming van mensgerichte waarden. Deze ethische insteek is goed maar moet niet uitmonden in een protectionistisch 'leven achter digitale dijken'. Er is ook wat voor te zeggen om de potentie van AI op een proactieve manier te exploreren ten gunste van een mensgerichte, veilige samenleving.

5.1. Innovatie 'versus' waarden

De EU wil inzetten op regie en de bescherming van waarden, maar ook met behoud van innovatiekracht. Ook in de Nederlandse strategie zet men in op gereguleerde vrijheid en duidelijke voorwaarden waaronder innovatie moet kunnen plaatsvinden.[150] Dreigingen ontstaan in het perspectief van Nederland met name wanneer regulering achterblijft en AI zich ongehinderd door ethische principes zou kunnen ontwikkelen. Nederland volgt de Europese wet- en regelgeving daarom nauwgezet. Daarnaast past ook de regulatieve toon van de EU goed bij het Nederland dat door de Toeslagenaffaire (en andere gevoelige zaken rondom overheidstoepassing van algoritmen) voorzichtig is geworden om het broze politieke vertrouwen niet verder op het spel te zetten. Nederland heeft als een van de weinige landen

[150] Rijksoverheid, Kabinet presenteert visie op generatieve AI, 2024.

de ervaringswijsheid opgebouwd dat ondoordachte algoritmen negatief kunnen uitpakken voor het vertrouwen in de democratische rechtsorde. Een van de grote uitdagingen voor Nederland (en tevens voor Europa) is om **AI-gedreven innovatie te ondersteunen, maar tegelijk 'Europese' waarden en rechten te waarborgen**. Dit kan deels door randvoorwaarden in Europese wetten en regels vast te leggen, waarbij de individuele lidstaten snel volgen. Maar daarbij doemen meerdere grote uitdagingen op:

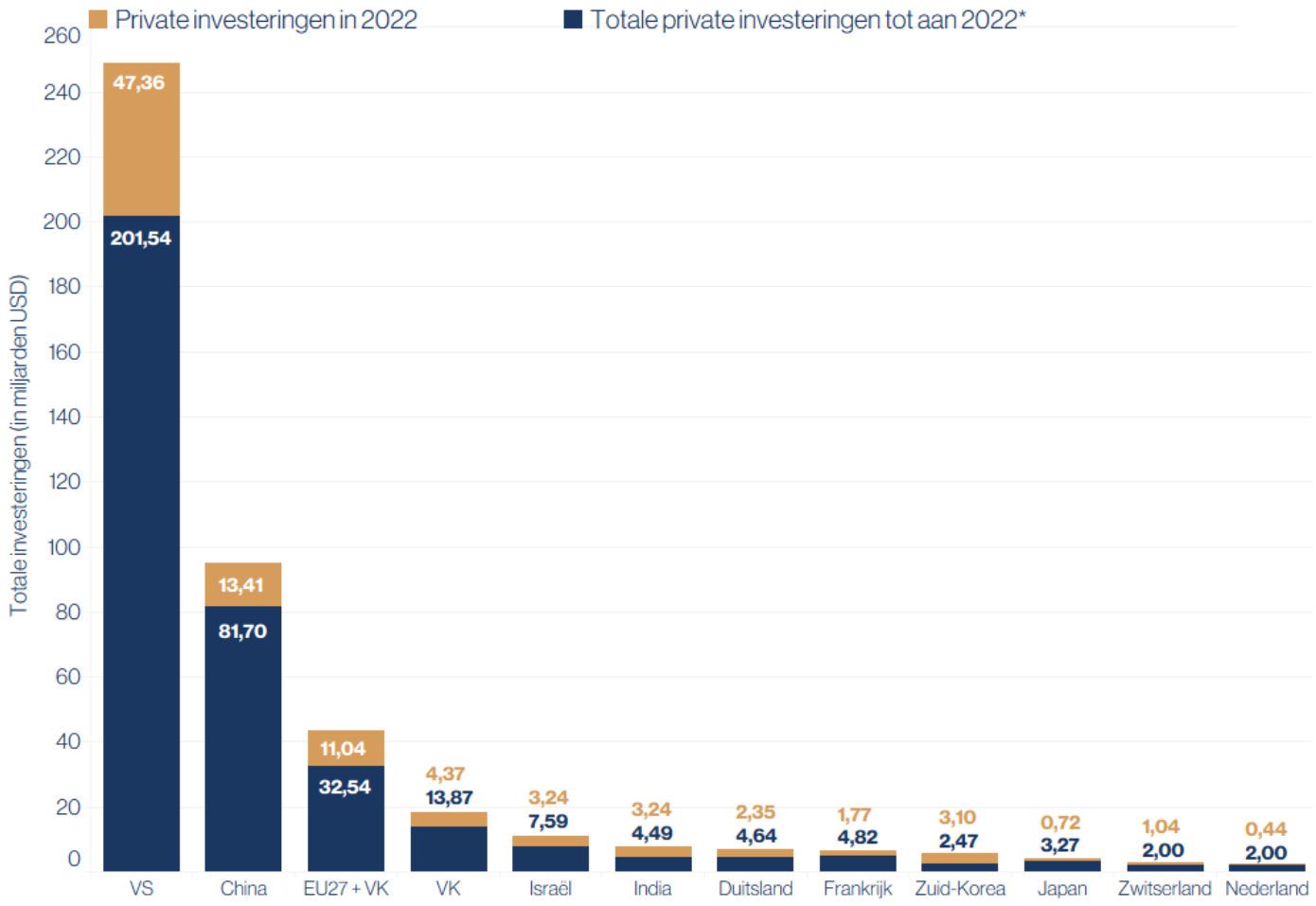
1. **Regulering kan de innovatie niet bijbenen.** De ontwikkelingen op AI-gebied gaan heel snel (zie ook 3.1). De reeds aangehaalde **AI Power Paradox**[57] vertelt ons dat de enorme ontwikkelingssnelheid van AI het nagenoeg onmogelijk maakt om beleid up-to-date te houden. Regulering, zeker wanneer deze sterk gefragmenteerd blijft tussen landen en grootmachten onderling, kan dan de beheersing van AI mogelijk zelfs tegenwerken. Dit heeft alles te maken met het feit dat er geen ontkoppeling mogelijk is van AI-technologie die wordt beheerst door alomtegenwoordige private buitenlandse partijen, en dat er geen gereguleerde safe zone kan bestaan waar de impact niet zal reiken. Zeker niet wanneer het regulatieve lichaam (de EU) is afgesneden van het technologisch ontwikkelingslichaam (met name de VS en China). Het is bovendien vrijwel onmogelijk om regelgeving en het bijbehorende handhavingsapparaat in de pas te laten lopen met ontwikkelingen. Voortdurend zullen overal grijze gebieden blijven bestaan die ook de handhaving zeer zal bemoeilijken. De EU heeft wel (hinder-) macht om de innovatie van Amerikaanse bedrijven te dwarsbomen maar het heeft te weinig zeggenschap om de bepalende stappen die in AI gezet worden te beïnvloeden.
2. **De innovatie in het AI-domein vindt vooral buiten Europa plaats.** Een minder vriendelijke verklaring voor de strategische behoedzaamheid en de focus op waardengedreven innovatie, is dat Europese landen niet veel anders kunnen dan toekijken en de digitale dijken ophogen. Een van de grote hindernissen die ook al in de voorgaande analyse wordt benoemd is dat uiteindelijk slechts een handvol grote Amerikaanse sterbedrijven uit Silicon Valley er tot nu toe in slaagt om, geholpen door forse, jarenlange investeringen, AI aan te bieden als een tool voor de (mondiale) massa. Waar de EU de wettelijke infrastructuur in tientallen jaren heeft kunnen optuigen in reactie op de assertieve digitale aanwezigheid van Google, Facebook, Microsoft e.a., hebben de Amerikaanse Big Techbedrijven jarenlang start-ups kunnen kopen en dankzij astronomische winsten kunnen investeren in ambitieuze R&D-projecten in de eigen achtertuin. Graag zouden de EU en Nederland onder het mom van 'strategische autonomie' onafhankelijker willen zijn van deze Amerikaanse online giganten, maar deze achterstand is niet met enkele impulsen of binnen enkele jaren opgelost. In de praktijk betekent dit dat de nodige macht over de koers van AI voor een deel ligt bij tech CEO's en niet bij staten.
3. **Europa (Nederland incluis) investeren niet genoeg om het innovatieproces te kunnen regisseren.** In het 2021 Coordinated Plan on Artificial Intelligence geeft de Europese Commissie aan AI te willen aanjagen, maar met behoud van een goede bescherming van rechten en EU-waarden. Om dit te bereiken worden investeringsfondsen opgetuigd. De Horizon Europe en Digital Europe programma's zouden samen goed moeten zijn voor een jaarlijkse Europese investeringsimpuls van circa €1 miljard.[151] Maar dat is nog maar een fractie van de totale innovatiegelden die in Europa in AI worden geïnvesteerd. Op nationaal niveau hebben alle landen weer hun eigen investeringsprogramma's. Maar, zoals in de grafiek en in het kader staat weergegeven, zowel de Europese als binnenlandse investeringen verbleken bij de investeringen waar de VS en Amerikaanse tech sector goed voor zijn. Het is daarom wijs om na te denken hoe voor deze verloren concurrentiekracht kan worden gecompenseerd en of het misschien beter is om selectiever op niches van AI-ontwikkeling in te zetten.

[151] Europese Commissie, A European approach to artificial intelligence, 2024.

Nederlandse investeringen in AI-innovatie. Volgens ramingen van ING Sector Research ("AI vindt zijn weg naar alle sectoren: Artificiële intelligentie biedt meeste waarde voor de IT sector", 2020) verdubbelen de AI-bestedingen in Nederland van naar schatting €1,6 miljard naar €3,2 miljard tussen 2020 en 2025. Dat komt neer op een groei van €0,32 miljard per jaar, waarbij wordt uitgegaan van een gestage en lineaire investering in AI op basis van de groei in 2020. Als onderdeel van het Nationaal Groeifonds 2021 werd de Nederlandse AI Coalitie (**AINed**), een publiek-private partnership bestaande uit meer dan 250 deelnemers, een investering van €276 miljoen toegekend. Dit bestaat uit €44 miljoen directe toekenning, €44 miljoen voorwaardelijke toekenning en een reservering van €188 miljoen. In 2021 kon de Nederlandse AI-sector dus op een publiek-private investering van €88 miljoen rekenen. Daarnaast werd voor 2022 €116,5 miljoen van de reservering toegekend. Uitgaande van een groei van €0,32 miljard per jaar (op basis van de groei van de AI-bestedingen in 2020, zie onderzoek ING), met daarbovenop de AINED investeringen van €88 miljoen in 2021 en €116,5 miljoen in 2022, komen onze schattingen uit op een investering van €0,44 miljard)in 2022 en een totaal aan investeringen in de AI-sector van €2,44 miljard in 2022. Daarbij moet opgemerkt worden dat 2021 een uitzonderlijk jaar was voor AI-investeringen: in de VS, maar ook zeker in de Europese Unie, kwamen (private) investeringen een stuk hoger uit in vergelijking tot 2020. In 2023 lijken de investeringen weer de 'normale' groei van 2020 aan te nemen. De schatting van de totale investeringen in 2022 valt dus vermoedelijk lager uit dan in werkelijkheid.

