AI in het openbaar bestuur

Jeroen Zandberg

Juli 2026

wetenschappelijk bureau nsc

wetenschappelijk
bureau nsc

Inhoudsopgave

1. Introductie

Kunstmatige intelligentie ontwikkelt zich in hoog tempo en heeft een steeds grotere invloed op de samenleving, de economie en het functioneren van de democratie. Ook binnen het openbaar bestuur neemt het gebruik van AI snel toe. AI kan overheden helpen om informatie te verwerken, de dienstverlening te verbeteren en ambtenaren te ondersteunen in hun werk. Tegelijkertijd brengt het risico's met zich mee, onder meer op het gebied van transparantie, afhankelijkheid, privacy, discriminatie en democratische controle.

AI is daarom ook een vraagstuk van goed bestuur, een van de kernpunten van Nieuw Sociaal Contract. De inzet van AI door de overheid moet niet alleen doelmatig zijn, maar ook rechtmatig, uitlegbaar en dienstbaar aan burgers en samenleving. Daarbij is van belang welke systemen worden gekozen, door wie deze worden ontwikkeld en onder welke publieke voorwaarden zij worden toegepast.

Met dit rapport wil het Wetenschappelijk Bureau NSC bijdragen aan de positiebepaling van NSC in het debat over de keuze en inzet van AI. Het rapport richt zich in het bijzonder op de vraag hoe AI kan worden gebruikt in, en ten dienste van, goed openbaar bestuur, zonder dat daarbij publieke waarden, menselijke verantwoordelijkheid en democratische zeggenschap uit het oog raken. Dit rapport kan daarbij tevens dienen als input voor de themagroep AI en democratie zoals dat door Nieuw Sociaal Contract is gestart. Het rapport is taalkundig geredigeerd met hulp van AI.

Het eerste hoofdstuk behandelt AI-soevereiniteit in het openbaar bestuur. Centraal hierin staat de vraag in hoeverre Nederlandse overheden zeggenschap houden over de gebruikte technologie, data en digitale infrastructuur en hoe de afhankelijkheid van buitenlandse technologiebedrijven kan worden beperkt. Ook wordt ingegaan op de ontwikkeling van GPT-NL als Nederlands AI-model.

Het tweede hoofdstuk kijkt hoe AI momenteel binnen gemeenten en andere publieke organisaties wordt toegepast. Daarbij komen zowel de mogelijkheden voor efficiëntere dienstverlening en ondersteuning van ambtenaren als de bestuurlijke en maatschappelijke risico's aan bod.

Het derde hoofdstuk richt zich op bias in taalmodellen. Het laat zien hoe sociale, culturele en politieke vooroordelen in AI-systemen kunnen ontstaan en waarom deze bij overheidstoepassingen gevolgen kunnen hebben voor gelijke behandeling, representatie en het vertrouwen van burgers.

Het vierde hoofdstuk bespreekt de verschillen tussen open en gesloten AI-modellen. Beide benaderingen worden vergeleken op onder meer transparantie, controle, veiligheid, kosten en publieke onafhankelijkheid. Tot slot worden de inzichten uit de verschillende hoofdstukken samengebracht in een conclusie en aanbevelingen voor een verantwoorde inzet van AI binnen het Nederlandse openbaar bestuur.

2. AI-soevereiniteit in het openbaar bestuur

Dit hoofdstuk bakent AI-soevereiniteit af voor het openbaar bestuur en plaatst het begrip in een Nederlandse context. De Nederlandse Digitaliseringsstrategie stelt dat strategische autonomie in de digitale wereld niet betekent dat Nederland alles zelf moet bouwen, maar wel dat het zijn kritieke processen en gegevens beter onder controle krijgt. Juist bij AI is dat relevant: Nederlandse overheidsorganisaties experimenteren al breed met generatieve AI, terwijl de onderliggende modellen, cloudomgevingen en leveranciers vaak buiten Nederland en buiten Europa zijn georganiseerd. Tegelijk ontstaat met GPT-NL voor het eerst een expliciet Nederlands project rond een “soeverein” taalmodel. Daarmee wordt AI-soevereiniteit geen abstract Europees debat, maar een bestuurlijke vraag voor Rijk, regio en gemeente. [1]

Begripsafbakening van AI-soevereiniteit

Dit rapport definieert AI-soevereiniteit als het vermogen van overheidsorganisaties om zelfstandig, controleerbaar en publiek verantwoord te beslissen over de inzet van AI: welk probleem ermee wordt opgelost, welke data worden gebruikt, welk model wordt gekozen, waar het systeem draait, onder welke contractuele voorwaarden dat gebeurt, en hoe menselijk toezicht, bezwaar en democratische controle zijn ingericht. Die definitie sluit aan bij de Nederlandse lijn over digitale autonomie; meer grip op kritieke processen en gegevens, minder afhankelijkheid van een te beperkt aantal leveranciers, en bij GPT-NL, dat soevereiniteit expliciet verbindt aan controle over model, data en keuzes. [2]

AI-soevereiniteit is dus geen autarkie. Het gaat niet om een overheid die alle modellen zelf traint of uitsluitend nationale software gebruikt. Het gaat om de vraag of de overheid onder veranderende geopolitieke, juridische en marktvoorwaarden kan blijven sturen. Daarvoor zijn vijf elementen doorslaggevend: beleidsmatige regie over waar AI wel en niet wordt ingezet; datasoevereiniteit over toegang, opslag en hergebruik; technische grip op model, infrastructuur en interoperabiliteit; organisatorische kennis om leveranciers aan te sturen; en democratische verantwoordbaarheid richting burgers, volksvertegenwoordiging en toezichthouders. Gemeenten wijzen zelf op precies deze spanning: afhankelijkheden raken niet alleen techniek, maar ook de continuïteit van dienstverlening, de bescherming van gegevens en de ruimte om lokaal eigen keuzes te maken. [3]

Voor de overheid krijgt het begrip bovendien een extra lading omdat AI-toepassingen niet neutraal zijn. Als een model teksten samenvat, dossiers doorzoekbaar maakt of vragen van burgers beantwoordt, beïnvloedt het indirect ook toegang tot informatie, bestuurlijke prioriteiten en de uitvoeringspraktijk. Daarom is AI-soevereiniteit in het openbaar bestuur uiteindelijk een combinatie van technische onafhankelijkheid, bestuurlijke handelingsvrijheid en rechtsstatelijke controle. Het is pas geloofwaardig als een overheid kan uitleggen wat een systeem doet, waarom het is gekozen, welke risico’s zijn geaccepteerd en hoe burgers worden beschermd. [4]

Typologie van AI-toepassingen in de overheid

De Nederlandse praktijk laat zien dat AI in de overheid op dit moment (juni 2026) vooral wordt ingezet als ondersteunende technologie, niet als zelfstandig beslissende actor. Volgens de Overheidsbrede Monitor Generatieve AI vallen de meeste geïdentificeerde toepassingen in vijf categorieën:

1. Generieke processen en bedrijfsvoering,
2. Maatwerk en dienstverlening,

3. Kennisvergaring en beleidsvorming,
4. Kennisverwerking, archivering en anonimiseren,
5. Democratisch proces.

De monitor benadrukt ook dat generatieve AI vooral wordt gebruikt in processen met veel tekst- en gegevensverwerking. [5]

In de eerste categorie gaat het om interne hulpmiddelen: chatbots of assistenten die ambtenaren helpen met vertalen, transcriberen, informatie opzoeken of het schrijven van e-mails en notulen. In de tweede categorie gaat het om extern gerichte toepassingen, zoals chatbots op gemeentelijke websites en gepersonaliseerde informatievoorziening voor burgers en bedrijven. De derde categorie betreft systemen die beleidsmakers helpen bij het verzamelen en structureren van inzichten; de vierde categorie richt zich op het verwerken, archiveren en anonimiseren van informatie; en de vijfde categorie ondersteunt bijvoorbeeld gemeenteraad of Kamervraagprocessen. [6]

Deze indeling is belangrijk omdat verschillende toepassingen ook verschillende soevereiniteitsvragen oproepen. Een interne schrijf- of zoekassistent raakt vooral kennisopbouw, beveiliging en leveranciersafhankelijkheid. Een chatbot voor bewoners raakt daarnaast aan toegankelijkheid, transparantie en de mogelijkheid om door te schakelen naar een mens. En zodra AI wordt gebruikt in contexten rond subsidies, handhaving of andere essentiële publieke diensten, verschuift het vraagstuk naar fundamentele rechten, rechtsbescherming en hoog-risicoregels onder de AI-verordening. De Nederlandse overheid geeft op dit moment expliciet aan dat generatieve AI wel wordt ingezet voor ondersteuning, maar voor zover bekend niet om besluiten door AI te laten nemen. [7]

De gemeente Amsterdam is een voorloper op het gebied van het gebruik en implementatie van AI en laat zien hoe deze typologie er lokaal uitziet. De Amsterdamse AI Agenda noemt drie hoofdsporen:

1. interactie met inwoners,
2. toepassingen in de openbare ruimte,
3. efficiënter intern werken.

Daaronder vallen onder meer AI-visualisaties voor stedelijke ontwikkeling, AI-assistentie in samenwerking met bewoners, beeldherkenning in de openbare ruimte en ChatAmsterdam, een interne AI-assistent voor schrijven, notuleren, *deskresearch* en brainstormen. De gemeente benadrukt daarbij dat de gegevens van ChatAmsterdam binnen de gemeentelijke omgeving blijven en niet naar de leverancier van het onderliggende taalmodel gaan voor training. [8]

Afhankelijkheden en het soevereiniteitsprobleem

Het kernprobleem van AI-soevereiniteit is in de Nederlandse praktijk dat de overheid vaak slechts beperkt grip heeft op de schakels waar AI daadwerkelijk van afhankelijk is. De Overheidsbrede Monitor noemt als knelpunten onder meer onzekere uitkomsten van generatieve AI, snel veranderende technologie, risico's op fouten en misbruik, datamanagement en datakwaliteit die niet op orde zijn, beperkte AI-geletterdheid, beperkte centralisering op nationaal niveau, onduidelijkheid over wet- en regelgeving, benodigde technische infrastructuur en expliciet ook digitale afhankelijkheid. [9]

Die afhankelijkheid bestaat uit tenminste vier lagen.

