Verantwoord vooruit

AP-visie op generatieve AI

AP | bescherming in een digitale wereld

Consultatieversie mei 2025

Inhoud

Consultatie

De Autoriteit Persoonsgegevens (AP) nodigt u uit om te reageren op dit document via een e-mail naar **genai-loket@autoriteitpersoonsgegevens.nl**. Dat kan tot en met 27 juni 2025. De reacties die we ontvangen, vatten we samen in één document. We noemen daarbij geen namen, organisaties of contactgegevens. De samenvatting publiceren we op de website van de AP en gebruiken we om de versie te verbeteren. De definitieve visie volgt later in 2025.

Managementsamenvatting

1. **Generatieve AI is een impactvolle technologie die veel zal raken.** We staan aan het begin van een maatschappelijke transformatie. Er liggen grote welvaarts- en welzijnsverhogende kansen in bijvoorbeeld de zorg, het onderwijs, onderzoek en innovatie binnen het bedrijfsleven. Daarbij is het cruciaal om generatieve AI zo in te zetten dat het bijdraagt aan het beschermen en versterken van grondrechten en publieke waarden. Onderdeel daarvan is ook de maatschappijbrede dialoog en consensus over hoe generatieve AI zich verhoudt tot onze grondrechten en publieke waarden.

2. **Waar generatieve AI tot op heden tekortschiet in rechtmatigheid, ziet de AP dat toewerken naar een rechtmatige ontwikkeling en inzet van generatieve AI mogelijk is.** De AP heeft enkele onrechtmatigheden en onzekerheden ten aanzien van de AVG blootgelegd in een juridische analyse. Vervolgens kan de ontwikkeling en inzet van deze technologie worden ingericht om die onrechtmatigheden en onzekerheden te voorkomen. Deze voorwaarden en mogelijkheden in het licht van de AVG bespreken we in het stuk 'AVG-Randvoorwaarden voor generatieve AI.'

3. **Op toepassingsniveau – daar waar generatieve AI door mensen en organisaties in praktische zin wordt gebruikt – geldt een aantal randvoorwaarden voor verantwoorde inzet.** Deze toepassingen moeten een helder doel hebben en worden voor dat doel gebruikt na een gedegen risico-afweging met passende (AVG-) waarborgen. Andere randvoorwaarden zijn bijvoorbeeld voldoende AI-geletterdheid bij de inzet en beheersing van systemen en data.

4. **Op samenlevingsniveau identificeert de AP een aantal uitgangspunten om verantwoorde generatieve AI te kunnen omarmen.** Hieronder vallen Europese digitale autonomie, maatschappelijke weerbaarheid, democratische sturing, een goed functionerende markt voor verantwoorde oplossingen en het vermogen om door de AI-keten heen te corrigeren. Aan al deze uitgangspunten kan actief gewerkt worden door de inzet van maatschappelijke instrumenten zoals het investeren in Europese aanbieders van AI-modellen en het aanjagen van transparantie via regelgeving en publiekelijke modelkaarten. Maar ook door te werken aan vereisten op het gebied van auditeren en risicomanagement voor AI-modellen en de AI-systemen die daarop zijn gebaseerd.

5. **Om te sturen op de verantwoorde inzet van generatieve AI in de samenleving is een breed maatschappelijk en politiek debat van belang.** Keuzes zoals het al dan niet de voorkeur geven aan open-source modellen zijn nu relevant. Dit vereist betrokkenheid van alle lagen van de samenleving: van onderwijs tot ontwikkelaar, van burger tot bestuurder en van consument tot CEO.

6. **Een situatie waarin effectieve regulering bijdraagt aan de ontwikkeling en inzet van verantwoorde generatieve AI-oplossingen is voor de AP het gewenste toekomstscenario.** De AP zet de komende tijd concrete stappen om hier te komen: we openen een generatief AI-loket, we organiseren periodieke bijeenkomsten en gaan in gesprek over de juiste vormen van guidance op het onderwerp. Hiermee willen we bijdragen aan een gedeeld richtinggevend perspectief voor de rol van generatieve AI in de samenleving. De AP draagt in Europees verband bij aan verdere normering van de technologie.

1. Inleiding

Generatieve AI heeft de afgelopen tweeënhalf jaar een periode van snelle ontwikkeling en maatschappelijke adoptie doorgemaakt. De technologie is onderdeel geworden van de digitale basisvoorzieningen van onze samenleving. Dat brengt een hele nieuwe set aan kansen en uitdagingen met zich mee op economisch, technologisch, juridisch en maatschappelijk vlak.

Een impactvolle ontwikkeling zoals de opkomst van generatieve AI verdient de volle aandacht van de samenleving. De AP stuurt op een toekomst waarin digitale technologie in dienst staat van de samenleving en het maatschappelijk belang. Bijvoorbeeld door bij te dragen aan een brede waaier van grondrechten zoals het recht op onderwijs, de vrijheid van ondernemerschap, het recht op sociale zekerheid en sociale bijstand of het recht op gezondheidsbescherming. Een uitgekiende en verantwoorde omarming van generatieve AI maakt het mogelijk om op al deze terreinen stappen te zetten.

Tegelijkertijd maakt de technologische werking van generatieve AI dat grondrechten juist extra waarborging en bescherming verdienen. Denk aan het recht op non-discriminatie en het recht op een eerlijk proces, het recht op eigendom zoals auteursrechten of het recht op de bescherming van persoonsgegevens. Op het gebied van gegevensbescherming schat de AP in dat het overgrote deel van beschikbare foundationmodellen tekortschiet qua rechtmatigheid vanwege het scrapen van (bijzondere) persoonsgegevens. De EDPB heeft vastgesteld dat de inzet van deze foundationmodellen door Nederlandse en Europese partijen niet per definitie onrechtmatig is. De AP acht de rechtmatige ontwikkeling en inzet van generatieve AI mogelijk, maar ziet ook uitdagingen op andere gebieden dan gegevensbescherming.

In deze visie kijkt de AP met een toekomstgerichte blik naar generatieve AI. Wat verstaan wij eronder, welke kansen biedt de technologie ons, met welke risico's gaat het gepaard en wat moet er gebeuren om generatieve AI verantwoord te omarmen? We kijken hierbij vanuit een helikopterperspectief naar de bescherming van onze grondrechten. Dit doen we niet alleen vanuit onze rol als toezichthouder op de AVG en RGR, maar ook als coördinerend AI- en algoritmetoezichthouder en om een bijdrage te leveren aan de voorbereiding van het toezicht op de AI-verordening.

Deze visie is in de eerste plaats gericht op professionals die in hun werk te maken krijgen met generatieve AI. Dit bedoelen we in brede zin: ontwikkelaars die dagelijks beslissingen maken over wat generatieve AI wel en niet kan, maar ook bestuurders en managers die moeten besluiten over de inzet en het gebruik van generatieve AI in hun organisatie.

De AP beoogt met deze visie bij te dragen aan het maatschappelijk debat over generatieve AI en de contouren te schetsen van wat er nodig is voor een veilige en verantwoorde inzet. Daarbij realiseert de AP zich dat technologische ontwikkelingen elkaar razendsnel opvolgen, en de kennis van morgen alweer anders is dan de kennis van vandaag. Deze visie is daarom geenszins een statische analyse, maar moet worden gezien als een uitnodiging om hierover met elkaar in gesprek te gaan. Wij moeten als samenleving vorm geven aan de rol voor generatieve AI.

2. Generatieve AI: beeld van de technologie

Generatieve AI is niet meer weg te denken uit ons dagelijks leven. Maar wat we precies verstaan onder generatieve AI kan per gesprek verschillen, waardoor soms verwarring ontstaat. Dit hoofdstuk belicht hoe generatieve AI in deze visie wordt benaderd en schetst een beeld van de laatste technologische ontwikkelingen.

Wat is generatieve AI?

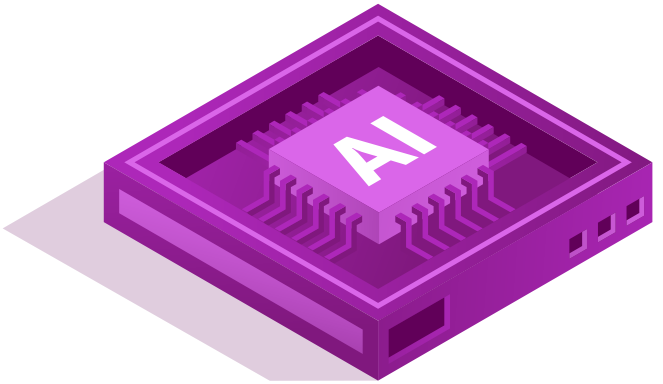
Generatieve artificiële intelligentie (generatieve AI) is een vorm van AI die in staat is om nieuwe data te genereren. De meest populaire toepassingen maken teksten en afbeeldingen waar niet meer aan af is te zien dat deze door AI gegenereerd zijn. Deze visie gaat over AI-modellen die in staat zijn realistische data te genereren en over de systemen en applicaties waar die modellen onderdeel van zijn.

Generatieve AI-modellen kunnen allerlei verschillende vormen van output genereren en daarin gestuurd worden. Hiermee kunnen deze modellen als fundament dienen voor veel verschillende specialistische toepassingen en ingezet worden voor allerlei doeleinden. Dit type modellen wordt vaak foundation models of general purpose AI-models genoemd. In de gekozen benadering vallen modellen met die noemers onder generatieve AI.