Behoedzaamheid van Nederland heeft een prijs. Nederland doet er in principe goed aan om behoedzaam om te gaan met AI-toepassingen. Maar het eerlijke verhaal is dat behoedzaamheid ook een prijs heeft. Nederland en de EU kunnen door (op zich substantiële) investeringen iets meer grip proberen te krijgen op de koers van innovatie binnenshuis maar het zal lastig zijn om op dit punt te wedijveren met de Amerikaanse tech corporates en de diepe zakken van de Chinese staatsgeleide innovatieprogramma's. Een van de belangrijke aandachtspunten voor zowel Nederland als de EU is dat innovatiegelden goed zouden moeten worden gericht, met inachtneming van de strategische belangen van de Europese toekomst.

Figuur 9: Private investeringen in AI.



Bron: Stanford University, ING en Rijksoverheid
*De totale private investeringen is een som van de jaarlijkse investeringen (2013 -2021) per regio of land, met uitzondering van Nederland. Voor de private investeringen van Nederland is een schatting gemaakt op basis van gecombineerde gegevens uit de genoemde bronnen.

Juist met een kleine portefeuille is het een risico dat investeringen te veel gefragmenteerd zijn en blijven. Tegen deze achtergrond, en de forse nadruk op regulering, is er juist ook behoefte aan meer strategisch denken en meer visie – een 'grand plan' – op welke AI-gebieden Nederland en EU precies leiderschap willen claimen. Het moet ook voor de bevolking duidelijk zijn dat AI geen technisch politiek dossier is maar een onderwerp waarmee veel maatschappelijk op het spel staat. Thema's als bestaanszekerheid staan nu nog hoog op de politieke agenda maar worden weinig in het licht geplaatst van de toekomstige impact van AI. Innovatiefondsen voor AI zijn daarom geen politiek 'extraatje' maar bittere noodzaak om Nederland concurrerend te houden en invloed te kunnen blijven houden op de ontwikkelingsagenda.

5.2. **Buitenhouden 'versus' meebuigen**

Goed, Europese landen, inclusief Nederland, zitten niet op de stoel van de chauffeur en investeren waarschijnlijk ook onvoldoende om nog op een leidende positie te komen. Welke mogelijkheden zijn er dan toch om druk uit te oefenen en te voorkomen dat de toekomst teveel 'Wild West' wordt?[152]

Een reactieve houding? Zoals we zagen in hoofdstuk 4 is AI-technologie moeilijk te controleren om intrinsieke redenen: broncodes zijn vaak niet open source, plus er is het probleem van AI als black box: 'onder de motorkap' is het niet zomaar duidelijk wat er allemaal gebeurt en welke effecten worden gesorteerd. Er zijn verder extrinsieke factoren die maken dat het gericht inspelen en reageren op AI-technologie ingewikkeld is: Europese landen zitten qua innovatie niet in de 'driving seat', zoals duidelijk werd in voorgaande paragraaf, waardoor men tot een afwachtende, reactieve houding wordt gedwongen. De macht van de EU komt voornamelijk voort uit marktontzegging op basis van een juridische rationale van consumentenbescherming. Daarmee ontstaat weliswaar een intern gevrijwaard gebied voor als risicovol beoordeelde toepassingen maar zonder daarbij grip te hebben op de vaart der volkeren om ons heen. We zien al dat bepaalde AI-features niet in Europa worden gelanceerd en wel elders.[153] Het is bovendien lastig in regels te vangen wat je nog niet kunt voorzien.

Daar komt nog bij: ook de nieuwe WAI, welke bedoeld is als het kardinale regulerend instrument van de EU, is zeker niet waterdicht. Hoewel er wel criteria zijn voor kwalijke of risicovolle AI-toepassingen zijn er veel grijsgebieden en moet jurisprudentie en ervaring met categorisering nog opgebouwd worden. Voor bepaalde toepassingscategorieën ligt de bewijslast voor ethische verantwoording bij de producent of dienstverlener zelf wat allerlei openingen biedt voor exploitatie van onduidelijkheden of het verzwijgen van functionaliteiten. Dit alles bij elkaar maakt dat de reactieve insteek wellicht nuttige effecten oplevert op het punt van protectionisme en consumentenbescherming maar of de EU met wettelijke kaders mondiaal ethische aanpassingsprocessen kan afdwingen, zoals het dat doet met het zetten van allerlei andere marktstandaarden, dat is nog maar zeer de vraag. Wel is men het erover eens dat de EU-landen voor het moment ten minste de minimale stap hebben gezet om grenzen te stellen en de meest risicovolle experimenten te verbieden.

Meeveren en aanpassen? Zoals de WRR nog maar eens benadrukt: AI blijft een 'systeemtechnologie' die de manier van werken, leren en creëren verandert, die andere innovaties mogelijk en 'gemakkelijk schaalbaar maakt' en allerlei positieve maatschappelijke bijdragen kan leveren.[154] Andersom: door teveel een slot te zetten op de ontwikkeling van AI zullen we ook stagnatie riskeren op allerlei andere fronten en in bestaande sectoren. We kunnen het ons in wezen niet permitteren om alleen op een reactieve manier om te gaan met AI-ontwikkelingen. Dit inzicht is ook in zeer belangrijke mate van toepassing voor publieke toepassingen van AI. Aan de ene kant ligt er altijd de angst voor een herhaling van de Toeslagenaffaire, maar aan de andere kant kan AI, goed toegepast, juist ook veel bestuurlijke of technische fouten voorkomen. AI kan er ook voor zorgen dat processen minder bureaucratisch of afstandelijk zijn.

Voor heel veel AI-toepassingen blijft een dual use waarschijnlijk. Zowel de eigenschappen ten goede, als die ten kwade zullen naar voren treden in het iteratieve proces van ontwikkeling.

---

152 Redactioneel commentaar, Met nieuwe AI-wet EU wordt de toekomst minder wild-west, NRC, 2024.

153 Linnea Ahlgren, Google's Gemini AI won't be available in Europe — for now, 2023.

154 WRR, Opgave AI. De nieuwe systeemtechnologie, 2021.

Dat impliceert dat potentieel omstreden AI-toepassingen niet direct moeten worden afgebroken omdat alleen de omstreden effecten in beeld komen. Het is dan verstandiger om door te ontwikkelen maar met het willen inbouwen van de nodige **vangrails** die de kwalijke effecten kunnen verhinderen of kanaliseren. Een hulpmethode om waardengedreven AI governance in de praktijk te doen is de ethische discipline van **value sensitive design.** Op dit gebied van ethisch verantwoord technologisch design heeft Nederland veel kennis in huis aan o.a. de TU Delft en Wageningen Universiteit en ligt een leading rol en zelfs mondiale voorbeeldfunctie voor de hand. Ook is de Theorie van Technologische bemiddeling, van Peter Paul Verbeek internationaal invloedrijk. Een voorbeeld van een project waarin AI-gestuurde veiligheidsprocessen en value sensitive design-methoden succesvol zijn toegepast is het TRESSPASS-project voor risicogestuurde grenscontroles.[155]

Value Sensitive Design (VSD) is een theoretisch onderbouwde benadering voor het ontwerpen van ethisch verantwoorde technologie, waarbij waarden worden verankerd in de loop van het technologisch ontwikkelproces. VSD houdt daarbij rekening met menselijke waarden gedurende het hele ontwerpproces dat bestaat uit drie fasen: 1) de conceptuele fase – wie zijn de belanghebbenden en welke waarden verdienen het om centraal te staan?; 2) de empirische fase – kwalitatieve of kwantitatieve studies dienen om meer inzicht te krijgen in gebruikscontext; 3) de technische fase – de analyse van het design en de mate waarin het de waarden integreert en ondersteunt. Zoals de verschillende fasen al aangeven is de goede integratie van ethiek in het designproces een zorgvuldig en daarmee vaak tijdrovend proces. Tegelijk zijn dergelijke ethische analyseprocedures onontbeerlijk om technologische ontwikkeling te bereiken met de waarden vooraf op het netvlies zonder dat ethiek vooral een reflectieproces achteraf hoeft te zijn wanneer de technologie al bestaat.

Ook als Nederland niet de ambitie heeft (noch kan waarmaken) om voorop te lopen, laat het dan tenminste maximaal inzetten op **innovatie op het gebied van responsible AI.** Overheden kunnen hulp aanbieden om de juiste vangrails te ontwerpen en de technologie verantwoord te maken. De Nederlandse overheid, in casu het Ministerie voor Infrastructuur en Waterstaat, heeft bijivoorbeeld een dergelijke assessmentinstrument ontwikkeld waarmee AI getoetst kan worden aan de relevante grondrechten en ethische principes.

Mogelijk kunnen meer omstreden experimentele projecten op het gebied van AI veel beter onder een van de samenleving afgeschermd klimaat binnenshuis worden gefaciliteerd dan dat Europese landen op een defensieve manier handelen ten aanzien van zaken die buiten het invloedsgebied ontwikkeld worden. Door de Europese waardenagenda in het hart te plaatsen van innovatiesupport kan deze strategie succesvol uitpakken mits er ook een uitstekend ecosysteem ontstaat voor het AI-experiment.

155 TRESSPASS, Onderzoek over risicogestuurde grenscontroles, TNO, 2023.

5.3. Defensief mensgericht 'versus' offensief mensgericht

Ook al ziet de Rijksoverheid Nederland graag als koploper[156], in wezen is Nederland vooral een toeschouwer van het mondiale web van AI-initiatieven dat wordt geweven door met elkaar concurrerende krachten op andere plekken in de wereld. Gezien alle problemen van controle die AI omgeven, zou het mooi zijn, zoals de VN ook voorstelt in haar visie, dat er op mondiale schaal AI-regulerende instituties ontstaan die (net zoals het IMF dat op macro-economisch niveau doet) universele risico's vast kunnen stellen en een neutraal toezichts- en handhavingsregime instellen (geïnspireerd op bijvoorbeeld het Internationaal Atoomenergieagentschap). De kans dat dergelijke mondiale governance-structuren op korte termijn worden opgetuigd is echter zeer klein, precies omdat China en de VS verwikkeld zijn in de hegemonie over de AI-industrie in de wereld – een strijd die, zeker in de ogen van deze beide grootmachten, een belangrijk winner-takes-all-karakter heeft. Nederland, en Europa, kunnen en moeten harder bepleiten dat AI-governance in alleen de EU uiteindelijk onvoldoende is. Een dergelijk agentschap zou niet misstaan in het Nederlandse en Haagse profiel van voorvechter internationale vrede en veiligheid.

Misschien zou Nederland zich een al te bescheiden rol op het domein van AI ook niet moeten permitteren. Nederland vervult met ASML als enige Europese speler een cruciale schakel in de mondiale productieketen die AI ondersteunt.[157] Dat betekent niet dat Nederland daarmee een hefboom heeft om extra verschil te maken in de toepassingswereld van AI. Maar wel: 'noblesse oblige' – laat deze geopolitieke sleutelrol ook een reden zijn om de urgentie en de 'stakes' van de AI-transformatie over te brengen aan de Brusselse of binnenlandse politieke gesprekstafel. Een punt van aandacht, ook in de Nationale Technologiestrategie, is dat de ambities op andere technologische terreinen vaak samenhangen met de wereldwijde AI-boom. Dit betekent in die zin dat een groot deel van de toekomstige brede welvaart afhangt van allerlei high-techsectoren (optische systemen, quantumtechnologie, imagingtechnologie) die op een vernetwerkte manier met elkaar samenhangen en met AI. Het idee 'jammer, maar voor de toepassing van AI hebben we de boot even gemist, we zetten ons geld in op iets anders', kan een valse voorstelling van zaken blijken en een vorm van defaitisme die economische ondermijnend werkt, alsmede ondermijnend is aan de geopolitieke rol van Nederland.

