

# **Verkenning ChatGPT, overwegingen voor verantwoord gebruik:**

**advies van de interprovinciale ethische  
commissie ten aanzien van LLM's  
binnen de provincies**

**Interprovinciaal Overleg**

Gezamenlijke provincies

[WWW.IPO.NL](http://www.ipo.nl)

---

# Inhoud

|                                                           |    |
|-----------------------------------------------------------|----|
| 1. Advies toepassing LLM's binnen de provincies .....     | 3  |
| 2. Inleiding .....                                        | 5  |
| 3. Samenvatting ethische analyse .....                    | 6  |
| 4. Ethische analyse .....                                 | 9  |
| 5. Verantwoording .....                                   | 18 |
| 6. Interprovinciale Ethische Commissie .....              | 18 |
| 7. Colofon .....                                          | 19 |
| Bijlage 1. Toepassingen ChatGPT binnen de provincie ..... | 20 |
| Bronnen .....                                             | 22 |

---

## 1. Advies toepassing LLM's binnen de provincies

De afgelopen maanden is de wereld in de ban van de snelheid waarmee Large Language Models (LLM's) zich ontwikkelen. Dat hoeft niet te verbazen omdat een LLM-toepassing als ChatGPT ook is gemaakt om tot onze verbeelding te spreken. De samenleving wordt opnieuw het laboratorium van de grote technologiebedrijven, met de wereldburger als proefkonijn. Overheden voelen de druk deze laatste ontwikkelingen in het domein van Artificiële Intelligentie te omarmen.

De interprovinciale ethische commissie heeft daarom de afgelopen periode gewerkt aan een verkenning, die probeert een breed scala aan relevante ethische overwegingen op een rij te zetten ten aanzien van het gebruik van LLM's in het algemeen en van ChatGPT in het bijzonder. Wij willen hiermee laten zien waar provincies afwegingsruimte hebben, en waar men het risico loopt wet- en regelgeving te overtreden, wanneer men zou besluiten deze tools te gebruiken.

Daarbij dient aangetekend te worden dat de ontwikkelingen razendsnel gaan en dit advies slechts het resultaat is van een momentopname. Naast de stappen die tech-bedrijven zetten, wordt er ook op Europees niveau op hoog tempo gewerkt aan nieuwe wetten en gedragscodes, maar ook aan technologische infrastructuur en verantwoorde innovaties. Het is daarom onmogelijk om in deze fase te zien welke veranderingen ethisch het meest problematisch zullen blijken te zijn. Net als bij de maatschappelijke ontwrichtende effecten van sociale media kan een dergelijk eindoordeel over maatschappelijke kosten en baten langer op zich laten wachten.

Allereerst willen we provincies oproepen om de rust te bewaren. Wij begrijpen de behoefte om in proeftuinen te experimenteren, maar er is geen reden om overhaast mee te gaan in overspannen verwachtingen die deze technologie wereldwijd oproept. Deze verkenning geeft overwegingen mee om keuzes te maken op basis van publieke waarden, maar wij erkennen tevens dat provincies dit steeds moeten wegen tegenover een veelheid van belangen in een bestuurlijk en politiek dynamische context.

Dit vraagt van provincies een reflectieve en adaptieve houding. Deze verkenning laat zien dat LLM-technologie momenteel niet transparant is, er grote onzekerheid bestaat of aan alle wettelijke kaders kan worden voldaan, en dat de negatieve externaliteiten bovendien niet volledig te overzien zijn. Wij adviseren daarom in ieder geval ChatGPT voorlopig niet in te zetten wanneer het gaat om privacygevoelige, bedrijfs-kritische, of geheime informatie. Of wanneer er auteursrechtelijke vragen in het spel zijn of wanneer de kwaliteit van output – in welke vorm dan ook – niet door mensen kan worden gecontroleerd of gegarandeerd.

Wij roepen provincies daarnaast op alleen van LLM's gebruik te maken wanneer de volgende overwegingen in acht zijn genomen: Welk belangrijk probleem wordt daadwerkelijk opgelost met de inzet van deze technologie, dat niet op andere wijze kan worden opgelost (noodzaak en proportionaliteit)? Worden er met

---

het gebruik geen andere problemen gecreëerd? Zijn er alternatieven onderzocht die minder impact hebben (subsidiariteit)? Is er een plan hoe om te gaan met de beperkingen van de technologie (zorgvuldigheid)? Is er een duidelijke exit-strategie zodat het mogelijk is snel te stoppen wanneer er problemen optreden (veiligheid)?

De verwevenheid van de Nederlandse overheden met Microsoft vormt daarbij een belangrijk dilemma. Zij zullen vanaf nu Copilot gaan integreren in hun software. De overwegingen uit de ethische analyse zetten echter vraagtekens bij deze ontwikkeling. Wij adviseren de provincies hier niet 'klakkeloos' in mee te gaan, maar gezamenlijk op te trekken in de keuzes die ze hierin kunnen en willen maken vanuit de publieke waarden die daarbij in het geding zijn.

Als commissie zien we dat de samenleving opnieuw in de ban en verblind is door de mogelijkheden van de laatste technologische ontwikkeling en onvoldoende nadenkt over de vraag hoe deze technologie de menselijke verhoudingen en instituties kan veranderen. Wij blijven de ontwikkelingen op de voet volgen, en zullen ons advies bijstellen wanneer daar aanleiding toe is. Ook zullen wij het gesprek met relevante partijen en het IPO-bestuur faciliteren om gezamenlijk na te gaan hoe wij verantwoord kunnen innoveren met generatieve AI in het publiek belang.

Namens de interprovinciale ethische commissie,

Voorzitter, Jeroen van den Hoven

---

## 2. Inleiding

ChatGPT heeft door de gratis gebruiksvaariant van OpenAI sinds november 2022 een groot publiek bereikt en veel mensen kennis laten maken met de mogelijkheden van generatieve AI. Andere partijen zijn ook bezig met het ontwikkelen van vergelijkbare Large Language Models (LLM's), zoals Google's Bard en LLaMa van Meta. Daarnaast heeft Microsoft in maart van dit jaar Copilot aangekondigd. Dit is een AI-applicatie die gebaseerd is op de modellen van OpenAI en zal in Microsoft 365-diensten worden geïntegreerd.<sup>1 2 3</sup>

ChatGPT is een toepassing van generatieve AI. Generatieve AI is een vorm van kunstmatige intelligentie die in staat is om op basis van bestaande datasets en Machine Learning algoritmes nieuwe output te genereren. ChatGPT genereert als Large Language Model *tekst*. Andere vormen van generatieve AI kunnen ook video's, illustraties en spraak genereren. ChatGPT is door OpenAI gratis beschikbaar gesteld en wordt door velen op z'n minst één keer geprobeerd.

Deze verkenning is geschreven om provincies overzicht en overwegingen te bieden voor het nemen van beslissingen over de toepassing van LLM's binnen de provincies. In deze verkenning staan de ethische aspecten van het gebruik van LLM's centraal die op moment van schrijven bekend zijn. Als voorbeeld wordt in dit document ingezoomd op het gebruik van ChatGPT-applicatie van OpenAI in de context van de provincie, om het concreet te maken. De Ethische Commissie heeft hierboven een advies geformuleerd ter ondersteuning van het maken van een keuze over de inzet van LLM's op basis van de nu beschikbare informatie. Het advies is tot stand gekomen op basis van een ethische analyse van de vraagstukken die hieronder uiteen worden gezet.

---

### 3. Samenvatting ethische analyse

ChatGPT van OpenAI staat sinds de lancering in november 2022 in de belangstelling van vele individuele gebruikers. Dit model is gelanceerd in november 2022 voor een groot publiek en roept naast enthousiasme, veel vragen op. Dit vraagt om een analyse van de ethische aspecten van het gebruik van LLM's in het algemeen bij de provincies, en ChatGPT in het bijzonder.

#### Technische aspecten

Large Language Models (LLM's) zijn getraind op omvangrijke datasets die grootschalig gebruik maken van openbare bronnen. ChatGPT is een LLM, en kan grote hoeveelheden tekst verwerken en patronen herkennen én voorspellen. Het kan daardoor teksten genereren waar onderling tussen de tekstdelen samenhang is, en het kan ook ingevoerde teksten verwerken. Aan de voorkant heeft de gebruiker te maken met iets wat lijkt op een geavanceerde chatbot. Op basis van een opdracht ('prompt') genereert ChatGPT *output* in de vorm van tekst. Daarover kan de gebruiker met het systeem verder over in interactie gaan.

De **toepassingsmogelijkheden** van ChatGPT in de provincie zijn velerlei. Het biedt de mogelijkheid om werk met een talig karakter te automatiseren en te versnellen en zou daarbij ondersteunend kunnen werken. De toepassingsmogelijkheden gaan van het genereren van tekst of code tot het structureren of leesbaar maken van teksten. Denk aan het samenvatten van verslaglegging of het genereren van (delen van) beleidsnotities.