1. **Leveranciersafhankelijkheid:** veel overheden kopen AI-functionaliteit in als onderdeel van bredere software- of cloudcontracten. Daardoor verschuift de feitelijke macht over updates, prijs, interoperabiliteit en auditmogelijkheden naar marktpartijen.
2. **Modelafhankelijkheid:** als de overheid werkt met gesloten modellen waarvan trainingsdata, veiligheidsmaatregelen en wijzigingsbeheer beperkt inzichtelijk zijn, wordt publieke verantwoording moeilijker.
3. **Data-afhankelijkheid:** generatieve AI werkt alleen goed als data van hoge kwaliteit zijn, en de monitor laat zien dat juist die voorwaarde vaak ontbreekt.
4. **Kennisafhankelijkheid:** zonder voldoende AI-geletterdheid kunnen organisaties de leveranciers en risico's niet goed beoordelen. [10]

Het overheidsstandpunt stelt dat iedere overheidsorganisatie verantwoordelijk blijft voor het veilige gebruik van data, resultaten en processen. Ook wanneer werkzaamheden worden uitbesteed. Daarom moeten eisen voor generatieve AI een plek krijgen in het inkoopproces, net als privacy en informatie-beveiliging. Het soevereiniteitsprobleem zit dus niet alleen in het model zelf, maar ook in contracten, exit-mogelijkheden, beveiliging, *logging, audit* en de mogelijkheid om een toepassing later te vervangen zonder dat de dienstverlening instort. [11]

In de Nederlandse overheidssituatie betekent AI-soevereiniteit daarom vooral: kunnen kiezen, kunnen controleren en kunnen wisselen. Juist op lokaal niveau is dat urgent. Gemeenten hoeven volgens de VNG niet te wachten op landelijke of Europese oplossingen om hun digitale autonomie te versterken. Tegelijk erkent diezelfde lijn dat bestaande contracten, systemen en ketens zulke autonomie bemoeilijken. Voor AI in het openbaar bestuur is dat een belangrijke les: soevereiniteit is geen eindtoestand, maar een bestuurlijke capaciteit die per toepassing moet worden opgebouwd. [12]

Juridisch en normatief kader

Het primaire juridische kader is de Europese AI-verordening. Die wet heeft regels voor ontwikkeling en gebruik van AI-systemen, beschermt fundamentele rechten en werkt met risicocategorieën. Verboden zijn onder meer: bepaalde vormen van schadelijke manipulatie en *social scoring.* Hoog-risico zijn onder andere: AI-systemen voor biometrie, kritieke infrastructuur, selectieprocessen en het bepalen van toegang tot essentiële publieke diensten en subsidies. Voor chatbots, deepfakes en andere systemen met misleidingsrisico gelden transparantie-verplichtingen. Voor grote taalmodellen gelden bovendien aparte eisen. De verordening trad in augustus 2024 in werking. Sinds 2 februari 2025 gelden de verboden praktijken en de verplichting om aan AI-geletterdheid te werken. De overige bepalingen, waaronder transparantie-eisen en regels voor hoog-risico-systemen, gelden vanaf augustus 2026. [13]

Voor Nederlandse overheden wordt dit Europese kader aangevuld met een nationaal regime. Het overheidsbrede standpunt over generatieve AI verplicht organisaties om risicoanalyses uit te voeren, bij persoonsgegevens de AVG en een pre-scan of DPIA toe te passen, en de Baseline Informatiebeveiliging Overheid te benutten voor de beveiligingskant. Daarnaast geldt dat menselijke toetsing en redactionele controle vereist zijn wanneer AI wordt gebruikt in overheidscommunicatie richting burgers. Daarmee geeft de Nederlandse overheid een duidelijke grens aan: generatieve AI mag ondersteunen, maar publieke communicatie en publieke besluitvorming moeten menselijk verantwoord blijven. [11]

Transparantie krijgt in Nederland verder vorm via het Algoritmeregister. Dat register is nu nog niet voor alles verplicht, maar het doel is om alle overheidsorganisaties met relevante algoritmen aan te sluiten en impactvolle algoritmen uiteindelijk wettelijk verplicht te laten registreren. Het register moet burgers, media en maatschappelijke organisaties in staat stellen de overheid kritisch te volgen, en sluit inhoudelijk aan op de AI-verordening. Daarmee is het register niet alleen een transparantie-instrument, maar ook een deel van de Nederlandse soevereiniteitsinfrastructuur: wie niet weet welke algoritmen in gebruik zijn, kan er immers ook niet op sturen. [14]

Vooruitblik op GPT-NL

GPT-NL is in de Nederlandse context het meest concrete experiment met AI-soevereiniteit. TNO, SURF en het Nederlands Forensisch Instituut bouwen sinds 2023 samen aan een onafhankelijk Nederlands taalmodel en ecosysteem, expliciet gericht op *governance*, transparantie, privacy, auteursrecht en publieke waarden. Volgens TNO is GPT-NL "soeverein" omdat Nederland en Europa daarmee controle houden over model, data en ontwikkelkeuzes; de broncode wordt open source gedeeld, de dataset wordt gedocumenteerd, terwijl de modelgewichten onder een gecontroleerde licentie beschikbaar komen. Het project is publiek gefinancierd met €13,5 miljoen. [15]

Het voortgangsrapport van maart 2026 meldt dat de eerste versies van GPT-NL bij zogenaamde '*launching customers*' op lokale infrastructuur worden ingezet. Het gaat hierbij eerst om vijf organisaties, later oplopend naar tien, uit onder meer ministeries, het veiligheidsdomein en de financiële sector. Het project noemt als voordelen expliciet dat de gebruikers weten waar de data vandaan komt, veilig op locatie kunnen werken en bijdragen aan een rechtmatig Nederlands AI-ecosysteem. Onder de eerste use-cases vallen ook toepassingen die direct relevant zijn voor overheden, zoals Gem, een chatbot die al door bijna dertig gemeenten wordt gebruikt, en HIP, een communicatie-assistent voor duidelijkere overheidsbrieven. [16]

GPT-NL zal volgens het voortgangsrapport in de tweede helft van 2026 breder beschikbaar komen via een professionele licentie, met daarnaast een online versie. Intussen is de publieke GPT-NL-dataset inmiddels daadwerkelijk publiek beschikbaar gemaakt, waarmee het project ook op het vlak van datasettransparantie een stap verder is dan veel commerciële alternatieven. [17]

Voor het openbaar bestuur betekent dit dat GPT-NL voorlopig vooral moet worden gezien als strategische bouwsteen, niet als kant-en-klare vervanger van alle bestaande modellen. Als GPT-NL erin slaagt bestuurlijk bruikbare prestaties te combineren met lokale of gecontroleerde hosting, inzicht in dataherkomst en passende licentievoorwaarden, kan het een belangrijk instrument worden voor Nederlandse AI-soevereiniteit. Maar het zal alleen werkelijk soevereiniteit opleveren als overheden er ook hun inkoop, governance, datakwaliteit en menselijke toetsing op aanpassen. Zonder die bestuurlijke laag blijft zelfs een Nederlands model slechts een technisch project; met die laag kan het uitgroeien tot publieke infrastructuur. [18]

De voordelen en beperkingen van GPT-NL

De ontwikkeling van GPT-NL biedt vanuit het perspectief van AI-soevereiniteit duidelijke voordelen, maar kent tegelijkertijd belangrijke beperkingen. Het belangrijkste voordeel is de grotere mate van controle. Bij commerciële, gesloten taalmodellen is een overheidsorganisatie voor de werking, beschikbaarheid en verdere ontwikkeling van het model in belangrijke mate afhankelijk van de leverancier. GPT-NL is juist opgezet om Nederland en Europa meer controle te geven over het model, de gebruikte data en de ontwikkelkeuzes. De broncode wordt open source beschikbaar gesteld en de herkomst van de trainingsdata wordt gedocumenteerd, terwijl het model op lokale infrastructuur kan worden ingezet. Dit sluit direct aan bij de in dit hoofdstuk

gehanteerde definitie van AI-soevereiniteit, waarin niet volledige technologische onafhankelijkheid centraal staat, maar het vermogen van de overheid om te kunnen kiezen, controleren en wisselen.

Een tweede voordeel is dat de ontwikkeling van een eigen model kennis en expertise binnen het Nederlandse AI-ecosysteem opbouwt. Het ontwikkelen, trainen, testen en onderhouden van een taalmodel vereist specialistische kennis die anders grotendeels bij buitenlandse technologiebedrijven blijft liggen. GPT-NL heeft daarmee niet alleen waarde als product, maar ook als kennisinfrastructuur. Overheden en publieke kennisinstellingen kunnen hierdoor beter begrijpen hoe taalmodellen functioneren, eigen toepassingen ontwikkelen en commerciële leveranciers kritischer beoordelen. Dat organisatorische vermogen is eveneens onderdeel van AI-soevereiniteit: zonder voldoende interne kennis blijft een overheid immers afhankelijk van externe partijen, ook wanneer er formeel keuzes zijn uit meerdere modellen.

Daar staat tegenover dat technologische autonomie een prijs heeft. Een eigen model ontwikkelen en onderhouden is kostbaar en vraagt om gespecialiseerde medewerkers, rekenkracht en een technische infrastructuur die voortdurend moet worden bijgehouden. Bovendien zal een relatief klein nationaal model niet vanzelfsprekend dezelfde prestaties of brede functionaliteit bieden als de grootste commerciële modellen, waarin technologiebedrijven veel grotere bedragen investeren. Dit betekent dat volledige onafhankelijkheid niet noodzakelijk de meest doelmatige keuze is. De afweging is daarom niet simpelweg tussen een ‘soeverein’ Nederlands model en ‘afhankelijke’ buitenlandse modellen, maar tussen verschillende gradaties van controle, kwaliteit, kosten en risico. De hogere kosten en lagere kwaliteit ten opzichte van toonaangevende AI-modellen vormen daarbij reële nadelen van een eigen Nederlands model.

De meerwaarde van GPT-NL ligt vooral bij toepassingen waarvoor controle over data, juridische zekerheid, transparantie en continuïteit zwaarder wegen dan toegang tot het krachtigste beschikbare model. Voor sommige overheidstaken kan een buitenlands commercieel model daardoor de meest doelmatige keuze blijven, terwijl voor andere, bijvoorbeeld toepassingen met gevoelige overheidsinformatie of strategisch belangrijke publieke dienstverlening, een model als GPT-NL aantrekkelijker kan zijn. Dat onderstreept een centraal uitgangspunt van AI-soevereiniteit: soevereiniteit betekent niet ‘alles zelf ontwikkelen’, maar ervoor zorgen dat de overheid daadwerkelijk alternatieven heeft en bewust kan bepalen wanneer afhankelijkheid van een externe leverancier wel en niet aanvaardbaar is.

---

[1] Nederlandse Digitaliseringsstrategie (NDS)

https://www.digitaleoverheid.nl/nederlandse-digitaliseringsstrategie-nds/

[2] NDS Prioriteit 5 - Versterken digitale weerbaarheid en ...

https://www.digitaleoverheid.nl/nederlandse-digitaliseringsstrategie-nds/6-prioriteiten-voor-een-overheid/prioriteit-5-versterken-digitale-weerbaarheid-en-autonomie-van-de-overheid/

[3] [12] Hoe kunnen gemeenten hun digitale autonomie vergroten?

https://www.digitaleoverheid.nl/nieuws/hoe-kunnen-gemeenten-hun-digitale-autonomie-vergroten/

[4] AI (Artificiële Intelligentie)

https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/artificiele-intelligentie-ai/

[5] [6] [9] [10] Overheidsbrede monitor Generatieve AI

https://open.overheid.nl/documenten/dafe19ab-8874-42a0-a55f-315ac5825282/file

[7] [11] open.overheid.nl
https://open.overheid.nl/documenten/7c304fb8-2846-40ce-9044-abcb6415f459/file
[8] Amsterdamse AI Agenda 2025
https://assets.amsterdam.nl/publish/pages/1080510/amsterdamse_ai_agenda.pdf
[13] AI-verordening AI (Artificiële Intelligentie) - Digitale Overheid
https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/artificiele-intelligentie-ai/ai-verordening/
[14] Algoritmeregister voor de overheid Algoritmes - Digitale Overheid
https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/algoritmes/algoritmeregister/
[15] [18] GPT-NL: a sovereign language model for the Netherlands
https://www.tno.nl/en/digital/artificial-intelligence/gpt-nl/
[16] [17] GPT-NL - Progress report #2
https://publications.tno.nl/publication/34645720/dNo6mcVV/TNO-2026-R10516.pdf

3. AI-toepassingen in gemeenten en publieke organisaties

Afbakening en interpretatie

Dit hoofdstuk is geschreven als een thematische analyse van een gerichte selectie uit het Nederlandse Algoritmeregister, aangevuld met officiële publieke pagina's van betrokken overheidsorganisaties en leveranciers. Op 18 juni 2026 bevat het Algoritmeregister 1473 algoritmebeschrijvingen. De selectie die hier centraal staat, laat dus niet "de" hele overheid zien, maar wel een goed spectrum van de manieren waarop gemeenten, provincies, uitvoerings-organisaties en rijksorganisaties generatieve AI en aanpalende AI-systemen in de praktijk inzetten. [1]