De wat ingehouden ambitie blijkt ook uit de Overheidsbrede visie generatieve AI (2024) die rust op zes pijlers: 'samenwerken', 'het volgen van de ontwikkelingen', 'het vormgeven en toepassen van wet- en regelgeving', 'het vergroten van kennis', 'het innoveren met AI', en ten slotte, 'toezicht en handhaving'.[158] De nadruk op samenwerken, waarnemen, volgen, past wellicht ook bij een land dat zich in de context van de EU als een hogere middenmoter ziet, en in relatie tot de VS als een 'transatlantische volger.' Tegelijkertijd behoort Nederland tot de koplopers in Europa als het gaat om de digitale economie. Circa 13% van de Nederlandse bedrijven heeft in 2023 AI geïntegreerd in hun organisatiemodel en moet daarin alleen Denemarken, Portugal en Finland boven zich dulden. Op de AI-readiness index, opgesteld door Oxford Insights[159], scoort Nederland eveneens tamelijk goed: op basis van verschillende maatschappelijke en bestuurlijke variabelen is Nederland een goede omgeving om te kunnen accommoderen naar nieuwe AI-technologie en de maatschappelijke verandering die ze kan meebrengen.

156 Rijksoverheid, De Nationale Technologiestrategie, 2024.

157 Henry Ren en Bloomberg, Dutch semiconductor giant ASML emerges as Europe's AI champion after hitting record high, Fortune Europe, 2024.

158 Rijksoverheid, Overheidsbrede visie Generatieve AI, 2024.

159 Oxford Insights. AI Readiness Index, 2023.

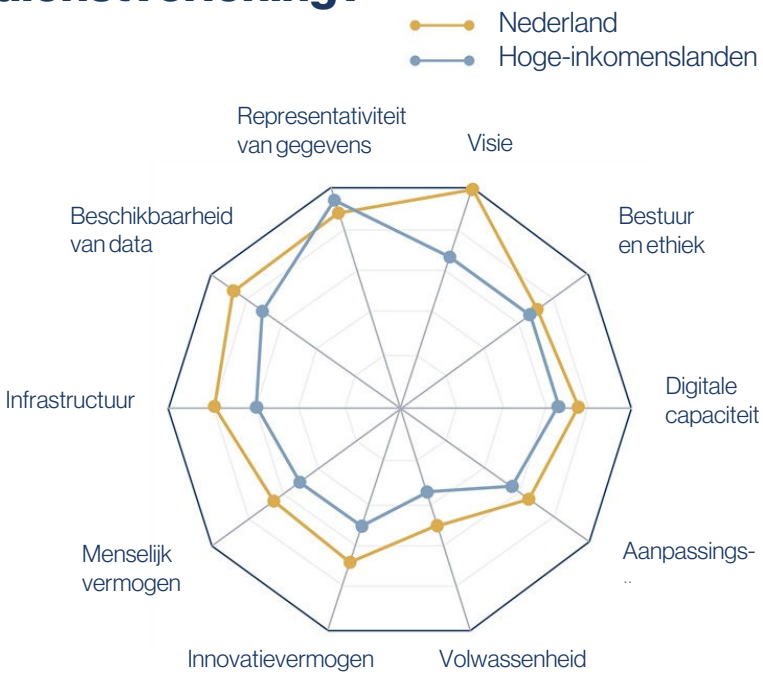
Figuur 10: indicatoren voor het kunnen omgaan met de impact van AI
(bron: Oxford Insights)

Hoe klaar is de Nederlandse overheid om AI te integreren in de publieke dienstverlening?

	Nederland	Hoge inkomens landen
Visie	####	66,39
Bestuur en ethiek	72,41	67,97
Digitale capaciteit	79,80	66,24
Aanpassingsvermogen	63,37	58,30
Volwassenheid	51,41	37,17
Innovatievermogen	69,25	52,96
Menselijk vermogen	65,22	52,91
Infrastructuur	80,29	60,82
Beschikbaarheid van data	83,29	72,38
Representativiteit van gegevens	84,09	87,26

Totale index score	74,47
Rank t.o.v. rest van de wereld	10

Bron: Oxford Insights



De Nederlandse bevolking is een ver bovengemiddeld digitaal vaardige bevolking wat de ontvankelijkheid voor positieve toepassingen van AI groter maakt. Kortom, de voorwaarden voor een groter aandeel in de AI-economie lijken prima in orde. Binnen Europese context zou het gerechtvaardigd zijn wanneer Nederland een minder voorzichtige rol zou pakken op het gebied van ideevorming over hoe AI zowel strategisch als verantwoord kan worden ingepast. Toch overheerst de behoedzaamheid, en misschien ook niet zonder reden. Nederland, binnen de Europese context lijkt bewust te kiezen voor een defensieve oriëntatie, misschien ook 'schuw' geworden door de Toeslagenaffaire.

Een meer 'offensief mensgerichte' oriëntatie op het gebied van de ontwikkeling van (responsible) AI-toepassingen zou bovendien bittere noodzaak kunnen blijken. De hoge mate van digitale volwassenheid van de Nederlandse samenleving werkt ook door in de democratisering van allerlei toepassingsvormen die door mensen, groepen en dus ook door anti-institutionelen, hackers en criminele organisaties zullen worden ontwikkeld en gebruikt. Het bewaken van 'mensgerichte AI' is erg ingestoken vanuit een protectionistische gedachte tegen de boze (met name Amerikaanse) buitenwereld en neemt weinig mee dat mensgerichte AI ook begrepen kan worden als een uitnodiging om AI te ontwikkelen *ter bescherming van* deze mensgerichtheid: Hoe kan AI ingezet kan worden tégen allerlei exploitatievormen door middel van AI? Of, hoe kan Nederland een voorhoederol pakken in onderzoek naar een goede en 'timely' aansluiting van de maatschappelijke situatie (zoals juridische of bestuurlijke vraagstukken) op AI-ontwikkelingen die aan het arriveren zijn?

Bijlage 1. Beknopte lijst met AI-termen en AI-taxonomie

Machine Learning. De meeste AI van dit moment werkt op basis van **machine learning** (ML) principes. ML is een op neurale netwerken geïnspireerde techniek. ML-systemen worden al breed toegepast, bijvoorbeeld voor gezichtsherkenning, als spamfilter en voor allerlei marketingtoepassingen. ML-systemen zijn in staat om op basis van ervaringsinput tot een schatting van een juist antwoord te komen. Vaak worden ze in één adem genoemd met de term 'big data', omdat ze veelal omvangrijke databases vergen. Een voorbeeld van een ML-systeem is het afstellen van stoplichten door het analyseren van de verkeersdoorstroom onder verschillende tijdsduren van rood licht. Hoe meer data verzameld wordt door het systeem, hoe effectiever en accurater het kan werken.

Deep learning (DL) – is een subset van Machine Learning (ML).[160] DL-architectuur maakt het uitermate geschikt voor het herkennen van patronen in zeer complexe semigestructureerde verschijnselen zoals menselijke taal. Normaliter werkt een ML-systeem met sorteringen en gestructureerde data. Op basis van vooraf bepaalde kenmerken wordt een matchingsproces uitgevoerd. Een DL-systeem kan de sortering (deels) op eigen kracht herkennen omdat hij patronen in informatie omzet naar een eigen sortering en tot een steeds complexer algoritme-ontwerp komt. Een DL-algoritme kan zo leren bepalen hoe groot de kans is dat, bijvoorbeeld, lettercombinaties telkens in een vaste volgorde worden gebruikt. Op die manier ontstaat een statistisch beeld van welke woorden in de menselijke taal kunnen worden gevormd. Vervolgens kunnen die woorden statistisch gerelateerd worden aan hoe vaak, en in welke combinaties ze verschijnen in teksten. Dit bouwt een beeld van correcte zinsbouw in verschillende contexten.[161] DL-algoritmes produceren, gegeven de input en het algoritme zelf, de statistisch meest logische uitkomst, wat niet per se ook juiste of betekenisvolle informatie inhoudt. De kwaliteit van de inputdata is cruciaal.

Generatieve AI. Is een verzamelnamen voor de capaciteiten van algoritmen die in staat om volledig nieuwe kunstmatige output te creëren op basis van een trainingsset. Dit kunnen teksten, afbeeldingen, muziekstukken of zelfs video's zijn. In tegenstelling tot computationele AI-systemen, bootst generatieve AI menselijke creativiteit na. Het kan unieke perspectieven bieden, zoals surrealistische stadsmontages en zeer gedetailleerde scènes. Hoewel de kwaliteit van gegenereerde content varieert, geeft generatieve AI ook nieuwe gezichtspunten die minder snel in het verlengde liggen van het menselijke creatieve proces.

---

160 IBM Data and AI Team, AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the Difference?, 2023.

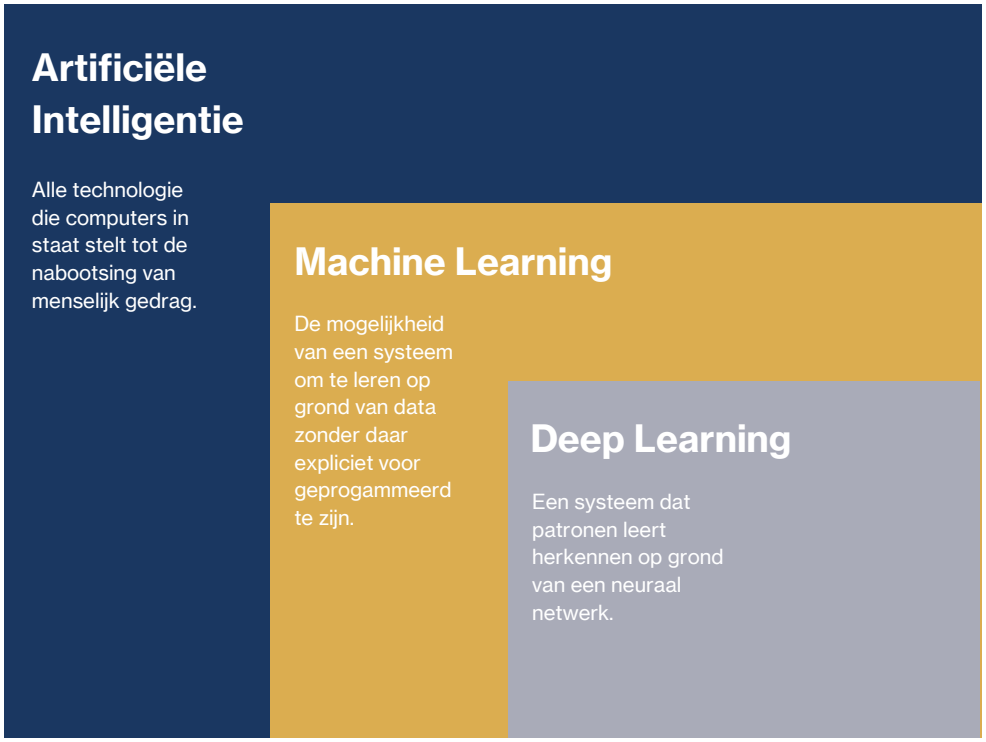
161 Adam C. and Richard Carter, Large Language Models and Intelligence Analysis, 2023.

Strong versus narrow AI. AI, opgevat als het vermogen van een computer om een breed scala aan geautomatiseerde taken uit te voeren vergelijkbaar met menselijke capaciteiten, kan op het hoogste niveau worden onderverdeeld tussen **Narrow AI** dat taken even goed of beter dan mensen kan uitvoeren binnen een specifiek toepassingsgebied en reeks van criteria; en **Strong AI** dat theoretisch het hele scala aan menselijke capaciteiten kan uitvoeren. Strong AI (ook wel aangeduid als deep AI) bestaat momenteel niet of verkeert hoogstens nog in experimentele staat. Alle huidige AI-toepassingen vallen in principe onder Narrow AI.