Het model kent ook beperkingen. Het belangrijkste kenmerk is de **onbetrouwbaarheid** van ChatGPT. Het model is ten eerste *niet neutraal*, doordat het model waardengeladen is (door de makers en trainers van het model) en bias kent. Ten tweede is er sprake van zogenaamde '*hallucinaties*'. ChatGPT als LLM heeft geen model van de werkelijkheid en produceert taal, geen kennis. Ten derde zijn het model, de trainingsdata en de keuzes in de totstandkoming van het model *intransparant*. De intransparantie is grotendeels een gevolg van de complexiteit van het model en de wijze waarop het is ontworpen. Daardoor is het moeilijk te achterhalen hoe en op basis waarvan het model tot output komt. Deze beperkingen brengen risico's met zich mee naarmate er meer 'autonomie' aan ChatGPT wordt gegeven.

#### Wettelijke kaders

De wettelijke kaders waarbinnen ChatGPT gebruikt zou kunnen worden leveren spanningen op. In verschillende Europese landen loopt er onderzoek naar de schending van de **Algemene Verordening Gegevensbescherming (AVG)** door OpenAI. Reeds geïdentificeerde privacyrisico's bij dit model zijn de zogenoemde prompt leaks en het gebruik van privégegevens (verkregen uit openbare bronnen en door het gebruik van ChatGPT) als trainingsdata. In het kader van de provincie moeten deze privacyrisico's breder worden getrokken naar bedrijfs(gevoelige) gegevens.

Generatieve AI in het algemeen zet de aankomende **Europese AI verordening** onder druk waardoor er onlangs bepaald is dat er aanvullende eisen aan Foundation Models (waaronder ChatGPT) worden gesteld. Er is bij ChatGPT sprake van vraagstukken rondom **Copyrightschending**. Bronnen die openbaar toegankelijk zijn, zijn zonder bronvermelding of toestemming gebruikt als trainingsdata en worden gebruikt om tot output te komen.

---

Het gebruik van ChatGPT moet in de provincie aan de **Provinciewet** en de **ambtseed** voldoen, die beide om een zorgvuldige omgang met gegevens vragen. De vraag is of dit haalbaar is als ChatGPT wordt gebruikt.

De bestaande wettelijke kaders roepen diverse vraagstukken op hoe de provincies zich hiertoe zullen verhouden bij het gebruiken van ChatGPT. Ondertussen zijn er ook oproepen tot nog strengere regulatie en beleid, in Nederland, maar ook daarbuiten.

### Ethische aspecten bij het gebruik van ChatGPT

Het gebruik van ChatGPT bij de provincie betekent iets voor de ambtenaar, de maatschappij en de provincie als organisatie. Het zal impact hebben op de *vaardigheden*, het *vakmanschap* en de *autonomie* van de **ambtenaar**. Er is bijvoorbeeld een risico dat de verantwoordelijkheid (ongemerkt) verlegd wordt en de ambtenaar te veel op ChatGPT gegenereerde teksten of mogelijkheden van de technologie gaat leunen.

De **maatschappij** heeft behoefte aan en baat bij een betrouwbare overheid. Bij het maken van rechtvaardige keuzes speelt de invloed van juiste informatie een rol. De onbetrouwbaarheid van het model, dat geen informatie maar *tekst* genereert, en de beperkte mogelijkheid tot verantwoording bij het gebruik van dit model bevat risico's voor de informatiekwaliteit en het effect daarvan op publieke waarden en mensenrechten. Dit kan leiden tot onbetrouwbaar handelen van **de provincie** en heeft gevolgen voor het in de provincie gestelde vertrouwen. Een ander effect is de mogelijke invloed op de arbeidsinzet en de verandering van rollen en roept vragen op over de toekomstbestendigheid van de organisaties. Een belangrijke ethische vraag is in wiens belang het inzetten van deze technologie is en of dit niet conflicteert met andere (morele) verplichtingen en wettelijke taken die de provincie heeft.

### Verborgene kosten

Er zijn ook ethische aspecten die gepaard gaan met de *inzet* en de ontwikkeling van deze modellen. De reeds genoemde ethische vraagstukken kunnen al worden gezien als 'kosten' voor de samenleving. Daarnaast is er sprake van andere 'verborgene kosten'. Het gebruik van ChatGPT heeft een flinke **klimaat**impact die conflicterend is met de provinciale plicht tot natuurbescherming. Het gebruik van ChatGPT draagt bovendien bij aan een ongewenste **machts**concentratie bij Big Tech en heeft door het technische ontwerp risico's voor onze publieke waarden. Tot slot maakt het gebruik van en versterkt het economische en sociale **ongelijkheden** door het model te trainen en bij te laten sturen door *clickworkers* die onder erbarmelijke omstandigheden dit werk doen. Op basis hiervan zou men kunnen stellen dat er sprake is van een wens tot efficiëntie, maar de vraag is of de efficiëntieslag een illusie is. De last vermindert niet, maar wordt doorgeschoven en komt in andere vormen en bij anderen terug.

### Technologische alternatieven

Er zijn alternatieve **open source** modellen beschikbaar, die in verschillende fasen van ontwikkeling zijn. Deze open source modellen zijn doorgaans transparanter over zaken als trainingsdata en klimaatimpact. De huidige open source modellen zijn op dit moment nog sterk afhankelijk van de toegang die de grote commerciële modellen verlenen aan hun modellen. De prestatie- en inzetmogelijkheden van de alternatieve open source modellen zijn op dit moment niet op het niveau van ChatGPT. Ze bieden bijvoorbeeld nog geen gebruiksvriendelijke *user interface*, vereisen lokale installatie en hebben meer data nodig om ze verder

---

te *finetunen*. Redelijk succesvol zijn kleinere modellen die getraind worden om één specifieke taak uit te voeren. Deze zijn wel gebonden aan scopebepaling en licentiemodellen. De ontwikkelingen gaan snel en mogelijk worden open source modellen steeds geschikter voor grootschalige inzet. Bij de verkenning van een alternatief, zou de overlap met ethische vraagstukken zoals verder in deze analyse geschetst zijn, verder moeten worden onderzocht.

---

## 4. Ethische analyse

In dit hoofdstuk wordt er een beschrijving gegeven de verschillende ethische aspecten van LLM's en ChatGPT in het bijzonder. Allereerst wordt in 4.1 de werking van LLM's beschreven en welke eerste ethische vraagstukken daarmee gepaard gaan. In 4.2 volgt een beschrijving van de wettelijke kaders en context en de ethische vraagstukken die daaruit voortkomen. In 4.3 komen de ethische aspecten bij het *gebruik* van ChatGPT aan bod en in 4.4 de verborgen kosten bij de inzet van ChatGPT.

### 4.1. Technische aspecten

Voordat de ethische aspecten van het gebruik van LLM's worden benoemd, eerst een korte technische toelichting hoe LLM's werken, welke toepassingsmogelijkheden er zijn en welke eerste ethische vraagstukken daarmee gepaard gaan.