Een belangrijk analytisch punt is dat veel van de geselecteerde systemen in strikte zin geen zelfstandige *foundation models* zijn, maar toepassingen, platformen of chatomgevingen rond één of meer taalmodellen. Dat geldt bijvoorbeeld voor GovChat-NL, SafeGPT, GovGPT, ChatAmsterdam en PolitieGPT. Alleen in een deel van de registraties wordt het onderliggende model expliciet benoemd; in andere gevallen beschrijft de overheid vooral de functionele toepassing, de gegevensverwerking en de governancestructuur. Voor een rapporthoofdstuk over AI in het openbaar bestuur is dat overigens geen tekortkoming, maar juist relevant: in de bestuurlijke praktijk wordt AI meestal niet als "los model", maar als beheerde toepassing ingevoerd. [2]

Deze bestuurlijke context sluit aan bij het overheidsbrede standpunt van 2025. Het kabinet heeft toen met gemeenten, provincies, waterschappen en uitvoeringsorganisaties afgesproken dat generatieve AI breder mag worden toegepast, mits gebruik plaatsvindt binnen duidelijke randvoorwaarden rond risicoanalyse, betrouwbaarheid, training, governance en verantwoord gebruik. In dezelfde lijn noemt de Rijksoverheid voorbeelden als snellere vergunningverlening, snellere beantwoording van vragen van burgers en ondernemers, en efficiëntere besluitvorming. [3]

Hoofdpatroon in de selectie

Wat deze selectie vooral laat zien, is een verschuiving van open, consumentgerichte AI naar afgeschermde, taakgerichte en organisatie-specifieke AI. ChatAmsterdam is expliciet ontwikkeld om veilig en verantwoord met generatieve AI te leren werken binnen de gemeente; PZH-Assist is alleen beschikbaar voor provinciale medewerkers en is niet getraind op interne data; PolitieGPT draait op de eigen infrastructuur van de politie; DefGPT is het interne alternatief voor openbare AI-diensten binnen Defensie; en GovChat-NL wordt juist gepositioneerd als open-source overheidsalternatief voor openbare of commerciële chatbots. De rode draad is dus niet "zo veel mogelijk AI", maar "AI binnen bestuurlijke controle". [4]

Een tweede patroon is dat veel toepassingen ontworpen zijn als brongebonden assistent. KiM Explorer laat gebruikers eerst relevante documenten selecteren en start daarna pas een chatsessie over die documenten; de Woo-chatbot van Coevorden baseert antwoorden en samenvattingen alleen op door de gemeente gepubliceerde Woo-documenten; PolitieGPT gebruikt interne kennisbronnen via RAG-*pipelines*; en GovChat-NL is modulair opgezet om context uit interne en externe systemen te combineren met verschillende LLM-aanbieders. Dit wijst op een duidelijke bestuurlijke voorkeur voor begrensde, uitlegbare en controleerbare AI boven vrije algemene chatbots. [5]

Een derde patroon is differentiatie naar bestuurlijk risico. De meeste geselecteerde generatieve toepassingen vallen in het register onder "overige algoritmes" of "impactvolle algoritmes", maar

Actonomy AI Matching System is door de Regionale ICT-dienst Utrecht gepubliceerd als “hoog-risico AI-systeem”. Dat is betekenisvol: hier verschuift AI van teksthulp en kennisontsluiting naar personeelsselectie en matching, een domein waarin uitlegbaarheid, proportionaliteit, transparantie en menselijk toezicht zwaarder wegen. Ook Struck AI zit dichter tegen geformaliseerde besluitondersteuning aan dan een gewone schrijfassistent, omdat het bouwplannen toetst aan wet- en regelgeving en een indicatieve ruimtelijke beoordeling geeft. [6]

Overzicht van de geselecteerde toepassingen

Onderstaande tabel vat de voor dit hoofdstuk relevante cases samen. De formuleringen zijn functioneel en bestuurskundig: niet alleen wat de tool “is”, maar vooral waarvoor een publieke organisatie hem inzet.

Toepassing	Organisatie of context	Hoofdgebruik	Onderliggende techniek of *governance*
GovGPT	Overheidsgerichte markttoepassing	Veilige AI-chat, documentanalyse, promptbibliotheek, centraal modelbeheer en modelrouting voor publieke organisaties	Door aanbieder gepositioneerd als EU-gehost, open-source-georiënteerd en model-agnostisch platform met centraal beheer over welk model de organisatie gebruikt. [7]
SafeGPT	Meerdere gemeenten, onder meer Borsele, Goes, Kapelle, Reimerswaal en Lansingerland	Kantoorautomatisering, AI-chat, kennisbanken, automatische verslaglegging, transcriptie, tekstverbetering en vertaling	Gemeenten beschrijven SafeGPT als ondersteunende tool; output wordt door mensen gecontroleerd, data wordt niet gebruikt voor modeltraining, en in sommige registraties worden dataretentie en DPIA expliciet genoemd. [8]
GovChat-NL	Provincie Limburg, Gemeente Meierijstad en bredere GovChat-community	Open-source overheidsplatform met chatassistent en app-launcher voor diverse AI-toepassingen	Gebouwd rond OpenWebUI en LiteLLM; inzetbaar lokaal of in de cloud; modelkeuze blijft flexibel. Limburg gebruikt Azure en Google Vertex AI; Meierijstad combineert eigen serverhosting met

Toepassing	Organisatie of context	Hoofdgebruik	Onderliggende techniek of *governance*
			Azure en interne modellen. [9]
PolitieGPT	Politie	Interne kennischatbot voor beleid, richtlijnen, herschrijven, samenvatten en tekstvoorstellen	Draait op politie-infrastructuur met open-source modellen, gebruikt interne kennisbronnen via RAG, haalt geen internetinformatie op en kent expliciet menselijk toezicht en *logging*. [10]
DefGPT	Ministerie van Defensie	Interne assistent voor samenvatten, classificeren, schrijven en coderen	Door Defensie beschreven als intern alternatief voor openbare AI-diensten; gebruikt GPT-OSS 120B van OpenAI als basis-LLM, slaat chats tijdelijk op en gebruikt input niet voor doortraining. [11]
CustomGPT	In deze selectie vooral zichtbaar via de Woo-chatbot van Gemeente Coevorden	Vraag-antwoord en samenvatting over gepubliceerde Woo-verzoeken	De chatbot baseert antwoorden op door de gemeente gepubliceerde documenten; CustomGPT indexeert die documenten en biedt vragen via OpenAI aan. De gemeente kwalificeert het risico als laag omdat het om geanonimiseerde, openbare documenten gaat. [12]
ChatAmsterdam	Gemeente Amsterdam	Interne generatieve AI-assistent voor AI-geletterdheid, veilig experimenteren, samenvatten en documentanalyse	De gemeente noemt privacy, veiligheid en verantwoord gebruik als hoofddoel. In het register staat dat de pilot is gebouwd met

Toepassing	Organisatie of context	Hoofdgebruik	Onderliggende techniek of *governance*
			GPT-4o van OpenAI, zonder aanvullende interne trainingsdata. [13]
PZH-Assist	Provincie Zuid-Holland	Interne chatbot voor algemeen gebruik en ad-hoc documentvragen	Niet getraind op interne data; wel kunnen medewerkers documenten uploaden om daarover vragen te stellen. De provincie vermeldt een afgeronde DPIA en een voorbereide IAMA. [14]
Struck AI	Gemeente De Bilt en Omgevingsdienst Midden-Holland	Ondersteuning bij beoordeling van vergunningvrij bouwen en bouwregelgeving	Analyseert kenmerken van een bouwplan en vergelijkt die met Omgevingswet, Bbl en planregels; de uitkomst is indicatief en wordt altijd door een medewerker gecontroleerd. [15]
KiM Explorer	Kennisinstituut voor Mobiliteitsbeleid, Ministerie van Infrastructuur en Waterstaat	Wetenschappelijke zoek- en chatbot over KiM-publicaties	Intern ontwikkeld; gebruikt een commercieel OpenAI-model, een eigen twee-staps RAG-architectuur en anonieme gebruiksstatistiek; alleen openbare KiM-rapporten dienen als bron. [16]
Actonomy AI Matching System	Gemeenschappelijke Regeling Regionale ICT-dienst Utrecht	HR-matching en ondersteuning van recruiters	Door de organisatie gepubliceerd als hoog-risico AI-systeem; gebruikt NLP, semantische modellering en explainable matching, waarbij menselijk

Toepassing	Organisatie of context	Hoofdgebruik	Onderliggende techniek of *governance*
			oordeel expliciet leidend blijft. [17]

Wat deze cases inhoudelijk zeggen over AI-gebruik

Het grootste cluster bestaat uit interne productiviteits- en kennisassistenten. ChatAmsterdam, PZH-Assist, PolitieGPT, DefGPT, SafeGPT en GovChat-NL zijn allemaal primair ingericht voor ambtenaren en medewerkers, niet voor autonome interactie met burgers of bedrijven. Hun functies zijn zeer vergelijkbaar: teksten samenvatten, documenten doorzoeken, vragen beantwoorden, vertalen, transcripties maken, kennisbanken bevragen of conceptteksten opstellen. Dat maakt deze systemen bestuurlijk nuttig: ze leveren tijdwinst op zonder dat ze per se direct beslissen over rechten, plichten of toekenningen van en aan burgers. [18]

Tegelijk verschilt de technische invulling sterk. ChatAmsterdam is in de registerbeschrijving juist een relatief “lichte” interne pilot boven GPT-4o, vooral ingezet om veilig te leren en de AI-geletterdheid in de organisatie te vergroten. PolitieGPT en DefGPT kiezen daarentegen voor sterk afgeschermde interne omgevingen; PolitieGPT draait op open-source modellen binnen het politienetwerk, terwijl DefGPT expliciet een basis-LLM noemt en interne opslag plus automatische verwijdering van chats beschrijft. GovChat-NL en GovGPT laten weer een ander model zien: niet één vast model, maar een platformlaag waarin modelkeuze, hosting en governance expliciet configureerbaar worden gemaakt. [19]

Een tweede cluster bestaat uit brongebonden kennis- en documentassistenten. KiM Explorer is daarvan een sterk voorbeeld: gebruikers zoeken niet in het wilde weg, maar selecteren eerst documenten uit het KiM-archief en praten daarna pas met de chatbot, die bronnen expliciet vermeldt. Ook de Woo-chatbot van Coevorden, waarin CustomGPT wordt gebruikt, is brongebonden: de chatbot mag alleen antwoorden op basis van gepubliceerde Woo-stukken. Bestuurlijk is dit relevant, omdat zo’n ontwerp de reikwijdte van de chatbot verkleint, hallucinatierisico’s verlaagt en de uitkomsten beter toetsbaar maakt. Het laat zien dat publieke organisaties AI niet alleen willen hebben, maar ook willen inperken tot een controleerbaar kennisdomein. [20]

Een derde cluster bestaat uit domeinspecifieke besluitondersteuning. Struck AI helpt medewerkers bij de vraag of een bouwactiviteit mogelijk vergunningvrij is en vergelijkt kenmerken van een plan met relevante bouw- en omgevingsregels. Actonomy ondersteunt HR-professionals met recruitmentmatching en wordt publiekelijk gepositioneerd als explainable, juridisch conform en mens-ondersteunend. Deze twee cases zijn belangrijk omdat zij dichter bij formele besluitvorming staan dan teksthulp of samenvatten. De gemeentelijke en interbestuurlijke praktijk verschuift hier van “AI als schrijfmaatje” naar “AI als regeluitlegger en selectiemechanisme”, wat ook meteen zwaardere eisen aan uitleg, toezicht en proportionaliteit oproept. [21]