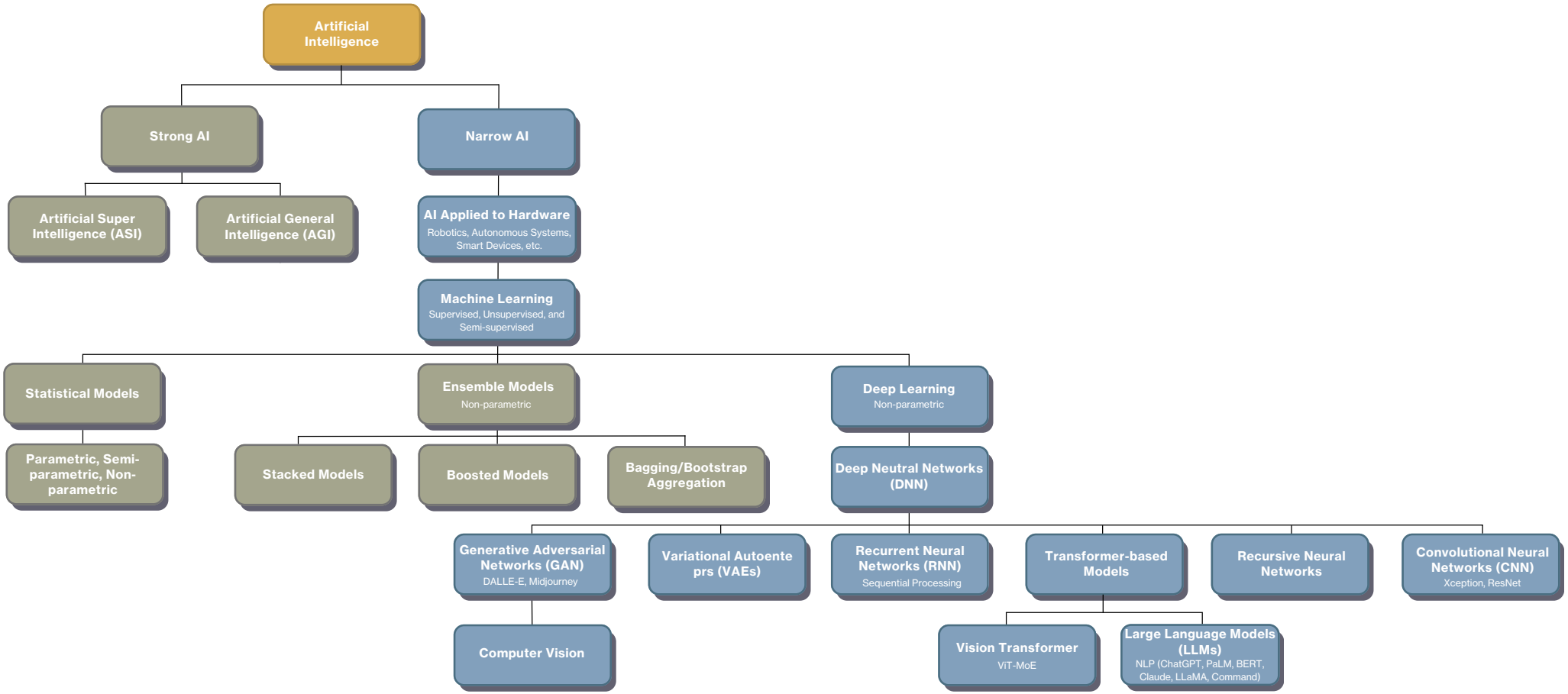
Superintelligentie – ook wel bekend als hyperintelligentie, verwijst naar een hypothetisch wezen dat menselijke intelligentie ver overtreft. In de praktijk verwijst het begrip met name naar de gebeurtenis dat kunstmatige intelligentie de menselijke cognitie voorbijstreeft. Hieronder liggen allerlei wetenschappelijke en filosofische discussies over wat intelligentie in essentie is. In relatie tot maatschappelijke stabiliteit is het vooral van belang dat de impact onbeheersbaar zal kunnen zijn op het moment dat de redeneringsvermogen het menselijk begrip zal overstijgen. Het idee van superintelligentie is verwant aan de notie van 'strong AI'.

Taxonomie van verschillende vormen van AI. Figuur 11 geeft een overzicht van de verschillende vormen van AI. Op het hoogste niveau wordt onderscheid gemaakt tussen Narrow AI en Strong AI. Zoals in §3.2 reeds aangegeven heeft het eerste betrekking op de uitvoering van menselijke taken binnen een afgebakend toepassingskader (het spelen van een schaakspel of het identificeren van tumoren op een longfoto); terwijl het tweede slaat op AI dat in staat is tot het uitvoeren van het hele scala aan menselijke capaciteiten en, net als de mens, zich zelf kan leren nieuwe problemen op te lossen. Strong AI bestaat momenteel niet of verkeert hoogstens nog in experimentele staat. Alle huidige AI-toepassingen vallen in principe onder Narrow AI.

Figuur 11: Hiërarchie in verschillende begrippen



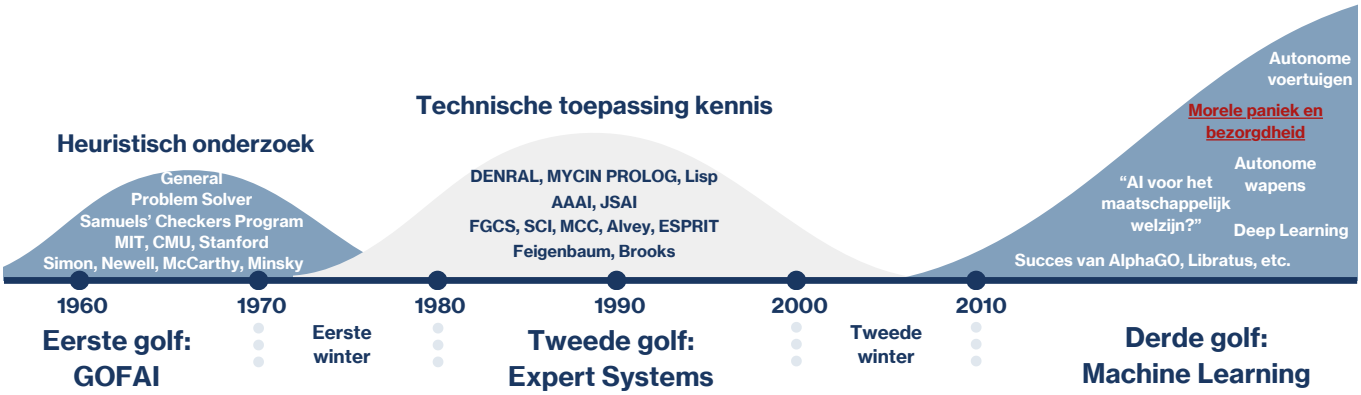
Figuur 12: Een niet-limitatieve taxonomie van verschillende technische vormen van AI en hun verwantschappen (HCSS). De grijze categorieën zijn gebieden van machine learning die voor deze studie buiten beschouwing zijn gelaten omdat het zwaartepunt ligt op de verandering die deep learning-systemen hebben gebracht.



Bijlage 2. **Een beknopte geschiedenis van 'AI-booms'**

De hausse over ChatGPT doet soms vergeten dat AI al een lang ontwikkelingstraject van pieken en dalen kent.[162] De eerste golf, van 1950 tot 1970 (ook wel bekend als de tijd van GOFAI - Good Old Fashioned AI), werd aangevoerd door wetenschappers als Alan Newell en Marvin Minsky. Toen al claimde men dat computers binnen tien jaar schaakkampioen zouden kunnen worden. Tegen de achtergrond van de Koude Oorlog zag men militaire toepasbaarheden en werden grote projecten gefinancierd door DARPA (Defense Advanced Research Projects Agency). De verwachtingen rondom AI raakten opgeklopt toen er steeds meer publicaties verschenen over 'cybernation', een samenvoeging van cybernetics en automation. Zo ontstond de eerste 'morele paniek' (in de sociologische betekenis van buitensporige, gemediatiseerde en maatschappelijke angstreactie) over AI over computers die alle banen zouden overnemen. Toen deze voorspellingen niet uitkwamen, verdwenen de aandacht en het geld en was de 'eerste AI-winter' een feit.

Figuur 13: De drie AI-golven.[163] We zien in de derde golf een verhoogd publiek bewustzijn, gepaard gaand met de nodige bezorgdheid, omdat de impact van AI-toepassingen in het heden reële contouren begint te krijgen in het dagelijks leven van iedereen.



[162] Europese Commissie et al., AI Watch, Historical Evolution of Artificial Intelligence: Analysis of the Three Main Paradigm Shifts in AI, 2020.

[163] Gebaseerd op: Europese Commissie et al., AI Watch, Historical Evolution of Artificial Intelligence: Analysis of the Three Main Paradigm Shifts in AI, 2020.

De tweede opleving volgde rond de jaren tachtig en had ook een geopolitieke context. Gedreven door het dreigingsbeeld dat Japan westerse economieën zou kunnen verdringen, kwamen met name in de VS AI-projecten weer op de agenda. Deze golf kwam ten einde toen de Japanse economie een terugval beleefde. Sinds enkele jaren verkeren we in de derde golf van AI. De financiering is wel verschillend: de derde golf is niet primair in gang gezet dankzij militair-academische allianties maar aangejaagd door bedrijven die (slimme) algoritmen al tijden hebben ontdekt als verdienmodel. Dit betekent overigens niet dat de Amerikaanse overheid niet nauw betrokken is bij geldstromen. Opvallend lijkt Europa niet goed mee te komen in de AI-wedloop die op gang gekomen is.

Bijlage 3. **Ethische en wettelijke kaders voor AI in Europa**

Responsible AI. De Europese Commissie heft de volgende ethische principes als leidend gepresenteerd voor de ontwikkeling en ethische toetsing van AI-toepassingen en -producten:

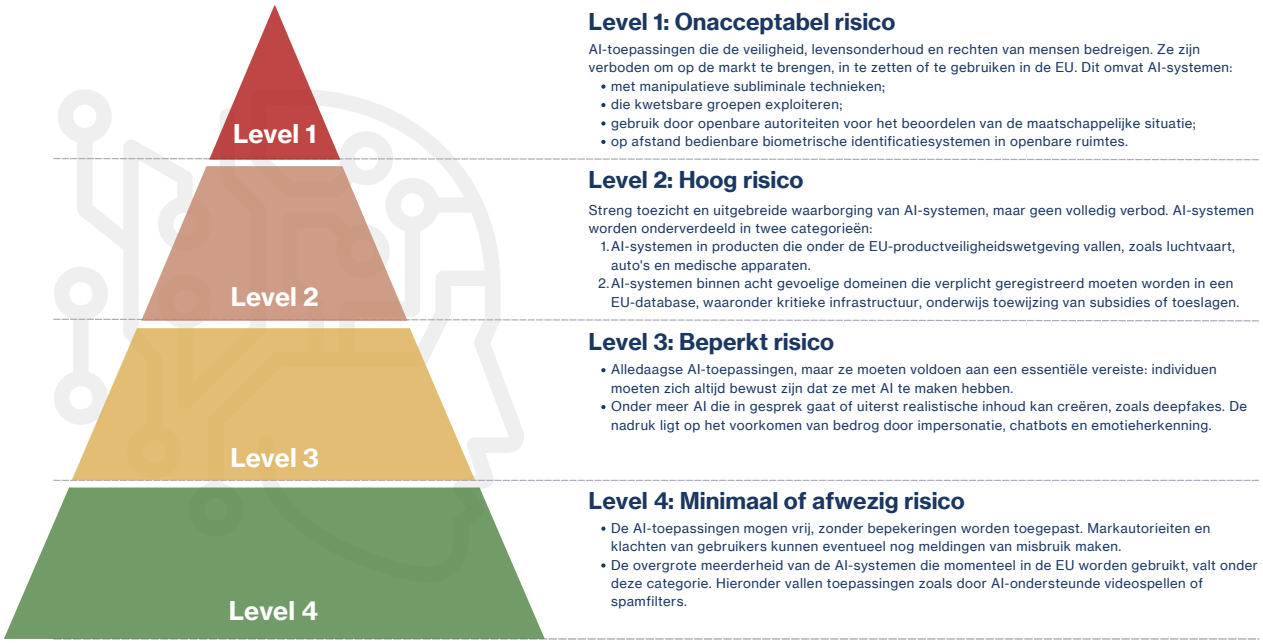
- **Menselijke controle en menselijk toezicht**. Systemen moeten kunnen garanderen dat menselijke controle en toezicht mogelijk is en blijft.
- **Technische robuustheid en veiligheid**. AI-systemen moeten technisch veilig zijn om te voorkomen dat deze systemen onbedoelde of ongewenste resultaten opleveren. Dit omvat "weerbaarheid tegen aanvallen en beveiliging, een uitwijkplan en algemene veiligheid, nauwkeurigheid, betrouwbaarheid en reproduceerbaarheid van data."[164]
- **Privacy en data governance**. Respect voor privacy moet worden aangetoond. AI-systemen moeten zorg dragen voor de integriteit en traceerbaarheid van (persoonlijke) gegevens.
- **Transparantie**. Het is belangrijk dat gebruikers begrijpen hoe AI-systemen werken. Uitkomsten moeten traceerbaar zijn, verklaarbaar en communiceerbaar.
- **Diversiteit, non-discriminatie en rechtvaardigheid**. AI-systemen moeten discriminatie voorkomen. Het ontwerp van applicaties moet zo zijn dat ze geen gebruikersgroepen uitsluiten of benadelen. In het ontwikkelproces zijn ethische stakeholders betrokken en geraadpleegd.
- **Maatschappelijk en milieuwelzijn**. Bijvoorbeeld de toeleveringsketen of het gebruik van AI-systemen moeten worden gecontroleerd. AI-systemen moeten ook negatieve sociale impact voorkomen en bijvoorbeeld geen verslaving in de hand werken. Systemen moeten maatschappelijk en in democratische processen kunnen worden gewogen.
- **Verantwoording**. Het AI-systeem, hoe het werkt en waar componenten van afkomstig zijn, moet onafhankelijk controleerbaar zijn. Ongelukken en negatieve ervaringen moeten worden geregistreerd. Het moet duidelijk zijn wie aansprakelijk is.

De Europese AI Act. Op 9 december 2023 werd een akkoord bereikt over de Europese AI Act. De wet is bedoeld om te komen tot harmonisering van regels voor AI binnen de EU en vormt de leidraad voor nationale wet- en regelgeving. De benadering van AI door de EU is vergelijkbaar met hoe andere innovaties gereguleerd worden: AI wordt feitelijk behandeld als een product dat moet voldoen aan specifieke richtlijnen om toegelaten te worden tot de Europese markt. Een aantal kenmerken:

[164] Kenniscentrum Data & Maatschappij, Ethisch principe 2: technische robuustheid en veiligheid, n.d.

- De wet reguleert AI producten en stelt eisen aan de industrie, inclusief distribiteurs en buitenlandse techbedrijven, met het oog op de bescherming van de gebruiker. Het is geen wet die AI-technologie als zodanig beperkt of in banen leidt, maar wel de AI-toepassingen.
- Er wordt in de wet een voorbehoud gemaakt voor de toepassing van AI-systemen voor aangelegenheden van Europese en nationale veiligheid.
- De wet verbiedt een aantal soorten AI-systemen waaronder discriminerende systemen, social sorting-systemen (denk aan het Chinese sociaal kredietsysteem) of systemen die leiden tot (heimelijke) gedragsmanipulatie.
- De wet maakt, naast de producten die een onacceptabel risico vormen, onderscheid tussen risicocategorieën, zie Figuur 11. De hoog-risicosystemen worden onderverdeeld in twee groepen. De eerste groep beschrijft AI-toepassingen waarvoor een impact assessment verplicht wordt gesteld. Producenten of leveranciers van AI producten zijn verantwoordelijk om zelf een assessment uit te voeren en aan te tonen dat de systemen niet schadelijk zijn ten aanzien van fundamentele rechten. De tweede groep betreft AI-toepassingen die door een externe instantie getoetst moeten worden. Het gaat dan bijvoorbeeld om producten waar extra oplettendheid vereist is omdat ze worden geïmplementeerd in de kritische infrastructuur. Hoe de precieze indeling zal uitwerken is nog niet geheel duidelijk. Een impactstudie naar circa honderd bestaande commerciële AI-systemen wees uit dat 18% duidelijk een hoog risico betrof, 42% een beperkt risico en voor 40% was het onduidelijk of het hoog of beperkt zou moeten zijn.[165]

Figuur 14: Risicoprofielen binnen de AI act (bron: Europese Commissie)



[165] Initiative for Applied Artificial Intelligence, AI Act: Risk Classification of AI Systems from a Practical Perspective, 2023.

- Beperkt-risicosystemen moeten aan transparantievereisten voldoen zodat duidelijk is voor de gebruiker dat er bijvoorbeeld sprake is van een gegenereerd beeld als het gaat om generatieve beeldmanipulatiesoftware.
- Overtredingen, bijvoorbeeld het ontwikkelen van verboden AI-systemen, kan worden bestraft met boetes van maximaal €40 miljoen of 7% van de wereldwijde omzet.

Naast de AI Act zijn er enkele andere relevante wetten die van invloed zijn op AI en die een belangrijke rol spelen om misbruik te voorkomen. Deze wetten zijn echter niet specifiek gericht op AI. Hierdoor ontbreekt soms het specifieke maatwerk dat de AI Act beoogt te ondervangen. Maar ze zijn niettemin van belang voor de bescherming van privacy en burgerrechten, niet in de laatste plaats omdat deze wetten al in een veel verder stadium van implementatie zijn in de Europese landen en hun nationale wetgevingskaders. Bedrijven begrijpen wat wel en niet mag en hebben de regels doorgevoerd in hun compliance-beleid. Zonder de bestaande wetten zou ook de EU als enige leidende partij op het gebied van AI-regulering op dit moment weinig hebben kunnen waarmaken.

De Algemene Verordening Gegevensbescherming (AVG), gebaseerd op de Europese General Data Protection Regulation (GDPR), speelt een cruciale rol bij de regulering van AI-systemen met betrekking tot de verwerking van persoonsgegevens. Ook de AVG is een voorbeeld van geharmoniseerde Europese regelgeving, bedoeld om consistente regels en handhaving te organiseren tussen nationale partijen. De AVG is in principe van kracht voor alle AI-systemen die persoonsgegevens verwerken. Meerdere soorten gegevens- en gegevensdragers vallen hieronder, van geschreven data tot beeldmateriaal en biometrische data of locatiegegevens. De AVG stelt basisvereisten aan gegevensverwerkers. Bijvoorbeeld:

- Gegevens moeten rechtmatig verkregen worden en gebruikers moeten op de hoogte kunnen zijn van het gebruik van AI en hoe gegevens worden verwerkt.
- Er is een verantwoordingsplicht die stelt beheerders van AI-systemen moeten kunnen aantonen te voldoen aan de AVG.
- Verwerkingsprocessen moeten worden geregistreerd en achterhaalbaar zijn.
- De AVG stelt restricties aan geautomatiseerde besluitvormingssystemen zoals bijvoorbeeld voor de toekenning van belastingtoeslagen.
- De AVG stelt restricties voor het verzamelen van bijzondere persoonsgegevens zoals huidskleur, religie, afkomst en dergelijke.
- Specifieke normen voor (uitsluitend) geautomatiseerde besluitvorming.

Algemene Productveiligheid Richtlijn is conform de Europese General Product Safety Directive (GPSD) en regelt allerlei zaken op het gebied van productveiligheidsvereisten en consumentenbescherming. Ook AI-systemen die als product op de markt verschijnen moeten zich aan deze kaders conformeren. Zo zijn in 2022 nog verscherpte richtlijnen vastgesteld en is het voor consumenten eenvoudiger gemaakt AI-producenten aansprakelijk te stellen en een vergoeding te ontvangen voor geleden schade door verkeerde toepassing van AI of gerobotiseerde dienstverlening.[166] Andere zaken die de wet regelt hebben te maken met verplichte informatievoorziening rondom AI producten.

Verordening digitale diensten, bekend als de Digital Services Act (DSA) heeft ook aanzienlijk gezag over AI-instrumenten, vooral wanneer deze worden toegepast in de online dienstverlening en sociale mediaplatforms. De verordening reguleert online diensten die betrekking

[166] Ministerie van Buitenlandse Zaken, Europese Commissie presenteert nieuwe aansprakelijkheidsregels voor producten en artificiële intelligentie, 2022.

hebben op de doorgifte en het opslaan van informatie afkomstig van de gebruikers van die diensten, zoals videodeelplatforms, onlinemarktplaatsen, bloggingwebsites, sociale media en internetaanbieders. De DSA beoogt illegale online-inhoud die via dergelijke 'tussenhandeldiensten' wordt verspreid beter te bestrijden en innovatie te bevorderen, maar zeker zo belangrijk is het streven gebruikers beter te beschermen in hun relatie met de aanbieders van deze diensten."[167]

Verordening digitale markten, of de Digital Markets Act (DMA), is met name gericht op het creëren van een eerlijk speelveld in relatie tot digitale marktpartijen.[168] De wet wijst op basis van criteria ook bedrijven aan die als poortwachters dienen als het gaat om digitale diensten. Deze bedrijven hebben een bijzondere status en kunnen rekenen op extra verplichtingen om hun marktmacht in te perken en consumentenrechten te beschermen. Daarnaast mogen ze zich niet schuldig maken aan concurrentievervalsing en kleinere bedrijven benadelen die afhankelijk zijn van hun infrastructuur. De wet raakt bijvoorbeeld aan vragen over de standaardintegratie van AI-assistenten in browsers van dezelfde dienstverlener.

De ethische en wettelijke kaders vanuit Europa bij elkaar opgeteld biedt stevige sturing en regulering van AI-innovaties in Europa. De inspanningen op dit gebied werpen dan ook hun vruchten af, zeker nu de Europese regulering mondiaal als benchmark begint te gelden. Ook is het een belangrijk signaal: het cowboygedrag van Amerikaanse AI-bedrijven wordt op Europese grond niet getolereerd zonder dat de rechten en vrijheden van burgers zijn gewaarborgd. Tegelijkertijd moet erkend worden dat ook EU-regulering veel *niet* oplost. AI-toepassingen houden zich niet aan grenzen en soevereiniteit. Bovendien lopen Europese landen door regulering het risico de handrem te zetten op de innovatiekracht waardoor ze de slag om AI dreigen te verliezen aan de VS en China.

