AUTORITEIT
PERSOONSGEGEVENS

Autoriteit Persoonsgegevens
Postbus 93374, 2509 AJ Den Haag
Hoge Nieuwstraat 8, 2514 EL Den Haag
T 070 8888 500 - F 070 8888 501
autoriteitpersoonsgegevens.nl

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

Datum
11 oktober 2023

Ons kenmerk
volgt

Contactpersoon
Programma
persoonsgegevens en
algoritmes/AI
070 8888 500

Onderwerp
Concept-standpunt gebruik online generatieve AI Rijksoverheid

Geachte mevrouw Van Huffelen,

Uw Directie CIO Rijk heeft (onder meer) de AP verzocht te adviseren op een in te nemen standpunt ten aanzien van het gebruik van online generatieve AI binnen de Rijksoverheid.[1] Daarnaast werkt u aan een visie op generatieve AI. Het ambtelijk concept-standpunt ten aanzien van het gebruik van online generatieve AI binnen de Rijksoverheid luidt als volgt: "U wordt geadviseerd in te stemmen met het standpunt om het gebruik van online generatieve AI bij de Rijksoverheid voorlopig te ontraden, tenzij aanbieders voldoen aan de wettelijke vereisten." De AP merkt hierover het volgende op.

Algemeen
De AP verwacht dat generatieve AI de komende periode steeds meer zal worden toegepast in de maatschappij. In de vorm van op zichzelf staande toepassingen, zoals ChatGPT of Midjourney, maar ook geïntegreerd in bestaande software, zoals de zoekmachine Bing van Microsoft of de inzet van Google Bard in Gmail. We verwachten ook dat steeds meer content afkomstig zal zijn van generatieve AI (denk aan nieuwsberichten, reviews, illustraties) en dat steeds meer burgers en bedrijven gegenereerde content in hun (werk)processen opnemen. De AP maakt zich ernstige zorgen over de omgang met persoonsgegevens bij organisaties die gebruik maken van generatieve AI. Overigens zijn er mogelijk ook andere wettelijke vereisten in het geding, maar de AP zal zich in deze brief beperken tot de omgang met persoonsgegevens. De AP ziet bij de toepassing van generatieve AI de volgende mogelijke risico's:

[1] De AP neemt aan dat met de term online (beschikbare) generatieve AI wordt gedoeld op toepassingen die door derden via internet worden aangeboden, zoals ChatGPT, Google Bard en Midjourney.

- *Bron van de trainingsdata mogelijk onrechtmatig verkregen*
  Veel voorbeelden waarmee de AI wordt getraind komen van het internet (scraping). Waar de voorbeelden persoonsgegevens bevatten is het onwaarschijnlijk dat daar een grondslag voor is.

- *Trainingsdata wordt bij verzameling niet geverifieerd*
  Bij het verzamelen van (tekst)voorbeelden van het internet wordt niet ingegaan op de inhoudelijke juistheid van de tekst. Dit maakt het mogelijk dat persoonsgegevens worden verwerkt die onjuist zijn.

- *Verwijdering en rectificatie moeilijk/onmogelijk*
  Het gebrek aan expliciete koppeling tussen persoon en gegeven maakt dat het onduidelijk is of het verwijderen of rectificeren van persoonsgegevens mogelijk is en of goed invulling kan worden gegeven aan andere rechten van betrokkenen

- *Geen zicht op verwerkingen bij gebruik van de dienst*
  Modellen die tekst of afbeeldingen genereren maken gebruik van opdrachten (prompts) van de gebruiker. Niets belet de gebruiker persoonsgegevens in te voeren dan wel op te vragen. Er is weinig zicht op de verwerking die hierdoor mogelijk ontstaat.

Voorgaande risico's worden versterkt door het feit dat meerdere modellen vrij beschikbaar zijn. Als deze modellen onrechtmatig persoonsgegevens verwerken dan zal dit op vele plekken voorkomen. Een eventuele update/rectificatie zal wellicht niet alle afnemers bereiken.

De AP onderstreept dat AI, waaronder ook generatieve AI, vooroordelen kan bevestigen en discriminatie in de hand kan werken. Daarom is het zo belangrijk dat generatieve AI verantwoord wordt ontwikkeld en ingezet, in lijn met de wettelijke vereisten, zoals voortvloeiend uit de privacywetgeving. De inzet van algoritmes en AI maakt processen mogelijk sneller, efficiënter en goedkoper; maar kan er ook voor zorgen dat mensen, zonder het te weten, bijvoorbeeld worden ingedeeld in categorieën en dat hun gedrag op basis van – mogelijk inaccurate en niet verifieerbare – historische data wordt voorspeld, gemanipuleerd en voorgeschreven. Denk hierbij aan digitale assistenten wiens aanbevelingen klakkeloos worden overgenomen. Een HR medewerker zou bijvoorbeeld een chatbot kunnen vragen om CV's en motivatiebrieven van sollicitanten door te nemen en samen te vatten wat de voor- en nadelen zijn per kandidaat. Zelfs neutraal ogende antwoorden kunnen gebaseerd zijn op vooroordelen of onvolkomenheden die in trainingsdata zaten. Het zou bijvoorbeeld kunnen dat bepaalde namen niet in contexten van bepaalde beroepen voorkomen. Daarbij komt dat het onduidelijk is hoe er verantwoording afgelegd zou kunnen worden over keuzes die worden gemaakt met behulp van generatieve AI. Het recht om in rechte op te kunnen komen tegen beslissingen die (mede) op deze wijze zijn gegenereerd kan daardoor feitelijk illusoir worden.

