

Privacyverklaring

GPT-NL Project

TNO vindt het belangrijk dat uw privacy goed wordt gewaarborgd. In deze privacyverklaring leggen we uit hoe we persoonsgegevens verzamelen en verwerken in het kader van het GPT-NL Project.

Het GPT-NL Project

TNO (*de Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek*) heeft van de Nederlandse overheid een onderzoeks- en innovatie subsidie ontvangen om in het nationaal belang een soeverein Nederlands taalmodel te ontwikkelen dat voldoet aan Europese en Nederlandse regels en publieke waarden (het “**GPT-NL Model**”).

Het GPT-NL Model is bedoeld om toepassingen van *large language models* (LLMs) in Nederland mogelijk te maken die veilig, transparant en betrouwbaar zijn. Denk bijvoorbeeld aan het automatisch samenvatten van Nederlandse teksten, het beantwoorden van vragen of het herschrijven van teksten.

Digitale soevereiniteit. Met de ontwikkeling van een eigen GPT-NL Model wil de overheid bevorderen dat Nederland minder afhankelijk is van buitenlandse LLM-modellen. Het Nederlands is verder een relatief klein taalgebied en daarmee geen prioriteit voor buitenlandse aanbieders. Doordat het GPT-NL Model wordt getraind op originele Nederlandstalige teksten van hoge kwaliteit, kunnen de finesses van de Nederlandse taal en cultuur beter worden geborgd.

Wetgeving en publieke waarden. Het GPT-NL Model wordt ontwikkeld in lijn met Europese en Nederlandse regels en publieke waarden, in het bijzonder de privacyregels en de rechten en belangen van auteursrechthebbenden. Bijvoorbeeld wordt het GPT-NL Model uitsluitend getraind op teksten waarvoor een geldige licentie is verkregen of die vrij zijn van auteursrecht. Waar auteursrechtelijk beschermde content wordt gebruikt wordt geborgd dat de rechthebbenden ook een eerlijke vergoeding voor hun bijdragen verkrijgen.

Samenwerking. Voor het trainen van het GPT-NL Model zijn veel Nederlandstalige teksten nodig. Om voldoende originele Nederlandstalige teksten van hoge kwaliteit te verzamelen werkt TNO samen met partijen die over zulke teksten beschikken, zoals overheidsinstanties, bibliotheken, archieven en uitgevers (de **Deelnemers**).

Privacybescherming. Alvorens het GPT-NL Model te trainen, bewerkt TNO de trainingsteksten om deze daarvoor geschikt te maken. Onderdeel daarvan is dat schadelijke en ongepaste inhoud en alle irrelevante persoonsgegevens worden verwijderd. Het restant van de persoonsgegevens wordt vergaand geanonimiseerd. Voor publieke personen met een eigen Wikipedia-pagina, gelden bijzondere regels zodat het GPT-NL Model bijvoorbeeld antwoord kan geven op vragen zoals: “*Wie is de koning van Nederland?*”

Op deze wijze wordt het GPT-NL Model op een verantwoorde manier getraind en kan het model veilig worden ingezet bij de overheid, het onderwijs en de wetenschap.

Wie zijn verantwoordelijk voor het verwerken van mijn persoonsgegevens?

TNO is verantwoordelijk voor het GPT-NL Project en de verantwoordelijke voor de verwerking van persoonsgegevens die TNO verzamelt en verwerkt in het kader van het trainen van het GPT-NL Model.

Deelnemers zijn verantwoordelijk voor hun selectie van de trainingsteksten en het veilig aanleveren van deze teksten aan TNO.

Als tegenprestatie voor het bijdragen van hun trainingsteksten aan het GPT-NL Project ontvangen Deelnemers een kopie van hun door TNO opgeschoonde trainingsteksten (zie *Hoe worden mijn persoonsgegevens bewerkt?*), een korting op hun licentievergoeding voor gebruik van het GPT-NL Model en een deel uit de opbrengsten van licentievergoedingen. De Deelnemers zijn zelf verantwoordelijk voor het verdere gebruik van deze opgeschoonde trainingsteksten en het gebruik van het GPT-NL Model als licentienemer.

Wat is het doel van de verwerking?