Een ander punt is de handhaving. Wetgeving werkt normerend en veel bedrijven of personen zullen op den duur gewend raken aan de wettelijke voorwaarden die gelden. Maar op dit moment is nog veel onduidelijkheid over hoe, en hoe intensief precies zal worden toegezien op wettelijke overtredingen. In het kader van AI Act is er veel discussie over de vraag of elk land een enkele of meerdere autoriteiten zou moeten hebben voor toezicht. Veel zal afhangen van de capaciteit die nationale toezichtsautoriteiten krijgen om controle te kunnen uitvoeren. Europese wetgeving lijkt ook meer ontworpen voor het aanpakken van grote niet-Europese bedrijven dan voor het toezicht op kleinere initiatieven binnenshuis die wellicht ook gevaarlijk zijn maar onder de radar blijven of niet duidelijk onder een risicoprofiel onder te brengen zijn.

---

167 Folkert Wilman, De Digital Services Act (DSA): een belangrijke stap naar betere regulering van onlinedienstverlening, 2022.

168 Simoon Hermus, De nieuwe Europese techwet: dit gaat u er vanaf nu van merken, de Volkskrant, 2024.

Bijlage 4. **Tabellen 'maatschappelijke impact op toepassingsniveau'**

In Tabel 6 t/m Tabel 10 hieronder hebben we de maatschappelijke impact van AI geschat voor de verschillende toepassingsgebieden die we identificeerden in hoofdstuk 2. We hanteren daarbij drie niveaus van impact (L = laag urgent, M = middel, H = hoog urgent) op basis van drie (gecombineerde) categorieën, **1) transformatieve kracht, 2) inperking door wet- en regelgeving, 3) maatschappelijke bloot-stelling**. De verschillende wegingen zijn indicatief en gebaseerd op de zienswijze van de auteurs, waar mogelijk ondersteund door voorbeelden en literatuur. De impactweging van AI-toepassingsgebieden (§ 4.3), bestaan uit onderstaande categorieën en zijn als volgt gedefinieerd:

- **Transformatieve kracht**, de mate waarin en wijze waarop de betreffende toepassing het domein gaat veranderen. In de weging is meegenomen hoe AI de status quo voor individuen en groepen nu al verandert en speculaties over welke veranderingen we gaan zien. Bij de weging (H/M/L) is uitgegaan van de te verwachten veranderingen zonder dat er een stevige rem op wordt gezet door regulering. Het achterliggende idee is dat verandering gezond is maar te veel verandering in te korte tijd schadelijk omdat het maatschappelijk evenwicht verstoord raakt.
- **Regulering**, de mate waarin en wijze waarop norm- en regelgeving het betreffende toepassingsgebied in banen leidt c.q. inperkt. Regulering werkt als een rem op de ontwikkeling en kan dus de principiële transformatieve kracht in de praktijk afzwakken; met de kanttekening dat wetten en regels moeten worden gehandhaafd. Sommige sectorale toepassingen zijn nu al omstreden waardoor de kans op inperking groot zal zijn. Naast ethische bezwaren zijn er de bestaande wettelijke kaders waaraan AI toepassingen onderworpen worden, zoals de recent aangenomen Europese AI Act. In de weging zijn de risicoprofielen van de AI Act behulpzaam geweest (zie bijlage 4).
- **Maatschappelijke blootstelling**, de mate waarin en wijze waarop (een groter of kleiner deel van) de bevolking directe of indirecte gevolgen zal ondervinden in het dagelijks leven. Dit nuanceert de transformatieve kracht: een grote disruptie voor een kleine groep kan betekenen dat de samenleving als geheel nauwelijks wordt geraakt. Omgekeerd kan het gebruik van algoritmen door bijvoorbeeld de Belastingdienst vrijwel iedereen raken. De weging is gebaseerd op drie vragen (1) welke dwarsdoorsnede van de bevolking zou getroffen worden; (2) zijn de getroffenen representatief voor de 'gemiddelde Nederlander'; en (3) zijn de effecten direct (= hogere impact) of eerder indirect.

Tabel 6: Impact van AI op toepassingsgebieden in het publieke domein

Publieke domein

Toepassings-gebied	Transformatieve kracht	Inperking door wet- en regelgeving	Maatschappelijke blootstelling
H Door AI ondersteunde beleidsvormingsprocessen	H Afhankelijk van het beleidsniveau. Het aggregeren en analyseren van gegevens om beleid op te baseren gebeurt nu al volop[169] en zal in toenemende mate door machines worden gedaan omdat er sprake is van outperformance van algoritmen boven menselijke analyse.[170]	M impact afhankelijk op welk risiconiveau een specifiek beleidsvormingsproces wordt gescoord. Voor ondersteunende doeleinden die niet raken aan de rechtsbescherming van het individu zijn er weinig tot geen beperkingen. De bestaande wettelijke kaders voor gegevensverzameling zijn wel van kracht.	H Wisselt per beleid en beleidsniveau maar uiteindelijk raak de indirecte invloed van AI iedereen. Op AI gefundeerde keuzes op het gebied van bijvoorbeeld de volksgezondheid kunnen iedereen raken. In algemene zin drijft de digitale bureaucratie processen van 'dataficatie'[171] van het publieke leefdomein en toenemende gevoelens van onpersoonlijkheid.
M Door AI ondersteunde wetgevingsprocessen	M de kans bestaat dat AI indirect wetgevingsprocessen beïnvloedt door zgn. 'microlegislation'. Maar dat AI zelf wetsvoorstellen ontwerpt en indient is nog sterk experimenteel.[172] Wel werd er door D66 al een door AI geschreven motie ingediend in de Tweede Kamer.[173] Er is sprake van experimenten maar nog niet van grootschalige verandering van processen.	L Directe beïnvloeding van politieke processen ligt onder een vergrootglas, maar het is moeilijk te controleren in hoeverre AI achter de schermen meeschrijft. Indirecte effecten die ontstaan door AI gedreven strategische lobby zijn moeilijk te controleren omdat de politieke lobby beperkt gereguleerd wordt. AI-toepassingen op het gebied van rechtsbedeling merkt de WAI aan als hoog risico, maar er is geen regulering voor de toepassing in wetsontwerp.[174]	M De impact van de strategische lobby (waardoor AI indirect 'meeschrijft' aan wetgeving) kan hoog zijn voor veel burgers. Maar deze impact blijft wel indirect. Zeker wanneer bulkwetgeving (zoals de omgevingswet) de norm blijft is er veel ruimte voor incrementele invloed van AI.
H Geautomatiseerde en gerobotiseerde publieke dienstverlening	H Hoewel er ethische reserves zijn, is de verwachting dat eGovernment-toepassingen vanwege de vele schaal- en efficiencyvoordelen zullen toenemen.	H AI systemen die deelname beoordelen voor mensen aan essentiële particuliere of publieke diensten en uitkeringen worden aangemerkt als hoog risico.[175] Een toenemend probleem zal de benodigde capaciteit en kunde vormen die de eerlijkheid en rechtvaardigheid van publieke algoritmen evalueert.	H Raakt iedereen die in contact wil staan met de overheid en heeft daarom een hoge impact voor de brede samenleving.

169 Annefleur van Wanroij, Peiling: Gemeenteambtenaren Gebruiken Volop ChatGPT, Ondanks Risico's, 2023.

170 Daan Kolkman, The Usefulness of Algorithmic Models in Policy Making, 2020

171 Marc Schuilenburg, Making surveillance public, Inaugurele rede Erasmus Universiteit Rotterdam, 2023.

172 Nathan E. Sanders and Bruce Schneier, How AI Could Write Our Laws, 2023.

173 BNNVARA, D66 dient motie in geschreven door AI, 2023.

174 Europese Commissie, Regulatory Framework Proposal on Artificial Intelligence, 2023, p30.

175 AI Act, art. 37. "Een ander gebied waarop het gebruik van AI-systemen bijzondere aandacht verdient, is de toegang tot en het gebruik van bepaalde essentiële particuliere en openbare diensten en uitkeringen die noodzakelijk zijn voor de volledige deelname van personen aan de samenleving of voor het verbeteren van de levensstandaard. In het bijzonder moeten AI-systemen die worden gebruikt om de kredietscore of de kredietwaardigheid van natuurlijke personen te evalueren, worden geclassificeerd als AI-systemen met een hoog risico."

Tabel 6: Impact van AI op toepassingsgebieden in het publieke domein (voortgezet)

Publieke domein

H AI gedreven maatschappelijke monitoring en voorspelling	H Hoewel ethische en juridische grenzen herhaaldelijk zijn overschreden in de publieke sector (bijvoorbeeld het SyRI systeem)[176], lijkt dit geen beletsel te vormen voor verdere ontwikkeling en implementatie van AI-systemen op het gebied van burgerlijke monitoring en surveillance, bijvoorbeeld op het gebied van fraudebestrijding.[177] Sentimentanalyses voor de monitoring van de maatschappelijke stabiliteit of andere publieksdoelen worden toegepast door de overheid.	M Gegevensverzameling moet voldoen aan de richtlijnen van de AVG. Het is in de meeste gevallen verboden om bijzondere persoonsgegevens te verwerken en om gegevenssystemen aan elkaar te koppelen. Met de nodige privacywaarborgen wordt sentimentanalyse echter wel toegepast door overheden.[178] Dat geldt ook voor Risico identificatiesystemen – deze mogen worden toegepast op voorwaarde dat ze aan de AVG voldoen. Ook de AI Act reguleert toepassingen, hoewel er uitzonderingen zijn voor toepassingen rondom nationale veiligheid. Faalkansen nemen echter toe wanneer kennis, toezicht en handhaving achterblijven, zoals in het geval van de Toeslagenaffaire.	H Iedereen heeft hier mee te maken omdat alle burgerlijke gegevens reeds worden gebruikt voor profilering en risicoidentificatiesystemen. Burgers kunnen slachtoffer worden van fouten of van illegale opsporingssystemen. Daarnaast zorgen verkeerd toegepaste identificatiesystemen voor wantrouwen en ontmenselijking van het contact tussen burger en overheid. De ondervonden gevolgen kunnen extreem hoog zijn wanneer monitoringsystemen worden gekoppeld aan sturingsmechanismen of een sociaal krediet-systeem.[179] Maar ook zonder die mechanismen is er al sprake van 'chilling effecten.'
H AI gedreven (campagnevoering) politieke partijen en publieke sector marketing	H De toepassing van algoritmen in publiekstargeting is een feit. Voor AI fundaising bestaan in de VS al gespecialiseerde bedrijven.[180] De kans dat AI greep krijgt op publieke beïnvloeding in verkiezingstijd zien we als hoog omdat politieke partijen hebben laten zien digitale strategieën te hebben waarbij in sommige gevallen ook profielen worden ingekocht van dataminingbedrijven. In Denemarken werd de eerste door AI geleide politieke partij opgericht. Symbolisch maar tekenend voor wat er mogelijk is.[181]	M In Nederland is microtargeting op grond van de AVG verboden voor zover daar bijzondere persoonsgegevens voor verwerkt worden.[182] Er is echter sprake van een groot grijs gebied en veel grote partijen beheren allerlei soorten gedetailleerde gegevens over hun potentiële electoraat.[183] AI gebaseerde microtargeting zal nog sterker gereguleerd gaan worden met de invoering van de Wet op de Politieke Partijen.[184] Digitale beïnvloeding vanuit het buitenland is echter slecht te reguleren en handhaven.	H Verkiezingsbeïnvloeding is een hoog-impactthema omdat het kan leiden tot manipulatie en oneerlijke bevoordeling en daarmee kan raken aan de maatschappelijke stabiliteit. Het ligt daarom onder een vergrootglas maar dit voorkomt niet dat politieke partijen hun digitale strategieën zullen opgeven. Ook buitenlandse inmenging laat zich op social media gelden.
M Toepassing van AI in instrumenten van directe democratische inspraak	M Disruptie is gekoppeld aan de ruimte die wordt ervaren op het vlak van burgerinitiatief. Hoewel de verwachtingen hoog gespannen zijn wordt de toepassing van AI voor democratisering beperkt door de ruimte voor directe politieke inspraak binnen het bestel. Op lokaal niveau zal wel meer ruimte zijn voor initiatief in het kader van de o.a. de participatiewet.	M AI systemen die gebruikt worden voor de bekrachtiging van het particulier initiatief moeten voldoen aan de wet- en regelgeving voor de verwerking van gegevens. Handhaving daarvan is extra ingewikkeld omdat deze initiatieven niet automatisch onder toezicht staan door hun particuliere karakter.	M De potentie is hoog. Bij brede toepassing op landelijke thema's heeft een democratische toepassing van AI een hoog bereik. Tot op heden zijn er opvallend weinig cases.
M Toepassing van AI in ethische of compliance gerichte (zelf-) beoordelingsprocessen	M De volledige vervanging van menselijke beoordelingsprocessen op het gebied van ethiek zal niet snel de norm worden i.v.m. dreigingen op controleverlies. Maar op het gebied van compliance wordt AI al wel ingezet als assistent bij bedrijven. De verwachting is dat algoritmen zo complex worden om ethisch door te lichten dat in de toekomst ook op het gebied van governance gebruik moet worden gemaakt van AI assistentie. Het zal ook normaler worden dat AI systemen hun eigen werkingsprincipes moeten toelichten en evalueren.	H Op dit moment is AI-geassisteerde ethische (zelf-)beoordeling nog een grijs gebied omdat het nog niet breed wordt toegepast. Ethisch gezien levert het buitengewone dreigingen op om AI zichzelf te laten beoordelen of evalueren.[185] De kans dat volledig autonome beoordelingstaken worden verboden of beperkt is dan ook groot. Ook zal de aansprakelijkheid niet naar de beoordelingssystemen zelf kunnen verhuizen waardoor menselijke evaluatie noodzakelijk blijft.	M Wanneer AI andere algoritmen moet controleren ontstaan er grotere problemen op het gebied van transparantie en teruglopende menselijke beschikking over AI architectuur. Vrijwel iedereen kan daar indirect gevolgen van ondervinden maar het betreft hier waarschijnlijk wel dreigingen voor de wat langere termijn.