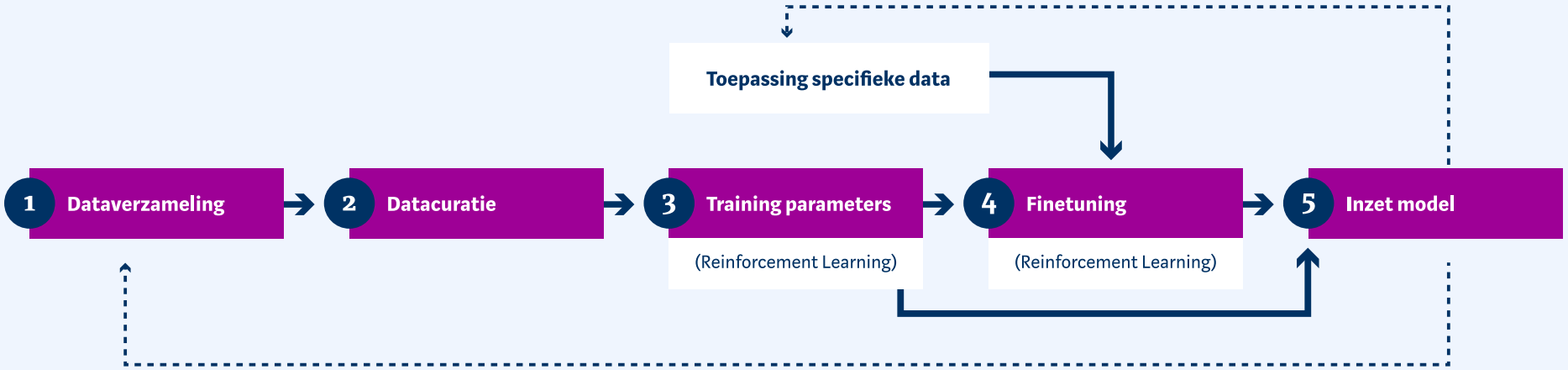
Buiten de scope van deze visie liggen tal van andere vormen van AI. AI-technologie is al lange tijd in ontwikkeling en wordt bijvoorbeeld veel gebruikt voor classificatie. De AI is dan alleen getraind om patronen te herkennen, in plaats van ook te genereren.

Training van een generatief AI-model

Simpel gezegd kan het trainen van een generatief AI-model worden gezien als een set opeenvolgende stappen. Omdat de gegeneerde data in veel gevallen weer wordt verzameld voor training is er sprake van een feedbackloop.



FIGUUR 1 | SCHEMATISCHE WEERGAVE GENERATIEVE AI-KETEN



1. **Dataverzameling.** Voorbeelddata wordt verzameld, vaak door middel van *scraping*.
2. **Datacuratie.** Ongewenste voorbeelden (zoals persoonsgegevens of haatdragende inhoud) worden verwijderd in een curatiestap.
3. **Training parameters.** De parameters worden getraind op basis van patronen in voorbeelddata en mogelijk met aanscherping door *reinforcement learning*.
4. **Finetuning.** Door *finetuning* wordt het model afgestemd op een specifieke toepassing of begrenzing.
5. **Inzet van het model.** De uitkomst van deze stappen is een getraind generatief AI-model dat wordt ingezet in een toepassing.

Technologische ontwikkelingen

De technologie achter generatieve AI is nog volop in ontwikkeling. In onderstaande figuur geven we een overzicht van een aantal belangrijke ontwikkelingen. Dit overzicht is niet compleet, maar geeft een indruk van ontwikkelingen binnen verschillende lagen van de technologie. Daarbij maken we onderscheid tussen ontwikkelingen in de modellen zelf (modellaag), ontwikkelingen met betrekking tot de systemen waarin de modellen worden ingezet (systeemlaag), en de toepassingen waarlangs die systemen contact maken met de omgeving (toepassingslaag).

In alle drie de lagen van de technologie zien we de afgelopen jaren veel ontwikkelingen. Binnen de modellaag, die uiteindelijk de kern van de technologie vormt, experimenteren wetenschappers en ontwikkelaars met verbeteringen in trainingsalgoritmes en aanpassingen op modelarchitectuur.[1][2][3] Doordat veel – ook geavanceerde – modellen openbaar beschikbaar zijn om te downloaden, kan iedereen wereldwijd hierop voortbouwen. Binnen de systeemlaag dragen innovaties bij om de modelcapaciteiten verder te benutten. Het verbinden van een model met een database kan er bijvoorbeeld voor zorgen dat een systeem ook geheugen gebruikt, in plaats van alleen te reageren op de laatste input[4][5]. In de toepassingslaag zien we veel

FIGUUR 2 | SCHEMATISCHE WEERGAVE VAN DE TECHNOLOGIE

	Modellaag	Systeemlaag	Toepassingslaag
Omschrijving	Het getrainde model is een verzameling bestanden waarin de geleerde patronen en de model vorm staan opgeslagen.	Het model wordt ingezet en gedraaid binnen een syteem. Input en output worden verwerkt en mogelijk aangepast.	Het systeem maakt contact met de omgeving via een toepassing. Bijvoorbeeld een chatapp.
Technologische ontwikkelingen	**Model grootte** Het aantal parameters wordt steeds groter.	**RAG architecture** Communicatie met database voor expliciete informatie.	**Multimodale interactie** Naast tekst is ook beeld en geluid interactiemedium.
	Multimodaal Modellen kunnen naast tekst omgaan met beeld en geluid.	**Chain-of-thought** Stapsgewijs genereren door meermaal aanroepen model.	**Ingebed bestaande tools** Interactieve assistentie in tools zoals een code, tekst of afbeelding editor.
	Mixture of experts De vorm stimuleert specifieke expertiseopbouw.	**Local deployment** Het systeem draait op eigen hardware i.p.v. centrale server.	**AI-agents** Autonoom uitvoeren van acties op instructie. Bijvoorbeeld code genereren en deze vervolgens ook draaien.
	Input vorm en grootte Model staat steeds grotere input toe, bijv. context.	**Outputfilters** Filters zodat ongewenste output niet doorkomt.	
	Reinforcement Learning Verbeteren o.b.v. menselijke feedback of zelfreflectie.	**Model Context Protocol** Aanpak om modelinput met context aan te vullen.	

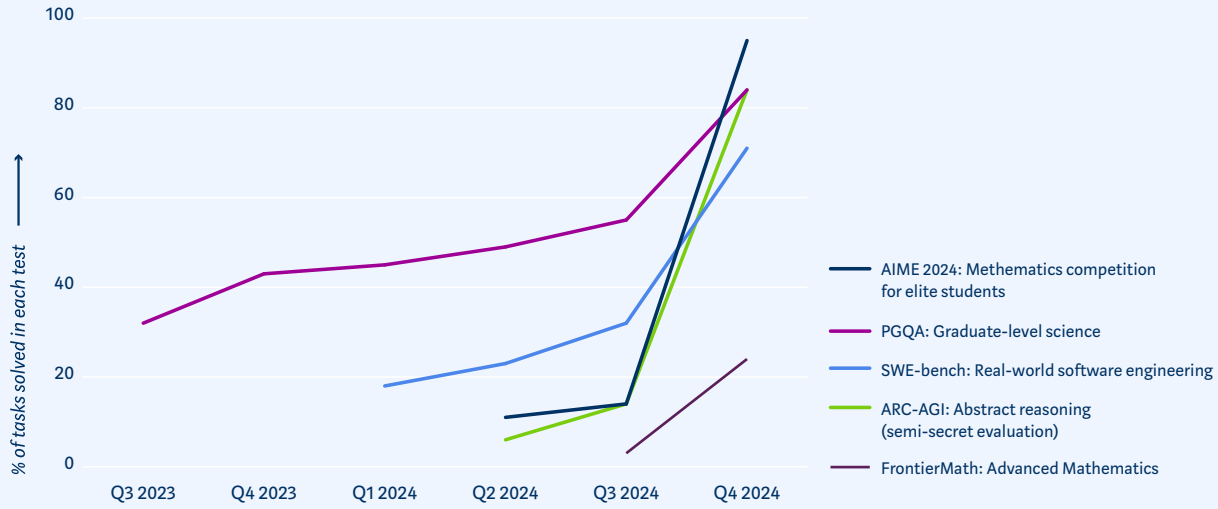
1. Mistral AI (December 2023). "Mixtral of experts"
2. Meta (December 2024). "Large Concept Models: Language Modeling in a Sentence Representation Space"
3. Deepseek (Januari 2025). "DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning"
4. Cohere. "Retrieval Augmented Generation (RAG)"
5. OpenAI (September 2024). "Learning to reason with LLMs"

vernieuwing in de manier waarop de systemen interacteren met de omgeving. Zo is er veel aandacht voor autonome *AI Agents*[6], een type applicatie dat autonoom taken uitvoert, waarover meer in de beschreven trends in hoofdstuk 4.

Kwaliteiten van generatieve AI

Als gevolg van de bovengenoemde ontwikkelingen is de kwaliteit van generatieve AI-output snel toegenomen. Om dit te meten worden modellen en systemen onderworpen aan allerlei tests en benchmarks, bijvoorbeeld met multiple-choice vragen over biologie en natuurkunde[7]. In 2024 leek de vooruitgang in generatieve AI-technologie te vertragen. Een steeds groter deel van de informatie op het internet is lage kwaliteit output van generatieve AI. Dit belemmert het verzamelen van hoge kwaliteit trainingsdata. Sinds december 2024 is er juist weer een versnelling geweest in de toename van kwaliteit. Het toepassen van reinforcement learning is hier belangrijk onderdeel van. Hierbij krijgt het model de vrijheid om antwoordstrategieën te ontwikkelen, en een prikkel om voort te bouwen op strategieën die goede antwoorden opleveren. Een model kan op deze manier bijvoorbeeld leren door terug te kijken op eigen antwoorden en die mogelijk ook te verbeteren[8].

FIGUUR 3 | KEY BENCHMARKS SCORES OVER TIME



BRON: INTERNATIONAL AI SAFETY REPORT (BENGIO ET AL. 2025)

Het is de vraag in welke mate deze ontwikkelingen zich de komende jaren doorvertalen in AI-modellen die tot nog veel meer taken in staat zijn. Er bestaat nog grote onzekerheid over het verder verloop van de kwaliteiten van deze technologie[9]. We kunnen wel aannemen dat, ongeacht de vooruitgang in de modellaag, er op basis van de bestaande modellen veel nieuwe toepassingen bijkomen de komende jaren. Verdere vooruitgang in de modellaag zal dit dus alleen versterken.

6. Anthropic (December 2024). "Building effective agents"
7. Rein, D., Hou, B. L., Stickland, A. C., Petty, J., Pang, R. Y., Dirani, J., ... & Bowman, S. R. (2024). Gpqa: A graduate-level google-proof q&a benchmark. In First Conference on Language Modeling.
8. Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., ... & He, Y. (2025). Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.
9. Bengio, Y., Mindermann, S., Privitera, D., Besiroglu, T., Bommasani, R., Casper, S., ... & Zeng, Y. (2025). International AI Safety Report. *arXiv preprint arXiv:2501.17805*.