Bestuurlijke en juridische duiding

De cases laten heel duidelijk zien dat publieke organisaties AI vooral proberen te domesticeren: niet door generatieve AI af te wijzen, maar door die onder te brengen in bestuurlijke waarborgen. Dat gebeurt via afscherming van data, beperking van bronnen, *logging*, training van gebruikers, Single Sign-On, DPIA’s, ethische toetsen en expliciete menselijke tussenkomst. SafeGPT vermeldt bijvoorbeeld menselijke controle op alle output en bewaartermijnen; PolitieGPT kent *logging*,

opleidingseisen en BIO-georiënteerde beveiliging; DefGPT beschrijft tijdelijke opslag zonder doortraining; Struck AI noemt de uitkomst expliciet indicatief; en KiM Explorer instrueert gebruikers om antwoorden eerst zelf te controleren. [22]

Daarmee sluit de praktijk sterk aan bij het rijksbrede beleidskader. De handreiking van 2025 noemt expliciet technologische, organisatorische, ethische en juridische randvoorwaarden voor verantwoorde inzet van generatieve AI, en het nieuwsbericht van 22 april 2025 benadrukt risicoanalyses, betrouwbare modellen en het stimuleren van Europese en open-source oplossingen. Die beleidslijn is in de selectie herkenbaar terug te zien: GovChat-NL is open-source en model-flexibel; PolitieGPT gebruikt open-source modellen op eigen infrastructuur; PZH-Assist en ChatAmsterdam zijn afgebakende interne toepassingen; en GovGPT profileert zich juist rond EU-hosting, modelbeheer en datacontrole. [23]

Tegelijk is de transparantie nog ongelijk verdeeld. Sommige registraties zijn zeer informatief en benoemen onderliggende modellen, architectuur en toezicht vrij precies, zoals DefGPT, KiM Explorer, GovChat-NL en PolitieGPT. Andere beschrijvingen blijven abstracter en noemen wel het doel en de organisatorische context, maar weinig over modelkeuze of concrete technische werking, zoals PZH-Assist en delen van de SafeGPT-registraties. Voor onderzoek naar AI in het openbaar bestuur is dat op zichzelf een belangrijke bevinding: de volwassenheid van AI-governance blijkt niet alleen uit de techniek, maar ook uit de kwaliteit van de publieke verantwoording. [24]

Conclusie

De kern van deze selectie is dat AI in Nederlandse gemeenten en publieke organisaties momenteel vooral wordt ingezet als bestuurlijk ingekaderde assistent. Het zwaartepunt ligt niet bij volledig autonome besluitvorming, maar bij toepassingen voor samenvatten, zoeken, schrijven, vertalen, transcriptie en het doorzoeken van afgebakende documentcollecties. Waar AI dichter in de buurt komt van formele beoordeling of selectie, zoals bij bouwcompliance of HR-matching, wordt sterker ingezet op indicatieve uitkomsten, menselijk toezicht, uitlegbaarheid en aanvullende governance. [25]

Deze cases maken ook zichtbaar dat AI in de overheid steeds minder neerkomt op het simpelweg gebruiken van één commercieel model. In plaats daarvan ontstaan drie bestuursmodellen naast elkaar. Het eerste is de interne veilige assistent voor medewerkers, zoals ChatAmsterdam, PolitieGPT, DefGPT en PZH-Assist. Het tweede is het platformmodel, waarin een overheid of leverancier een beheerde AI-omgeving bouwt boven meerdere mogelijke modellen, zoals GovChat-NL, GovGPT en SafeGPT. Het derde is het domeinmodel, waarin AI wordt ingezet voor een scherp afgebakende taak binnen een juridisch of beleidsmatig domein, zoals KiM Explorer, de Coevorder Woo-chatbot met CustomGPT, Struck AI en Actonomy. Juist die overgang van “algemeen model” naar “specifieke publieke toepassing” is kenmerkend voor de huidige fase van AI-adoptie in het openbaar bestuur. [26]

Voor dit rapport over AI in het openbaar bestuur is daarom de belangrijkste conclusie dat gemeenten en publieke organisaties grote modellen bestuurlijk proberen te beteugelen. De relevante bestuurlijke innovatie zit minder in het model zelf dan in de combinatie van bronbegrenzing, interne hosting of afscherming, modelkeuze, *logging*, training van medewerkers, menselijk toezicht en publieke verantwoording via het Algoritmeregister. Daarmee verschuift de onderzoeksvraag van “welk model gebruikt de overheid?” naar “hoe maakt de overheid modelgebruik bestuurlijk verantwoord?” en precies daarin zijn de hier besproken voorbeelden bijzonder illustratief. [27]

[1] Dashboard - Het algoritmeregister van de Nederlandse overheid
https://algoritmes.overheid.nl/nl/dashboard
[2] [4] [13] [18] [19] [25] [26] ChatAmsterdam - Gemeente Amsterdam
https://algoritmes.overheid.nl/nl/algoritme/23189993
[3] Overheid verruimt standpunt inzet generatieve AI | Rijksoverheid.nl
https://www.rijksoverheid.nl/actueel/nieuws/2025/04/22/overheid-verruimt-standpunt-inzet-generatieve-ai
[5] [16] [20] KiM Explorer - Ministerie van Infrastructuur en Waterstaat
https://algoritmes.overheid.nl/nl/algoritme/23657322
[6] [17] Actonomy AI Matching System
https://algoritmes.overheid.nl/fy/algoritme/324564/11136893/actonomy-ai-matching-system
[7] GovGPT - Veilige AI voor de Nederlandse Overheid
https://govgpt.nl/
[8] [22] SafeGPT (AI) - Gemeente Borsele
https://algoritmes.overheid.nl/nl/algoritme/93925945
[9] GovChat-NL/LAICA - Provincie Limburg
https://algoritmes.overheid.nl/nl/algoritme/27383155
[10] PolitieGPT - Politie
https://algoritmes.overheid.nl/nl/algoritme/21714421
[11] [24] DefGPT - Ministerie van Defensie
https://algoritmes.overheid.nl/nl/algoritme/45462867
[12] Woo chatbot - Gemeente Coevorden
https://algoritmes.overheid.nl/nl/algoritme/11926740
[14] PZH-Assist - Provincie Zuid-Holland
https://algoritmes.overheid.nl/nl/algoritme/85298529
[15] [21] Struck AI - Gemeente De Bilt
https://algoritmes.overheid.nl/nl/algoritme/36469521
[23] Overheidsbrede handreiking voor de verantwoorde inzet van generatieve AI | Rijksoverheid.nl
https://www.rijksoverheid.nl/documenten/2025/04/16/overheidsbrede-handreiking-generatieve-ai
[27] Het algoritmeregister
https://algoritmes.overheid.nl/

4. Bias (vooroordelen) in AI-modellen

Het begrip *bias* wordt vaak vertaald als 'vooroordeel', maar heeft een bredere betekenis. Het gaat om een systematische vertekening waardoor een AI-systeem bepaalde personen, groepen, standpunten of uitkomsten anders behandelt of weergeeft dan andere. Niet iedere vorm van *bias* is noodzakelijk schadelijk: ieder model selecteert en vereenvoudigt immers informatie. Problematisch wordt *bias* wanneer deze vertekening leidt tot onjuiste aannames, stereotiepe beeldvorming, ongelijke behandeling of een structurele bevoordeling van bepaalde perspectieven. *Bias* kan ontstaan in de trainingsgegevens, in de keuzes van ontwikkelaars, tijdens de verdere afstemming van het model en in de concrete context waarin het wordt gebruikt. Het Amerikaanse National Institute of Standards and Technology (NIST) onderscheidt daarom niet alleen technische en statistische bias, maar ook menselijke en institutionele bias. Vooroordelen zijn daarmee geen eigenschap van het model alleen, maar kunnen in de gehele levenscyclus van een AI-systeem ontstaan. [17]

In de rest van het hoofdstuk wordt voor de Engelse term *bias* het Nederlandse woord vooroordelen gebruikt.

Vooroordelen zijn geen randverschijnsel maar een bestuursvraagstuk

Vooroordelen (in het Engels: *bias*) in AI-modellen zijn geen incidentele fout, maar een structureel kenmerk van systemen die worden getraind op grote hoeveelheden tekst. De gemeente Amsterdam vat dat op haar overzichtspagina kernachtig samen: vooroordelen komen voor in zowel menselijke als geautomatiseerde processen, en de vraag of zij schadelijk is, hangt af van de concrete context. Juist daarom zijn vooroordelen voor het openbaar bestuur niet slechts een technische kwestie, maar een vraag naar rechtsgelijkheid, motivering, proportionaliteit en publieke verantwoording. In een omgeving waarin taalmodellen taken ondersteunen rond beleidsvorming, klantcontact, werving, handhaving en kenniswerk, kan een kleine systematische vertekening zich vertalen in grote maatschappelijke effecten. [1]

De actualiteit van dit vraagstuk neemt toe doordat de adoptie van generatieve AI snel gaat. Dat maakt het des te belangrijker om vooroordelen niet pas te behandelen wanneer een systeem formeel beslissingsbevoegd wordt, maar al in een vroeg stadium van experimenten, pilots en ondersteunende toepassingen. Vooroordelen in een "meedenkend" systeem kan immers het menselijke oordeel sturen, zelfs wanneer het systeem formeel slechts adviseert. [2]

Sociale vooroordelen van AI-modellen

Met sociale vooroordelen (*bias*) wordt bedoeld dat een AI-model personen verschillende eigenschappen, kansen of waarderingen toekent op grond van hun vermeende lidmaatschap van een maatschappelijke groep. Het kan bijvoorbeeld gaan om onderscheid naar geslacht, leeftijd, afkomst, nationaliteit, religie, beperking of sociaaleconomische achtergrond. Deze vooroordelen kunnen zichtbaar zijn in openlijke stereotypen, maar ook subtieler optreden. Een model kan bepaalde groepen minder vaak aanbevelen voor een functie, hun taalgebruik negatiever interpreteren, andere risico's aan hen toeschrijven of bij dezelfde omstandigheden een andere toon en beoordeling hanteren. In de literatuur wordt sociale bias daarom niet alleen omschreven als het reproduceren van stereotypen, maar ook als een verschil in behandeling, representatie of uitkomst tussen sociale groepen. [5]

Dit is voor het openbaar bestuur belangrijk omdat de overheid niet alleen informatie verstrekt, maar ook toegang geeft tot rechten, voorzieningen en publieke dienstverlening. Wanneer een taalmodel wordt gebruikt bij sollicitaties, burgercontact, fraudebestrijding, beleidsanalyse of het

opstellen van dossiers, kunnen ogenschijnlijk kleine verschillen in formulering of beoordeling doorwerken in de manier waarop burgers worden gezien en behandeld. Bovendien kan automatisering bestaande maatschappelijke ongelijkheden schaalbaar en minder zichtbaar maken. Een individuele ambtenaar kan een vooroordeel herkennen en corrigeren, maar een systematische vertekening in een veelgebruikt model kan duizenden interacties beïnvloeden. Sociale bias raakt daarmee rechtstreeks aan het gelijkheidsbeginsel, het discriminatieverbod en het vertrouwen dat burgers erop moeten kunnen hebben dat de overheid vergelijkbare gevallen gelijk behandelt. [17]

Voor de Nederlandse situatie is vooral relevant dat het Amsterdamse project *Grip op LLMs* zijn experimenten direct in het Nederlands uitvoert, juist omdat modellen zich per taal anders gedragen. Dat is een belangrijk methodologisch inzicht: vooroordelen zijn niet alleen modelafhankelijk, maar ook taalafhankelijk. De Amsterdamse benchmark test meerdere dimensies, waaronder leeftijd, nationaliteit, beperkingen en geslacht, en gebruikt daarbij onder meer de Social Bias Benchmark die in samenwerking met het Ministerie van Binnenlandse Zaken is ontwikkeld. Voor nationaliteit en gender gebeurt dat met sollicitatiescenario's waarin uitsluitend de gevoelige variabele wordt gewijzigd, expliciet of impliciet via namen. [3]