176 Rechtbank Den Haag, SyRI-wetgeving in strijd met het Europees Verdrag voor de Rechten voor de Mens, 2020.

177 Een voorbeeld op gemeentelijk niveau is de ontwikkeling van een AI bijstandsfraude opsporingsinstrument in de gemeente Nissewaard: Magazine, Gemeente Nissewaard Spoort Bijstandsfraude Op Met AI, n.d.

178 Ministerie van Volksgezondheid, Welzijn en Sport, COM - Uitvoeren sentimentanalyse CBG, n.d.; Een ander voorbeeld van toegepaste sentimentanalyses door de EU is het European Observatory for Online Hate.

179 Zeyi Yang, China Just Announced a New Social Credit Law. Here's What It Means, 2022.

180 BWF, How to Unlock the Power of AI Fundraising: A Complete Guide, 2023.

181 Chloe Xiang, This Danish Political Party Is Led by an AI, 2022.

182 Autoriteit Persoonsgegevens, Brief microtargeting verkiezingen, 2023.

183 Thomas Mulder, Zo proberen politieke partijen jou als kiezer binnen te hengelen... met je eigen data, 2020.

184 Parlement, Wet op de politieke partijen, 2023.

185 Hans de Bruijn, Martijn Warnier, and Marijn Janssen, The Perils and Pitfalls of Explainable AI: Strategies for Explaining Algorithmic Decision-Making, 2022.

Tabel 7: Impact van AI op toepassingsgebieden in het veiligheidsdomein

Veiligheidsdomein

Toepassings-gebied	Transformatieve kracht	Inperking door wet- en regelgeving	Maatschappelijke blootstelling
H AI-tooling in fraude en cybercriminaliteit	H AI-enabled crime maakt de modus operandi van cybercriminelen effectiever en meer 'high-tech' waardoor grotere uitdagingen voor opsporing en criminaliteits-bestrijding staan te wachten.[186] Ook de AI criminele 'dienstverleningssector' is bloeiende. AI-enabled crime verlaagt bovendien de drempel om cyber crime uit te voeren omdat minder technische kennis nodig is dankzij AI-toepassingen die criminele taken automatiseren.	H AI-enabled crime is steeds lastiger aan te pakken zonder hulp van AI ('automated cyber security'[187]). Cybercriminelen hebben het voordeel dat zij niet gebonden zijn aan ethische en juridische kaders waar de opsporingscapaciteit dat wel is. AI processen zullen soms moeilijk te attribueren zijn aan menselijke actoren door 'chained AI'. De scheiding tussen legale en illegale toepassing van AI kan moeilijk te bepalen zijn en er zijn soms nog geen juridische precedenten. Het grijze gebied tussen legale en illegale toepassingen kan crimineel worden geëxploiteerd.	H Er is reeds sprake van een grote toename van slachtofferschap cyber crime en e-fraude.[188] Door AI mogelijk gemaakte mass-targeting zal waarschijnlijk leiden tot verdere toename en verdere verschuiving van het slachtofferschap van traditionele criminaliteit naar dat van digitale criminaliteit.
M AI ondersteuning bij handhaving openbare orde en radicaliserings-/ criminaliteitspreventie	M/H Afhankelijkheid van nieuwe instrumenten verandert de manier waarop menselijke verantwoordelijkheden worden waargenomen en uitgevoerd.	M Privacybescherming is afhankelijk van de wettelijke kaders.[189] Wettelijke bevoegdheden zijn vaak nog te onduidelijk zijn waardoor profiling-systemen zijn ontwikkeld die illegaal blijken.[190] De innovatie kan vooruitlopen op de ethische en juridische toetsing van systemen, wat een risico vormt van ongevalideerde AI-systemen.[191] De verwachting is dat deze spanning tussen bevoegdheden en systeemontwikkeling relevant blijft. Daar komt bij dat de WAI ambiguïteit laat bestaan over de toepassing van AI systemen voor het doel van de nationale veiligheid.	M Massa surveillance toepassingen hebben verschillende directe en indirecte effecten op de bevolking (privacyschending, chilling effecten, vertrouwenscrises) Maar het is de vraag in hoeverre dit soort toepassingen daadwerkelijk wettelijk de ruimte krijgen. Het is echter mogelijk dat meer mensen op een waarschuwingsradar worden geplaatst, afhankelijk van hoe de AI-tools worden geïmplementeerd en door privacyregels worden beperkt.
M AI ondersteuning bij opsporing, inlichtingenvergaring, criminaliteitsbestrijding en (digitale) recherche	H De opsporingspraktijk is stevig aan het veranderen als gevolg van dataverwerking. Een gevolg daarvan is dat de grens tussen opsporings- en inlichtingendiensten aan het vervagen is. Voortdurend loopt de politie daarnaast aan tegen de grenzen van legale toepassingen van digitale middelen. Ook het domein van opsporing verandert rap: criminaliteit verhuist in toenemende mate naar het (semi-)virtuele domein waardoor veiligheidswerk een steeds meer digitaal profiel zal krijgen.	M De politie mag AI vooral gebruiken om forensisch onderzoek te ondersteunen maar is gebonden aan de Politiewet en het Wetboek van Strafvordering die beperkingen opleggen. Juridische kaders zijn echter vaak nog niet toegerust op de vragen die ontstaan rond (niet-) strafvorderlijke datavergaring en verwerking.[192]	M De impact raakt met name de werkwijze van de politie- en inlichtingenfunctie. Het brede publiek kan geraakt worden door mass surveillance toepassingen, bijvoorbeeld in het kader van bulkdata-analyse door inlichtingendiensten.
M Ondersteuning van AI in strafrechtelijke beoordelingsprocessen en rechtsbedeling	H Volgens Amerikaans onderzoek kan 44% van de juridische taken worden overgenomen door AI.[193] Hoewel AI systemen niet snel zullen rechtspreken, kunnen systemen zeer veel juridisch ondersteunend werk overnemen en de dossierdruk verlichten. Voor de rechtspraak is bovendien de kennis over AI van belang om te kunnen oordelen over zaken met een technische dimensie.	H AI systemen op dit gebied worden aangemerkt door de AI act als 'hoog risico'.[194] Robotrechters zijn echter voorlopig nog niet aan de orde. De juridische toepassing van AI leidt echter tot dreigingen omdat AI bias en gehallucineerde data in gerechtelijke dossiers, pleidooien en uitspraken terecht kunnen komen.	L De gemiddelde persoon zal geen directe gevolgen hebben van dit soort besluitvorming. Omdat door AI ondersteunde juridische beslissingen echter referentiepunten zullen worden voor toekomstige jurisprudentie, kan dit ook de manier waarop wetten in de praktijk worden geïmplementeerd beïnvloeden. Er is een potentieel cumulatief effect dat uiteindelijk meer mensen zal treffen.

186 Derek Manky, Threat Predictions for 2024: Chained AI and CaaS Operations Give Attackers More 'Easy' Buttons Than Ever, 2023.

187 Fortinet, How Artificial Intelligence (AI) Can Help With Cybersecurity Threats, n.d.

188 Centraal Bureau voor de Statistiek, Veiligheidsmonitor 2021, 2022.

189 Inlichtingendiensten en politie hebben relatief veel ruimte in het kader van bijzondere wettelijke kaders zoals de WIV en de Tijdelijke wet cyberoperaties.

190 Andreas Kouwenhoven et al., NCTV volgt heimelijk burgers op sociale media, 2021; David Davidson, Politie Stopt Met Gewraakt Algoritme Dat 'Voorspelt' Wie in de Toekomst Geweld Gebruikt, 2023.

191 Atlantic Council, Experts React: The EU Made a Deal on AI Rules. But Can Regulators Move at the Speed of Tech?, 2023.

192 Marianne Hirsch Ballin and Jan-Jaap Oerlemans, Datagedreven Opsporing Verzet de Bakens in Het Toezicht Op Strafvorderlijk Optreden, 2023.

193 Zsuzsa Czobor, Generative AI Could Radically Alter the Practice of Law, 2023.

194 AI Act, art. 40. Bepaalde AI-systemen die bedoeld zijn voor de rechtsbedeling en democratische processen moeten als systemen met een hoog risico worden geclassificeerd gezien hun mogelijk aanzienlijke effecten op de democratie, de rechtsstaat, de individuele vrijheden en het recht op een doeltreffende voorziening in rechte en op een onpartijdig gerecht.