3. Inzet: markt & toepassingen

Generatieve AI vindt razendsnel ingang in zowel de private als publieke sector. Ook voor persoonlijke doeleinden maken Nederlanders steeds vaker gebruik van toepassingen die gebaseerd zijn op generatieve AI. De belofte is groot en de verwachtingen hooggespannen, waardoor organisaties over de gehele linie zich voorbereiden op een verdere gebruiks-toename in de nabij toekomst.

De markt voor generatieve AI heeft sinds eind 2022 een explosieve groei doorgemaakt. Met de lancering van ChatGPT door OpenAI is een periode van ongekend snelle maatschappelijke adoptie ingezet waarvan het einde nog niet in zicht is. Dit heeft ook geleid tot een stevige concurrentiestrijd tussen technologiebedrijven wereld-wijd, waardoor de ontwikkeling en inzet van generatieve AI inmiddels nauw verweven is geraakt met geopolitieke ontwikkelingen.

Veel sectoren werken en experimenteren inmiddels met generatieve AI. Het is belangrijk om dit gebruik in kaart te brengen om een goede inschatting te kunnen maken van de risico's en veiligheid van generatieve AI. Modellen worden doorgaans uitvoerig getest voordat ze op de markt worden gebracht, maar het gebruik in de praktijk (en de toepassingen die op de modellen voort-bouwen) kan nooit volledig worden nagebootst – terwijl de risico's en incidenten zich juist daar manifesteren. Het is daarom van belang dat we hier goed zicht op krijgen, zodat we kunnen leren wat er nodig is voor verantwoord gebruik en hoe we de voornaamste risico's kunnen mitigeren

Private sector

Binnen de private sector wordt generatieve AI al ingezet in allerlei bedrijfsprocessen. De private sector loopt voorop als het gaat om de inzet van generatieve AI. *In The State of AI 2024* van McKinsey geeft 65% van de ondervraagde organisaties aan regelmatig gebruik te maken van generatieve AI: een verdubbeling ten opzichte van hetzelfde onderzoek een jaar eerder.[10] Capgemini schetst een vergelijkbaar beeld in haar onderzoek naar het gebruik van generatieve AI in verschillende sectoren: dit is in een jaar in alle domeinen tijd sterk gegroeid.[11] Organisaties zetten generatieve AI voor een grote verscheidenheid aan bedrijfsprocessen in, vaak met als doel om de efficiëntie en productiviteit te vergroten – van IT tot logistiek en van HR tot marketing. Het illustreert de veelzijdigheid van generatieve AI en laat ook zien dat mensen in heel verschillende functies in hun werk te maken zullen krijgen met een vorm van generatieve AI.

Publieke sector

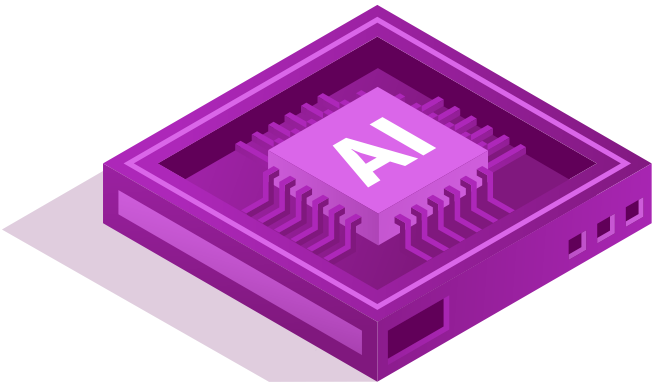
De publieke sector experimenteert met generatieve AI, maar dit staat vaak nog in de kinderschoenen. Ook de publieke sector verkent de mogelijkheden van (genera-tieve) AI. De laatste quickscan van TNO uit juli 2024 laat zien dat in de publieke dienstverlening in allerlei domei-nen steeds meer gebruik wordt gemaakt van AI.[12] Het aandeel generatieve AI was hierin ten tijde van het onderzoek nog laag, maar dat kunnen we deels verklaren door het feit dat veel experimenten nog in de kinder-schoenen stonden. Ook was het individuele gebruik van ambtenaren hier niet in meegenomen. TNO wijst hierbij ook op het risico van 'schaduw-IT': technologie of software

10. McKinsey (Mei 2024). "The state of AI in early 2024: Gen AI adoption spikes and starts to generate value"

11. Capgemini Research Institute (2024). "Generative AI in organizations in 2024"

12. TNO (Juni 2024). "Steeds meer kunstmatige intelligentie ingezet door overheid"

die door individuele medewerkers binnen een organisatie wordt gebruikt zonder formele goedkeuring van de IT-afdeling. Het overheidsbrede standpunt (april 2025) over het gebruik van generatieve AI binnen overheidsorganisaties is in deze context ook van belang: dat stelt dat het gebruik van generatieve AI wordt aangemoedigd, mits hierover eigen afspraken worden gemaakt met leveranciers en dit niet onder consumentenvoorwaarden is.[13] Op basis van dit standpunt is het waarschijnlijk dat het gebruik van generatieve AI in de overheidssector hierdoor de komende tijd toeneemt.



Persoonlijk gebruik

Wat betreft het persoonlijk gebruik van generatieve AI is een generatiekloof zichtbaar. Inmiddels maakt ruim één op de drie Nederlanders regelmatig gebruik van generatieve AI[14]. De verschillen tussen leeftijdsgroepen zijn daarbij groot: zo geeft meer dan 50% van de ondervraagden tussen de 18 en 34 jaar aan dat zij generatieve AI gebruiken, terwijl dit in de leeftijdscategorie 65 tot 75 jaar slechts 10% is. Dit generatieverschil is niet vreemd: jongeren van nu zijn opgegroeid met het internet en hebben daardoor vaak meer vertrouwen om te experimenteren met nieuwe technologieën in hun dagelijks leven[15]. Mensen gebruiken generatieve AI-tools vooral voor het zoeken en verzamelen van informatie en het genereren van ideeën. Een ruime meerderheid van de Nederlanders gelooft dat generatieve AI hun werk de komende twee jaar gemakkelijker (77%) en leuker (77%) zal maken[16].

Praktijkvoorbeelden

In de praktijk heeft de inzet van generatieve AI allerlei verschillende uitingsvormen. De fictieve voorbeelden hieronder zijn gebaseerd op bestaande toepassingen en bedoeld om een indruk te geven van wat er op dit moment mogelijk is met generatieve AI in verschillende contexten, en hoe dit impact heeft op gebruikers en omgeving.

Accountancy

Financieel advies op maat

Accountantskantoor *Loukili & Laheij* heeft flinke stappen gezet sinds de aanschaf van fiscale software op basis van generatieve AI. Standaardrapportages en financiële prognoses kunnen nu in een paar minuten worden gegenereerd.
De accountants van *Loukili & Laheij* hoeven deze alleen nog te valideren, waardoor zij meer tijd kunnen besteden aan persoonlijk contact en financieel advies op maat. De software ondersteunt hen hier ook bij door op basis van de historische klantdata en de klantvraag suggesties te doen voor oplossingsrichtingen. Voor klanten is dit van grote toegevoegde waarde.

De software blijkt ook effectief bij foutdetectie. Tegelijkertijd schuilt hierin ook een risico voor *Loukili & Laheij*: het gevaar is dat medewerkers te veel gaan vertrouwen op de software en output onvoldoende kritisch beoordelen, terwijl dit juist bij jaarlijks wijzigende wetgeving en belastingsystemen van belang is.

13. Rijksoverheid (April 2025). "Het overheidsbrede standpunt voor de inzet van generatieve AI"

14. KPMG (2024). "Algoritme Vertrouwensmonitor 2024"
15. Algosoc (2024), "AI Opinion Monitor"
16. Deloitte (Oktober 2024). "Vertrouwen in generatieve AI: een Europees en Nederlands perspectief"

Huisarts

Chatbot als doktersassistent

Huisartsenpraktijk *ZorgModern* heeft vanaf de start gekozen voor een chatbot als doktersassistent: niet alleen om geld te besparen, maar ook omdat goed personeel lastig te vinden was. De chatbot is nu het eerste contactpunt voor patiënten van de praktijk. Via een app omschrijven patiënten hun klachten. De chatbot stelt een paar vervolgvragen en maakt vervolgens een samenvatting die naar de huisarts wordt gestuurd. De huisarts beoordeelt het bericht en stuurt een persoonlijk antwoord aan de patiënt: een advies of een afspraakverzoek.

Patiënten zijn over het algemeen tevreden met het korte lijntje tussen hen en de arts. Oudere patiënten voelen zich echter minder thuis bij deze praktijk: omdat zij onvoldoende digitaal vaardig zijn voelt de app als een obstakel voor goede zorg.

Privégebruik

Persoonlijke vakantieadviseur

Spanje, Duitsland, Franrijk, of toch een verre reis naar Thailand of Zuid-Afrika? Is het daar wel het goede seizoen voor? En zijn drie weken voldoende om genoeg te kunnen zien? Vragen waarvoor je eerst een paar avonden het internet afstruinde op zoek naar antwoorden, kun je nu in een paar minuten laten beantwoorden door een taalmodel. En met gebruik van de juiste prompts genereer je in een handomdraai een gedetailleerd reisplan dat aan al jouw wensen en eisen voldoet: van aanbevelingen voor hotels en restaurants tot tips voor natuurparken en culturele hoogtepunten. Of informatie over lokale eetgewoonten en noodnummers. En ook tijdens je reis staat de digitale persoonlijke vakantieadviseur altijd voor je klaar.

Onderwijs

Digitale leescoach

De 8-jarige Noor krijgt op basisschool De Ekster elke week les van een gepersonaliseerde digitale leescoach. Noor heeft dyslexie en komt daardoor moeilijk mee in de reguliere lessen over begrijpend lezen. Dat werkte frustrerend en demotiverend. Op de computer kan ze zich beter concentreren en krijgt ze teksten te lezen die speciaal op haar zijn afgestemd. Bijvoorbeeld een tekst over Disney, waar ze groot fan van is. De teksten die haar digitale leescoach genereert sluiten niet alleen aan bij haar interesses, maar bevatten ook specifieke woorden en spellingspatronen die op haar niveau zijn afgestemd. Voor Noor is het lezen zo niet alleen leuker geworden, maar ze gaat ook sneller vooruit.

