Ministerie van Binnenlandse Zaken
en Koninkrijksrelaties

Overheidsbreed standpunt voor de inzet van generatieve AI

Werken aan de basis. Van Nederland en ons Koninkrijk

Overheidsbreed standpunt voor de inzet van generatieve AI

De overheid wil generatieve AI in het voordeel van haar burgers en medewerkers inzetten. Deze technologie biedt volop kansen om maatschappelijke vraagstukken aan te pakken en de dienstverlening te verbeteren. Denk bijvoorbeeld aan het vereenvoudigen van administratieve processen, het versterken van data-analyse en het verbeteren van communicatie. Generatieve AI is een technologie die de overheid op tal van terreinen kansen biedt. Tegelijkertijd is de inzet ervan door overheden niet zonder risico's. Om als één overheid de kansen van deze technologie optimaal te benutten, is een duidelijk standpunt met betrekking tot de inzet ervan essentieel. Dit standpunt biedt duidelijkheid over de manier waarop overheidsorganisaties generatieve AI op een waardevolle, verantwoorde en ethische wijze kunnen inzetten. Het biedt een kader waarbinnen overheidsorganisaties kunnen handelen, met ruimte voor innovatie en aandacht voor de risico's. Hierbij is rekening gehouden met de huidige stand van de technologie. Dit standpunt kan worden gewijzigd wanneer nieuwe inzichten daarom vragen.

Dit standpunt is tot stand gekomen door actieve samenwerking tussen alle bestuurslagen, onder coördinatie van het ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK). Om overheidsorganisaties te ondersteunen bij de concrete doorvertaling van dit standpunt naar de uitvoeringspraktijk, is een handreiking opgesteld.[1]

Wat is het standpunt?

Als overheid stimuleren wij innovatie ten behoeve van betere dienstverlening met behulp van generatieve AI. Daarnaast worden overheidsorganisaties aangemoedigd om deze technologie te betrekken in het zoeken naar oplossingen voor problemen. Generatieve AI kan worden ingekocht, ontwikkeld of gebruikt wanneer:

- Er bij de inzet van generatieve AI voldaan wordt aan de bestaande wet- en regelgeving (zoals de AVG, de AI-verordening en de Grondwet).
- Voldoende duidelijk is gemaakt welk doel gediend wordt met de inzet van generatieve AI.
- Risicoanalyses zijn uitgevoerd om na te gaan of een specifieke generatieve AI-toepassing ook verantwoord en veilig is. Deze dienen te worden uitgevoerd met input van een zo divers mogelijke groep experts binnen of buiten de organisatie. Voor het kiezen van een passend toetsingsinstrument voor de risicoanalyse (zoals bijvoorbeeld een (pre- scan) DPIA in het geval er persoonsgegevens worden verwerkt), is het Algoritmekader van meerwaarde.[2] Belangrijke vragen om bij een risicoanalyse te overwegen zijn[3]:
  - Welk belangrijk probleem wordt opgelost met de inzet van deze technologie, dat niet op andere wijze kan worden opgelost (noodzaak en proportionaliteit)?

1 Zie ook de overheidsbrede handreiking verantwoorde inzet generatieve AI.

2 Het algoritmekader bevat wetten en regels, tips en hulpmiddelen voor verantwoord gebruik van algoritmes en AI. Het biedt daarmee een goed – maar niet uitputtend - overzicht van verschillende veelvoorkomende instrumenten. Zie: https://minbzk.github.io/Algoritmekader/.

3 Deze overwegingen zijn (deels) geïnspireerd op het IPO-advies over het gebruik van ChatGPT in provincies: https://www.ipo.nl/nieuws/advies-over-het-gebruik-van-chatgpt-in-provincies/.

  - Zijn de juiste randvoorwaarden (bv. AI-geletterdheid, heldere governance of informatiehuishouding) op orde?
  - Zijn er alternatieven onderzocht die minder impact hebben (subsidiariteit)?
  - Is er een plan hoe om te gaan met de beperkingen van de technologie (zorgvuldigheid)?
  - Is er een duidelijke exit-strategie zodat het mogelijk is snel te stoppen wanneer er problemen optreden (veiligheid)?
- De inzet van generatieve AI, met de bijbehorende risicoafwegingen, past binnen de desbetreffende beleidsmatige context. Dit betekent bijvoorbeeld dat binnen organisaties de juiste rollen betrokken worden bij de besluitvorming over generatieve AI-inzet en dit past binnen de beleidsmatige kaders.
- Bij het inkopen van generatieve AI-modellen en/of applicaties zijn (specifieke) afspraken gemaakt met de leveranciers. Belangrijke onderwerpen voor deze afspraken (en/of bedrijfsvoorwaarden) zijn datadeling, dataretentie en het al dan niet toestaan van het hertrainen van modellen met de input van gebruikers (zoals prompts).