Tabel 8: Impact van AI op toepassingsgebieden in het economische domein

Economisch domein

Toepassings-gebied	Transformatieve kracht	Inperking door wet- en regelgeving	Maatschappelijke blootstelling
M De opkomst van AI-producten en AI product design	M Een extra bron van concurrentie voor menselijke makers die rekening moeten houden met de toegenomen snelheid waarmee AI content kan produceren.	M/H De AI Act stelt dat toepassingen moeten vermelden dat inhoud door AI gegenereerd is. Echter is er nog onduidelijkheid over de relatie tussen auteursrecht en AI-gegenereerde inhoud. Ethisch gezien blijft het de vraag wat er overblijft van het belang van en de waardering voor menselijke bijdragen en creativiteit.	H Men zal steeds meer in contact komen met producten waarbij de ontwikkeling en productie door AI zijn beïnvloed.
M (Verdere) integratie van AI-analytics	M Wanneer toegang tot nieuwe databronnen beschikbaar worden voor bedrijven zullen deze gebruikt worden voor het verder analyseren en begrijpen van mensen en het beïnvloeden van hun besluiten	M Europese consumenten worden, buiten de AI act beschermd door de Digitaledienstenverordening (DSA) en de Digitalemarktenverordening (DMA). Online platforms mogen geen advertenties personaliseren op basis van bijzondere persoonsgegevens zoals geloof of sekse. Het is nog onduidelijk in hoeverre de wet daadwerkelijk effect gaat sorteren omdat veel afhangt van integratie in nationale wetgeving en handhaving.	H Men zal bij elk gebruik van het internet en sociale media in contact worden gebracht met technieken die steeds verder gaan in het analyseren van de data die zij bij het gebruik van deze media genereren
H AI-toepassingen voor het meten en verhogen van arbeidsprestaties.	H Grote bedrijven zoals Amazon maken reeds gebruik van algoritmen om HRM-taken over te nemen en personeel te ontslaan op basis van data, zonder tussenkomst van mensen.[195] Bedrijfscompetitie drijft het efficiencydenken en daarmee ook de behoefte aan dit soort toepassingen. Dit soort algoritmen hebben met name impact op repetitieve arbeid waar output eenvoudig te meten en waarderen is.	H Het toepassen van emotie-tracking op de werkvloer wordt verboden onder de AI act. Het verzamelen van gegevens over iemands output of workflow is echter niet verboden. Dit kan een rol spelen in de beslissing voor contractverlening of -beëindiging, maar zolang de beslissing door een mens wordt gemaakt is dit legaal.	M Een groot aantal menen werkt in sectoren waarbij hun 'output' kwantitatief gemeten kan worden, waardoor het meten van deze output steeds breder ingezet zal worden, en het streven naar het verhogen hiervan steeds uitgebreider zal worden door middel van AI optimalisatie.

[195] Spencer Soper, Fired by Bot at Amazon: 'It's You Against the Machine', 2021.

Tabel 9: Impact van AI op toepassingsgebieden in het onderwijsdomein

Onderwijsdomein

Toepassings-gebied	Transformatieve kracht	Inperking door wet- en regelgeving	Maatschappelijke blootstelling
H AI-toepassingen die gebruikt worden voor het plegen van fraude en plagiaat.	H Fraude en plagiaat vormen nu al een forse uitdaging voor onderwijsinstellingen.[196] AI neemt denk- en zoekwerk uit handen en werkt daarmee luiheid en afleiding in de hand. Een andere zorg van negatieve verandering is dat AI-leerassistentie de impuls wegneemt om de langere weg te bewandelen van zelfbekwaming. Dit is een urgent probleem omdat leer en leesprestaties in Nederland achteruitgaan.[197]	M Regels vanuit de overheid zijn nog niet in de maak omdat de langetermijngevolgen nog niet duidelijk zijn.[198] Het thema staat wel uitgebreid op de radar en er is veel debat over hoe onderwijsinstituties de kansen kunnen bevorderen en de dreigingen kunnen stoppen.	H AI-fraude heeft vermoedelijk het meeste consequenties voor het voortgezet onderwijs, het beroepsonderwijs en het hoger onderwijs. Opleidingen en gediplomeerden ondervinden er last van wanneer fraude uiteindelijk leidt tot inflatie van het niveau en het diploma.[199]
H AI verlegt de focus van competenties in kennisverwerving naar onderzoekende en creatieve competenties	H Het weg bewegen van traditionele kennisoverdracht stimuleert de behoefte aan andere werkvormen zoals projectgestuurd en probleemgestuurd onderwijs. De noodzaak tot adaptatie drijft zo vormen van onderwijsinnovatie. De manier waarop men leert, traint en toetst zal bovendien aangepast moeten worden aan een realiteit van permanent beschikbare informatie. Nieuwe soorten kennis en vaardigheden zullen ook de inhoud beïnvloeden waar bijvoorbeeld meer nadruk komt op mediawijsheid, prompt engineering.	M Veranderingen in didactiek zijn ten dele gebonden aan regels en accreditatieprocedures. Maar in beginsel is er veel speelruimte om te experimenteren met andere leer- en toetsvormen.	H Bijna iedereen zal regelmatig gebruik gaan maken van AI tools voor het opzoeken van informatie, waardoor kennis beschikbaarder wordt maar de vaardigheden om zelf grondige analyse te verrichten en kennis op te doen kunnen afnemen.[200] correlatie is tussen verminderd kennisonderwijs en leerprestaties, is er sprake van een urgent probleem. Er is bijvoorbeeld een negatieve trend als het gaat om kennis over de rechtsstaat en democratie. de leerprestaties gaan in Nederland achteruit.[201]
M AI-toepassingen kunnen de toegankelijkheid van het onderwijs verhogen	M Ook voor het onderwijs geldt dat AI-toepassingen een dual use zullen kennen: enerzijds kunnen ze leiden tot meer maatwerk en speelse, innovatieve leervormen, anderzijds tot uitsluiting.	M Niet alle instellingen zullen over de middelen of capaciteiten beschikken om AI op deze manier te gebruiken, waardoor het egaliserende effect van AI teniet wordt gedaan. De vraag rest of overheden moeten ingrijpen om AI-instrumenten te subsidiëren voor onderwijs.	M Persoonlijke assistentie zal breed toegankelijk worden en daarmee leven lang leren bereikbaar maken. Het is onduidelijk in hoeverre AR of VR in het onderwijs zouden worden geïntegreerd, afgezien van als ondersteunend hulpmiddel.
M Toenemende analyse en observatie van studentengedrag door middel van AI.	L Onderwijsprocessen zouden als zodanig niet veranderen, maar dit zou eerder een impact hebben op de instrumenten waarover leraren beschikken om beslissingen te nemen over de begeleiding van studenten. Algoritmen kunnen discriminerende biases ontwikkelen die schadelijk zijn voor de diversiteit van het onderwijs. Bovendien, wanneer systemen voor louter efficiencydoelen ingezet worden, zoals het verhogen van het opleidingsrendement, raakt de menselijke maat snel uit het oog.	H AI-systemen die worden ingezet om sociale scoring uit te voeren zijn gemarkeerd als hoog risico, en het toepassen van emotion tracking wordt verboden. Echter, het zal in het onderwijs waarschijnlijk niet snel komen tot dit soort toepassingen. Dit neemt niet weg dat dit soort technologie niet problematisch kan zijn vis-a-vis privacy en datavergaring. Studenten en kinderen zijn een kwetsbare groep, specifiek voor het gebruik van technologie voor gezicht- en emotieherkenning.	M Hoewel er allerlei soorten datagedreven leervolgsystemen zijn, zal het zal enige tijd duren voordat dit soort capaciteiten wordt getest en bovendien volledig in onderwijssystemen geïntegreerd.

196 NOS, Hoe AI dit jaar het onderwijs opschudde en de zorg werk uit handen nam, 2023.

197 Stichting Lezen, Leesprestaties Nederlandse middelbare scholieren gaan achteruit, december 2023.

198 Bright RTL, Kabinet: gevolgen AI in onderwijs nog onbekend, 'kansen en bedreigingen', 2023.

199 The Chronicle of Higher Education, AI Means Professors Need to Raise Their Grading Standards, 2023.

200 Marguerita Lane, The Impact of AI on the Labour Market: Is This Time Different?, 2021.

201 Stichting Lezen, Leesprestaties Nederlandse middelbare scholieren gaan achteruit, december 2023.

Tabel 10: Impact van AI op toepassingsgebieden in het sociale domein

Sociale domein

Toepassings-gebied	Transformatieve kracht	Inperking door wet- en regelgeving	Maatschappelijke blootstelling
M Invloed van AI op interpersoonlijke communicatie[202] en sociale vaardigheden	M A In generieke zin maken AI-toepassingen fysiek contact minder nodig,[203] terwijl ze digitale interactie vergemakkelijken en verpersoonlijken.[204] Live vertalingsapplicaties of live taakinstructies zijn allemaal denkbare toepassingen die werk en leven stevig kunnen veranderen. Daarnaast zal AI een deel van het denkwerk in sociale interacties kunnen overnemen, met als gevolg toenemend vertrouwen op AI toepassingen voor deze situaties[205] Ook AI-chatbots hebben impact en er zijn steeds meer gevallen waarin mensen affectieve gevoelens koesteren.	L Ethische en juridische voorwaarden zullen vooral gericht zijn op het afdwingen van transparantie van systemen waarmee gebruikersinformatie wordt uitgewisseld. Chatbots zullen moeten aantonen niet-discriminerend te zijn. Waar mensen schade kunnen leiden door verkeerde informatie zullen aansprakelijkheidsmechanismen van sociale AI-systemen een vereiste worden. Het zal echter in veel gevallen nog onduidelijk zijn of een systeem als hoog risico aan te merken is voor de Wet AI.	M De blootstelling hangt samen met de 'adoption rate' van AI-assistentie. Deze hangt weer samen met het gemak en de prijs van tools en interfaces. Met de huidige generatieve AI-toepassingen zijn communicatie- en vertaalmogelijkheden al beschikbaar. De integratie van AI-chatbots en AI-assistenten in communicatiemiddelen is nu nog geen gemeengoed behalve in de vorm van interactie met de browser en het scherm. De komst van nieuwe, draagbare virtuele tools, bijvoorbeeld in de vorm van AR-brillen of de integratie in HUD's zal de drempel snel verlagen.
M AI ondersteuning op het gebied van coaching en (mentale) gezondheid	M AI heeft de potentie om te ondersteunen in de (mentale) gezondheidszorg. Zo worden chatbots bijvoorbeeld gebruikt als 'mental coaches'.[206] Chatbots kunnen interactief advies verlenen over algemene gezondheidskwesties. Het is echter de vraag of AI specifieke diagnoses kan en mag stellen, zeker wanneer het uitzonderlijke en ernstige gezondheidsproblemen betreft.[207]	M Medische gegevens mogen over het algemeen alleen met expliciete toestemming van de patiënt worden gedeeld tussen zorgverleners. Privacy en het beheer van medische gegevens staat voorop, en misbruik van informatieaanvragen wordt bestraft.[208] De inzet van AI als zelfhulpmiddel of voor zelfdiagnose is in principe niet verboden en is aan te merken als een grijs gebied waarin datavergaring en -distributie ongereguleerd zijn. Er is daarnaast moeilijk op te handhaven omdat het zich vooral in de privé-omgeving afspeelt.	L/M Negatieve ervaring op het gebied van menselijke betekenisgeving in een kunstmatige wereld kan de brede samenleving raken en wellicht een toename van existentiële angst veroorzaken.[209] Zoals bij sociale media is gezien, kwamen de gevolgen van een technologie voor mentale gezondheid pas naar voren na meer dan tien jaar actief gebruik.

202 Eda Erensoy, How AI Is Changing Human Communication, 2021.

203 Pok Man Tang et al., No Person Is an Island: Unpacking the Work and after-Work Consequences of Interacting with Artificial Intelligence, 2023.

204 AIContentfy, The Role of AI in Content Creation for Immersive Technologies, 2023.

205 Jess Hohenstein et al., Artificial Intelligence in Communication Impacts Language and Social Relationships, 2023.

206 CNN, Analysis: Chatbots for mental health care are booming, but there's little proof that they help, 2023.

207 WHO, Artificial intelligence in mental health research: new WHO study on applications and challenges, 2023.

208 Consumentenbond, Medische gegevens delen: wat zijn je rechten?, n.d.

209 Nir Eisikovits, AI Is an Existential Threat--Just Not the Way You Think, 2023.

The Hague Centre
for Strategic Studies

HCSS
Lange Voorhout 1
2514 EA The Hague

Follow us on social media:
@hcssnl

The Hague Centre for Strategic Studies
Email: info@hcss.nl
Website: www.hcss.nl