TNO. TNO verwerkt persoonsgegevens om een Nederlandstalig taalmodel te trainen dat voldoet aan Europese en Nederlandse wetgeving en publieke waarden. Het doel van het GPT-NL is om teksten automatisch samen te vatten, vragen te beantwoorden en teksten te verkorten of herschrijven. Hiervoor is het niet nodig om beeld, video, of geluid te verwerken en het GPT-NL Model wordt alleen met teksten getraind.

Voor Deelnemers is het doel van de verwerking een bijdrage leveren aan het trainen van het GPT-NL Model, en vervolgens zelf ook het GPT-NL Model te kunnen gebruiken en mogelijk verder te trainen voor hun eigen doeleinden.

Uit welke bronnen worden trainingsteksten verzameld?

TNO ontvangt trainingsteksten van Deelnemers. Afhankelijk van de Deelnemer betreft dit teksten uit kranten, tijdschriften, in openbare archieven opgeslagen documenten, en informatie van (semi-)overheidsinstellingen.

TNO verzamelt daarnaast zogenaamde *open source* teksten van websites waarvan de rechthebbende heeft aangegeven dat ze door anderen vrijelijk mogen worden gebruikt, zoals de teksten op de websites van de overheid en de Europese Unie en verder teksten waar geen intellectuele eigendomsrechten op rusten, zoals teksten van wetten en rechterlijke uitspraken. TNO hanteert een streng protocol voor het selecteren van open source teksten, en publiceert een overzicht van alle gebruikte bronnen op de website: <https://gpt-nl.nl/>.

De door Deelnemers aangeleverde teksten noemen we samen met de *open source* teksten hierna de “ruwe content”.

Hoe bewerkt TNO de ruwe content voordat het GPT NL Model ermee wordt getraind?

Voordat de ruwe content kan worden gebruikt voor het trainen van het GPT-NL Model, past TNO een zorgvuldig opschoningsproces toe. Alle schadelijke, ongepaste, of irrelevante content wordt verwijderd, zoals informatie uit een roddelblad. Het doel van het GPT-NL Model is het

automatisch samenvatten, verbeteren of verkorten van teksten en beantwoorden van eenvoudige vragen. Daarvoor is niet nodig dat het GPT-NL Model wordt getraind met informatie over mensen die geen publieke rol hebben (“niet-publieke personen”). Van deze niet-publieke personen worden alle persoonsgegevens uit de ruwe content weg gefilterd of geanonimiseerd.

Om vragen te kunnen beantwoorden zoals “*Wie is de koning van Nederland?*”, is het wél nodig dat het GPT-NL Model informatie bevat over mensen met een publieke rol. Voor politici, artiesten of andere personen die een eigen Wikipedia-pagina hebben (“publieke personen”) geldt dat informatie die direct verband houdt met hun publieke rol wordt gebruikt om het GPT-NL Model te trainen. In bepaalde gevallen zal deze informatie ook speciale categorieën persoonsgegevens betreffen, zoals geloof, politieke overtuiging, of gezondheid. Dit betreft dan informatie die in publiek domein is (zoals het feit dat de Paus katholiek is of dat Mark Rutte lid is van de VVD). Andere gevoelige gegevens van publieke personen, zoals contactgegevens, locatiegegevens, of bankgegevens, worden altijd verwijderd of vervangen door willekeurige voorbeelden.

De opgeschoonde ruwe content noemen we hierna de “**Schone Dataset**”. Alleen de Schone Dataset wordt gebruikt voor het trainen van het GPT NL Model.

Een gedetailleerde omschrijving van het opschoningsproces van TNO is opgenomen in het Training Content Protocol, dat beschikbaar is via deze [link](#).

Wat is de grondslag voor verwerking van mijn persoonsgegevens?

TNO is verantwoordelijk voor het uitvoeren van het GPT-NL Project. Op grond van artikel 4 en 5 van de TNO-wet,¹ heeft TNO een wettelijke taak om toegepast technisch-wetenschappelijk onderzoek te verrichten in het algemeen belang. TNO heeft van de overheid een onderzoeks- en innovatie subsidie toegekend gekregen om in het nationaal belang een trainingsinfrastructuur te ontwikkelen en daarmee vervolgens het GPT-NL Model te trainen. De verwerking van persoonsgegevens in het kader van het GPT-NL Project vindt daarom plaats voor toegepast technisch-wetenschappelijk onderzoek in het algemeen belang.