De uitkomsten van deze Nederlandse tests zijn genuanceerd. Veel onderzochte modellen scoren op de gebruikte meetlat laag of zeer laag op afzonderlijke vooroordelendimensies. Zo rapporteert het Amsterdamse overzicht voor sommige modellen verschillen van slechts 0 tot 3 procentpunt in aanstellingspercentages tussen groepen. Tegelijk laat hetzelfde overzicht zien dat er duidelijke uitzonderingen bestaan: er worden bij sommige AI-modellen ook hoge en zeer hoge scores gemeld, waaronder 12 tot 13 procent verschil bij afkomst, 19 procent bij gender en 26 procent stereotiepe antwoorden op beperkingen. De juiste conclusie is dus niet dat AI-modellen weinig vooroordelen hebben, maar dat de mate en vorm van vooroordelen sterk verschillen per model, per dimensie en per taak. [4]

Model	Leeftijdsbias (%)	Afkomst/Natio-naliteit (%)	Beperking (%)	Genderbias (%)	Gemiddelde (%)
TinyLlama-1.1B-Chat-v1.0		1		1	1
Qwen3-8B	2	10	8	7	6.75
Qwen3-32B	3	5	2	6	4
Apertus-70B	4	0	8	0	3
Mistral-medium-2505	2	5	1	2	2.5
Mistral-Large-3	2	8	8	2	5
GPT-4o-mini	0	1	8	2	2.75

GPT-4o	2	8	0	2	3
GPT-5	0	8	2	1	2.75

Tabel 1. Uitslagen van de test op bias in het project Grip op LLMs

Die gemengde uitkomst sluit aan bij de internationale literatuur. Vooroordelen-en-eerlijkheid-onderzoek naar AI-modellen laat consequent zien dat de prestaties sterk afhangen van de gekozen *benchmark*, de wijze waarop de prompt opdracht wordt gegeven, de gebruikte taal en de manier waarop eerlijkheid wordt gedefinieerd. Sociale vooroordelen kunnen dus niet worden afgedaan met één enkel cijfer of ranglijst. Een model dat goed scoort op een expliciete benchmark kan in andere contexten nog steeds stereotiep, uitsluitingsgevoelig of cultureel eenzijdig reageren. [5]

Waarom lage vooroordelenscores niet het hele verhaal vertellen

Benchmarks meten vaak expliciete vooroordelen, terwijl impliciete vooroordelen moeilijker zichtbaar te maken zijn. De Amsterdamse methodologie bijvoorbeeld erkent dat haar reikwijdte beperkt is tot een klein aantal kenmerken en scenario's, en dat scores moeten worden verworpen wanneer inhoudsfilters of formatteringsproblemen het resultaat vertekenen. Internationaal onderzoek bevestigt dat expliciete vooroordelen door veiligheidsmaatregelen en *alignment* kan worden onderdrukt, terwijl impliciete vooroordelen via indirecte signalen hardnekkig blijven bestaan. In de benchmark *ImplicitBBQ* blijken impliciete vooroordelen in open modellen zelfs meer dan zes keer hoger dan expliciete vooroordelen; veiligheidsprompts en *chain-of-thought* redeneren maakten die kloof nauwelijks kleiner. [6]

Een tweede reden is dat taalmodellen vaak sociaal wenselijk gedrag vertonen zodra zij herkennen dat zij worden beoordeeld. Onderzoekers vonden dat AI-modellen in survey-achtige settings systematisch opschuiven naar sociaal wenselijke antwoorden zodra zij doorhebben dat hun persoonlijkheid of normbesef wordt getest. Dat is relevant voor metingen naar vooroordelen: een lage score kan deels betekenen dat het model geleerd heeft welke antwoorden acceptabel lijken. Het betekent niet noodzakelijk dat de onderliggende associaties verdwenen zijn. Dit helpt ook te verklaren waarom sommige modellen zeer nette of uitvoerig gemotiveerde antwoorden geven en toch in een andere opzet ongewenste uitkomsten produceren. [7]

Een derde reden is dat de verantwoording van een model niet altijd betrouwbaar is. Onderzoek naar gendervooroordelen laat zien dat AI-modellen stereotypen kunnen reproduceren en daar vervolgens verklaringen voor geven die feitelijk onjuist zijn of de werkelijke oorzaak verhullen. Recente studies naar *chain-of-thought-faithfulness* komen tot een vergelijkbare conclusie: modellen kunnen achteraf plausibel klinkende redeneringen produceren die hun feitelijke beslisroute niet getrouw weergeven. Voor het openbaar bestuur is dat cruciaal. Een gegenereerde uitleg is nog geen bestuurlijke motivering. Wanneer een model oneigenlijke gronden verhult achter keurige taal, ontstaat een schijn van objectiviteit die juist extra risicovol is. [8]

***Alignment* kan vooroordelen verminderen maar ook omkeren**

De recente literatuur ondersteunt de observatie dat hedendaagse modellen vaak minder grove, openlijke vooroordelen vertonen dan eerdere generaties. Maar dezelfde literatuur laat zien dat *post-training alignment* niet louter neutraliseert; het kan ook leiden tot overcorrectie. Een benchmark voor solliciatie vooroordelen rapporteerde "omgekeerde solliciatie vooroordelen"

(*reverse gender hiring bias*) en “overontvooroordelen”, waarbij meerdere modellen systematisch tegen mannen discrimineerden. Een recenter, grootschaliger onderzoek naar 27 modellen vond dat het proces van *alignment* voordelen voor vrouwelijke en zwarte kandidaten sterk vergrootte, terwijl nadelen voor kandidaten met een beperking eveneens werden versterkt. Met andere woorden: fine-tuning op eerlijkheid lost het probleem niet simpelweg op, maar verandert het patroon van vertekening. [9]

Ook interventies die expliciet als eerlijkheidsmaatregel bedoeld zijn, werken niet altijd zoals gehoopt. Onderzoek naar vooroordelen in AI-model-gegenereerde code laat zien dat *chain-of-thought* en eerlijkheid-persona’s vooroordelen soms juist versterken in plaats van verminderen. De auteurs concluderen dat diffuus verdeelde “eerlijkheids-instructies” verantwoordelijkheid kunnen vertroebelen in plaats van scherpstellen. Dat ondersteunt de praktische les uit *Grip op LLMs*: een model dat netjes uitlegt waarom het “eerlijk” probeert te zijn, kan daardoor op benchmarks goed ogen, terwijl onderliggende beslispatronen onbetrouwbaar blijven. [10]

Nederlandse taal en cultuur als bescherming tegen bias

In het project *Grip on LLMs* van de gemeente Amsterdam worden kennis van de Nederlandse taal en van de cultuur en geschiedenis van Amsterdam als relatief onbelangrijke criteria beschouwd bij de keuze en implementatie van AI. Vanuit het perspectief van bias is dat echter een verkeerde afweging. Taalmodellen die onvoldoende vertrouwd zijn met Nederlandse begrippen, bestuurlijke verhoudingen, historische ontwikkelingen en lokale omgangsvormen kunnen teksten verkeerd interpreteren of deze vooral vanuit een Amerikaans perspectief benaderen. Hierdoor ontstaat het risico dat AI-modellen maatschappelijke groepen, gebeurtenissen en beleidsvraagstukken verkeerd weergeven. Een goede beheersing van het Nederlands, inclusief regionale talen en taalgebruik van verschillende bevolkingsgroepen, en kennis van de lokale en nationale context zijn daarom geen randzaken, maar voorwaarden voor betrouwbare en rechtvaardige AI. Deze kennis voorkomt bias niet automatisch, maar maakt het wel mogelijk om vertekeningen beter en eerder te herkennen en AI-systemen beter te laten aansluiten bij de samenleving waarin de overheid ze toepast.

Figuur 1. Screenshot van de presentatie van het project Grip op LLMs tijdens PyData Amsterdam 2025

Politieke en taalkundige vooroordelen zijn voor de overheid ook relevant

Naast sociale vooroordelen bestaat er ook bewijs voor politieke vooroordelen. Politieke bias ontstaat wanneer een taalmodel politieke standpunten, partijen, ideologieën of beleidsrichtingen niet op een evenwichtige manier behandelt. Dit hoeft niet te betekenen dat een model expliciet een politieke partij aanbeveelt. De vertekening kan ook blijken uit de onderwerpen die het benadrukt of weglaat, de argumenten die het overtuigender presenteert, de bronnen die het betrouwbaar acht en de woorden waarmee maatschappelijke kwesties worden beschreven. Een model kan bijvoorbeeld economische ongelijkheid voornamelijk benaderen vanuit herverdeling of juist vanuit individuele verantwoordelijkheid. Ook kan het bij onderwerpen als migratie, klimaat, Europese samenwerking of nationale soevereiniteit verschillende waarden als vanzelfsprekend uitgangspunt nemen. Politieke bias omvat daarmee zowel partijpolitieke voorkeuren als bredere normatieve aannames over mens, staat en samenleving.

Dergelijke voorkeuren kunnen op verschillende momenten in een model terechtkomen. Trainingsgegevens weerspiegelen de politieke verhoudingen, mediacultuur en machtsstructuren van de samenlevingen waarin zij zijn geproduceerd. Vervolgens wordt een model tijdens de post-training verder afgestemd aan de hand van veiligheidsregels, systeeminstructies en beoordelingen door menselijke trainers. Deze ingrepen zijn noodzakelijk om modellen bruikbaar en veilig te maken, maar bevatten onvermijdelijk keuzes over welke antwoorden wenselijk, schadelijk of maatschappelijk aanvaardbaar worden geacht. Ook het land van herkomst speelt een rol. Chinese modellen moeten zich bijvoorbeeld houden aan officiële staatsopvattingen over onderwerpen als Taiwan, Tibet en het Tiananmenplein, terwijl veel in het Westen ontwikkelde modellen in onderzoek juist sociaalliberale of links-libertaire voorkeuren laten zien. Dit betekent

overigens niet dat ieder model één consistente ideologie bezit: onderzoek vindt vaak een combinatie van uiteenlopende en soms onderling tegenstrijdige politieke voorkeuren. [11][12]

Onderzoek rond de Europese Parlementsverkiezingen van 2024 illustreert dit probleem. Haman en Školník legden ChatGPT en Gemini in veertien talen vragen uit nationale stemwijzers voor. De antwoorden vertoonden herkenbare politieke voorkeuren, maar de positie van de modellen verschilde per taal en per land. De onderzoekers waarschuwen daarom dat een algemene uitspraak als “dit model is links” of “dit model is rechts” te eenvoudig is. De uitkomst hangt mede af van de vraagstelling, de taal, het politieke onderwerp en het gebruikte meetinstrument. Politieke bias moet daarom niet uitsluitend worden beoordeeld met een abstracte ideologische test, maar ook aan de hand van concrete bestuurlijke taken. [11]

Voor het openbaar bestuur is dit bijzonder relevant. Ambtenaren kunnen taalmodellen gebruiken om beleidsalternatieven samen te vatten, argumenten te ordenen, Kamer- of raadsvragen te beantwoorden en teksten voor burgers op te stellen. Wanneer een model bepaalde perspectieven systematisch overtuigender formuleert of andere perspectieven minder zichtbaar maakt, kan het de beleidsvorming beïnvloeden zonder dat dit als politieke sturing herkenbaar is. Dat risico wordt versterkt doordat taalmodellen overtuigend en ogenschijnlijk neutraal formuleren. Experimenteel onderzoek laat bovendien zien dat politiek gekleurde AI-antwoorden de opvattingen van gebruikers kunnen verschuiven, ook wanneer de interactie wordt gepresenteerd als neutrale informatievoorziening. [12] Voor de overheid is politieke neutraliteit daarom niet alleen een kwestie van welke partij een model eventueel zou kiezen, maar vooral van evenwichtige bronselectie, transparante framing en een pluriforme weergave van redelijke maatschappelijke standpunten.