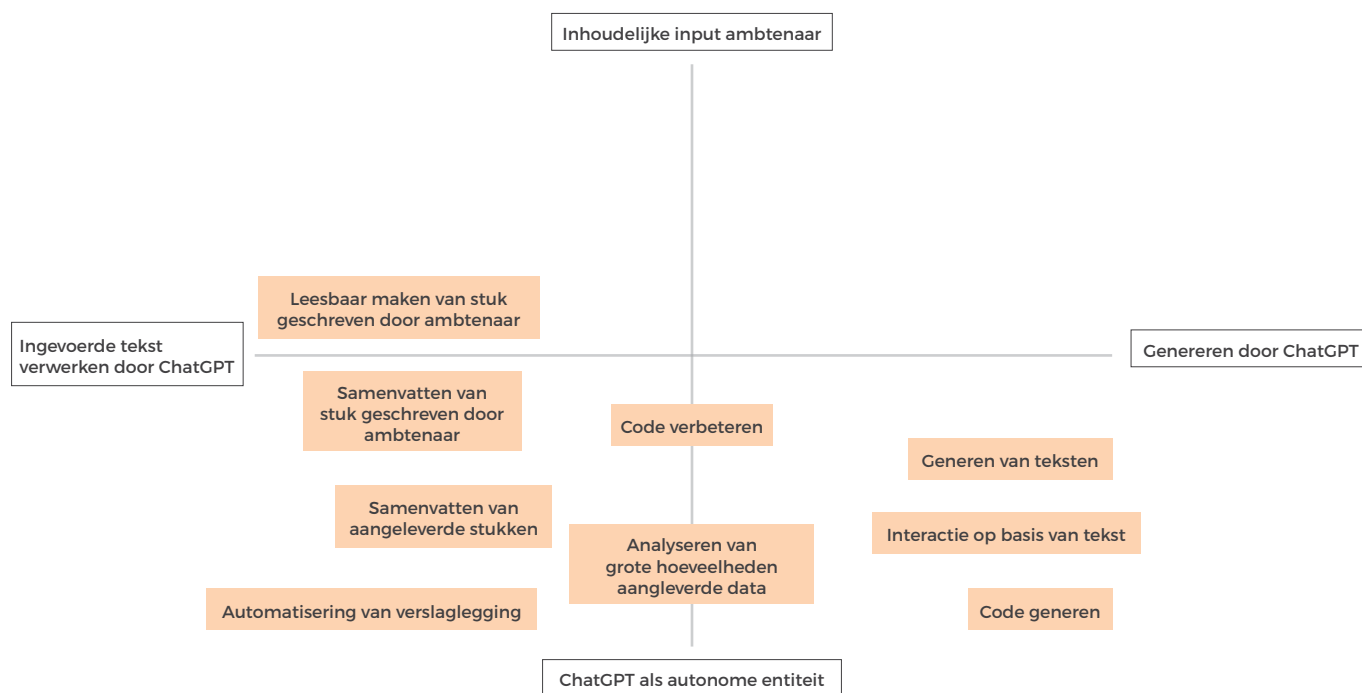
#### 4.1.1. Werking van LLM's

Hoe werken Large Language Models? Eerst de 'achterkant': LLM's zijn getraind op omvangrijke datasets die grootschalig gebruik maken van 'openbare' bronnen.<sup>4</sup> Dat wil zeggen, tekst die op het internet vrij toegankelijk is, maar niet per se 'vrijgegeven' is (verkregen via o.a. *web scraping*). ChatGPT is een LLM. GPT staat voor *Generative Pretrained Transformer*: dit systeem kan grote hoeveelheden tekst verwerken en patronen herkennen in taalgebruik. Het kan vervolgens berekenen welke woorden logisch zouden kunnen volgen en complexe redeneringen neerzetten. Vergelijk het met een autocorrectiesysteem:<sup>5</sup> dat past jouw woorden aan of suggereert een ander woord of spellingswijze. LLM's zijn in staat dit te doen in een meer voorspellende manier van wat er mogelijk zou kunnen volgen na een ander woord en in grotere sequenties, bovendien is het in staat om samenhang tussen meerdere tekstdelen en redeneringen te laten ontstaan. Veel LLM's, waaronder ChatGPT, maken gebruik van *Reinforcement Learning on Human Feedback* (RLHF). Dit betekent dat mensen het model bijsturen door fouten of onwenselijkheden aan te wijzen en ook door aan te geven hoe 'goed' ze de kwaliteit van de resultaten vinden. Deze feedback wordt in LLM's verwerkt om de modellen tot betere resultaten te laten komen. LLM's maken hierbij van grote hoeveelheden parameters gebruik om de relaties tussen woorden te kunnen leggen en ze samen te voegen tot een antwoord. De precieze hoeveelheid is niet bekend (geschat wordt tussen 117 miljoen en meer dan 100 biljoen).

Aan de 'voorkant' van ChatGPT heeft de gebruiker te maken met iets wat lijkt op een geavanceerde chatbot of zoekmachine. De gebruiker voert een 'prompt' in, de opdracht, waar ChatGPT vervolgens een tekst bij genereert. Over die tekst kan de gebruiker met ChatGPT verder een interactie aangaan. Dit is geavanceerder dan een 'gewone' chatbot, omdat ChatGPT teksten genereert in de plaats van voorgeprogrammeerde antwoorden geeft, zoals bij de tot dusver bekende chatbots.

#### 4.1.2. Toepassingsmogelijkheden ChatGPT

ChatGPT biedt meerdere toepassingsmogelijkheden binnen de werkzaamheden van de provincie. Het biedt de mogelijkheid om werk met een talig karakter te automatiseren en te versnellen en zou daarbij ondersteunend kunnen werken. Het is hierbij belangrijk om verschil te maken tussen het *genereren* van teksten en het *structureren* of leesbaar maken van teksten. Zie figuur 1.



Figuur 1: Mogelijke toepassingen in een kwadrant. Op de horizontale as de mate van het verwerken van teksten tot het volledig genereren van teksten; op de verticale as de mate waarin de ambtenaar een inhoudelijke rol heeft tot een meer autonome werking van ChatGPT bij het gebruik van ChatGPT. Voor toelichting en concrete toepassingen, zie Bijlage I.

#### 4.1.3. Technische beperkingen en ethische aspecten daarvan

Bij LLM's zijn er veel mogelijkheden voor toepassing, maar die toepassingen komen niet zonder beperkingen. Dat wil zeggen: een LLM kan veel, maar het kan ook veel *niet* of niet goed (genoeg). De gebruiker wordt het meest blootgesteld aan de beperkingen van het model en de ethische aspecten ervan wanneer die door ChatGPT (volledige) teksten laat genereren (kwadrant rechtsonder fig. 1). Figuur 2 laat zien dat de betrouwbaarheid van het model de belangrijkste beperking is. De ethische aspecten van de technische beperkingen komen vervolgens aan de orde.

| Betrouwbaarheid model |      |                                 |                |                                    |                           |
|-----------------------|------|---------------------------------|----------------|------------------------------------|---------------------------|
| Niet neutraal         |      | 'Hallucinaties'                 |                | Intransparant                      |                           |
| Waarden-geladen       | Bias | Geen model van de werkelijkheid | Taal als basis | Tot standkoming output = Black Box | Trainingsdata = Black Box |

Figuur 2: Ethische aspecten bij de technische werking van ChatGPT

Bij het genereren van teksten is ChatGPT niet betrouwbaar als model. Dat komt door een aantal aspecten. Ten eerste zijn de gegenereerde teksten van ChatGPT zijn **niet neutraal**. Het model is **waardengeladen**, doordat het model getraind is op basis van vele datasets waarin bepaalde (impliciete) waarden, mensen en wereldbeelden onbedoeld en impliciet naar voren komen (**bias**).<sup>6</sup> Daarnaast is het model door *RLHF* bijgestuurd, zodat minder waarschijnlijk wordt dat bepaalde aanstootgevende teksten (extremistische, discriminerende of beledigende) door ChatGPT kunnen worden gegenereerd (deze zijn overigens met bepaalde prompts wel weer te omzeilen).<sup>7</sup>

Ten tweede is er sprake van zogenaamde '**hallucinaties**' van het model. ChatGPT heeft geen 'model van de werkelijkheid' en baseert zich louter op reeds bestaande **taal**(bronnen). Of iets waar is of niet, heeft geen rol in de totstandkoming van teksten door ChatGPT. ChatGPT brengt met net zo veel stelligheid onjuiste informatie als dat het talige uitingen met feiten reproduceert. Het risico op grootschalige verspreiding van onjuiste informatie bij gebruik is reëel.<sup>8</sup>

Ten derde is een factor van onbetrouwbaarheid de **intransparantie** van het model en diens totstandkoming. Het is moeilijk, zo niet onmogelijk, om te achterhalen hoe en op basis waarvan het model tot de output is gekomen. Daarnaast is het voor een model als ChatGPT onduidelijk welke trainingsdata het model allemaal bevat.<sup>9</sup> De intransparantie is grotendeels een gevolg van de complexiteit van het model en de wijze waarop het is ontworpen. Er is nauwelijks tot geen zicht op de voorwaarden waaraan het systeem zou moeten voldoen.<sup>10</sup>

De risico's van het model komen met name tot uitdrukking naarmate het meer als zelfstandige entiteit mag functioneren en ('nieuwe') teksten mag genereren. Risico's zijn o.a. brede verspreiding van mis- en desinformatie, discriminatie en stigmatisering. Een verwerkende talige handeling op een reeds bestaande, door de ambtenaar ingevoerde, tekst is minder risicovol, maar zeker niet zonder risico's, in het licht van de hier genoemde beperkingen.