De grondslag hiervoor is artikel 6, lid 1, onder e, van de Algemene Verordening Gegevensbescherming (AVG).

Voor zover TNO *bijzondere categorieën van persoonsgegevens* (zoals politieke voorkeur, geloof of gezondheidsgegevens) verwerkt in het kader van het GPT-NL Project, gebeurt dit op basis van artikel 9, lid 2, onder g AVG, in combinatie met de TNO-wet. Deze verwerking is toegestaan vanwege het algemeen belang bij het ontwikkelen van het GPT-NL Model en deze vindt plaats onder strikte waarborgen ter bescherming van de rechten en vrijheden van betrokkenen.

Voor zover TNO *strafrechtelijke gegevens* verwerkt in het kader van het GPT-NL Project, gebeurt dit op basis van artikel 10 AVG in samenhang met artikelen 24 en 32 onder f van de Nederlandse Uitvoeringswet AVG (“**UAVG**”), waarin specifieke waarborgen zijn opgenomen voor wetenschappelijk onderzoek.

¹ Wet van 19 december 1985, houdende regeling van de Nederlandse Organisatie voor toegepast-natuurwetenschappelijk onderzoek TNO, available at: <https://wetten.overheid.nl/BWBR0003906>.

Een gedetailleerde omschrijving van de waarborgen ter bescherming van betrokkenen zijn opgenomen in het Training Content Protocol, dat beschikbaar is via deze [link](#). De specifieke waarborgen ter bescherming van betrokkenen zijn verder uitgewerkt in de gegevensbeschermingseffectbeoordeling voor het GPT-NL Project (ook wel *data protection impact assessment*, of *DPIA* genoemd), welke beschikbaar wordt gemaakt via de GPT-NL Website.

Deelnemers zijn zelf verantwoordelijk voor het selecteren van trainingsteksten en het overdragen daarvan aan TNO ten behoeve van de training van het GPT-NL Model. Door de vergaande waarborgen die zijn getroffen door TNO ter bescherming van de privacy (zie hiervoor) zal voor Deelnemers deze vertrekking een toegestane “verdere” verwerking van persoonsgegevens betreffen voor wetenschappelijk onderzoek. Artikel 5 lid 1 onder b AVG geeft aan dat deze verdere verwerking is toegestaan, mits voldaan wordt aan de waarborgen voor de verwerking van persoonsgegevens voor wetenschappelijke doeleinden omschreven in artikel 89 AVG, zoals het minimaliseren en anonimiseren van persoonsgegevens. Deze maatregelen zijn door TNO toegepast.

Licentienemers van het GPT NL Model zijn zelf verantwoordelijk voor het bepalen van de grondslag waarop zij persoonsgegevens verwerken. TNO heeft in het ontwerp en de ontwikkeling van het GPT-NL Model de waarborgen toegepast die Europese privacy-toezichthouders verwachten van de makers van LLMs. Door de specifieke waarborgen die TNO treft, borgt TNO dat GPT-NL in lijn met de beginselen van dataminimalisatie, privacy-by-design en transparantie wordt ontwikkeld. Hierdoor wordt het mogelijk gemaakt dat toekomstige licentienemers van GPT-NL zich voor gebruik daarvan kunnen baseren op de grondslag van gerechtvaardigd belang (artikel 6 lid 1 onder f AVG). De licentienemer zal betrokkenen verdere informatie geven over de grondslag en het betreffende gerechtvaardigde belang.

Hoe worden mijn persoonsgegevens beveiligd?

TNO verwerkt zowel de ruwe content als de Schone Dataset binnen een streng beveiligde infrastructuur. Deze infrastructuur is opgezet conform de geldende normen voor informatiebeveiliging, waaronder ISO 27001. Er worden zowel technische als organisatorische maatregelen genomen om de persoonsgegevens te beschermen tegen verlies, onbevoegde toegang, ongeoorloofde wijziging of andere vormen van onrechtmatige verwerking.

Zo heeft alleen een selecte groep geautoriseerde medewerkers van TNO die deel uitmaken van het team dat de opschoning van ruwe content verzorgt toegang tot deze ruwe content. Deze medewerkers hebben een geheimhoudingsplicht en krijgen alleen toegang tot de gegevens voor zover en zolang dat strikt noodzakelijk is voor hun werkzaamheden.