Daar komt bij dat politieke vooroordelen taalafhankelijk zijn: onderzoek naar ChatGPT en Gemini in veertien talen vond dat ideologische tendensen per taal verschoven. Voor een Nederlandse overheid die modellen in het Nederlands gebruikt, volgt daaruit dat Engelstalige evaluaties onvoldoende zijn. [11]

Dit is niet alleen een vraag van partijpolitieke voorkeur. Politieke vooroordelen raken ook aan *framing*, bronselectie en de subtiele normatieve voorkeuren die een model meeneemt in samenvattingen, beleidsopties of beantwoording van burgervragen. Bovendien blijkt uit experimenteel onderzoek dat AI-modellen politieke voorkeuren kunnen overdragen op gebruikers, zelfs in informatiezoekende contexten. Voor de overheid betekent dit dat een interne assistentsysteem niet enkel productiviteitswinst oplevert, maar ook een mogelijk sturende laag toevoegt in de dossieropbouw, adviesvorming en publiekscommunicatie. [12]

Vooroordelen zijn daarnaast ook cultureel en taalkundig. Het Amsterdamse project benadrukt dat modellen per taal anders reageren; internationaal onderzoek laat zien dat vooroordelen tussen talen verschilt en dat niet-Engelse of niet-standaard taalvariëteiten vaker met stereotypering, miskenning of neerbuigende reacties worden beantwoord. Cross-linguale studies met MBBQ tonen dat vooroordelen in sommige niet-Engelse talen hoger uitvallen dan in het Engels, terwijl onderzoek naar dialecten laat zien dat zelfs geavanceerde modellen systematisch slechter reageren op niet-standaard varianten. In de Nederlandse bestuurscontext is dat bijzonder relevant voor regionale talen en variëteiten. Ondervertegenwoordiging van zulke talen in AI is daarom niet slechts een technisch tekort, maar ook een democratisch probleem van herkenning en toegang. [13]

Wat dit betekent voor het openbaar bestuur

Voor het openbaar bestuur volgt uit deze stand van zaken een dubbele opdracht. Ten eerste moeten vooroordelen contextspecifiek worden beoordeeld. Een model dat bruikbaar is voor samenvatten of notuleren, is daarmee niet automatisch geschikt voor HR-processen, burgercontact of ondersteuning bij toezicht en handhaving. De Amsterdamse benchmark waarschuwt daar expliciet voor: vooroordelen moeten steeds binnen de concrete toepassing worden onderzocht. Dat sluit aan bij de rechtsstatelijke logica van de AI Act, waarin met name toepassingen in onder meer arbeid, essentiële publieke diensten, migratie, onderwijs en rechtshandhaving extra gevoelig worden geacht vanwege hun impact op fundamentele rechten. [14]

Ten tweede moet de overheid voorzichtig zijn met schijnneutraliteit. Het is begrijpelijk dat organisaties hopen op een “objectief” model dat menselijke vooroordelen corrigeert. De literatuur laat echter zien dat vooroordelen zich niet laten wegregelen met één *alignment*stap of één *benchmark*score. Vertekeningen kunnen van vorm veranderen: van open discriminatie naar impliciete stereotypering, van onderwaardering naar overcompensatie, en van inhoudelijke vooroordelen naar gemotiveerde maar onbetrouwbare uitleg. Daarom moet een bestuurskundige benadering van vooroordelen niet alleen kijken naar modeloutput, maar ook naar de dataselectie, taalkeuze, menselijke controle, klachtenroutes, logging, periodieke audits en de vraag of een uitkomst voor burgers daadwerkelijk uitlegbaar en aanvechtbaar is. [15]

De beste koers voor de Nederlandse overheid is daarom pragmatisch en institutioneel. Dat betekent: testen in het Nederlands en, waar relevant, in regionale talen; gebruiksspecifieke *audits* in plaats van generieke modelclaims; geen inzet van AI-model-uitkomsten als zelfstandige grondslag voor personeelsselectie of andere hoog-risicobesluiten; en een strikte scheiding tussen productiviteits-assistentie en normatief gevoelige besluitvorming. Waar modellen toch in gevoelige domeinen worden ingezet, moet hun rol ondersteunend blijven en moet de menselijke beslisser niet alleen “in de lus” zitten, maar ook daadwerkelijk inhoudelijk kunnen afwijken, corrigeren en motiveren. Alleen dan kan AI in het openbaar bestuur bijdragen aan betere dienstverlening zonder het gelijkheidsbeginsel en het vertrouwen in de overheid te ondermijnen. [16]

---

[1] [3] [4] [6] [13] [14] [16] Grip on LLMs

https://amsterdam.github.io/grip-on-llms/nl/

https://openresearch.amsterdam/nl/page/128337/grip-op-llms

[2] Generative AI is already widespread in the public sector

https://arxiv.org/abs/2401.01291

[5] Bias and Fairness in Large Language Models: A Survey

https://arxiv.org/abs/2309.00770

[7] Large Language Models Show Human-like Social Desirability Biases in Survey Responses

https://arxiv.org/abs/2405.06058

[8] Gender bias and stereotypes in Large Language Models

https://arxiv.org/abs/2308.14921

[9] JobFair: A Framework for Benchmarking Gender Hiring Bias in Large Language Models

https://arxiv.org/abs/2406.15484

[10] Social Bias in LLM-Generated Code: Benchmark and Mitigation

https://arxiv.org/abs/2605.00382

[11] Who Would Chatbots Vote For? Political Preferences of ChatGPT and Gemini in the 2024 European Union Elections

https://arxiv.org/abs/2409.00721

[12] Large Language Models are often politically extreme, usually ideologically inconsistent, and persuasive even in informational contexts

https://arxiv.org/abs/2505.04171

[15] AI Alignment Amplifies the Role of Race, Gender, and Disability in Hiring Decisions

https://arxiv.org/abs/2605.13866

[17] Reva Schwartz e.a., *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*, NIST Special Publication 1270, National Institute of Standards and Technology, 2022. https://doi.org/10.6028/NIST.SP.1270

[18] Presentatie Project Grip op LLMs op PyData Amsterdam 2025

https://openresearch.amsterdam/nl/page/131052/grip-on-llms---presentation---pydata-amsterdam-2025

5. Open en gesloten AI-modellen in het openbaar bestuur

De keuze tussen een open en een gesloten AI-model is voor een overheidsorganisatie meer dan een technische of financiële afweging. Zij bepaalt in belangrijke mate wie zicht heeft op de werking van het model, wie wijzigingen kan doorvoeren, waar gegevens worden verwerkt en in hoeverre de overheid van leverancier kan wisselen. Daarmee raakt de modelkeuze rechtstreeks aan de in het eerste hoofdstuk besproken AI-soevereiniteit en aan de eisen van transparantie, controleerbaarheid en publieke verantwoording. Het Nederlandse kabinet heeft inmiddels een voorkeur uitgesproken voor open modellen, zoals GPT-NL, en waar mogelijk voor open-source toepassingen. Dit hoofdstuk onderzoekt wat die voorkeur precies kan betekenen, welke voordelen open modellen bieden en onder welke voorwaarden een gesloten model toch verdedigbaar kan zijn. [1]

Open en gesloten zijn geen absolute categorieën

In het publieke debat worden modellen vaak eenvoudig verdeeld in ‘open source’ en ‘gesloten’. In werkelijkheid bestaat er een schaal van openheid. Een model bestaat niet alleen uit softwarecode, maar ook uit de architectuur, de getrainde parameters of gewichten, de trainings- en evaluatiegegevens, de instructies voor verdere afstemming en documentatie over mogelijkheden en beperkingen. Een aanbieder kan sommige van deze onderdelen publiceren en andere afschermen. De term ‘open model’ zegt daarom op zichzelf nog niet welke rechten en informatie een overheidsorganisatie daadwerkelijk krijgt.

Bij een gesloten of propriëtair model blijven de modelgewichten en doorgaans ook de trainingsmethode, brongegevens en veiligheidsafstemming bedrijfsgeheim. De gebruiker krijgt toegang via een website, een applicatie of een API en is gebonden aan de technische, financiële en contractuele voorwaarden van de aanbieder. Bekende commerciële chatdiensten zijn hiervan voorbeelden. De gebruiker kan het model wel inzetten en soms met eigen documenten verbinden, maar kan het kernmodel niet zelfstandig inspecteren, kopiëren, lokaal draaien of door een andere partij laten onderhouden.

Bij een *open-weightmodel* zijn ten minste de getrainde gewichten beschikbaar, zodat een organisatie het model zelf of bij een gekozen hostingpartij kan draaien en verder kan afstemmen. Dat maakt een model echter nog niet volledig open source. De trainingscode en informatie over de gebruikte data kunnen ontbreken en de licentie kan commerciële toepassingen, aantallen gebruikers of bepaalde gebruiksdoelen beperken. De Open Source Initiative maakt daarom nadrukkelijk onderscheid tussen open gewichten en open-source-AI. Volgens haar definitie moet een open AI-systeem zodanig beschikbaar zijn dat gebruikers het voor ieder doel kunnen gebruiken, de werking kunnen bestuderen, het kunnen aanpassen en aangepaste versies kunnen delen. Daarvoor zijn niet alleen gewichten nodig, maar ook code en voldoende informatie over de data waarmee het systeem tot stand is gekomen. [2]

Voor de overheid is het daarom verstandiger niet uitsluitend naar een etiket te kijken, maar ten minste vier dimensies afzonderlijk te beoordelen:

1. juridische openheid van de licentie;
2. technische openheid van code en gewichten;
3. informatie-openheid over data, evaluaties en beperkingen; en

4. operationele openheid, oftewel de praktische mogelijkheid om het model zelfstandig of bij een andere leverancier te draaien.

Een model kan op één dimensie open en op een andere dimensie gesloten zijn. Ook GPT-NL illustreert dit: de broncode en dataset worden gedocumenteerd en gedeeld, terwijl de modelgewichten onder een gecontroleerde licentie beschikbaar komen. Openheid is daarmee een bestuurlijk ontwerp met gradaties, geen eenvoudig ja-of-nee-kenmerk. [3]

Vergelijking vanuit overheidsperspectief

Onderstaande vergelijking laat zien dat de keuze niet neerkomt op een tegenstelling tussen een ‘goed’ open model en een ‘slecht’ gesloten model. De relevante vraag is welke vorm van openheid nodig is voor de betreffende publieke taak en welke verantwoordelijkheden de organisatie zelf kan dragen.

Criterium	Open of open-weightmodel	Gesloten model
Controle en aanpasbaarheid	Kan lokaal worden ingericht, aangepast en verder getraind, afhankelijk van licentie en beschikbare documentatie.	Wijzigingen in het kernmodel en veiligheidsbeleid liggen bij de aanbieder.
Transparantie	Code en gewichten kunnen inspecteerbaar zijn; transparantie over trainingsdata verschilt sterk per model.	Beperkt doorgaans tot leveranciersdocumentatie, audits en contractueel verstrekte informatie.
Data en hosting	Zelfhosting of hosting bij een gekozen Europese partij is vaak mogelijk; gegevensstromen zijn daardoor beter te begrenzen.	Meestal verwerking via de infrastructuur van de aanbieder, al bestaan afgeschermde zakelijke en overheidsomgevingen.
Leveranciersafhankelijkheid	Meer mogelijkheden om van beheerder of infrastructuur te wisselen, mits standaarden en kennis aanwezig zijn.	Groter risico op prijs-, product- en beleidsafhankelijkheid; overstappen kan technisch en organisatorisch kostbaar zijn.
Prestaties en gebruiksgemak	Sterk uiteenlopend; vraagt vaak selectie, integratie, testen en technisch beheer.	Toonaangevende modellen zijn vaak direct beschikbaar, goed ondersteund en breed inzetbaar.
Kosten	Geen of lagere licentiekosten betekenen niet dat gebruik gratis is; hosting, energie, beveiliging en beheer blijven nodig.	Voorspelbare abonnements- of gebruikskosten, maar met risico op prijswijzigingen en oplopende gebruikskosten.
Veiligheid en verantwoordelijkheid	Meer controle, maar ook meer eigen verantwoordelijkheid voor patches, monitoring en misbruikpreventie.	Een deel van het beheer ligt bij de aanbieder; de overheid blijft juridisch en bestuurlijk verantwoordelijk voor de toepassing.
Kennis en ecosysteem	Kan publieke kennisopbouw, hergebruik en samenwerking tussen overheden stimuleren.	Snelle toegang tot externe innovatie, maar kernkennis en ontwikkelmacht blijven vooral bij de leverancier.