In de meeste van de hierboven beschreven cases zal een verwerking van persoonsgegevens plaatsvinden. Als er persoonsgegevens worden verwerkt, is de AVG van toepassing en zal de verwerkingsverantwoordelijke over een grondslag moeten beschikken onder de AVG. Bovenstaande persoonsgegevens kunnen mogelijk op basis van toestemming zijn verwerkt. Het beschikken over een grondslag is een van de voorwaarden om generatieve AI verantwoord en rechtmatig in te kunnen zetten. Een overzicht van enkele AVG-randvoorwaarden staat in het guidance-document "AVG-Randvoorwaarden voor Generatieve AI."

4. Trends in generatieve AI: opkomende toepassingen

Generatieve AI heeft de potentie om te zorgen voor een fundamentele verandering in de manier waarop we werken, communiceren en informatie tot ons nemen. Van autonome uitvoerders (AI-agents) tot zoekmachines en nieuwe virtuele sociale actoren waarmee we in contact staan. De technologie biedt op al deze terreinen nieuwe mogelijkheden maar stelt ons ook voor uitdagingen en kan direct en indirect impact hebben op grondrechten en fundamentele waarden.

De impact van generatieve AI zien we in de praktijk en hangt samen met de vorm waarin het wordt ingezet door organisaties en individuen. Daarbij wordt nog volop geëxperimenteerd met nieuwe toepassingsmogelijkheden. In het hoofdstuk 'Beeld van de technologie' hebben we een omschrijving gegeven van de technologische ontwikkelingen op het gebied van generatieve AI. Deze ontwikkelingen liggen ten grondslag aan de toepassingsmogelijkheden in de praktijk. Maar op welke manier brengt generatieve AI nu verandering te weeg? Dat hangt af van de wijze waarop de technologie wordt ingezet. We behandelen hier een aantal trends in toepassingsvormen die we zien met betrekking tot generatieve AI.

Generatieve AI als autonome uitvoerder (AI-agents): het idee om AI-modellen zelfstandig acties te laten uitvoeren om taken te automatiseren is niet nieuw, maar generatieve AI schept hiervoor nieuwe mogelijkheden. Ten eerste kunnen opdrachten die in menselijke taal zijn omschreven beter worden geïnterpreteerd door grote taalmodellen. Doelen worden daarbij in brokjes opgedeeld en stapsgewijs uitgevoerd. Daarnaast kunnen deze modellen instructies genereren voor allerlei verschillende instrumenten: bijvoorbeeld om de muis en het toetsenbord van computers virtueel te bedienen[17]. Deze verbeteringen samen zorgen ervoor dat het veld dichter bij autonome uitvoerders (AI-agents) komt. Een enkele AI-ontwikkelaar stelt dat er eind 2027 modellen zijn die bijna alles kunnen wat de meeste mensen kunnen[18]. Deze autonome uitvoerders kunnen altijd aanstaan en op de achtergrond werken, waardoor ze zo goed als onzichtbaar zijn. Voor bedrijven kunnen autonome uitvoerders een uitkomst bieden, maar de inzet heeft mogelijk onder meer ook invloed op de werkgelegenheid.

Generatieve AI als 24/7 persoonlijke assistent: de integratie van generatieve AI in bestaande software en de combinatie van efficiëntere modellen met sterkere mobiele hardware en grotere internetbandbreedte zal het mogelijk maken om een (lokale) generatieve AI als assistent te gebruiken die continu aanstaat. Als mensen doorlopend aantekeningen willen laten maken gedurende de dag zou de AI-assistent continu mee moeten luisteren. Hierdoor zou het veel persoonlijke details mee kunnen krijgen, ook van gesprekspartners die geen besef hebben dat de assistent aanstaat of daar geen controle over uit kunnen oefenen.

Generatieve AI als sociale actor: de output die een generatieve AI genereert lijkt heel erg op iets dat door mensen is gemaakt. Je kunt dus het gevoel hebben dat je menselijk contact hebt als je interacteert met een virtuele (chat)bot. Dit geeft AI de kans om op een nieuw vlak te opereren. Sommigen ervaren het als virtuele vriend of geliefde in zogenaamde *companion apps*, of als digitale therapeut die helpt bij mentale problemen. Ook kan een generatieve AI als leraar een lesprogramma personaliseren voor een leerling of simpelweg helpen om een sollicitatie te oefenen, door een sociale setting te stimuleren.

17. Anthropic. "Computer use (beta)",
18. ArsTechnica (Januari 2025). "Anthropic chief says AI could surpass "almost all humans at almost everything" shortly after 2027"

Tenslotte worden chatbots steeds vaker gebruikt om front office-achtige taken over te nemen, zoals interactie met klanten en het maken van afspraken.

Generatieve AI als hulpmiddel voor optimalisatie en efficiëntie in organisaties: een van de beloftes van generatieve AI is dat deze routinematige cognitieve taken van werknemers over kan nemen. Bijvoorbeeld door (online) vergaderingen te transcriberen en samen te vatten, of door klantvragen over producten te beantwoorden. Dit zou potentieel de productiviteit kunnen verhogen, maar het is nog maar de vraag of dit daadwerkelijk zo is. Het uitvoeren van routinematige taken door de medewerker kan juist verlichting bieden waardoor cognitief zware taken, die nog niet geautomatiseerd (kunnen) worden, weer uitgevoerd kunnen worden. De routinematige taken dragen dus bij aan de totale productiviteit: als deze wegvallen is het niet evident dat meer tijd voor zware taken ook zorgt voor meer of betere uitvoering daarvan[19].

Generatieve AI als onderzoeker: generatieve AI draagt bij aan het bedenken, controleren en uitwerken van ideeën voor onderzoek. Hierdoor ontstaat mogelijk een versnelling in onderzoeksbevindingen in tal van richtingen, bijvoorbeeld in de natuurkunde, wiskunde en biologie. Generatieve AI is bijzonder efficiënt in het vinden van patronen en verbanden en kan tot op zekere hoogte eigen suggesties nakijken en herzien. Dit maakt het uitermate geschikt voor academische disciplines die verder bouwen op voorgaande bronnen. Tegelijkertijd valt niet per definitie te verwachten dat generatieve AI een paradigmaverschuiving veroorzaken, waar die juist nodig kan zijn voor grote wetenschappelijke doorbraken.

Generatieve AI als basis voor beeld en geluid: steeds meer online content wordt gegenereerd door generatieve AI-modellen en -toepassingen. Aangezien deze steeds beter worden, is deze content vaak moeilijk van echt te onderscheiden[20]. Een logisch gevolg hiervan is dat mensen steeds minder vertrouwen hebben in de echtheid van digitale informatie. Ook maakt het aanbieden van openbaar beschikbare modellen het relatief eenvoudig om misbruik te maken van de technologie en inhoud te genereren die misleidend is, zonder dat er direct bewijs voor is dat het om gegenereerde inhoud gaat. Want alle "nieuwe" inhoud die wordt gegenereerd, reproduceert de voorbeelddata waar het mee getraind is. Dat wil zeggen dat bestaande vooroordelen die in de voorbeelddata zitten ook terug zullen komen in de inhoud die gegenereerd wordt door de AI. Deze vooroordelen worden daardoor verder bestendigd in de maatschappij.

Generatieve AI als zoekmachine: generatieve AI wordt vaak ingezet als kennisbron. Dit is op het terrein van de meer traditionele zoekmachine, die op basis van zoektermen een lijst teruggeeft van relevante bestaande bronnen. Een belangrijk verschil is dat generatieve AI ook de inhoud van openbare bronnen verwerkt, of heeft verwerkt tijdens de training. Op basis daarvan wordt nieuwe informatie gegenereerd om de zoekopdracht te beantwoorden. Dit maakt het voor de gebruiker makkelijk om informatie te krijgen, maar in dat gemak schuilt ook een risico. Er is geen garantie dat die gegenereerde informatie correct is, en door de simpele presentatie is dit voor mensen minder intuïtief in te schatten of te achterhalen. Mede om deze reden zien we recent een ontwikkeling waarbij generatieve AI-zoekmachines meer expliciet verwijzen naar bestaande bronnen. Tegelijkertijd bewegen de traditionele zoekmachines ook richting generatieve AI, waarbij bovenaan de resultaten een gegenereerde samenvatting verschijnt van gevonden informatie.

19. TNO (Januari 2025). "Technologieradar gezond en veilig werken"

20. Miller, E. J., Steward, B. A., Witkower, Z., Sutherland, C. A., Krumhuber, E. G., & Dawel, A. (2023). AI hyperrealism: Why AI faces are perceived as more real than human ones. Psychological Science, 34(12), 1390-1403.

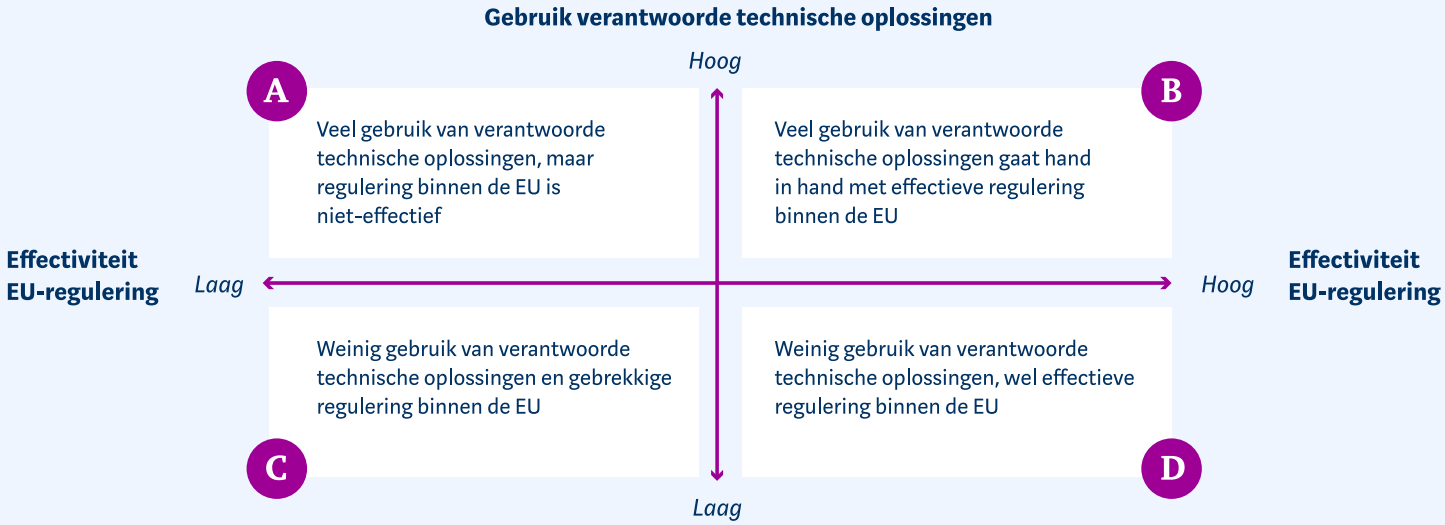
5. Toekomstscenario's: Generatieve AI in 2030

Met een technologie die zich zo snel ontwikkelt, is het moeilijk om te voorspellen hoe de toekomst eruit zal zien, zelfs op relatief korte termijn. Dit hoofdstuk verkent vier toekomstscenario's voor de wijze waarop generatieve AI in 2030 ingebed kan zijn in onze samenleving.

