

GPT-NL

De ontwikkeling van GPT-NL

Progress report #2 | 26 Februari 2026

Document ID: TNO 2026 R10516

Versie: 2

Datum: v1 26-2-2026; v2 18-3-2026

Financiering: Financiering
Faciliteiten Toegepast Onderzoek (EZ)

Contact

www.gpt-nl.nl

www.linkedin.com/company/gpt-nl/

info@gpt-nl.nl

TNO innovation
for life

SURF



Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

Voorwoord

Saskia Lensink

Na twee jaar hard werken zijn we aanbeland in een spannende fase: de eerste versies van GPT-NL worden momenteel bij een aantal launching customers ingezet op lokale infrastructuur om te valideren waar en hoe GPT-NL van waarde kan zijn. Hier draait het uiteindelijk allemaal om: samen concrete stappen zetten naar soevereine en waarde(n)volle generatieve AI binnen de kaders van onze waarden en normen, die op een veilige manier kan omgaan met hooggevoelige informatie. En wat een jaar hebben we achter de rug. Het is zondermeer pijnlijk duidelijk geworden dat geopolitieke verhoudingen aan het verschuiven zijn en dat Europa momenteel te weinig eigen grip heeft op sleuteltechnologieën zoals AI. De roep om eigen, soevereine Europese alternatieven voor big tech groeit, en als team GPT-NL ondervinden we dagelijks de druk en hooggespannen verwachtingen die dat oproept.

Natuurlijk weet iedereen dat we met ons startbudget van 13.5 miljoen euro keuzes moeten maken – en dat hebben we durven doen. Wat ben ik trots op ons team dat laat zien wat wél kan, het team dat constant de lat voor zichzelf hoger legt, het optimisme van het team dat data aan strenge kwaliteits- en compliance-eisen onderwierp en met een echte doenersmentaliteit op een eerlijke manier grote datasets wist te verzamelen, de innovatieve manier waarop het team innovatie en privacy heeft laten samenkomen, het team dat non-stop ook tijdens de zomermaanden een babysitschema had ingericht om de training zo vlekkeloos mogelijk te laten verlopen, het lef van partijen zoals de NDP om een nooit eerder geziene overeenkomst van een AI-initiatief te sluiten en tientallen jaren aan archiefmateriaal van de commerciële nieuwsmidia beschikbaar te stellen, de talloze experts die ons tijdens dit jaar kritisch hebben bevraagd en met ons meedachten. In dit progress report kun je er alles over lezen. We zijn trots, en werken vol energie door – want er is zoveel meer te doen!

Inhoudsopgave

Het tweede jaar GPT-NL	5
In gesprek met Saskia Lensink en Frank Brinkkemper	
Launching Customers	9
De eerste use cases van GPT-NL	
De urgentie van digitale autonomie	14
Lokke Moerel over AI-soevereiniteit, GPT-NL en een tweesporenbeleid voor Nederland en Europa	
Frequently Asked Questions	18
Over de plannen van GPT-NL in 2026	
De data acquisitie voor GPT-NL 1.0	19
Van pionieren naar partnerschap	
Dataproviders	23
Van GPT-NL v1.0	

Het tweede jaar GPT-NL

In gesprek met Saskia Lensink en Frank Brinkkemper

Een jaar geleden stond GPT-NL vooral in de steigers. Met de start van 2026 zijn we een nieuwe fase ingegaan. Waar staat GPT-NL nu precies? Welke mijlpalen werden het afgelopen jaar bereikt en wat bleek lastiger dan verwacht? Product Manager Saskia Lensink en R&D Manager Frank Brinkkemper maken de balans op en blikken vooruit naar de volgende, spannende fase.

Waar kijken jullie met de meeste trots op terug als jullie naar 2025 kijken?

Saskia: „Voor mij was het sluiten van de licentieovereenkomst met de commerciële nieuwsmedia, via brancheorganisatie NDP, echt een hoogtepunt. Dat was het moment waarop ik dacht: het gaat lukken met dataverzameling. De NDP leverde een grote, kwalitatief sterke dataset en gaf daarmee ook een belangrijk signaal af richting andere partijen.”

Die deal kwam niet vanzelf tot stand, vertelt ze. „Er speelden zorgen, vooral over wat er met data gebeurt zodra die eenmaal in een model zit. Dat waren goede gesprekken. Het mooie is dat we nu merken dat ook partijen buiten Nederland zeggen: die aanpak van GPT-NL is interessant.”

Frank: „En los daarvan: technisch hebben we enorm veel data verzameld. Veel meer dan we aan het begin dachten, en op een manier die juridisch en ethisch klopt. Dat is echt een teamprestatie.”

Waar staat GPT-NL nu technisch gezien?

Saskia: „We hebben zoals bekend 13,5 miljoen euro beschikbaar gekregen en zijn inmiddels klaar met de pre-training

van het model. Je ziet nu al dat het model op bepaalde taken, zoals samenvatten, beter presteert dan oudere modellen als ChatGPT-3. Dat is best bijzonder als je bedenkt met welk budget dat destijds is gebouwd.”