## 4.2. Wettelijke kaders en context

De inzet van LLM's, en ChatGPT in het bijzonder, raakt aan wettelijke kaders. Wereldwijd is zichtbaar hoe organisaties en overheden hiermee worstelen. Er is waarschijnlijk sprake van schending van de AVG en het auteursrecht, en de vraag is in hoeverre toepassingen van ChatGPT te verenigen zijn met de plichten waaraan een ambtenaar moet voldoen vanuit de Provinciewet en de ambtseed. In Figuur 3 zijn een aantal wettelijke kaders en vraagstukken samengevat, die vervolgens worden toegelicht.

| Wettelijke kaders                      |                                                 |                                                |                                              |                                     |                                                   |                                                            |                                 |                                                                  |                                                         |
|----------------------------------------|-------------------------------------------------|------------------------------------------------|----------------------------------------------|-------------------------------------|---------------------------------------------------|------------------------------------------------------------|---------------------------------|------------------------------------------------------------------|---------------------------------------------------------|
| AVG                                    |                                                 |                                                | AI verordening<br>(nog niet van kracht)      |                                     | Copyright                                         |                                                            | Provincie-<br>wet               | Ambtseed                                                         |                                                         |
| Privacy-<br>risico's voor<br>gebruiker | Privacy-<br>schending<br>in de<br>trainingsdata | Privacy-<br>risico bij<br>invoering<br>prompts | Voldoet<br>niet aan<br>transparantie-<br>eis | Voldoet<br>niet aan<br>labeling-eis | Copyright-<br>schending<br>in de<br>trainingsdata | Onduidelijk-<br>heid over<br>auteursrecht<br>van de output | Geheim-<br>houding<br>ambtenaar | Eis t.a.v.<br>zorgvuldig-<br>heid in<br>omgang met<br>informatie | Plicht om<br>geen valse<br>informatie te<br>verstrekken |

Figuur 3: Relevante wettelijke kaders

---

In april 2023 is het gebruik van ChatGPT in Italië tijdelijk geblokkeerd in afwachting van de beantwoording van de vragen van de Italiaanse toezichthouder GPDP aan OpenAI. Er is gelijktijdig een onderzoek opgestart naar de mogelijke overtreding op de **Algemene Verordening Gegevensbescherming (AVG)**.<sup>11</sup> Ook in Nederland heeft de Autoriteit Persoonsgegevens begin juni 2023 OpenAI om opheldering gevraagd over de omgang met persoonsgegevens bij het trainen van het model.<sup>12</sup> Google kondigde op 13 juni 2023 aan de Europese introductie van AI-Chatbot Bard uit te stellen na kritiek van Ierse privacytoezichthouder DPC over de omgang met persoonsgegevens.<sup>13</sup> Welke privacyrisico's zijn er al bekend? Het invoeren van persoonsgegevens via 'prompts' komt in het model terecht als trainingsdata.<sup>14</sup> Bij de training van het model zijn persoonsgegevens via openbare bronnen ongevraagd en zonder toestemming in het model terechtgekomen als trainingsdata. Mensen hebben hier geen zeggenschap, noch controle over.<sup>15 16 17</sup> Een aantal van deze risico's rondom de omgang met persoonsgegevens kan in het kader van de provincie breder worden getrokken naar bedrijfs(gevoelige)gegevens.

In het najaar van 2023 zal naar verwachting de **Europese AI Verordening** worden vastgesteld, nu het Europees Parlement en de Raad van de EU de laatste onderhandelingen zijn gestart over de eindtekst. In de concept AI Verordening wordt een risicoclassificatie gemaakt van AI-systemen. Afhankelijk van de geïdentificeerde risico's gelden er verboden, tot maatregelen of eisen aan transparantie. De komst van Generatieve AI zet deze risicogebaseerde aanpak onder druk.<sup>18</sup> De modellen zijn grootschalig en niet ontworpen voor één specifiek doel, waardoor de risico's lastig in te schatten zijn.<sup>19 20</sup> Het Europees Parlement wil daarom in de AI verordening aanvullende eisen opnemen voor zogenaamde Foundation Models (waar ChatGPT toe behoort). Dit betekent dat er eisen aan transparantie worden gesteld aan de makers van de modellen rondom de bescherming van 'health and safety', de democratische rechtsorde en het milieu. Verder wordt er gevraagd om gegenereerde content te gaan labelen die door AI gegenereerd is, zodat de uitkomst herleidbaar is voor gebruikers.<sup>21</sup> De Europese AI verordening zal op z'n vroegst in 2025 van kracht zijn.<sup>22</sup>

Er is sprake van vraagstukken rondom **Copyrightschending** (auteursrecht) doordat bronnen die openbaar toegankelijk zijn zonder bronvermelding of toestemming zijn gebruikt als trainingsdata.<sup>23 24</sup> Bij andere vormen van Generatieve AI zijn de eerste rechtszaken hierover al aangespannen, waarbij pogingen van ontwikkelaars om de rechtszaken te laten verwerpen beslist zijn in het voordeel van de eisers.<sup>25 26 27 28</sup> Tegelijkertijd staat het eigenaarschap en auteursrecht ter discussie, omdat het niet duidelijk is of een door de gebruiker gegenereerde tekst in ChatGPT daaronder valt. Men zou kunnen beargumenteren dat de output tot stand komt door de creatieve keuzes en 'originele' keuzes bij de invoering van de 'prompt' door een gebruiker.<sup>29 30</sup> OpenAI geeft aan dat er niet altijd sprake is van auteursrecht omdat bij een gelijksoortige 'prompt' door een andere gebruiker tot een vergelijkbare output kan worden gekomen.<sup>31</sup>

Twee aspecten die ook in ogenschouw moeten worden genomen zijn de verplichtingen van de provincie en die van de ambtenaar. In de **Provinciewet** staat dat de ambtenaar een geheimhoudingsplicht heeft en zorgvuldig met (gevoelige) gegevens moet omgaan en dat wat (nog) niet openbaar mag worden.<sup>32</sup> Door het gebruik van ChatGPT kan de ambtenaar ten onrechte de indruk hebben iets te doen wat in een beschermde omgeving blijft. De ingevoerde tekst wordt echter gebruikt om het model verder te

trainen en kan via ‘prompt leaks’ bovendien weer bij anderen terecht komen. Naast verplichtingen uit de Provinciewet, heeft de ambtenaar een **ambtseed** afgelegd. Daarin belooft de ambtenaar zorgvuldig om te gaan met informatie en heeft diegene een plicht om geen valse informatie te verstrekken. Zorgvuldigheid staat onder druk door het gebruik van ChatGPT en vraagt op z’n minst om controlemiddelen. Wanneer ChatGPT meer autonoom functioneert en teksten genereert is het tot stand komen van valse informatie een reële mogelijkheid (zie 4.1.3).

De bestaande wettelijke kaders geven aan dat het gebruik van ChatGPT problematisch is. Daarnaast zijn er op dit moment **oproepen tot (nog) strengere regulering en beleid** ten aanzien van AI. In maart 2023 publiceerde de Britse overheid een whitepaper voor AI met de oproep aan toezichthouders om betere regels te maken.<sup>33</sup> In juni 2023 lanceerde voormalig D66-Tweede Kamerlid Kees Verhoeven een vergelijkbaar initiatief door een petitie waarin hij de politiek oproept tot het maken van vaart achter de Europese richtlijnen en het opzetten van een nationaal actieplan.<sup>34</sup> In een rapport van de WRR van begin dit jaar worden aanbevelingen aan overheden gedaan om AI beter in te bedden – hieraan valt op dat vooral de nadruk wordt gelegd op het ontwikkelen van vaardigheden en kennis over AI en goede regulering om de controle over de techniek niet te verliezen.<sup>35</sup> Vanuit ethisch perspectief is het de vraag hoe de provincie zich tot deze kaders, ontwikkelingen en context wil verhouden.

### 4.3. Ethische aspecten bij het gebruik van ChatGPT

In 4.1 en 4.2 zijn al verschillende ethische vraagstukken aan bod gekomen. Er wordt nu ingezoomd op de ethische aspecten bij het *gebruik* van ChatGPT en wat dit betekent voor de ambtenaar, de maatschappij en de provincie. Deze aspecten zijn samengevat in Figuur 4 en worden daarna toegelicht.

| Gebruik      |             |           |                      |                  |            |                  |                |            |              |                |
|--------------|-------------|-----------|----------------------|------------------|------------|------------------|----------------|------------|--------------|----------------|
| Ambtenaar    |             |           |                      | Maatschappij     |            |                  |                | Provincie  |              |                |
| Vaardigheden | Vakmanschap | Autonomie | Verantwoordelijkheid | Rechtvaardigheid | Informatie | Publieke waarden | Verantwoording | Vertrouwen | Arbeidsinzet | Verplichtingen |

Figuur 4: Ethische aspecten bij het gebruik van ChatGPT

#### De ambtenaar

Het gebruik van ChatGPT betekent iets voor degenen die ermee gaan werken. Ambtenaren hebben vaardigheden nodig om het model zo nuttig mogelijk te kunnen gebruiken, denk aan training in de invoering van ‘prompts’ of het op waarde kunnen schatten van de output van ChatGPT. Hier is een risico dat er mogelijk een scheiding ontstaat tussen digivaardige en minder digivaardige ambtenaren.<sup>36</sup> Ook zou er aandacht kunnen zijn voor het trainen van *techno-morele vaardigheden*, zodat de ambtenaar een overwogen keuze kan maken voor het inzetten van technologie en wat dat betekent. Het risico dat gepaard gaat met het gebruik van ChatGPT is dat men bepaalde vaardigheden zal gaan verliezen (de-

---

*skilling*). Er kan een afhankelijkheid ontstaan van de technologie die nadelig kan zijn voor de kennis en kunde van de ambtenaar. Dit raakt aan diens **vakmanschap**, en de risico's hangen af van de mate waarin en hoe ChatGPT wordt toegepast (zie fig. 1). Het kritische denkproces dat bijvoorbeeld nodig is om een beleidsstuk tot stand te laten komen, is onderdeel van het vakmanschap van een beleidsmaker en is een proces om tot verstandige keuzes te komen. De ambtenaar moet diens keuzes kunnen afwegen en kunnen onderbouwen. Bij het genereren van teksten door ChatGPT is dat niet mogelijk.