De scenario's zijn opgesteld langs twee assen van kernonzekerheden op het gebied van regulering en technologie. De eerste kernonzekerheid gaat over het gebruik van verantwoorde technologische oplossingen. Lukt het onderzoekers om technologische oplossingen te creëren om de risico's die gepaard gaan met generatieve AI te beperken? En weten deze oplossingen zich succesvol naar de markt te bewegen? De tweede kernonzekerheid betreft de effectiviteit van regulering op de Europese markt. De komende jaren moet blijken of de recent aangenomen Europese wetten om (generatieve) AI in goede banen te leiden in de praktijk doeltreffend zijn, ook in combinatie met bestaande wetgeving.

Door de twee assen te combineren komen we tot vier scenario's die ieder een ander toekomstbeeld voorstellen. Hierna schetsen we op hoofdlijnen het toekomstbeeld per scenario met het jaar 2030 als horizon.

FIGUUR 4 | TOEKOMSTSCENARIO'S GENERATIEVE AI



A

Scenario A: "Het recht van de sterkste"
Veel gebruik van verantwoorde technische oplossingen, maar regulering binnen de EU is niet effectief

Er is een bloeiende markt voor verantwoorde generatieve AI-oplossingen. Dit komt niet zozeer voort uit wet- en regelgeving, maar door de aanhoudende vraag vanuit de markt. Dit is het directe gevolg van een toegenomen bewustzijn en kennis over de kansen en risico's van generatieve AI in de samenleving. Na enkele incidenten in de voorgaande jaren is men zich bewust geworden van het belang van technologie die de democratische waarden onderschrijft en de grondrechten van mensen beschermt. Het feit dat er veel gebruik wordt gemaakt van verantwoorde oplossingen is een sterke bottom-up beweging geweest, waarbij de ontwikkeling van Europese alternatieven vooral is voortgekomen uit het idealisme van enkele ondernemers en private financiers. Het zijn vooral bewuste burgers die hiervan profiteren. Tegelijkertijd komen er nog met regelmaat AI-gerelateerde incidenten voor: het gebruik van generatieve AI is snel toegenomen, maar door achterblijvende handhavingscapaciteit zijn burgers vogelvrij wanneer hun rechten worden geschonden door een AI-systeem. Burgers zijn afhankelijk van de welwillendheid van bedrijven en kunnen de rechten die zij op papier hebben niet afdwingen. De verantwoorde oplossingen blijken daardoor steeds vaker een wassen neus: als het toch misgaat, is onrecht moeilijk te herstellen.

B

Scenario B: "Waarden aan het werk"
Veel gebruik van verantwoorde technische oplossingen, met effectieve regulering binnen de EU.

Bedrijven, organisaties en consumenten maken volop gebruik van waardengedreven generatieve AI-toepassingen. De Europese regelgeving die bij de introductie enkele jaren geleden nog vele wenkbrauwen deed fronzen, heeft haar vruchten afgeworpen: mede door forse Europese investeringen is er een bloeiend AI-ecosysteem ontstaan. Kunstmatige intelligentie is een vast onderdeel van het curriculum van Europese lagere en middelbare scholen. Ook heeft interdisciplinair wetenschappelijk onderzoek naar AI een *boost* gekregen, waardoor het aantal patentaanvragen op AI-gedreven innovaties de afgelopen jaren sterk is gestegen. Generatieve AI-applicaties met persoonsgegevens kunnen AVG-*compliant* gebruikt worden, omdat deze gegevens weer verwijderd kunnen worden uit modellen. Toezichthouders spelen in dit alles een belangrijke faciliterende rol: onder meer door intensieve samenwerking op nationaal en Europees niveau is er een internationaal concurrerend *level playing field* gecreëerd, waarbij veel aandacht is voor harmonisatie en constante uitwisseling plaatsvindt tussen toezichthouders en de praktijk. Verantwoorde generatieve AI-toepassingen zijn hierdoor de norm – en best beschikbare producten – geworden. Heldere standaarden en frameworks maken ook dat risico's tijdig worden opgespoord en gemitigeerd. Er heerst een sterk lerende cultuur: incidenten worden systematisch benut om systemen en applicaties te verbeteren, waardoor het vertrouwen in generatieve AI in de samenleving groot is en burgers hun rechten weten uit te oefenen.

C

Scenario C: "Onder de radar"
Weinig gebruik van verantwoorde technische oplossingen, regulering binnen de EU is niet effectief

De regulering die tot doel had om van Europa een waardengedreven AI-continent te maken, heeft haar doel voorbijgeschoten. De wet- en regelgeving bleek in de praktijk te complex en publieke investeringen bleven achter, waardoor veel beloftevolle initiatieven voor de ontwikkeling van verantwoorde generatieve AI in de vroege ontwikkelfase strandden. Tegelijkertijd is het gebruik van generatieve AI niet afgenomen: er zijn nog altijd vele toepassingen op de markt waarvan niet duidelijk is of ze *compliant* zijn door tegenstrijdige wet- en regelgeving, maar door het gebrek aan effectieve controle en handhaving veelvuldig worden ingezet.
Dit leidt tot slecht functionerende systemen die veelal onder de radar blijven: door een gebrek aan transparantie en kennis weten burgers vaak niet dat er sprake is van onrechtmatige behandeling door een AI-systeem, of het lukt hen niet om foutieve persoonsgegevens uit modellen te laten verwijderen. Aanbieders leggen censurerende restricties op aan hun taalmodellen, waardoor burgers selectieve informatie voorgeschoteld krijgen over allerlei onderwerpen. Het lukt toezichthouders niet om burgers hiertegen te beschermen: het toezicht is versnipperd en de kenniskloof tussen beleidsmakers, toezichthouders en AI-ontwikkelaars is enorm. Het ontbreekt aan een gemeenschappelijke taal, waardoor burgers zijn overgeleverd aan de grillen van de markt en het vertrouwen in generatieve AI laag is. Incidenten lijken vooral kwetsbare groepen in de samenleving te raken: de toegenomen kloof tussen digitaal laag- en hooggeletterden is hier mede debet aan. De weerstand onder burgers groeit en de roep om menselijk contact klinkt almaar luider.

D

Scenario D: "Dirigent met een blinddoek"
Weinig gebruik van verantwoorde technische oplossingen, met effectieve regulering binnen de EU

De mismatch tussen regulering en marktontwikkelingen zorgt voor een lastige Europese situatie. De regulering van generatieve AI is effectief in de zin dat grondrechten en publieke waarden goed beschermd zijn: er zijn heldere standaarden, burgers worden op een duidelijke manier geïnformeerd wanneer er een generatief AI-systeem wordt ingezet en door robuuste samenwerking tussen toezichthouders is er goed zicht op risico's en blijven de gevolgen van incidenten voor burgers relatief beperkt. Tegelijkertijd is er weinig contact tussen toezichthouders en marktpartijen: wet- en regelgeving is steeds meer losgezongen geraakt van de wereldwijde technologische en economische ontwikkelingen. De relatief kleine markt die de EU vormt voor grote marktpartijen maakt dat zij hun innovaties in eerste instantie richten op markten elders. Bedrijven en organisaties zijn daarbij terughoudend met het inzetten van generatieve AI, uit angst voor boetes bij incidenten. Tegelijkertijd verliezen sectorale markten waar Europa oorspronkelijk marktleider was hun internationale positie doordat hun innovatiemogelijkheden beperkt worden. Het gat tussen de EU en andere landen groeit, tot frustratie van ondernemers en burgers die zien wat er buitenland wél mogelijk is. De belofte van digitale strategische autonomie blijkt moeilijk te worden ingelost. Dit heeft niet alleen economische consequenties voor de positie van Europa op het wereldtoneel, maar dit maakt ook dat onze verworven vrijheden en grondrechten steeds meer onder druk komen te staan.

Gewenste richting

Van de omschreven toekomstscenario's ziet de AP scenario B 'Waarden aan het werk' als het meest gewenste scenario. Hierin is regulering effectief en zorgen we samen dat waardengedreven innovatie tot stand komt. In het afsluitende hoofdstuk over verantwoorde inzet wordt dit gewenst toekomstperspectief verder uitgewerkt: Wat is er voor nodig om hier te komen en hoe draagt de AP hieraan bij?

Box 1: Impact van openbare modeltoegang op risico's

Voor de toekomstige ontwikkeling en inzet van generatieve AI is het al dan niet open-source beschikbaar komen van geavanceerde modellen een belangrijke en bepalende factor. Meer specifiek gaat het dan over of de zogenaamde modelparameters beschikbaar komen, oftewel *open-weight* modellen. Een open-weight model is vrij te downloaden op het internet, en met complete vrijheid in te zetten of aan te passen. Er zijn al honderdduizenden open-weight generatieve AI-modellen beschikbaar[21]. Het volgende overzicht benoemt een aantal relevante implicaties voor risico's als gevolg van het wel of niet openbaar maken van de modellen. Dit is een versimpelde weergave[22].

Uit deze vergelijking blijkt dat het beschikbaar maken van open-weight modellen andere risico's met zich meebrengt dan closed-weight modellen. Er komen risico's bij, terwijl andere risico's gemitigeerd worden. Op dit moment zijn beide varianten in omloop. Als maatschappij zullen we dan ook rekening moeten houden met de risico's van beide varianten.