Voor wie is dit standpunt opgesteld?

Dit standpunt is opgesteld voor alle overheidsorganisaties die generatieve AI inkopen, ontwikkelen en/of gebruiken.
In de praktijk krijgen veel overheidsmedewerkers – van data scientists tot inkoopadviseurs en van beleidsmedewerkers tot bestuurders – met dit standpunt te maken.

Wanneer is dit standpunt van toepassing?

- Dit standpunt is van toepassing op zowel individueel gebruik door (overheids)medewerkers als op het gebruik van generatieve AI op projectbasis. Het geldt voor alle typen werkzaamheden en alle fasen van een project. Dit omvat bijvoorbeeld onderzoek naar het verbeteren van interne processen met behulp van generatieve AI, experimenteel gebruik in pilots of proeftuinen en het uitvoeren van bedrijfsvoeringstaken.
- Daarnaast geldt het standpunt ook voor externe partijen die diensten leveren aan overheidsorganisaties en daarbij gebruikmaken van generatieve AI.
- Voor onvermijdbaar gebruik, zoals generatieve AI die is ingebouwd in zoekresultaten van zoekmachines of automatische machinevertalingen van websites, wordt van individuele medewerkers verwacht dat zij hier gewetensvol mee omgaan.

Welke organisatorische en maatschappelijke overwegingen spelen een rol?

De inzet van generatieve AI als overheidspartij maakt altijd deel uit van een bredere organisatorische en maatschappelijke context. Het is belangrijk om hiermee rekening te houden, zowel binnen de eigen organisatie als met de bredere maatschappelijke effecten die de inzet van generatieve AI kan hebben. Concreet betekent dit onder meer het volgende:

- Bij de inzet van generatieve AI moeten medewerkers voldoende worden geïnformeerd over het effectieve en verantwoorde gebruik van deze technologie. Dit houdt in dat er trainingen en richtsnoeren worden aangeboden voor de verantwoorde inzet, om zo te werken aan AI-geletterdheid[4]. Naast verantwoord gebruik zijn trainingen nodig om de optimale waarde van een generatieve AI-toepassing te ontsluiten.
- Voor overheidsmedewerkers is het gebruik van generatieve AI-toepassingen onder consumentenvoorwaarden, ook voor individueel gebruik, niet toegestaan omdat er met de leveranciers geen afspraken kunnen worden gemaakt[5]. Verkenningen van nieuwe (generatieve) AI-systemen zijn toegestaan in een gecontroleerde niet-productieomgeving, mits er geen gevoelige of vertrouwelijke data worden gebruikt. Er wordt overheidsbreed onderzocht welke lopende kansrijke experimenten opgeschaald kunnen worden. Daarnaast wordt er gewerkt aan betere afspraken met leveranciers voor de inkoop van generatieve AI-toepassingen.
- Omdat niet alle generatieve AI-applicaties even goed zijn in het verwerken van de Nederlandse taal en andere Europese talen, en om onwenselijke afhankelijkheden van landen buiten de EU te voorkomen, wordt het gebruik aan van in Europa ontwikkelde en beheerde applicaties aanbevolen. Hierbij is expliciet aandacht besteed aan Nederlandstalige bronnen tijdens het trainingsproces.

4 De Europese AI-verordening stelt dat organisaties die AI inzetten moeten zorgen voor een toereikend niveau van AI-geletterdheid bij medewerkers.

5 Zie ook het advies over de inzet van generatieve AI bij de overheid van de Autoriteit Persoonsgegevens van eind 2023. Hierin wordt onderstreept dat het gebruik van online beschikbare generatieve AI veel risico's met zich meebrengt rondom de omgang met persoonsgegevens.