Het gebruik van ChatGPT heeft invloed op de **autonomie** van de ambtenaar. Er kan worden beargumenteerd dat het autonomie versterkt, omdat bij een goede toepassing van een goed model, de ambtenaar mogelijk in staat is tot iets wat diegene eerder niet of minder snel kon. Tegelijk staat de autonomie onder druk. Door het gebruik van ChatGPT bestaat het risico dat de ambtenaar niet alleen een deel van zijn denken maar, door het genereren van teksten, ook **verantwoordelijkheid** (ongemerkt) gaat verleggen doordat diegene zich laat beïnvloeden door dat wat ChatGPT genereert en daar te veel op gaat vertrouwen ('overreliance').<sup>37 38</sup> Onderzoek laat zien dat het moeilijk blijkt om oordelen van anderen geheel terzijde te schuiven, dit is mogelijk ook van toepassing op teksten die gegenereerd zijn.<sup>39 40 41 42</sup> Een mogelijk risico ten aanzien van de autonomie, het vakmanschap en de verantwoordelijkheid van de ambtenaar is de mogelijke onverschilligheid en verlies aan arbeidsplezier door de mogelijke explosieve toename van teksten waar de ambtenaar zelf minder inhoudelijk betrokken bij is, maar wel 'iets mee moet.'

## De maatschappij

Vanuit de maatschappij is er behoefte aan een betrouwbare overheid waarin publieke waarden geborgd zijn.<sup>43</sup> Een aantal aspecten die raken aan de inzet van ChatGPT bij de provincie en hier van belang zijn om te noemen, zijn rechtvaardigheid, informatie en verantwoording.

Men verwacht door de provincie rechtvaardig te worden behandeld en dat **rechtvaardigheid** een rol speelt in de keuzes die er door de provincie worden gemaakt. Dit betekent dat er gebruik moet worden gemaakt van juiste **informatie**, om tot juiste afwegingen te komen. In 4.1.3 worden de beperkingen van het model zichtbaar en zijn de risico's over het verspreiden van onjuiste informatie aan het licht gekomen. De maatschappij kan de risico's van het gebruik van ChatGPT gaan ervaren, bijvoorbeeld doordat er sprake is van onjuiste informatieverstrekking of als gevolg van bias en de gerepresenteerde waarden in het model. Als de provincie door het gebruik van ChatGPT onjuiste informatie verstrekt of (ongemerkt) onjuiste of door het model 'waardengeladen' informatie gebruikt bij het maken van keuzes, zal de maatschappij of zullen groepen uit de maatschappij het vertrouwen verliezen en de provincie niet meer als betrouwbaar zien.<sup>44</sup> Dit brengt risico's mee voor functioneren van de provincie – en de overheid als geheel. De in 4.2 besproken privacyrisico's en inbreuken bij het gebruik van ChatGPT, zowel door de herkomst van trainingsdata als in de prompts, doen af aan **publieke waarden** en mogelijk ook aan mensenrechten. De maatschappij verwacht van de overheid in het algemeen, en de provincie in het bijzonder, dat er zorgvuldig met diens (persoons)gegevens of (bedrijfs)informatie wordt omgegaan.

**Verantwoording** heeft een belangrijke rol in de relatie tussen de overheid en de maatschappij: de provincie, en de individuele bestuurder in het bijzonder, moet enerzijds inhoudelijk kunnen beargumenteren waar keuzes, beleid, beslissingen op zijn gebaseerd, anderzijds moet de provincie kunnen uitleggen hoe het proces van besluitvorming is verlopen, met in achtneming van publieke waarden en grondrechten. Bij het gebruik van gegenereerde teksten is dit niet of slechts ten dele mogelijk.

## Provincie

Door het automatiseren van taken kunnen taken die nu door mensen worden uitgevoerd mogelijk gedeeltelijk of geheel worden overgenomen door een LLM's. Dit betekent mogelijk een verandering in de **arbeidsinzet** in de provincie. Het is onduidelijk of en met welke snelheid de arbeidsmarkt zal veranderen. Er wordt verwacht dat het proces niet voor alle banen op eenzelfde manier verlopen.<sup>45</sup> Rollen zullen veranderen, verdwijnen en mogelijk zullen er ook nieuwe rollen bij komen, passend bij de plek die AI krijgt toegedicht en wat dat van de organisatie en ambtenaren vraagt.<sup>46</sup> De komst van Generatieve AI roept vragen op over hoe toekomstbestendig de provincies als organisaties zijn. Wat vraagt het van de organisatie om voor te bereiden of zorgvuldig afwegingen te maken ten aanzien van verdere digitalisering en automatisering? De **zorgvuldigheid** is gerelateerd aan de publieke taak die gepaard gaat met initieel en essentieel vertrouwen door burgers en andere overheidsinstanties (zie hierboven). Dit vertrouwen brengt verantwoordelijkheid met zich mee om een zorgvuldige afweging te maken over de inzet van technologie in contact met bedrijven en burgers. Op een vertrouwenswaardige manier werken is met name van belang vanwege het informatiemonopolie en machtsmonopolie die de overheid, in dit geval de provincie, heeft ten opzichte van de burger. De maatschappij moet kunnen vertrouwen op het vakmanschap van (individuele) ambtenaren, de publieke taak die ze vervullen en de eed of belofte die zij hebben afgelegd. Het idee dat zij zullen handelen vanuit het algemeen belang en dat publieke waarden zullen worden beschermd. De inzet van ChatGPT door de provincie roept vragen op *in wiens belang* het inzetten is, en of dit niet conflicteert met andere (morele) **verplichtingen** die de provincie heeft (zie ook 4.4).

### 4.4. Verborgen kosten van LLM's

Er zijn ook ethische aspecten die niet direct een rol spelen bij het *gebruik* van LLM's, en ChatGPT in het bijzonder, door de ambtenaar of in de provincie, maar die gepaard gaan met de *inzet* en het bestaan van deze middelen. In eerdere hoofdstukken zijn al verschillende ethische aspecten genoemd die kunnen worden gezien als 'kosten' voor de samenleving (zoals privacy en security). In Figuur 5 worden verdere externaliteiten samengevat weergegeven, daaronder volgt een toelichting.

| Externaliteiten          |                |                  |                             |                          |                      |
|--------------------------|----------------|------------------|-----------------------------|--------------------------|----------------------|
| Klimaat                  |                | Macht            |                             | Ongelijkheid             |                      |
| Plicht tot natuur-behoud | Klimaat impact | Invloed Big Tech | Aantasting publieke waarden | Economische ongelijkheid | Sociale ongelijkheid |

Figuur 5: Externaliteiten, de verborgen kosten

Het trainen (en gebruiken) van ChatGPT zorgt voor een **impact op het klimaat** door flinke CO2-uitstoot en grote hoeveelheden water voor datacenters om servers te koelen.<sup>47 48</sup> Er wordt een groot beroep gedaan op het al overspannen elektriciteitsnetwerk, waarbij doorgaans geen gebruik wordt gemaakt van duurzaam

---

opgewekte elektriciteit. En, er gaat vervuiling gepaard met de productie van de hardware die nodig is voor de servers.<sup>49</sup> Deze servers nemen daarnaast veel ruimte in, waar andere 'functies' plaats voor moeten maken, zoals historische plekken of natuur.<sup>50</sup> Deze nadelige impact op het milieu is niet direct zichtbaar daar waar de modellen gebruikt worden en komt met name terecht bij meer kwetsbare gemeenschappen.<sup>51 52</sup> De provincie heeft de **plicht tot natuurbescherming**,<sup>53</sup> een wettelijke taak om planten en dieren te beschermen. De vraag is in hoeverre aan deze plicht kan worden voldaan met de inzet van een milieubelastend LLM, en of die plicht alleen geldt binnen de provinciegrenzen of ook daarbuiten.