Volgens de AP is dit het moment om risico's te beïnvloeden door het openbaar maken van modellen aan te moedigen of juist tegen te gaan. Het bepalen van de gewenste richting hierin vraagt om brede maatschappelijke discussie.

21. Hugging Face. "Models"
22. Deze weergave voorziet in een eerste indruk, maar de impact op risico's is onvolledig en ligt vaak gecompliceerder. Zo zijn er bijvoorbeeld diensten om closed-weight modellen in een gesloten omgeving te draaien en te finetunen, waardoor hybride oplossingen ontstaan met weer een ander risicobeeld.

FIGUUR 5 | VERGELIJKING OPEN-WEIGHT MODELLEN VS. CLOSED-WEIGHT MODELLEN

Risicocategorie →	*Open-weight modellen*		*Closed-weight modellen*
Privacy en gegevensbescherming	Eventuele persoonsgegevens in modellen raken wijdverspreid	⚠ ↔ ✓	Aanbieder kan modellen terugroepen/corrigeren
	Lokaal draaien mogelijk, minder verkeer van gegevens	✓ ↔ ⚠	Veel input en output verkeer naar aanbieder
Cybersecurity	GenAI cyber verdediging en aanvallen voor iedereen mogelijk	↔	GenAI cyber verdediging en aanvallen op basis van toegang
	Risico op backdoors, malware of verborgen acties in modellen	⚠ ↔ ✓	Aanbieder heeft controle over de veiligheid van het model
	Hele community kan bijdragen aan modelcontrole en testen	✓ ↔ ⚠	Communicatie tussen systemen introduceert kwetsbaarheden
Bias, stereotypering & discriminatie	Hele community kan bijdragen aan model alignment	✓ ↔ ⚠	Aanbieder heeft grote invloed op gebruikte normen en waarden
	Er ontstaan veel modellen met bias zonder aanspreekpunt	⚠ ↔ ✓	Aanbieder kan aangesproken worden op bias
Gebruik met slechte intenties	Het wordt makkelijker om misinformatie te genereren	⚠ ↔ ✓	Aanbieder kan toegang ontzeggen bij misbruik
	Krachtige AI kan gebruikt worden veel schade aan te richten	⚠	
Autonomie en marktmacht	Onafhankelijkheid groter, makkelijker toetreden markt	✓ ↔ ⚠	Afhankelijkheid van aanbieder die groeit door gebruik
	Macht en controle op applicatieniveau	✓ ↔ ⚠	Aanbieders krijgen machtspositie over applicaties
	Iedereen kan bijdragen aan ontwikkelingen; Meer divers en minder gestuurd	↔	Aanbieders bepalen richting van ontwikkelingen, mogelijk niet gewenste richting maar wel een aanspreekpunt

⚠ Hoog risico
✓ Laag risico

Box 2: Generatieve AI en de AVG

Wanneer generatieve AI-modellen worden getraind op (ongericht) *gescrapete* data, bevatten de datasets voor het trainen van die modellen dikwijls persoonsgegevens. Dat maakt dat de AVG van toepassing kan zijn vanaf de eerste stap in de generatieve AI-keten: de dataverzameling. Maar ook later in de generatieve AI-keten kan de AVG van toepassing zijn, bijvoorbeeld bij het finetunen van een foundation model of bij het implementeren of gebruik van een generatieve AI-toepassing. In onderstaand schema worden een aantal AVG-randvoorwaarden uiteengezet voor verantwoorde ontwikkeling en gebruik van generatieve AI-toepassingen waar die toepassingen de verwerking van persoonsgegevens betreffen. Deze randvoorwaarden zijn verder uitgewerkt in het document 'AVG-Randvoorwaarden voor generatieve AI'.

Uit deze randvoorwaarden kunnen we concluderen dat voldoen aan AVG-vereisten nog wel de nodige aandacht vraagt van verwerkingsverantwoordelijken. Het is aannemelijk dat bijzondere persoonsgegevens onrechtmatig zijn gescrapet voor het trainen van foundation modellen. Ook is het niet vanzelfsprekend dat scraping van persoonsgegevens altijd is toegestaan. Ontwikkelaars, finetuners en gebruikers van generatieve AI moeten de uitdagingen rondom het inwilligen van rechten van betrokkenen herkennen en hier bewust mee omgaan. **Deze normering is het standpunt van de AP op het moment van publicatie.** De AP volgt echter de Europese normering ten aanzien van generatieve AI op de voet en draagt hier actief aan bij. Dat betekent dat deze randvoorwaarden in een later stadium nog aangepast kunnen worden.

FIGUUR 6 | AVG-RANDVOORWAARDEN GENERATIEVE AI

Trainingsdata voor het trainen en finetunen van een model moet rechtmatig zijn verkregen.

Strengere randvoorwaarden voor de verzameling van bijzondere persoonsgegevens.

Trainingsdata moet rechtmatig en zorgvuldig gecureerd worden en zo goed als mogelijk worden ontdaan van (ongewenste) persoonsgegevens.

Verwerkingsverantwoordelijken hebben een systeem ingesteld om rechten van betrokkenen te faciliteren.

Doeleinden voor het meetrainen van persoonsgegevens in generatieve AI-modellen en voor het verwerken van persoonsgegevens in generatieve AI-toepassingen moeten voorafgaand worden bepaald en gecommuniceerd aan betrokkenen.

Generatieve AI-toepassingen genereren zo weinig mogelijk foutieve of ongewenste persoonsgegevens.

6. Juridisch kader en instrumenten

Generatieve AI is een technologische ontwikkeling die dwars door alle sectoren heengaat en daarom niet slechts onder één wet te scharen valt. Verschillende wettelijke kaders zijn van toepassing op de ontwikkeling en het gebruik van generatieve AI-modellen en toepassingen. Generatieve AI valt dus niet in een juridisch vacuüm, maar moet zich conformeren aan zowel technologie-specifieke regelgeving als domeinspecifieke wetgeving.

Afhankelijk van de sector waarin het generatieve AI-model wordt ontwikkeld of ingezet, kan verschillende wet- en regelgeving van toepassing zijn. Denk bijvoorbeeld aan financiële wetgeving, consumentenrecht, mediawetgeving, wetgeving voor de gezondheidszorg, bestuursrecht en auteursrechtenbescherming.

Nieuwe Europese regelgeving

Naast deze bestaande wet- en regelgeving is er recentelijk veel nieuwe Europese regelgeving ontwikkeld om de inzet en het gebruik van nieuwe technologieën te reguleren. Dat betreft in de eerste plaats de inwerkingtreding van de AVG (2016) ter bescherming van de persoonsgegevens van burgers, maar daarnaast zijn er de afgelopen jaren diverse wetten aangenomen om onder meer digitale platformen te reguleren (DSA), marktdominantie door Big Tech bedrijven tegen te gaan (DMA), cybersecurity te waarborgen (NIS2, CSR), eerlijke toegang tot data en de bescherming daarvan te waarborgen (DA, DGA), en te zorgen dat AI op een verantwoorde manier wordt ontwikkeld en ingezet (AI-verordening). Ook staan er nog verschillende andere wetgevingsinitiatieven op de Europese agenda die mogelijk ook invloed gaan hebben op generatieve AI, zoals de *Digital Fairness Act*.

Grondrechten van burgers

Verder raakt de ontwikkeling, de inzet en het gebruik van generatieve AI ook de grondrechten van burgers[23]. Grondrechten gaan enerzijds uit van een verticale werking. Dat wil zeggen dat zij burgers beschermen tegen overheidsoptreden. Anderzijds kennen grondrechten ook een horizontale werking. Dat houdt in dat burgers een beroep kunnen doen op bepaalde grondrechten ten opzichte van andere burgers (en bedrijven). Generatieve AI kan een bijdrage leveren aan de bevordering van grondrechten, maar kan ook risico's met zich meebrengen voor de bescherming van grondrechten.

23. Handvest van de Grondrechten van de Europese Unie. Beschikbaar via: <https://www.europarl.europa.eu/charter/pdf/text_nl.pdf>.

Risico's voor grondrechten

Generatieve AI-modellen en -toepassingen kunnen een inbreuk maken op grondrechten van burgers als ontwikkelaars en gebruikers deze grondrechten onvoldoende in acht nemen. Denk bijvoorbeeld aan het meetrainen van persoonsgegevens in een generatief AI-model of de verwerking daarvan in generatieve AI-toepassingen. Als de bescherming van persoonsgegevens niet genoeg gewaarborgd wordt, kan dat mogelijk een inbreuk vormen op het recht op bescherming van persoonsgegevens. Ook kan het ontwikkelen en gebruiken van generatieve AI-modellen en -toepassingen invloed hebben op het recht op bescherming van het intellectuele eigendomsrecht, bijvoorbeeld bij het genereren van afbeeldingen of video's via generatieve AI.

Grondrechten versterken

Generatieve AI kan bepaalde grondrechten juist ook versterken. Als generatieve AI-toepassingen baanbrekende veranderingen teweegbrengen binnen de gezondheidszorg of het onderwijs, kan dat mogelijk leiden tot een betere toegankelijkheid voor burgers binnen die sectoren. Op die manier zorgt generatieve AI indirect voor een betere inkleuring van het recht op onderwijs en het recht op gezondheidszorg.

Samen met internationale (niet-bindende) beleidsafspraken vormt deze wet- en regelgeving de kaders voor de verantwoorde ontwikkeling, inzet en het gebruik van generatieve AI[24]. De figuur op de volgende pagina dient als inspiratiebron voor de beheersing van de ontwikkeling, inzet en het gebruik van generatieve AI, zonder daarbij volledig te willen zijn.

Rechtmatigheid van dataverzameling

Bij de lancering van de eerste grote generatieve AI-applicaties zijn veel vragen opgekomen over de rechtmatigheid van de meegetrainde datasets voor grote generatieve AI-modellen. De dataverzameling voor deze grote generatieve AI-modellen vindt doorgaans namelijk plaats via ongerichte scraping. Bij het scrapen van informatie van het internet kunnen ook bijzondere persoonsgegevens worden gescrapet. De AVG schrijft een verwerkingsverbod voor op deze bijzondere persoonsgegevens, tenzij een beroep kan worden gedaan op een uitzonderingsgrond[25]. Omdat er doorgaans geen directe relatie bestaat tussen de scraper en de betrokkene, kan er vaak geen beroep worden gedaan op de uitzonderingsgrond toestemming. Er kan sprake zijn van kennelijke openbaarmaking van de bijzondere persoonsgegevens als deze door de betrokkene zelf op internet zijn geplaatst[26]. Soms kan het echter ook voorkomen dat iemand anders bijzondere persoonsgegevens van een betrokkene op het internet plaatst. Het valt dus niet uit te sluiten dat in beperkte gevallen ook bijzondere persoonsgegevens op onrechtmatige wijze in generatieve AI-modellen terecht zijn gekomen.