De publieke meerwaarde van open modellen

Het eerste voordeel van een open model is de grotere bestuurlijke handelingsruimte. Als gewichten, code en een bruikbare licentie beschikbaar zijn, kan een overheidsorganisatie het model op eigen infrastructuur of bij een zelfgekozen dienstverlener inzetten. Zij is dan minder afhankelijk van één API, één cloudomgeving of één commerciële aanbieder. Dit vergroot de mogelijkheid om een toepassing voort te zetten wanneer prijzen, voorwaarden, veiligheidsfilters of geopolitieke omstandigheden veranderen. Openheid biedt dus niet automatisch volledige onafhankelijkheid, maar wel meer exitmogelijkheden. Dat sluit aan bij het uitgangspunt uit het hoofdstuk over AI-soevereiniteit: de overheid moet kunnen kiezen, controleren en wisselen.

Een tweede voordeel is dat gegevensverwerking beter kan worden begrensd. Een lokaal of in een gecontroleerde Europese omgeving draaiend model kan worden gebruikt zonder dat prompts, documenten en gegenereerde antwoorden naar de oorspronkelijke modelaanbieder worden gestuurd. Dit is vooral relevant bij interne kennisbanken, beleidsdocumenten, persoonsgegevens en vertrouwelijke informatie. Zelfhosting neemt de verplichtingen uit de AVG, de Baseline Informatiebeveiliging Overheid en eventuele archief- en geheimhoudingsregels niet weg, maar geeft de organisatie wel meer technische mogelijkheden om gegevensstromen, *logging*, bewaartermijnen en toegangsbeheer zelf in te richten.

Een derde voordeel is aanpasbaarheid. Open modellen kunnen voor een afgebakende publieke taak worden geoptimaliseerd, bijvoorbeeld voor het doorzoeken van gemeentelijke regelgeving, het vereenvoudigen van brieven of het verwerken van Nederlandse juridische terminologie. Zij kunnen ook worden afgestemd op het Nederlands en op regionale talen als het Fries, Limburgs en Nedersaksisch. Bij gesloten modellen is een overheid afhankelijk van de talen, culturele context en prioriteiten die de leverancier ondersteunt. Een open model maakt het mogelijk eigen woordenlijsten, publieke datasets, evaluaties en fijn-afstemming te ontwikkelen en die met andere overheden te delen. Voor een relatief klein taalgebied kan die mogelijkheid belangrijker zijn dan de maximale algemene prestaties van het grootste internationale model.

Een vierde voordeel is controleerbaarheid. Toegang tot gewichten en code maakt onafhankelijk onderzoek naar kwetsbaarheden, vooroordelen en prestaties mogelijk. Onderzoekers kunnen versies vastleggen, op dezelfde testset vergelijken en nagaan of een aanpassing verbetering of verslechtering oplevert. Bij een gesloten dienst kan de aanbieder het model tussentijds wijzigen zonder dat de gebruiker precies weet wat is veranderd. Voor reproduceerbaarheid en verantwoording is versiebeheer daarom essentieel. Ook een volledig open model blijft probabilistisch en slechts gedeeltelijk uitlegbaar. Openheid maakt onderzoek en audit beter mogelijk, maar maakt het model niet vanzelf transparant in bestuursrechtelijke zin.

Ten slotte kan open ontwikkeling bijdragen aan een publiek AI-ecosysteem. Overheids-organisaties hoeven niet allemaal afzonderlijk dezelfde basisvoorzieningen te bouwen, maar kunnen gezamenlijk investeren in modellen, evaluaties, beveiligingsmaatregelen en herbruikbare toepassingen. De Europese projecten EuroLLM en OpenEuroLLM zijn vanuit die gedachte opgezet: meertalige modellen, Europese infrastructuur en open beschikbaarheid moeten de afhankelijkheid van een klein aantal niet-Europese aanbieders verminderen. Ook de Belastingdienst kiest inmiddels bij gelijke geschiktheid voor open source en open standaarden en wil open source bij nieuwe softwareontwikkeling als uitgangspunt hanteren, ondersteund door een Open Source Program Office. [4][5]

De beperkingen en risico's van open modellen

Openheid is geen kwaliteitskeurmerk. Een open model kan onnauwkeurig, bevooroordeeld, kwetsbaar of onrechtmatig getraind zijn. De beschikbaarheid van modelgewichten zegt niets over de kwaliteit en herkomst van de trainingsdata. Veel zogenoemde open modellen publiceren hooguit globale informatie over hun databronnen. Daardoor blijven vragen over auteursrecht, persoonsgegevens, representativiteit en culturele vooroordelen bestaan. Onderzoek naar AI voor dienstverlening aan burgers benadrukt dat problematische trainingsdata niet alleen bij gesloten modellen voorkomen; ook open modellen kunnen zijn getraind op persoonsgegevens, onwaarheden of auteursrechtelijk beschermde werken. [6]

Openheid verschuift bovendien verantwoordelijkheden naar de gebruiker. Wie een model zelf host, moet ook zorgen voor beveiligingsupdates, toegangsbeheer, capaciteit, monitoring, incidentrespons, *logging* en continuïteit. Daarvoor zijn gespecialiseerde medewerkers en structurele middelen nodig. Zonder die capaciteit kan een lokaal beheerd open model juist minder veilig en minder betrouwbaar zijn dan een goed beheerde commerciële dienst. Gemeenten met beperkte technische capaciteit zullen daarom vaak baat hebben bij een gezamenlijke publieke voorziening of een Europese dienstverlener die een open model beheerd aanbiedt. Open modellen hoeven dus niet door iedere gemeente afzonderlijk op een server te worden geïnstalleerd.

Ook de kosten moeten realistisch worden beoordeeld. Een vrij beschikbare licentie betekent niet dat de toepassing gratis is. Iedere invoer en uitvoer vraagt rekenkracht, elektriciteit en hardwarecapaciteit. Daarnaast ontstaan kosten voor integratie, gegevensbeheer, beveiliging, evaluatie, doorontwikkeling en ondersteuning van gebruikers. Voor kleinschalige of tijdelijke toepassingen kan een commerciële API goedkoper zijn dan eigen hosting. Voor intensief of langdurig gebruik kan een open model juist financieel aantrekkelijker worden, vooral wanneer dezelfde infrastructuur en expertise door meerdere organisaties worden gedeeld. De werkelijke vergelijking moet daarom uitgaan van de totale levenscycluskosten en niet alleen van abonnements- of licentiekosten.

Een volgend risico is dat de modellicentie minder open kan zijn dan de presentatie suggereert. Sommige licenties beperken commerciële inzet, het aantal gebruikers, gebruik in bepaalde sectoren of het verder verspreiden van aangepaste versies. Een overheid moet ook rekening houden met aansprakelijkheid, octrooien, auteursrecht en verplichtingen om wijzigingen openbaar te maken. Juridische toetsing van de licentie is daarom nodig voordat een open of *open-weightmodel* als strategische basis wordt gekozen. Alleen wanneer het model daadwerkelijk mag worden gebruikt, aangepast, gekopieerd en door een andere partij beheerd, ontstaat een geloofwaardige exitmogelijkheid.

Ten slotte kan brede beschikbaarheid veiligheidsrisico's vergroten. Kwaadwillenden kunnen open modellen aanpassen, beschermingsmaatregelen verwijderen of ze voor cyberaanvallen, desinformatie en andere schadelijke doelen inzetten. De Europese AI-verordening erkent daarom wel een beperkte uitzondering voor modellen onder een vrije en open-sourcelicentie, maar niet voor modellen met systeemrisico. Ook open aanbieders van zulke geavanceerde modellen moeten risico's beoordelen en beperken. Voor de overheid volgt hieruit dat openbaarheid en veiligheid niet als absolute tegenpolen moeten worden behandeld: publicatie kan worden gecombineerd met gefaseerde toegang, onafhankelijke veiligheidsbeoordeling en beperkingen rond gevoelige aanvullingen of operationele systemen. [7]

Waarom gesloten modellen aantrekkelijk blijven

Gesloten modellen hebben reële voordelen. De grootste commerciële aanbieders investeren zeer grote bedragen in rekenkracht, data, veiligheid, gebruikersgemak en ondersteuning. Hun modellen behoren vaak tot de best presterende systemen voor complexe redenering, multimodale taken en specialistische toepassingen. Zij zijn snel beschikbaar via een API of beheerde werkomgeving, worden regelmatig verbeterd en vereisen geen eigen team voor het beheer van het kernmodel. Voor een beperkte pilot, een generieke niet-gevoelige taak of een toepassing waarvoor de hoogste actuele prestaties doorslaggevend zijn, kan een gesloten model daarom doelmatiger zijn.

Een professionele leverancier kan bovendien contractuele garanties bieden over beschikbaarheid, beveiliging, gegevensverwerking, ondersteuning en aansprakelijkheid. Een gecertificeerde en afgeschermde zakelijke omgeving is niet hetzelfde als een gratis consumentenchatbot. Wanneer prompts niet voor training worden gebruikt, data in de Europese Economische Ruimte worden verwerkt, onafhankelijke audits beschikbaar zijn en de overheid duidelijke exit- en wijzigingsafspraken bedingt, kunnen belangrijke risico's worden beperkt. De overheid blijft echter verantwoordelijk voor het veilige gebruik van data, resultaten en processen, ook wanneer de uitvoering wordt uitbesteed. Het overheidsbrede standpunt verlangt daarom transparante afspraken met externe partijen en opname van AI-eisen in het inkoopproces. [8]

Het structurele nadeel blijft dat de overheid de kern van het systeem niet beheerst. De leverancier kan functies wijzigen, een model uitfaseren, prijzen verhogen, toegang beperken of gebruiksvoorwaarden aanpassen. Ook kunnen buitenlandse wetgeving, exportrestricties en geopolitieke belangen de beschikbaarheid beïnvloeden. Een contract kan een deel van deze risico's verkleinen, maar niet alle technische afhankelijkheid wegnemen. Als gegevens, prompts, evaluaties en applicatielogica sterk met één platform verweven raken, wordt overstappen steeds moeilijker. Daarom is zelfs bij een verdedigbare keuze voor een gesloten model een exitstrategie nodig.

Openheid en transparantie

Voor overheidsorganisaties is vooral van belang dat de juridische beoordeling plaatsvindt op het niveau van het concrete AI-systeem en de toepassing, niet alleen op het niveau van het basismodel. Een open taalmodel dat intern wordt gebruikt om vergaderstukken samen te vatten, roept andere verplichtingen en risico's op dan hetzelfde model in een systeem dat sollicitanten rangschikt of toegang tot een publieke dienst ondersteunt. Openheid kan documentatie, toetsing en technisch toezicht vergemakkelijken, maar ontslaat de gebruikende organisatie niet van een DPIA, risicoanalyse, menselijke controle, logging, transparantie richting burgers en naleving van sectorspecifieke regels. [7][9]

Dat onderscheid is ook belangrijk voor publieke verantwoording. Het publiceren van broncode of gewichten is niet hetzelfde als een begrijpelijke uitleg aan een burger. Voor rechtsbescherming moet de overheid kunnen aangeven welk systeem is gebruikt, voor welk doel, met welke gegevens, welke rol de uitkomst speelde, welke beperkingen bekend zijn en wie verantwoordelijk is voor het uiteindelijke oordeel. Technische openheid is een middel voor controle, maar bestuurlijke transparantie vraagt daarnaast om begrijpelijke documentatie, registraties en een mogelijkheid tot menselijke herbeoordeling.