Er wordt **macht** en invloed aan buitenstaanders gegeven door de prompts en teksten die worden ingevoerd. Als daar bedrijfsgegevens, persoonsgegevens of overheidsgegevens in terechtkomen, wordt die beschikbaar voor de ontwikkelaars van de modellen. Dit levert risico's op. Er zou kunnen worden gezegd dat overheden daarmee (onbedoeld of ongemerkt) macht verliezen of weggeven.<sup>54</sup> Via LLM's kan ook ongemerkt invloed worden uitgeoefend daar waar ze toegepast worden.<sup>55</sup> In de totstandkoming van strategie- en beleidsteksten kunnen waarden, mensbeelden en ideologieën naar voren komen die wellicht niet aansluiten bij dat wat in de provincie belangrijk is en bij **publieke waarden**. De vraag is in hoeverre deze mogelijke subtiele beïnvloeding wordt opgemerkt en kan worden tegengegaan.<sup>56</sup>

De totstandkoming van LLM's gebeurt door bestaande **economische en sociale ongelijkheden** te gebruiken en versterken. Clickworkers in 'lage lonen-landen' worden ingezet om het model verder te trainen (RLHF), en ook in Europa en Amerika zijn veel clickworkers actief. Overal ter wereld werken zij tegen extreem lage tarieven. Dit werk is mentaal erg zwaar en soms zelfs traumatiserend.<sup>57 58 59</sup> Naast de inzet van clickworkers tegen lage lonen, waarschuwde het Rathenau Instituut al dat burgers (en ook ambtenaren) zelf gratis de modellen van Silicon Valley trainen. Daarbij is het belangrijk, zo stelt het Rathenau, om voor ogen te houden waartoe AI ontwikkeld wordt en wie daar belang bij heeft.<sup>60</sup> Ontwikkelaars maken nu gratis gebruik van de input van gebruikers om de systemen te verfijnen om daarmee hun marktpositie en kapitaal te vergroten.<sup>61</sup>

Een belangrijk argument vóór het gebruik van ChatGPT is de mogelijkheid tot efficiënter werken. Op basis van de analyse van de verborgen kosten zou men kunnen zeggen dat er sprake is van het verschijnsel van de 'efficiëntie mythe',<sup>62</sup> waarbij de toepassing de last niet vermindert, maar die doorschuift naar andere (onzichtbare) groepen en/of in andere vormen of op andere momenten terugkomt.

## 4.5. Technologische alternatieven

In deze analyse is gekeken naar LLM's in brede context en ChatGPT in het bijzonder. Er zijn alternatieve *open source* modellen beschikbaar, die in verschillende fasen van ontwikkeling zijn.<sup>63</sup>

### Wat is open source?

'Open source' wil zeggen dat zowel de software als de broncode achter de software beschikbaar is voor gebruikers om te bestuderen, maar ook te veranderen en verder te verspreiden. Het kabinetsbeleid stelt dat software die de overheid maakt of laat maken zoveel mogelijk open source moet zijn.<sup>64</sup> Het open source beleid van de overheid richt zich momenteel voornamelijk op software (algoritmes en AI) die de

---

overheid op applicatieniveau aanschaf of laat ontwikkelen. De overheid heeft een inspanningsverplichting om broncode uit eigen beweging openbaar te maken.<sup>65</sup> Open source software is een belangrijk onderdeel van de ontwikkelingen die plaatsvinden in de industrie. Veel bedrijven maken in hun systemen gebruik van (delen van) open source software en sommige toepassingen, zoals Linux en Kubernetes, zijn volledig open source.

### Open source LLM's

Ook op het gebied van AI vindt open source ontwikkeling plaats. Naast grote *proprietary* taalmodellen (gesloten modellen in bezit van grote bedrijven) zoals ChatGPT en Bard, zijn er grote en kleinere open source LLM's te vinden, die transparanter zijn over zaken als trainingsdata en klimaatimpact. Een groter model dat open source beschikbaar is gemaakt, is het LLaMA model van Meta. Dit model was in eerste instantie een *proprietary* model. Veel andere grotere open source modellen zijn gefinetuned op LLaMa, zoals Alpaca, en indirect Vicuna.<sup>66</sup> Momenteel wordt in Duitsland gewerkt aan een model naar Europese waarden, openGPT-X.<sup>67</sup> Dit model bevindt zich nog in de ontwerpfase en kan niet direct ingezet worden. Op het gebied van modellen die getraind zijn op de Nederlandse taal is er RobBERT, een taalmodel gebaseerd op het RoBERTa model.<sup>68</sup>

Hoewel de verwachting is dat open source modellen steeds beter zullen worden,<sup>69</sup> zijn de huidige versies momenteel afhankelijk van de openheid van grote bedrijven over hun *proprietary* modellen, zoals informatie van OpenAI over GPT-3 en toegang die Meta verschaft tot LLaMA. De toegang tot deze informatie wordt steeds meer beperkt in de huidige strijd om het bouwen van het grootste en beste model.<sup>70</sup>

De prestatie- en inzetmogelijkheden van deze modellen zijn op dit moment niet op het niveau van ChatGPT. Veel van de bekende open source modellen bieden (nog) geen goede gebruiksvriendelijke *user interface* en vereisen lokale installatie. Op sites als Hackernews<sup>71</sup> en Reddit komen geluiden vanuit de *developer community* dat gebruik van open-source modellen in praktijk tegenvalt. Oorzaken hiervan zijn zowel *underperformance* t.o.v. *proprietary* modellen en de moeite die het kost om ze te laten draaien op *consumer hardware*. Om een model te krijgen dat vergelijkbaar werkt als ChatGPT, is er veel meer data nodig dan waar de kleinere open source modellen nu op getraind worden.<sup>72</sup>

Kleinere modellen kunnen redelijk succesvol gefinetuned worden om één specifieke taak goed uit te voeren. Vanwege het open karakter van deze modellen is er meestal een duidelijk scope bepaald en zijn bijbehorende licenties toegevoegd. Zo geeft het model BLOOM nadrukkelijk aan dat het verboden is om dit model te gebruiken binnen "*biomedical domains, political and legal domains and financial domains*"<sup>73</sup> en geven zowel Meta als Stanford aan dat LLaMA en Alpaca (gefinetuned op het LLaMA model) uitsluitend bedoeld zijn voor onderzoeksdoeleinden.<sup>74</sup>

De ontwikkelingen gaan snel, en open source modellen worden mogelijk in de (nabije) toekomst steeds geschikter voor grootschalige inzet. Bij de verkenning van een dergelijk technologisch alternatief, zou de overlap met ethische vraagstukken zoals verder in deze analyse geschetst zijn, moeten worden onderzocht.

---

## 5. Verantwoording

Deze verkenning is tot stand gekomen in opdracht van het Interprovinciaal Overleg. Piek Visser-Knijff, Isis Hazewindus en Joeri Hazelaar van Filosofie in actie hebben in juni 2023 vooronderzoek verricht en de ethische analyse geschreven. Ze zijn hierin ondersteund door Jannes Burger (Provincie Zuid Holland), Joas van Ham en Lucas Luitjes. De analyse is tot stand gekomen door middel van desk research, contact met verschillende provincies, het voorleggen van teksten aan de interprovinciale werkgroep ethiek, en door interviews te houden met de leden van de Ethische Commissie en hun kritische meelezen en feedback. Op de vergadering van 20 juni jl. is de ethische analyse besproken. Vervolgens is door de ethische commissie, onder leiding van voorzitter Jeroen van de Hoven, het advies geformuleerd aan provincies.

## 6. Interprovinciale Ethische Commissie

### Over de commissie

De interprovinciale ethische commissie is gericht op het soort vraagstukken dat alle provincies gelijklijk raakt, zoals rond de opkomst van een nieuwe technologie als Artificial Intelligence (AI), die mogelijk van waarde kan zijn voor de uitvoering van provincietaken, of het gebruik van een specifieke technologie binnen de context van een specifiek maatschappelijk vraagstuk, zoals voor slimme oplossingen voor de hierboven genoemde energietransitie.

De werkzaamheden van de commissie zijn nadrukkelijk van *onderzoekende* aard, en dus niet gericht op besluitvorming. Door vraagstukken te onderzoeken kan een interprovinciale ethische commissie adviezen verstrekken die handelingsperspectief bieden aan de provincies bij het maken van hun eigen afwegingen en besluitvorming. De commissie draagt op die manier bij aan het sterker borgen van kernwaarden van het publiek domein in de werkzaamheden van provincies.

### Samenstelling

Jeroen van den Hoven (*voorzitter*)

Anne-Marie Spierings

Erna Ruijter

Esther Keymolen

Nelleke Groen

Roel Dobbe

### Contact

[ethischecommissie@ipo.nl](mailto:ethischecommissie@ipo.nl)