Ze benadrukt dat er veel wordt getest. „We gebruiken standaard benchmarks. We geven het model bijvoorbeeld langere teksten en laten het samenvattingen maken, die we vergelijken met geverifieerde referenties.”

Frank: „Die benchmarks zijn inmiddels ook gangbaar geworden in de markt, zoals EuroEval. Niet alles is perfect afgestemd op het Nederlands, maar daar doen we wel correcties op. Zo kunnen we prestaties goed vergelijken.”

Wat waren het afgelopen jaar de grootste uitdagingen?

Saskia: „Het blijft simpelweg een complex project. We werken met een relatief klein team, ongeveer 25 mensen verspreid over meerdere organisaties, en met een beperkt budget. Tegelijkertijd wil je het juridisch en ethisch goed doen.”

Daar is wel vooruitgang in geboekt. „De wettelijke kaders zijn helderder geworden. We hebben goede juristen betrokken en een soort werkwijze gevonden die rust brengt: zo verzamelen we data, zo leggen we het vast. Dat helpt enorm.”

We gaan een nieuwe fase in. Wat staat er nu te gebeuren?

Saskia: „De volgende stap is spannend: het model de praktijk in brengen. We gaan werken met een selecte groep launching customers. Niet iedereen kan tegelijk beginnen, dus we kiezen bewust voor een beperkte groep om te leren hoe het model zich in de praktijk houdt.”

Frank: „Ik heb veel contact met die partijen. We starten nu met vijf launching customers en werken toe naar de tien. Met hen doen we eerst een ‘feasibility study’: hoe draait GPT-NL in hun context, op hun hardware en voor hun specifieke use cases. Die partijen komen uit uiteenlopende domeinen. Denk aan ministeries, het veiligheidsdomein, grote financiële instellingen en corporates waar privacy en compliance cruciaal zijn.”

Waarom willen organisaties bij die eerste tien horen?

Frank: „We noemen altijd vijf redenen. Eén: ze bouwen mee aan soevereine AI in Nederland en Europa. Twee: ze weten precies waar de data vandaan komt en verkleinen risico’s. Drie: ze kunnen volledig secure werken, bijvoorbeeld on premise. Vier: hun investering komt terecht bij rechtmatige partijen, zoals dataleveranciers. En vijf, misschien wel de spannendste, ze helpen mee om de kwaliteit van Nederlandstalige use cases te verbeteren.”

Saskia: „We hoeven niet direct gelijk te zijn aan de allerbeste modellen wereldwijd, maar we moeten wel een product maken dat waarde toevoegt in de praktijk. Dat lukt alleen door echte gebruikersfeedback.”

Wat heeft jullie het meest verrast het afgelopen jaar?

Saskia: „Hoeveel interesse er is vanuit partijen waarvan je het niet direct verwacht. Bijvoorbeeld uit de financiële dienstverlening. En ook dat het voor veel mensen pas later duidelijk werd dat een eerste versie bouwen niet betekent dat je klaar bent. Dit is kapitaalintensief en vraagt langdurige commitment.”

Frank: „Wat mij blijft verbazen, is het talent dat bij TNO, SURF, NFI en onze partners werkt. Het idee dat Nederland niet genoeg mensen zou hebben om dit te bouwen, herken ik totaal niet.”

Welke vragen krijgen jullie nu het vaakst?

Saskia: „Mensen vragen wanneer GPT-NL ‘af’ is, wat het gaat kosten en of ze bij de launching customers kunnen horen. En ik hoor ook zorgen over soevereiniteit: wordt dit straks geen Amerikaans bedrijf? Dat kunnen we echt borgen. GPT-NL is verankerd in publieke organisaties; TNO zal nooit Amerikaans worden.”

Frank: „Ik krijg vooral vragen over kwaliteit, aansluiting bij Europese initiatieven en natuurlijk: wat kost het concreet voor klanten?“

Tot slot: wat kunnen we de komende tijd verwachten?

Frank: „In Q1 starten we dus met de launching customers, die krijgen een voorsprong. In het najaar volgt een bredere uitrol, mogelijk ook via hostingproviders.“

Saskia: „En we onderzoeken de mogelijkheden om een krachtige opvolger van GPT-NL te bouwen die meerdere talen en functionaliteiten ondersteunt. Op korte termijn gaan we al aan de slag met verbeterde RAG functionaliteiten en het ondersteunen van spraak.“

Saskia Lensink
Product Manager

Frank Brinkkemper
R&D Manager



Launching customers

De eerste use cases voor GPT-NL

In februari zijn we gestart met partijen die willen investeren in de toekomst van GPT-NL. We voeren met deze organisaties zogeheten 'Feasibility Studies' uit: een onderzoek naar de deployment van GPT-NL binnen een specifieke use case én binnen de organisatie. De financiers van deze studies zien het belang in van de doorontwikkeling en de investering in kennis en kunde in Nederland.