Acties AP
Als toezichthouder op de privacywetgeving onderneemt de AP de komende tijd verschillende acties. Zo heeft de AP software-ontwikkelaar OpenAI om opheldering gevraagd over ChatGPT. Dit onderzoek loopt. Het samenwerkingsverband van Europese privacytoezichthouders, de EDPB, heeft een Europese Taskforce ChatGPT ingesteld. Deze taskforce is inmiddels meerdere malen bijeen gekomen. Afgesproken is acties vanuit de Europese toezichthouders zoveel mogelijk te harmoniseren en antwoorden en scope te coördineren in de taskforce. Ook heeft de AP een techbedrijf om verantwoording gevraagd over de verwerking van persoonsgegevens via de chatbot die werd geïntegreerd in hun app die populair is bij kinderen.

Inhoudelijke reactie concept-standpunt ambtelijk BZK t.b.v. staatssecretaris
De AP heeft met instemming kennisgenomen van de brief d.d. 7 juli jl. waarin u de Tweede Kamer heeft geïnformeerd over de voortgang van het visietraject over generatieve AI. De kern is te waarborgen dat deze technologie op een verantwoorde manier in onze samenleving wordt ingebed. De verantwoorde ontwikkeling en inzet is van groot belang. Verantwoord betekent voor ons onder meer en in ieder geval in lijn met de grondrechten in het Handvest van de grondrechten van de EU en de beginselen in de AVG.

De AP herkent het belang van de richting van het concept-standpunt om het gebruik van online generatieve AI door de Rijksoverheid te ontraden, tenzij aanbieders voldoen aan de wettelijke vereisten. Wel laat dit standpunt (te veel) ruimte voor interpretatie. De AP benadrukt dat zij het innemen van een duidelijk standpunt t.a.v. het gebruik van (online) generatieve AI door de Rijksoverheid aanmoedigt.

- In de eerste plaats dient duidelijk te zijn wat bedoeld wordt met online generatieve AI. Hiermee lijkt gedoeld te worden op door derden ontwikkelde en online aangeboden diensten. Dit laat ruimte open voor vragen. Bijvoorbeeld hoe om te gaan met (ook) offline te gebruiken modellen, zoals LLAMA2 en Stable Diffusion. Ten aanzien van deze modellen gelden dezelfde wettelijke vereisten, inclusief ten aanzien van de wijze waarop ze getraind worden, ook als ze lokaal worden gedraaid of van een eigen server worden aangeboden. Ook wordt niet ingegaan op eventuele door de Rijksoverheid zelf ontwikkelde modellen, die al dan niet online en al dan niet binnen de Rijksoverheid worden aangeboden. Uiteraard dient de Rijksoverheid als aanbieder ook te voldoen aan de wettelijke vereisten.

- In de tweede plaats benadrukt de AP dat ook experimenten, bijvoorbeeld in de zin van pilots of proeftuinen, aan het Handvest van de grondrechten van de EU en de AVG moeten voldoen. Het standpunt dient zich derhalve ook kenbaar uit te strekken tot het experimenteel gebruik van generatieve AI.

- In de derde plaats dient te worden uitgelegd waarom de scope beperkt is tot de Rijksoverheid, terwijl dezelfde zorgen ook daarbuiten gelden, bijvoorbeeld ten aanzien van het gebruik door

decentrale overheden. Met andere woorden, is er in de visie van BZK een verschil met de manier waarop anderen dan actoren binnen of werkend in opdracht van de Rijksoverheid om zouden kunnen/moeten/mogen maken van generatieve AI?

- In de vierde plaats acht de AP 'ontraden, tenzij' van het gebruik van generatieve AI niet duidelijk genoeg. Het woord ontraden lijkt op zichzelf ruimte te bieden voor de Rijksoverheid om van het standpunt af te wijken. De AP stelt vast dat BZK op grond van het Coördinatiebesluit organisatie, bedrijfsvoering en informatiesystemen rijksdienst bevoegd is om het standpunt als dwingend kader vorm te geven. De AP adviseert in een dergelijk dwingend kader te expliciteren dat het gebruik van online generatieve AI niet is toegestaan, en dat dit slechts anders zou kunnen zijn, indien door alle betrokken partijen met de juiste invulling van de in de AVG opgenomen verantwoordingsplicht wordt gehandeld.

- In de vijfde plaats merkt de AP op dat het standpunt alleen lijkt te verwijzen naar de op dit moment geldende wettelijke vereisten, terwijl op Europees niveau gewerkt wordt aan nieuwe wetgevende kaders, zoals de AI-verordening, die additionele eisen kan stellen aan het gebruik van generatieve AI. Het innemen van een standpunt kan niet wachten totdat de onderhandelingen zijn afgerond, maar deze ontwikkeling is wel relevant voor de inhoud van het standpunt. Bovendien wordt in deze kaders rekenschap gegeven van risico's dat AI, waaronder ook generatieve AI, bijvoorbeeld vooroordelen kan bevestigen en discriminatie in de hand kan werken. Die risico's zouden ook moeten worden meegewogen in de standpuntbepaling door BZK en aanleiding moeten geven tot meer voorzichtigheid.

Hoogachtend,
Autoriteit Persoonsgegevens,

Aleid Wolfsen
voorzitter