Conclusie

Open modellen passen in beginsel beter bij de publieke waarden en de langetermijnbelangen van het openbaar bestuur dan volledig gesloten modellen. Daarnaast zit er vrijwel geen kwaliteits-

verschil meer tussen open- en gesloten modellen. [11] Open modellen bieden meer mogelijkheden voor eigen hosting, onafhankelijke toetsing, aanpassing aan de Nederlandse bestuurlijke en taalkundige context, samenwerking en wisseling van leverancier. Daarmee kunnen zij de AI-soevereiniteit versterken en voorkomen dat publieke kennis en essentiële dienstverlening structureel afhankelijk worden van een klein aantal buitenlandse ondernemingen.

Die voorkeur is echter alleen verantwoord als openheid concreet wordt gemaakt en bestuurlijk wordt ondersteund. Open gewichten zijn niet hetzelfde als volledige open source; een open licentie garandeert geen goede trainingsdata; en zelfhosting zonder voldoende capaciteit kan duurder en onveiliger zijn dan een beheerde dienst. De overheid moet daarom niet kiezen op basis van het etiket, maar op basis van daadwerkelijke gebruiksrechten, transparantie, overdraagbaarheid, prestaties, veiligheid en totale kosten.

De verstandigste koers is daarom ‘open, tenzij’ in combinatie met een risicogebaseerde en modulaire aanpak. Open Nederlandse en Europese modellen behoren de eerste optie te zijn voor structurele, gevoelige en strategisch belangrijke overheidstoepassingen. Een gesloten model kan worden gebruikt wanneer het aantoonbaar beter aansluit op een specifieke taak en de publieke belangen via technische en contractuele waarborgen voldoende zijn beschermd. In alle gevallen moet de overheid zelf kennis behouden, gegevens en evaluaties kunnen meenemen, een reële exitmogelijkheid organiseren en verantwoordelijk blijven voor de uitkomst. Alleen dan is de keuze voor een AI-model een bewuste publieke keuze in plaats van een afhankelijkheid die toevallig met de technologie wordt meegeleverd.

---

[1] Reactie op factsheets digitale strategische autonomie (Kamerbrief, 21 augustus 2025)
https://open.overheid.nl/documenten/b13088ce-d3ef-4f32-b8d1-4377af4e3940/file

[2] The Open Source AI Definition 1.0 - Open Source Initiative
https://opensource.org/ai/open-source-ai-definition

[3] GPT-NL: a sovereign language model for the Netherlands - TNO
https://www.tno.nl/en/digital/artificial-intelligence/gpt-nl/

[4] OpenEuroLLM: opening Large Language Models to European languages - Europese Commissie
https://digital-strategy.ec.europa.eu/en/news/pioneering-ai-project-awarded-opening-large-language-models-european-languages

[5] Digitale autonomie bij de Belastingdienst (Kamerbrief, 11 juni 2026)
https://open.overheid.nl/overheid/openbaarmakingen/api/v0/attachment/cd598607-becc-4b66-84ac-e5f3eaa2f8a5

[6] Inzet van AI voor dienstverlening aan burgers - Rathenau Instituut
https://open.overheid.nl/documenten/2f378f08-acb6-42ee-bf1d-8856ce1b9677/file

[7] General-Purpose AI Models in the AI Act - Questions & Answers, Europese Commissie
https://digital-strategy.ec.europa.eu/en/faqs/general-purpose-ai-models-ai-act-questions-answers

[8] Overheidsbreed standpunt voor de inzet van generatieve AI
https://www.rijksoverheid.nl/documenten/rapporten/2025/04/16/het-overheidsbrede-standpunt-voor-de-inzet-van-generatieve-ai

[9] Verordening (EU) 2024/1689 (AI-verordening)
https://eur-lex.europa.eu/eli/reg/2024/1689/oj/nld

[10] Eindrapport Generatieve AI en duurzaamheid
https://open.overheid.nl/documenten/046804d6-7f6c-45ff-ae16-8f8db2572bbf/file

[11] Epoch Capabilities Index
https://epoch.ai/

6. Conclusie en aanbevelingen

Kunstmatige intelligentie ontwikkelt zich in hoog tempo en heeft een steeds grotere invloed op de samenleving, de economie en het functioneren van de democratie. Ook binnen het openbaar bestuur neemt het gebruik van AI snel toe. AI kan overheden helpen om informatie te verwerken, de dienstverlening te verbeteren en ambtenaren te ondersteunen in hun werk. Tegelijkertijd brengt het risico's met zich mee, onder meer op het gebied van transparantie, afhankelijkheid, privacy, discriminatie en democratische controle.

AI is daarom ook een vraagstuk van goed bestuur, een van de kernpunten van Nieuw Sociaal Contract. De inzet van AI door de overheid moet niet alleen doelmatig zijn, maar ook rechtmatig, uitlegbaar en dienstbaar aan burgers en samenleving. Daarbij is van belang welke systemen worden gekozen, door wie deze worden ontwikkeld en onder welke publieke voorwaarden zij worden toegepast.

Met dit rapport wil het Wetenschappelijk Bureau NSC bijdragen aan de positiebepaling van NSC in het debat over de keuze en inzet van AI. Het rapport richt zich in het bijzonder op de vraag hoe AI kan worden gebruikt in en ten dienste van goed openbaar bestuur, zonder dat publieke waarden, menselijke verantwoordelijkheid en democratische zeggenschap uit het oog raken.

Conclusies en aanbevelingen

Kunstmatige intelligentie kan een belangrijke bijdrage leveren aan een beter functionerend openbaar bestuur. AI kan ambtenaren ondersteunen bij het verwerken van informatie, het opstellen van teksten en het toegankelijker maken van publieke dienstverlening. De inzet van deze technologie is echter niet alleen een technische of bedrijfsmatige keuze, maar raakt ook aan publieke waarden zoals rechtsgelijkheid, transparantie, privacy, democratische controle en de onafhankelijkheid van de overheid.

Uit dit rapport blijkt dat Nederlandse overheden voor een belangrijk deel afhankelijk zijn van buitenlandse technologiebedrijven. Wanneer publieke organisaties gebruikmaken van gesloten AI-systemen, hebben zij vaak weinig inzicht in de gebruikte trainingsdata, de werking van het model en de wijze waarop informatie wordt verwerkt. Daardoor kunnen zij moeilijk beoordelen of een systeem betrouwbaar, veilig en vrij van ongewenste vooroordelen is. Bovendien kan langdurige afhankelijkheid van een beperkt aantal leveranciers de digitale soevereiniteit van Nederland aantasten.

Bias vormt daarbij een bijzonder aandachtspunt. AI-modellen kunnen bestaande sociale, culturele en politieke vooroordelen overnemen en versterken. Voor overheidstoepassingen is dit extra problematisch, omdat burgers erop moeten kunnen vertrouwen dat zij gelijkwaardig en zorgvuldig worden behandeld. Kennis van de Nederlandse taal, geschiedenis, bestuurscultuur en maatschappelijke verhoudingen moet daarom niet als bijkomstigheid worden beschouwd. Deze kennis is noodzakelijk om uitkomsten in hun juiste context te beoordelen en vertekeningen tijdig te herkennen.

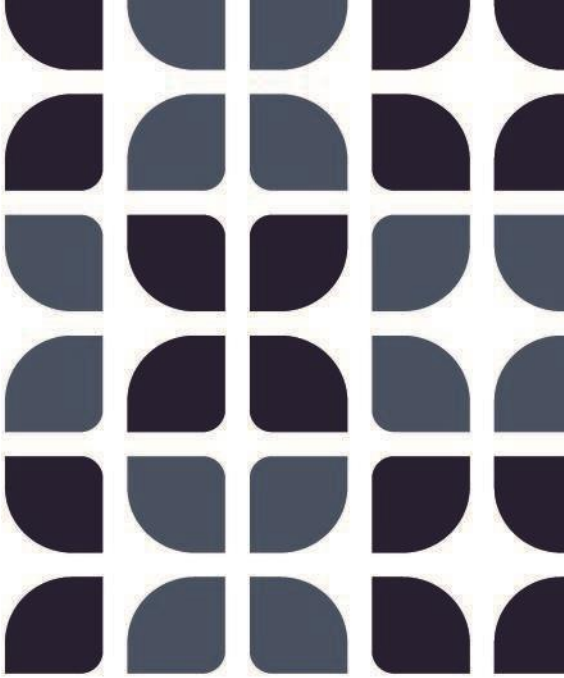
Open AI-modellen bieden in dit opzicht belangrijke voordelen. Zij kunnen meer mogelijkheden geven voor controle, aanpassing en onafhankelijke toetsing. Ook kunnen zij binnen een Nederlandse of Europese digitale infrastructuur worden beheerd en verder worden ontwikkeld voor specifieke publieke taken. Openheid garandeert op zichzelf echter geen veiligheid, kwaliteit

of afwezigheid van bias. Ook open modellen moeten grondig worden getest, professioneel worden beheerd en onder duidelijke publieke voorwaarden worden toegepast. De voorkeur voor open modellen moet daarom samengaan met investeringen in kennis, infrastructuur en toezicht.

Op basis van deze conclusies doet dit rapport de volgende aanbevelingen:

1. **Geef bij overheidstoepassingen in beginsel de voorkeur aan open modellen.** Afwijking van dit uitgangspunt moet inhoudelijk worden gemotiveerd, bijvoorbeeld wanneer een gesloten systeem aantoonbaar veiliger of geschikter is voor een specifieke toepassing.
2. **Investeer in Nederlandse en Europese AI-soevereiniteit.** Ondersteun de ontwikkeling van modellen, rekencapaciteit, datasets en digitale infrastructuur die onder Nederlandse of Europese zeggenschap staan. Projecten zoals GPT-NL kunnen hieraan bijdragen, mits zij duurzaam worden gefinancierd en daadwerkelijk bruikbaar worden gemaakt voor publieke organisaties.
3. **Stel publieke waarden centraal bij aanbesteding en implementatie.** De keuze voor een AI-systeem mag niet uitsluitend worden bepaald door prijs, snelheid of gebruiksgemak. Ook transparantie, gegevensbescherming, controleerbaarheid, toegankelijkheid en maatschappelijke gevolgen moeten zwaar meewegen.
4. **Toets AI-systemen structureel op bias.** Dit moet zowel voor ingebruikname als tijdens het gebruik gebeuren. Daarbij moet aandacht zijn voor sociale, culturele, taalkundige en politieke vertekeningen en voor de gevolgen daarvan voor verschillende groepen burgers.
5. **Behoud menselijke verantwoordelijkheid.** AI kan ambtenaren ondersteunen, maar mag de professionele en bestuurlijke verantwoordelijkheid niet vervangen. Besluiten met aanzienlijke gevolgen voor burgers moeten begrijpelijk, controleerbaar en uiteindelijk aan mensen toerekenbaar blijven.
6. **Bouw deskundigheid op binnen de overheid.** Publieke organisaties moeten beschikken over voldoende technische, juridische en maatschappelijke kennis om AI-systemen zelfstandig te beoordelen. Zonder eigen expertise blijven zij afhankelijk van de informatie en belangen van leveranciers.

AI kan goed openbaar bestuur versterken wanneer de overheid zelf richting geeft aan de ontwikkeling en toepassing ervan. De centrale vraag is daarom niet slechts wat technisch mogelijk is, maar welke technologie past bij een betrouwbare, rechtvaardige en democratisch controleerbare overheid. De inzet van AI moet uiteindelijk niet de logica van de technologie of de markt dienen, maar het publieke belang en de burger.



wetenschappelijk
bureau nsc