Hoe lang worden mijn persoonsgegevens bewaard?

De ruwe content wordt verwijderd zodra de Schone Dataset is samengesteld en de pre-training fase van het GPT-NL Model is afgerond. De Schone Dataset wordt zeven jaar bewaard. Dit is de wettelijke bewaartermijn die voor TNO geldt met betrekking tot door de overheid gesubsidieerd onderzoek.

Rol van SURF

SURF is de IT-partner van TNO in het GPT-NL Project. SURF stelt de rekenkracht en IT-capaciteit ter beschikking die nodig is om het GPT NL Model te trainen (zie: [Snellius: de Nationale Supercomputer | SURF.nl](#)). SURF verwerkt gegevens uitsluiten ten behoeve van TNO als een ‘verwerker’. TNO heeft hiervoor een contract afgesloten met SURF, waarin waarborgen zijn opgenomen om ervoor te zorgen dat de gegevens goed worden beveiligd.

Rechten van betrokkenen

Onder de AVG hebben betrokkenen het recht op inzage in de verwerking van hun persoonsgegevens, en verder in bepaalde gevallen het recht op rectificatie, verwijdering, beperking, data portabiliteit, en bezwaar. Voor TNO als instelling van wetenschappelijk onderzoek gelden wettelijke uitzonderingen op deze rechten. Hieronder is aangegeven welke rechten u kunt uitoefenen in de **trainingsfase** van het GPT-NL Model en welke rechten u kunt uitoefenen in de **gebruiksfase** van het GPT-NL Model.

Trainingsfase

Tijdens de Trainingsfase verwerkt TNO persoonsgegevens in het kader van wetenschappelijk onderzoek in het algemeen belang. Voor dergelijke verwerkingen gelden wettelijke uitzonderingen op de rechten van betrokkenen, om te voorkomen dat deze het onderzoek onmogelijk maken of ernstig belemmeren. Dit geldt voor alle rechten van betrokkenen. De wettelijke uitzonderingen zijn hieronder aangegeven:

- **Inzage, rectificatie, en beperking:** de Nederlandse wetgever heeft in artikel 44 UAVG opgenomen dat deze rechten niet van toepassing zijn op de verwerking van persoonsgegevens voor wetenschappelijk onderzoek.
- **Verwijdering:** dit recht is onder artikel 17 lid 3 onder d AVG niet van toepassing op de verwerking van persoonsgegevens voor wetenschappelijk onderzoek onder artikel 89 lid 1 AVG, als dit deze wetenschappelijke doeleinden onmogelijk zouden maken of ernstig zouden belemmeren, wat het geval is voor de onderzoeksfase van het GPT-NL Project.
- **Data operabiliteit:** dit recht is onder artikel 20 lid 1 AVG niet van toepassing wanneer de verwerking van persoonsgegevens noodzakelijk is voor de vervulling van een taak van algemeen belang, zoals hier het geval is (zie hierboven onder “*Wat is de grondslag voor verwerking van mijn persoonsgegevens?*”)
- **Bezwaar:** dit recht is onder artikel 21 lid 6 AVG niet van toepassing op de verwerking van persoonsgegevens voor wetenschappelijk onderzoek in het algemeen belang, zoals het geval is voor het GPT-NL Project.

Als uw verzoek ziet op de vertrekking van ruwe content door Deelnemers aan TNO, dan kunt u uw verzoek bij de betreffende Deelnemer indienen.

Gebruiksfase

In de gebruiksfase kunt u uw rechten uitoefenen ten aanzien van het GPT-NL Model zelf. Zo kunt u bijvoorbeeld een rectificatieverzoek doen als het GPT-NL Model onjuiste antwoorden geeft op vragen over u. Deze rechten zijn verschillend als u een *publiek persoon* bent (een persoon met een eigen Wikipedia-pagina) of een *niet-publiek persoon*.