De AP neemt in acht dat het scrapen van deze bijzondere persoonsgegevens mogelijk een secundair karakter kent ten opzichte van het onrechtmatig online plaatsen ervan. Verder betekenen onrechtmatigheden in een vroeg stadium van de generatieve AI-keten niet altijd dat de inzet van die modellen door andere partijen ook onrechtmatig is. We zijn ons echter bewust van de risico's die het verzamelen van bijzondere persoonsgegevens voor het trainen van generatieve AI-modellen met zich meebrengt. Daarom richt de AP zich in dit visiestuk op de verantwoorde inzet van generatieve AI in Nederland. Verder draagt de AP in Europees verband bij aan verdere normering van generatieve AI. Op die manier leveren we een bijdrage aan het maatschappelijk verantwoorde teweegbrengen van generatieve AI-toepassingen.

24. Voorbeelden van internationale afspraken die raken aan het verantwoord inzetten van generatieve AI zijn OECD Principles on Artificial Intelligence(2019), UNESCO *Recommendation on the Ethics of Artificial Intelligence* (2021) en de *EU Code of Practice on Disinformation* (2022).
25. Zie artikel 9 van de AVG.
26. Zie voor een verdere uitleg over de kennelijke openbaarmaking bij scraping de Handreiking scraping door particulieren en private organisaties van de AP.

FIGUUR 7 | BEHEERSINGSKADER VERANTWOORDE GENERATIEVE AI

	GenAI model	GenAI systeem	GenAI toepassing
Risicomanagement	• Impact-assessments (bv. DPIA, FRIA) (AVG, AIV) • Security by design/secure development lifecycle (NIS2, ISO 27001) • Neutrale data intermediairs (DGA)	• Impact-assessments (bv. DPIA, FRIA) • Gegevenskwaliteitsbeoordeling (AVG, AIV) • Conformiteitsbeoordeling (AIV)	• Impact-assessments (bv. DPIA, FRIA) • Transparantieverplichtingen (bv. deepfakes) (AIV)
Technische beheersmaatregelen	• Modelkaart en technische documentatie (AIV) • Datakwaliteits- en interoperabiliteitseisen (DA) • Pseudonimisering/anonimisering (AVG) • Fairness & bias detectie	• Fairness & bias detectie • Privacy Enhancing Technologies (bv. differential privacy) (AVG) • Interoperabiliteit (DA) • Technische documentatie (AIV)	• Fairness & bias detectie • Monitoring & logging • Explainability & interpretability
Organisatorische beheersmaatregelen	• Privacy by design, Privacy by default (AVG) • Aanwijzen EU vertegenwoordiger (AIV)	• Informatiebeveiligingsplan • AI-governance • Privacy by design, Privacy by default (AVG)	• Training en bewustwording (AI-geletterdheid) (AIV) • Menselijke tussenkomst (AVG) • Privacy by design, Privacy by default (AVG)
Juridische beheersmaatregelen	• Register van gegevensbemiddelingsdiensten (DGA) • Regulatory sandbox (AIV) • Voorafgaande raadpleging (AVG) • Bescherming van intellectueel eigendomsrecht, bedrijfsgeheimen en vertrouwelijke bedrijfsinformatie (AIV)	• Verwerkingsregister (AVG) • Algoritmische audits • Onafhankelijke audits (DSA)	• Incident reportage (bv. datalek) (AVG) • Post-market monitoring (AIV) • Meldpunten/klachtmechanismen (DSA, AVG, AIV)

7. Waarden aan het werk: richting de verantwoorde inzet van generatieve AI

De AP spant zich in voor een toekomst waarin de samenleving de vruchten kan plukken van verantwoorde vormen van generatieve AI op basis van effectieve regulering (toekomstscenario B). In dit hoofdstuk werken we uit hoe in dit toekomstbeeld fundamentele waarden en grondrechten zijn beschermd en de technologie in dienst staat van de samenleving en het maatschappelijke belang.

Dit toekomstbeeld kenmerkt zich enerzijds door de volgende overkoepelende criteria voor het ecosysteem van generatieve AI (op systeemniveau): Europese digitale autonomie, maatschappelijke weerbaarheid, democratische sturing, een functionerende markt voor verantwoorde oplossingen en het vermogen om door de AI-keten heen te corrigeren. Anderzijds voldoen individuele AI-toepassingen in dit toekomstbeeld aan de volgende kenmerken (op toepassingsniveau): inzichtelijk en transparant, de risico's in kaart en afgewogen, een heldere doelomschrijving, systemen en data in gecontroleerd beheer, rechtmatigheid.

In de voorgaande hoofdstukken is vanuit een helikopterview gekeken naar generatieve AI. We bespraken wat de technologie is (hoofdstuk 2), hoe deze wordt gebruikt (hoofdstuk 3) en welke trends op dit moment spelen (hoofdstuk 4). Daarnaast zijn op basis van scenario's mogelijke toekomstbeelden geschetst (hoofdstuk 5). In dit hoofdstuk werken we uit hoe in het gewenste toekomstbeeld ("waarden aan het werk",

FIGUUR 8 | WAARDEN AAN HET WERK: UITGANGSPUNT VOOR VERANTWOORDE INZET

Kenmerken van een situatie waarin fundamentele waarden en grondrechten zijn beschermd en de technologie voor generatieve AI in dienst staat van de samenleving en het maatschappelijke belang

Systeemkenmerken

Waardengedreven generatieve AI is een motor van innovatie door...

- Europese digitale autonomie
- Maatschappelijke weerbaarheid
- Democratische sturing
- Goed functionerende markt voor verantwoorde oplossingen
- Vermogen om te corrigeren door de AI-keten heen

Toepassingskenmerken

Verantwoorde inzet van generatieve AI is mogelijk door...

- Inzichtelijkheid en transparantie
- Risico's in kaart, afgewogen en gemitigeerd
- Heldere doelomschrijving
- Systemen en data in gecontroleerd beheer
- Rechtmatigheid

scenario B) fundamentele waarden en grondrechten zijn beschermd en de technologie in dienst staat van de samenleving en het maatschappelijke belang.

We beschrijven de vormgeving van scenario B ("waarden aan het werk") aan de hand van kenmerken. Deze kenmerken maken sturing mogelijk. Want, door deze te realiseren, creëren we de maatschappelijke basis om verantwoord vooruit te kunnen met generatieve AI en het welvaarts- en welzijnsverhogende potentieel te benutten. We kijken daarbij eerst naar kenmerken op het (maatschappelijk) systeemniveau en vervolgens naar kenmerken op toepassingsniveau. De visie sluit af met acties die de AP neemt om aan de gewenste koers bij te dragen.

Kenmerken op systeemniveau: de rol van generatieve AI in de samenleving

In het toekomstbeeld "waarden aan het werk" vormt waardengedreven generatieve AI de motor voor innovatie. Een aantal systeemkenmerken draagt hieraan bij:

- **Systeemkenmerk 1: Europese digitale autonomie**

  Het nastreven van dit kenmerk zorgt dat de EU in staat is te sturen en in te grijpen op de manier waarop de technologie de samenleving raakt. Hierbij wordt meegenomen dat een grote afhankelijkheid van een organisatie de mogelijkheid op correctie moeilijker kan maken. Binnen de EU worden activiteiten getoetst aan Europese regels, onafhankelijk van druk op de regulering van buiten de unie. In het Draghi-rapport wordt een mate van autonomie en concurrentievermogen als fundament beschreven om grondrechten te kunnen blijven beschermen in de toekomst[27].

  Mogelijk instrument: stimuleren van opkomst Europese aanbieders generatieve AI.

- **Systeemkenmerk 2: Maatschappelijke weerbaarheid**

  Het nastreven van dit kenmerk zorgt dat mensen in de gehele samenleving in de basis weten hoe generatieve AI werkt en daardoor realistische verwachtingen van de technologie hebben. Ook kunnen mensen kritisch kijken naar de ontwikkeling en inzet ervan[28]. Dit stelt individuen in staat om gebruiksrisico's in te schatten, maar ook te doorzien hoe aspecten in het dagelijks leven beïnvloed worden door generatieve AI, zoals onze informatievoorziening.

  Mogelijke instrumenten: werken aan hoge AI-geletterdheid door maatschappij en periodieke externe controle van generatieve AI-systemen.

- **Systeemkenmerk 3: Democratische sturing**

  Het nastreven van dit kenmerk waarborgt dat via het democratisch proces de wensen en zorgen van het brede publiek effectief opgenomen worden in beleid[29]. Brede maatschappelijke discussie onderbouwt hierbij de richting, bijvoorbeeld op de vraag of open-weight modellen de voorkeur hebben boven closed-weight modellen. Publieksbijdragen kunnen ook bestaan uit testen of benchmarks voor generatieve AI, om vanuit diverse perspectieven bij te dragen aan beheerste impact.

  Mogelijke instrumenten: werken aan adequate expertise over AI-systemen binnen volksvertegenwoordiging, laagdrempelige platforms voor specialistische benchmarks, aanjagen van maatschappelijke dialoog over generatieve AI.

- **Systeemkenmerk 4: Een goed functionerende markt voor verantwoorde oplossingen**

  Het nastreven van dit kenmerk zorgt ervoor dat de samenleving niet afhankelijk is van één of enkele partijen. In plaats daarvan bestaat er een markt met meerdere opties door de hele keten, van het trainen tot toepassen van generatieve AI. De regulering creëert een *level playing field* waarin verantwoorde oplossingen niet afdoen aan kwaliteit. Daarnaast is er sprake van interoperabiliteit tussen toepassingen:

27. Europese Commissie (September 2024). "The Draghi report: In-depth analysis and recommendations"
28. AP (Januari 2025). "Aan de slag met AI-geletterdheid"
29. AP (Maart 2025). "Position paper AP democratische beheersing algoritmes en AI"

het is voor gebruikers eenvoudig om tussen vergelijkbare toepassingen te wisselen. Dit betekent ook dat de gebruikersdata of het opgebouwde profiel van gebruikers op zo'n manier beschikbaar is dat een 'lock-in' positie voorkomen wordt.