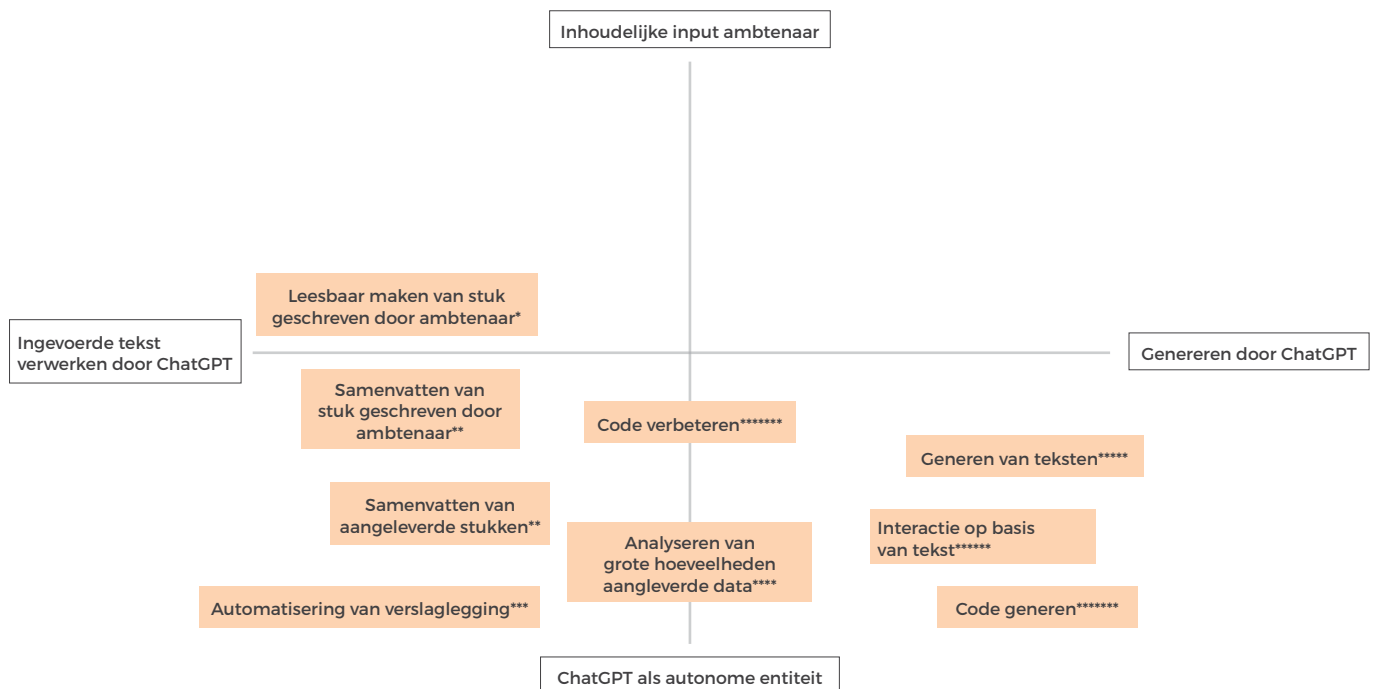
---

## 7. Colofon

De interprovinciale ethische commissie richt zich op morele vraagstukken waarvoor provincies gesteld staan bij de toepassing van digitale technologie op hun maatschappelijke opgaven. Zowel digitale technologie als maatschappelijke opgaven en de context waarin provincies werken zijn voortdurend in beweging. Dit maakt elk advies een momentopname gegeven de technologie, opgaven en context van dat moment. Dit betekent dat de interprovinciale commissie periodiek publicaties en adviezen kan herzien en bijwerken naar de laatste stand van zaken. Daarom zijn verkenningen, handreikingen en adviezen voorzien van versienummers en een datum. De actuele versie van een document zal telkens op de website van het IPO te vinden zijn.

Dit is **“Verkenning ChatGPT, overwegingen voor verantwoord gebruik**: Advies van de interprovinciale ethische commissie ten aanzien van LLM's binnen de provincies”. Versie 1.0, d.d. 3-7-2023

## Bijlage 1. Toepassingen ChatGPT binnen de provincie



Figuur 1 met toelichting. Mogelijke toepassingen in een kwadrant: waarin op de horizontale as de mate van het verwerken van teksten tot het volledig genereren van teksten wordt weergegeven en op de verticale as de mate waarin de ambtenaar een inhoudelijke rol heeft tot een meer autonome werking van ChatGPT.

- \* Leesbaar maken van stuk geschreven door ambtenaar: De ambtenaar zet ChatGPT in om een stuk tekst dat die heeft geschreven te stroomlijnen. Denk hierbij aan het toevoegen van het aanpassen van taalgebruik om de tekst beter leesbaar te maken voor een specifieke doelgroep of het highlighten van belangrijke elementen in de tekst.
- \*\* Samenvatten van stuk geschreven door ambtenaar: ChatGPT kan ingezet worden om stukken samen te vatten. Dit kan zowel gelden voor een individuele ambtenaar die een *eigen geschreven stuk* wil samenvatten voor bijvoorbeeld verspreiding als memo, of over een ambtenaar die *door anderen aangeleverde stukken* wil laten samenvatten om efficiënter te kunnen werken.
- \*\*\* Automatiseren van verslaglegging: OpenAI heeft een open-source spraakherkenningsmodel, Whisper,<sup>75</sup> dat kan worden gebruikt voor het transcriberen van gesproken tekst. Het is niet ondenkbaar dat Microsoft in het Copilot pakket voor Microsoft 365 een optie voor transcriberen van vergaderingen via Teams opneemt. Daarnaast is het mogelijk om notulen van vergaderingen te laten uitwerken of samenvatten door ChatGPT.
- \*\*\*\* Analyseren van grote hoeveelheden data: ChatGPT kan grote hoeveelheden data analyseren en er patronen in te herkennen, de data ordenen, structureren of er bepaalde kenmerken uit filteren (gebruik voor het genereren van zogenaamde “boolean search strings” om de meest geschikte cv’s te selecteren. Hierbij zit overlap met genereren van code).

- 
- \*\*\*\* Genereren van tekst: Het is mogelijk om door ChatGPT allerlei teksten te laten genereren, zoals beleids- of visiestukken, interne communicatie (e-mails, memo's, etc.), externe communicatie (zoals persberichten, (standaard)brieven of toespraken. Daarnaast kan ChatGPT ingezet worden bij het rekruteringsproces van nieuwe werknemers, bijvoorbeeld door vacatureteksten, e-mails en vragen aan sollicitanten te genereren of planningen te maken voor sollicitatiegesprekken.
- \*\*\*\*\* Interactie op basis van tekst: Via de ChatGPT API is het mogelijk een eigen chatbot te bouwen die geïntegreerd kan worden in bijvoorbeeld websites als onderdeel van klantenservice – daaraan gekoppeld kan het helpen om afspraken te plannen. Hier zijn mogelijkheden voor inzet bij een vergunningaanvraag en vergunningverlening, waarbij een uitwisseling van informatie plaatsvindt tussen de vergunningsaanvrager en het systeem. Ook is het eventueel mogelijk om ChatGPT te gebruiken voor het opstellen van gepersonaliseerde adviezen aan burgers of bedrijven.
- \*\*\*\*\* Code genereren of verbeteren: ChatGPT kan overweg met de meeste programmeertalen en zodoende ingezet worden om code zowel te verbeteren als genereren. Bestaande stukken code kunnen ingevoerd worden met het verzoek om deze te debuggen, foutjes te corrigeren (denk aan zaken als vergeten komma's toevoegen of verkeerd gebruik van 'quotation marks' herkennen) of aan te passen om efficiënter te zijn, maar er kunnen ook prompts geschreven worden om het model hele "nieuwe" code te laten genereren (code "snippets").

---

## Bronnen

- 1 <https://blogs.microsoft.com/blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work/>
- 2 <https://www.computerworld.com/article/3691109/microsoft-365-copilot-chatbot-is-the-ai-based-future-of-work.html#:~:text=Microsoft%20365%20Copilot%20is%20generative,the%20wildly%20popular%20ChatGPT%20chatbot>
- 3 <https://www.theverge.com/2023/5/23/23733306/microsoft-ai-365-copilot-edge-build>
- 4 <https://arxiv.org/abs/2201.08239>
- 5 <https://www.wired.com/story/how-chatgpt-works-large-language-model/>
- 6 <https://bioethicstoday.org/blog/chatgpt-ethics-temptations-of-progress/>  
<https://www.sciencedirect.com/science/article/pii/S266734522300024X#sec7>
- 7 <https://www.universiteitleiden.nl/nieuws/2023/03/chatgpt-heeft-linkse-voorkeur-bij-stemwijzer>
- 8 <https://www.nytimes.com/2023/05/01/business/ai-chatbots-hallucination.html>  
<https://www.bloomberg.com/news/newsletters/2023-04-03/chatgpt-bing-and-bard-don-t-hallucinate-they-fabricate>
- 9 <https://towardsdatascience.com/why-we-will-never-open-deep-learning-black-box-4c27cd335118>
- 10 <https://arxiv.org/pdf/2202.09292.pdf>, pg. 11-12
- 11 <https://www.bbc.com/news/technology-65139406>
- 12 <https://autoriteitpersoonsgegevens.nl/actueel/ap-vraagt-om-opheldering-over-chatgpt>
- 13 <https://nos.nl/artikel/2478817-google-stelt-introductie-ai-chatbot-in-europese-unie-uit-na-kritiek-ierse-privacywaakhond>
- 14 <https://www.forbes.com/sites/lanceeliot/2023/01/27/generative-ai-chatgpt-can-disturbingly-gobble-up-your-private-and-confidential-data-forewarns-ai-ethics-and-ai-law/?sh=5d7dd7ce7fdb>
- 15 <https://www.wired.com/story/italy-ban-chatgpt-privacy-gdpr/>  
<https://theconversation.com/chatgpt-is-a-data-privacy-nightmare-if-youve-ever-posted-online-you-ought-to-be-concerned-199283>
- 16 <https://journals.sagepub.com/doi/10.1177/0263775816633195>
- 17 <https://waxy.org/2022/09/ai-data-launders-how-academic-and-nonprofit-researchers-shield-tech-companies-from-accountability/>
- 18 <https://crfm.stanford.edu/2023/06/15/eu-ai-act.html>
- 19 <https://policyreview.info/essay/chatgpt-and-ai-act>
- 20 <https://policyreview.info/essay/chatgpt-and-ai-act>
- 21 <https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence> en  
<https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence>
- 22 <https://artificialintelligenceact.eu/standard-setting/>
- 23 <https://www.groene.nl/artikel/dat-zijn-toch-gewoon-al-onze-artikelen>
- 24 <https://decorrespondent.nl/14592/mijn-boek-is-waarschijnlijk-trainingsmateriaal-voor-chatgpt-mag-dat/6107d0a0-62f0-047a-316e-4169e8a74884>
- 25 <https://www.theverge.com/2022/11/8/23446821/microsoft-openai-github-copilot-class-action-lawsuit-ai-copyright-violation-training-data>
- 26 <https://www.theverge.com/2023/1/16/23557098/generative-ai-art-copyright-legal-lawsuit-stable-diffusion-midjourney-deviantart>
- 27 <https://theconversation.com/chatgpt-is-a-data-privacy-nightmare-if-youve-ever-posted-online-you-ought-to-be-concerned-199283>