- *U bent een niet-publiek persoon*
Onderdeel van de trainingsfase is dat TNO uw persoonsgegevens zo snel mogelijk verwijderd of anonimiseert en deze worden niet gebruikt voor het trainen van het GPT-

NL Model (zie voor meer informatie hierboven onder “*Hoe bewerkt TNO de ruwe content voordat het GPT NL Model ermee wordt getraind?*”). Dit betekent dat uw persoonsgegevens in de gebruiksfase van het GPT-NL Model in beginsel niet meer worden verwerkt. TNO is niet in staat om uw persoonsgegevens in de Schone Dataset of in het GPT-NL Model te identificeren. Indien u toch een verzoek om inzage indient zal TNO u laten weten dat zij niet aan uw verzoek kan voldoen (dit conform artikel 11 lid 2 AVG).

- *U bent een publiek persoon*

Het GPT-NL Model is getraind om vragen te beantwoorden die direct verband houden met uw publieke rol. Als u gebruikt wilt maken van uw rechten onder de AVG, zoals uw recht op rectificatie, dan kunt u uw verzoek richten tot TNO of tot de licentienemer die het GPT-NL Model gebruikt (welke het verzoek met TNO zal delen).

TNO en de licentienemers zullen uw verzoek vervolgens in behandeling nemen in overeenstemming met de AVG en de UAVG. Zij kunnen u vragen om meer informatie te verstrekken over uw verzoek voordat zij op uw verzoek reageren. Zo kunnen zij u bijvoorbeeld vragen om aan te geven op basis van welke vraag het GPT-NL Model het desbetreffende antwoord genereert en waarom u van mening bent dat het antwoorde onjuiste informatie over u bevat.

Als uw verzoek wordt gehonoreerd dan zal TNO alle licentienemers instrueren tijdelijke maatregelen nemen om de desbetreffende antwoorden te voorkomen (zoals opgenomen in het [Data Protection Protocol](#)). Het is niet mogelijk om het GPT-NL Model na ieder gehonoreerd verzoek opnieuw te trainen, omdat een dergelijke trainingssessie een lang en uitgebreid proces is. TNO zal uw verzoek definitief verwerken in de volgende versie van het GPT-NL Model. Deze aanpak is in overeenstemming met de richtlijnen van het consortium van EU privacy-toezichthouders (de “EDPB”) en de Franse privacy-toezichthouder (“CNIL”) over het uitoefenen van rechten onder de AVG met betrekking tot AI modellen.

Gebruik van het GPT NL Model door licentienemers

Als uw verzoek niet ziet op het GPT NL Model zelf maar specifiek op gebruik van het GPT-NL *door licentienemers*, dient u uw verzoek te richten tot deze licentienemer. TNO is namelijk niet betrokken bij het gebruik van het GPT NL Model door gebruikers, en kan uw verzoek daarom niet voor u doorsturen naar de juiste gebruiker.

Verdere vragen of klachten

Als u verdere vragen heeft over de verwerking van uw persoonsgegevens in het kader van het GPT-NL Project, dan kunt u contact opnemen met TNO via de onderstaande contactgegevens. Daarnaast heeft u altijd het recht om een klacht in te dienen bij de Autoriteit Persoonsgegevens.

Contact

Heeft u vragen over deze privacyverklaring of over de verwerking van persoonsgegevens in het kader van het GPT-NL project?

Neem dan contact op met: privacy@gpt-nl.nl, of schriftelijk via Anna van Buerenplein 1, 2595 DA Den Haag, onder vermelding van het GPT-NL project.

Verdere informatie

Het GPT-NL Project is in detail uitgewerkt in meer documenten die publiekelijk beschikbaar zijn of worden gemaakt via de GPT-NL website: <https://gpt-nl.nl/>. Zie specifiek:

- Het [Data Protection Protocol](#), waarin wordt bepaald welke taken en verantwoordelijkheden TNO, de Deelnemers, en licentienemers van het GPT NL Model hebben met betrekking tot de verwerking van persoonsgegevens;
- Het [Training Content Protocol](#), waarin wordt bepaald hoe (persoons)gegevens worden opgeschoond en gebruikt om het GPT NL Model te trainen;
- Het [Data Protection Impact Assessment \(DPIA\)](#), waarin de verwerking van persoonsgegevens in het kader van het GPT-NL Project in detail wordt omschreven, en waarbij de maatregelen worden beschreven die zijn getroffen om de impact op de privacy van betrokkenen te beperken; en
- Het [Bronnenoverzicht](#), waarin een overzicht is gegeven van alle bronnen van door TNO verzamelde trainingsteksten voor het GPT-NL Project.