Mogelijke instrumenten: het nastreven van een gevarieerd aanbod van beschikbare modellen, stimuleren van interoperabiliteit tussen systemen, aanjagen van vergelijkingsplatforms.

- **Systeemkenmerk 5: Het vermogen om te corrigeren door de AI-keten heen**

Het nastreven van dit kenmerk zorgt dat de inbedding van generatieve AI-modellen in digitale systemen zo is opgezet dat kwetsbaarheden of onrechtmatigheden achteraf gecorrigeerd kunnen worden. Als een model persoonsgegevens bevat of ongewenste uitkomsten produceert, heeft iedere partij in de keten een systeem in werking gesteld om die persoonsgegevens of ongewenste uitkomsten te corrigeren.

Mogelijke instrumenten: correctie-methoden zoals machine unlearning, registratie van modelversies binnen hoog-risico-toepassingen.

Kenmerken op toepassingsniveau

Er zijn verschillende kenmerken van een verantwoorde inzet van generatieve AI op toepassingsniveau. Dit gaat dus over hoe een organisatie in het toekomstbeeld de inzet van individuele systemen benadert die gebruikmaken van generatieve AI:

- **Toepassingskenmerk 1: Inzichtelijk en transparant**

Het nastreven van dit kenmerk waarborgt dat generatieve AI-toepassingen duidelijk herkenbaar zijn en zich lenen zich voor verdere analyse waar gewenst. Bijvoorbeeld door op aanvraag (uitkomsten van) diverse beoordelingscriteria te delen. Verregaande mate van transparantie vormt de basis voor vertrouwen en stelt de gebruiker in staat zich weerbaar op te stellen en zijn rechten uit te oefenen.

Mogelijke instrumenten: waarborgen dat gebruikers zien dat ze te maken hebben met AI, het beschikbaar stellen van publiekelijke modelkaarten[30].

- **Toepassingskenmerk 2: Risico's in kaart, afgewogen en gemitigeerd**

Het nastreven van dit kenmerk waarborgt dat een gedegen risicoafweging voorafgaat aan de inzet van generatieve AI-toepassingen. Hierin worden juridische en ethische afwegingen expliciet gemaakt en gedocumenteerd. De toepassing bevat geen onacceptabele risico's en is transparant over verhoogde risico's. Diversiteit van perspectieven is in deze analyse en afweging essentieel.

Mogelijke instrumenten: het inrichten van risicomanagementraamwerken voor AI-modellen, monitoring van de risico's van AI-modellen in de praktijk via steekproeven.

- **Toepassingskenmerk 3: Heldere doelomschrijving**

Het nastreven van dit kenmerk waarborgt dat het doel en gebruik van generatieve AI zo precies mogelijk wordt geformuleerd en gedocumenteerd. Dit is belangrijk omdat generatieve AI inherent voor vele doelen kan worden ingezet. De toepassing wordt ingericht om het gebruik binnen de geplande kaders te houden. Na ingebruikname wordt dit gemonitord. De werknemers die betrokken zijn bij de inzet van het systeem zijn goed op de hoogte van de werking van het systeem en de mogelijke risico's en weten hoe ze deze in de praktijk kunnen herkennen.

- **Toepassingskenmerk 4: Systemen en data in gecontroleerd beheer**

Het nastreven van dit kenmerk waarborgt dat er controle is over de omgeving waar het generatieve AI-systeem draait en de data die verwerkt worden. Hierbij is de afhankelijkheid van derde partijen beperkt voor belangrijke processen. Daarnaast is het

30. Hugging Face. "Model Cards"

helder en transparant waar de gebruikersdata zich bevinden en voor wie die beschikbaar zijn.

Mogelijke instrumenten: gebruik van open-weight modellen op eigen hardware, gebruik van private cloudoplossingen.

- **Toepassingskenmerk 5: Rechtmatigheid**

  Het nastreven van dit kenmerk waarborgt dat wordt voldaan aan de wettelijke vereisten van de relevante wetgeving, waaronder de AVG en de AI-verordening en eventuele sectorale wetgeving. Dit geldt voor het hele proces, van het verzamelen van data tot aan het inzetten van de toepassing. Modellen van derde partijen hebben een acceptabel niveau van anonimiteit, zoals beschreven door de EDPB[31].

Inzet vanuit de AP

De AP gaat zich komende periode inzetten om een bijdrage te leveren aan het gewenste toekomstbeeld 'Waarden aan het werk'. Dit doen we enerzijds door extra aandacht te hebben voor generatieve AI binnen onze bestaande werkzaamheden en anderzijds door een aantal nieuwe activiteiten te starten die een positieve bijdrage leveren aan de verantwoorde inzet van generatieve AI in onze samenleving.

Vanuit het reguliere werk levert de AP een bijdrage aan heldere en realistische werkmethoden, onder andere door actief mee te schrijven aan standaarden en opinies. Bijvoorbeeld de EDPB-opinie uit december en de standaarden voor hoog-risico-systemen uit de AI-verordening. Daarnaast werkt de AP aan digitale weerbaarheid, onder andere door te investeren in het kennisniveau van de samenleving. Bijvoorbeeld door het bieden van guidance rondom AI-geletterdheid, begrijpelijke informatie op onze website en het organiseren van seminars. Ook risicosignalering vormt een integraal onderdeel van de werkzaamheden van de AP. Bijvoorbeeld via het overzicht aan AVG-risico's voor generatieve AI en de halfjaarlijkse risicorapportages over algoritmes en AI. Als laatste is de AP beschikbaar voor voorafgaande raadpleging en richt het een sandboxtraject in volgens de AI-verordening.

De AP zet ook een aantal extra stappen om verantwoord vooruit te gaan met generatieve AI. Organisaties kunnen ons bereiken op een nieuw loket voor vragen over generatieve AI. We beginnen met het in kaart brengen welke vragen en uitdagingen er liggen rondom de verantwoorde ontwikkeling en inzet van generatieve AI. Hiervoor organiseren we een aantal bijeenkomsten en start de AP een online loket voor vragen en ideeën over generatieve AI. De informatie die we hiermee ophalen geeft inzicht in bijvoorbeeld de compliance-vragen en issues die 'top of mind' zijn binnen organisaties. Dit vormt de basis voor verdere periodieke dialoog over de verantwoorde ontwikkeling en inzet van generatieve AI. En het helpt om de focus te zetten op die vraagstukken die het meest urgent zijn.

Ook gaan we aan de slag met concrete instrumenten en handvatten die bijdragen aan het beschermen van fundamentele waarden bij de ontwikkeling en inzet van generatieve AI. Denk bijvoorbeeld aan Europese richtlijnen, het stimuleren van methodes voor het anonimiseren of verwijderen van persoonsgegevens uit modellen, handreikingen voor AI-geletterdheid en uitgangspunten voor transparantie. En samen met andere toezichthouders in het digitale domein zetten we ons daarnaast in voor gezamenlijke normuitleg, zodat we zoveel mogelijk duidelijkheid en daarmee rechtszekerheid creëren voor organisaties die aan de slag willen met generatieve AI. Positieve voorbeelden en use cases geven we een podium om daarmee te laten zien en te inspireren hoe verantwoorde generatieve AI er in de praktijk uit kan zien. Vanzelfsprekend blijft de AP ook toezichthouden op onrechtmatigheden die zich voordoen in het speelveld van generatieve AI.

Zo maken we samen de verantwoorde inzet van generatieve AI mogelijk.

31. European Data Protection Board (December 2024). "Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models"

Bijlage 1. Begrippenlijst

Hieronder volgt een korte (niet-juridische) omschrijving van de gehanteerde termen en begrippen:

Begrip	Omschrijving
AI *(Artificial Intelligence, Artificiële Intelligentie, Kunstmatige Intelligentie)*	Het nabootsen van menselijke vaardigheden met een computersysteem, zoals leren, plannen, redeneren, anticiperen en zelfstandig beslissen zonder tussenkomst van menselijke intelligentie.
AI-model	Een softwarecomponent waarin kennis ligt besloten om op basis van invoer tot uitkomsten (zoals voorspellingen of classificaties) te komen. Modellen leren patronen uit voorbeeld data (voorbeeld invoer met een correcte uitkomst) en verwerken nieuwe invoer op basis hiervan tot een zo goed mogelijk uitkomst.
AI-systeem	Een systeem dat is ontworpen om met verschillende niveaus van autonomie te werken en dat gebruikmaakt van een AI-model.
Generatieve AI	Een vorm van AI die in staat is om nieuwe data te genereren. De meest populaire toepassingen maken teksten en afbeeldingen waar niet meer aan af is te zien dat deze door AI gegenereerd zijn. Een invoer van een gebruiker geeft vaak aan wat te genereren.
Machine learning	Methodes om patronen uit data op te nemen in een AI-model.
Neuraal netwerk	Een type AI-model waarin een abstractie van informatie kan worden opgeslagen in de vorm van parameters. Deze parameters zijn getallen die worden gekozen door te trainen op voorbeelddata. Een neuraal netwerk bepaalt op basis van een input en de parameters een output.
Deep learning	Een methode binnen machine learning. Deze methode maakt gebruikt van neurale netwerken met veel parameters.
Open-source model	Een AI-model dat openbaar beschikbaar is. In dat geval is het bijvoorbeeld mogelijk de besliscriteria van een beslisboom of de parameters van een neuraal netwerk te downloaden.

Begrip	Omschrijving
Closed-source model	Een AI-model dat alleen beschikbaar wordt gesteld op een manier waarbij je output terugkrijgt op basis van de input. De exacte berekening van de beslissing is dus niet openbaar.
Foundation model *(ook wel: General Purpose model)*	Een AI-model dat op basis van veel data patronen heeft opgeslagen in grote neurale netwerken. Het model kan hierdoor worden ingezet in veel verschillende toepassingen. Daarnaast kan het model als 'fundament' dienen voor modellen voor specifieke doeleinden, waarbij het foundation model gespecialiseerd wordt.
Large Language Model (LLM)	Een AI-model dat op basis van veel tekstdata patronen heeft opgeslagen in grote neurale netwerken. Het model kan tekst genereren en kan bijvoorbeeld worden toegepast in vraag-antwoord-situaties.