- Bij voorkeur worden open modellen gebruikt en, waar mogelijk, open source(-applicaties), of wordt hiernaartoe gewerkt[6]. Deze modellen bieden meer inzicht, bijvoorbeeld op het gebied van transparantie. Hierbij geldt wel dat transparantie niet ten koste mag gaan van de veiligheid van generatieve AI-modellen.

Technische maatregelen en ontwikkelingen

- Generatieve AI-toepassingen kunnen zeer behulpzaam zijn, maar soms zijn ze gebaseerd op oudere of onjuiste informatie en maken ze fouten. Omdat ze zijn ontworpen om aannemelijke antwoorden te geven in plaats van volledig nauwkeurige, kunnen deze toepassingen ook hallucineren[7]. Dit probleem speelt vooral als ze worden ingezet in onbekende of zeer gespecialiseerde situaties. Hier is het vaak lastig om precies te achterhalen waar de informatie vandaan komt. Het is goed om hier rekening mee te houden in de ontwikkeling of inkoop. Dat kan door ervoor te zorgen dat resultaten gebaseerd zijn op actuele, betrouwbare en context-specifieke informatie. Er bestaan technieken die de kans op fouten verkleinen en de antwoorden beter afstemmen op de specifieke context waarin generatieve AI wordt ingezet.
- Het is verstandig om per toepassing te onderzoeken welk AI-model de beste resultaten oplevert (-leveren). Soms zijn kleinere modellen beter geschikt, vooral bij specialistische en/of complexe onderwerpen, omdat ze eenvoudiger en beter kunnen worden aangepast aan de context. Voor overheidsorganisaties en andere partijen die binnen hetzelfde vakgebied actief zijn, is samenwerking bij het aanpassen van modellen aan te raden. Dit draagt bij aan betere traceerbaarheid, consistentie, efficiëntie en betrouwbaarheid van de technologie.

Een uitgebreidere toelichting van organisatorische en technische maatregelen die concreet kunnen worden genomen om de verantwoorde inzet van generatieve AI binnen (overheids)organisaties te borgen, is opgenomen in de overheidsbrede handreiking verantwoorde inzet van generatieve AI.

Samenhang van dit standpunt en de Europese AI-verordening

De Europese AI-verordening bevat eisen aan AI-modellen die in staat zijn om veel verschillende taken uit te voeren en in veel verschillende AI-systemen kunnen worden geïntegreerd (vaak is hier sprake van generatieve AI). In de AI-verordening worden deze aangeduid als 'AI-modellen voor algemene doeleinden' (GPAI-modellen)'. Aanbieders van deze modellen zijn verplicht om informatie over de werking van het model te delen met ontwikkelaars die op dit model voortbouwen. Ook moet duidelijk worden hoe het auteursrecht wordt gerespecteerd en moet bij alle door AI gegenereerde content duidelijk worden gemaakt dat het dit het geval is. Deze eisen dragen eraan bij dat beter inzichtelijk wordt welke content tot stand is gekomen met AI en mogelijk onzeker, fout of ongeverifieerd is. In veel gevallen zullen overheidspartijen aanbieder en/of gebruiksverantwoordelijke zijn van een (generatief) AI-systeem dat voortbouwt op een GPAI-model. Dit betekent dat overheidsorganisaties bij de inzet van generatieve AI veelal te maken zullen krijgen met vereisten (bijvoorbeeld rondom transparantie of hoog-risico) uit de AI-verordening. Naast de AI-verordening kunnen ook andere wettelijke kaders van toepassing zijn op de inzet van generatieve AI. In de overheidsbrede handreiking is hier meer over opgenomen.

6 Zie voor een open source definitie van AI ook de definitie van de Open Source Initiative: https://opensource.org/ai/open-source-ai-definition.

7 Door een 'large language model' (LLM) gegenereerde informatie die feitelijk onjuist is. Het model genereert antwoorden die niet gebaseerd zijn op de gegeven input of op feitelijke informatie uit de trainingsdata. Een hallucinatie kan worden veroorzaakt door verschillende redenen, zoals een gebrek aan specifieke informatie in de trainingsdata, een gebrek aan context of foutieve en inconsistente informatie in de trainingsdata.

Ministerie van Binnenlandse Zaken en Koninkrijksrelaties
Postbus 20011
2500 EA Den Haag

Datum
30 januari 2025