- 
- 28 <https://decorrespondent.nl/14529/kunstmatige-intelligentie-is-voortuitgang-beloven-ai-bedrijven-maar-voor-wie/1858845b-2099-022b-3610-f403316d2e43>
- 29 <https://openai.com/policies/terms-of-use>
- 30 [https://intellectual-property-helpdesk.ec.europa.eu/news-events/news/intellectual-property-chatgpt-2023-02-20\\_en](https://intellectual-property-helpdesk.ec.europa.eu/news-events/news/intellectual-property-chatgpt-2023-02-20_en)
- 31 <https://openai.com/policies/terms-of-use>
- 32 Artikel 23 <https://wetten.overheid.nl/BWBR0005645/2023-04-01>
- 33 <https://www.cnn.com/2023/03/29/with-chatgpt-hype-swinging-uk-government-urges-regulators-to-come-up-with-rules-for-ai.html>
- 34 [www.Aipetitie.nl](http://www.Aipetitie.nl)
- 35 <https://english.wrr.nl/topics/artificial-intelligence/documents/reports/2023/01/31/mission-ai-the-new-system-technology>
- 36 <https://www.utwente.nl/en/news/2023/5/673501/how-digitalisation-is-exacerbating-social-inequality#intelligent-devices-but-not-the-right-decisions>
- 37 [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3452052](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3452052)
- 38 <https://hai.stanford.edu/news/ai-overreliance-problem-has-explanations-solution>
- 39 <https://www.nature.com/articles/s41598-023-31341-0>
- 40 Proefschrift "Mores prudentie en het leren van organisaties over morele vraagstukken" Henriette Bout (9 juni 2023)
- 41 [https://www.coe.int/en/web/bioethics/common-ethical-challenges-in-ai#\(%22123745781%22:\[4\]\)](https://www.coe.int/en/web/bioethics/common-ethical-challenges-in-ai#(%22123745781%22:[4]))
- 42 <https://www.uu.nl/sites/default/files/Sankaranetal2020.pdf>
- 43 <https://www.rathenau.nl/nl/digitalisering/publieke-waarden-en-digitalisering>
- 44 <https://direct.mit.edu/daed/article/151/2/309/110609/Distrust-of-Artificial-Intelligence-Sources-and>
- 45 <https://www.washingtonpost.com/opinions/interactive/2023/ai-artificial-intelligence-jobs-impact-research/>
- 46 <https://www.volkskrant.nl/kijkverder/v/2023/wat-verandert-ai-op-de-arbeidsmarkt-deze-machine-is-echt-veel-te-slim-v712524/>
- 47 1) <https://hai.stanford.edu/news/2023-state-ai-14-charts>  
2) <https://www.wired.com/story/the-generative-ai-search-race-has-a-dirty-secret/>
- 48 1) <https://arxiv.org/pdf/2304.03271.pdf>  
2) <https://www.nbcnews.com/tech/internet/drought-stricken-communities-push-back-against-data-centers-n1271344>
- 49 <https://left.eu/content/uploads/2023/02/ENVI-Study-Data-centers-web.pdf>
- 50 1) <https://www.npca.org/advocacy/102-keep-massive-industrial-data-centers-away-from-our-national-parks>  
2) <https://www.theguardian.com/us-news/2023/jun/05/virginia-historic-preservation-data-center-development>
- 51 <https://www.nbcnews.com/tech/internet/drought-stricken-communities-push-back-against-data-centers-n1271344>
- 52 <https://arxiv.org/pdf/2305.05733.pdf> pg. 3-4
- 53 <https://wetten.overheid.nl/BWBR0037552/2021-07-01>
- 54 <https://www.rathenau.nl/nl/digitalisering/ai-bedrijven-gebruiken-burgers-als-proefkonijnen>
- 55 <https://arxiv.org/pdf/2304.08968.pdf>
- 56 [https://www.technologyreview.com/2023/03/14/1069717/how-ai-could-write-our-laws/?utm\\_source=LinkedIn&utm\\_campaign=site\\_visitor.unpaid.engagement&utm\\_medium=tr\\_social](https://www.technologyreview.com/2023/03/14/1069717/how-ai-could-write-our-laws/?utm_source=LinkedIn&utm_campaign=site_visitor.unpaid.engagement&utm_medium=tr_social)
- 57 1) <https://time.com/6247678/openai-chatgpt-kenya-workers/>  
2) <https://www.linkedin.com/pulse/he-helped-train-chatgpt-traumatized-him-alex-kantrowitz%3FtrackingId=1xpu7c1JQ121dbBuawEYNQ%253D%253D/?trackingId=1xpu7c1JQ121dbBuawEYNQ%3D%3D>
- 58 <https://www.volkskrant.nl/nieuws-achtergrond/eindeloos-stoplichten-taggen-en-praten-met-alexa-wie-zijn-de-mensen-achter-ai-bd971b009/>
- 59 <https://www.vpro.nl/programmas/tegenlicht/kijk/afleringen/2023-2024/de-prijs-van-ai.html>
- 60 <https://www.rathenau.nl/nl/digitalisering/ai-bedrijven-gebruiken-burgers-als-proefkonijnen>
-

- 
- 61 1) <https://www.technologyreview.com/2022/04/19/1049378/ai-inequality-problem/>  
2) <https://theconversation.com/ai-will-increase-inequality-and-raise-tough-questions-about-humanity-economists-warn-203056>
- 62 Murray N. Rothbard, "The Myth of Efficiency." In Time, Uncertainty, and Disequilibrium, ed. Mario J. Rizzo (Lexington, Mass.: Lexington Books, 1979), pp. 92–98. see also in The Logic of Action One, pp. 266–73.
- 63 <https://lmsys.org/blog>
- 64 <https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/open-source/beleid/>
- 65 <https://www.rijksoverheid.nl/documenten/rapporten/2022/09/13/opensourcewerken-de-vrijblijvendheid-voorbij>
- 66 <https://bdtechtalks.com/2023/04/17/open-source-chatgpt-alternatives/> Een overzicht van alle open-source modellen is te vinden op hosting platform [Hugging Face](https://huggingface.co/), een leaderboard van modellen is te vinden via [LMSYS](https://lmsys.org/).
- 67 <https://opengpt-x.de/en/project/>, <https://scads.ai/education/summer-schools/summer-school-2023/developing-multilingual-open-and-european-large-language-models/>
- 68 <https://pieter.ai/robbert/>
- 69 <https://thenewstack.io/open-source-movement-emerging-in-ai-to-counter-greed/>
- 70 <https://www.technologyreview.com/2023/05/12/1072950/open-source-ai-google-openai-eleuther-meta/>, <https://venturebeat.com/ai/with-a-wave-of-new-llms-open-source-ai-is-having-a-moment-and-a-red-hot-debate/>, <https://www.technologyreview.com/2023/03/14/1069823/gpt-4-is-bigger-and-better-chatgpt-openai/>
- 71 Voorbeeld: <https://news.ycombinator.com/item?id=35141446> N.B. vanwege een protest bij Reddit zijn de kanalen momenteel niet zichtbaar voor niet-gebruikers. Zie ook: <https://analyticsindiamag.com/why-bloom-didnt-blossom/>, <https://arxiv.org/pdf/2303.01911.pdf>
- 72 <https://arxiv.org/pdf/2305.15717.pdf> (pg. 4-7)
- 73 <https://huggingface.co/bigscience/bloom> (kopje "How to use")
- 74 <https://crfm.stanford.edu/2023/03/13/alpaca.html>, <https://huggingface.co/decapoda-research/llama-7b-hf> (Kopje "Intended Use")
- 75 <https://openai.com/research/whisper>





Telefoon: 070 888 1212  
E-mail: [Secretariaat-IDA@ipo.nl](mailto:Secretariaat-IDA@ipo.nl)



Den Haag  
Herengracht 23