Financier

Ministerie van Binnenlandse Zaken en
Koninkrijksrelaties

In uitvoering met



Chatbot 'Gem' is een chatbot die zo'n 6 jaar bestaat en inmiddels door bijna dertig gemeenten op dagelijkse basis wordt gebruikt. In 2024 voerde Gem bijna 70.000 gesprekken en rapporteerde over de ervaringen. Eerder is al de inzet van GenAI voor de chatbot al verkend. In deze use case testen we de meerwaarde van GPT-NL voor Gem. We zetten het model in voor het **beantwoorden van gebruikersvragen**, en evalueren de kwaliteit in de context van de vragen die voor de gemeentes relevant zijn.

Financier

Ministerie van Binnenlandse Zaken en
Koninkrijksrelaties

In uitvoering met



Het is voor burgers en ondernemers vaak lastig om alle relevante informatie over dienstverlening van de overheid te vinden omdat deze verspreid is over diverse websites van verschillende overheidsinstanties. Het ministerie van Binnenlandse Zaken realiseert daarom onder andere **een digitale assistent die antwoorden geeft op basis van betrouwbare informatie** van Overheid.nl en in een later stadium ook andere informatiebronnen van de overheid. In deze use case vergelijken we de antwoorden van GPT-NL met die van het huidige, commerciële model.

Financier

Ministerie van Binnenlandse Zaken en
Koninkrijksrelaties

In uitvoering met



Brieven van de overheid zijn voor burgers en ondernemers niet altijd begrijpelijk. Het gaat vaak om belangrijke communicatie en dienstverlening, zoals schulden en toeslagen. Daarom is de communicatie-assistent HIP (Helder, Intelligent en Productief) ontwikkeld: één van de tools die overheidsmedewerkers ondersteunt in hun werk. HIP helpt om brieven op te stellen en te controleren op begrijpelijke taal en **ondersteunt zo ambtenaren bij het creëren van duidelijke, toegankelijke en effectieve communicatie**. In deze use case vergelijken we het best presterende model dat momenteel wordt ingezet voor HIP met GPT-NL.

GPT-NL Partner



In veel projecten van TNO wordt gewerkt met DV+ geclassificeerde documenten, privacygevoelige data of betreft het werk dat hoge eisen stelt aan privacy, security en compliance. De adoptie van GenAI, waar TNO op inzet, gaat daarom volgens het interne adagium ‘Copilot, tenzij’, en laat nu precies die “tenzij” hetgeen zijn waarmee TNO intern aan de slag gaat. De Feasibility Study richt zich op de on-premise inzet van GPT-NL (software en data zijn geïnstalleerd en beheerd op eigen servers) om de meerwaarde te toetsen. De use case waarin GPT-NL wordt toegepast, wordt tijdens de Feasibility Study verder gedefinieerd.

GPT-NL Partner



Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

Taalmodellen zijn onmisbaar voor opsporingswerk, dat is al jaren zo. De analyse van terabytes aan bewijsmateriaal is een gigantische opgave. Het is voor forensisch experts praktisch onmogelijk zulke grote datavolumes snel te doorgronden. Daarom zet het NFI AI-modellen in. In deze use case met GPT-NL wordt het model **gefinetuned op forensische data en wordt onderzocht in hoeverre het gefinetunede model het beter doet in classificatietaken** ten opzichte van de huidige modellen die het NFI nu inzet.

**Vijf redenen
om je aan te
sluiten bij
GPT-NL**



1



Bouw mee aan **soevereine AI** in Europa en versterk zo de onafhankelijkheid van je organisatie.

2



Verlaag je risico's door te **voldoen aan de wet** en te weten waar de data vandaan komt.

3



Investeer in kennis en kunde in Nederland waarbij het geld terecht komt bij de rechtmatige eigenaren.

4



Heb garantie van een toekomst-vaste architectuur door GPT-NL on-premise te hosten en werk zo **volledig secure**, ook bij vertrouwelijke classificaties.

5



Ontvang ondersteuning voor **goede performance en kwaliteit** van de usecases.

De urgentie van digitale soevereiniteit

**Lokke Moerel over AI-
soevereiniteit, GPT-
NL en een
tweesporenbeleid
voor Nederland en
Europa**

De ontwikkeling van GPT-NL had niet gekund zonder de juridische expertise van Lokke Moerel. Als expert op het snijvlak van technologie en wetgeving ziet zij dagelijks hoe groot de afhankelijkheid van Europa is. Overheden, bedrijven en maatschappelijke organisaties draaien grotendeels op niet-Europese cloudplatforms, kantoorsoftware en AI-tools. Dat schuurt steeds vaker, niet alleen vanwege privacy of compliance, maar ook door geopolitieke onzekerheid. Moerel: “Je kan klagen over afhankelijkheid van Big Tech en dat hun LLMs niet aan onze Europese wetgeving en publieke waarden voldoen, of je bouwt een goed alternatief. Juist als niemand zich meer kan voorstellen dat het ook anders kan, moet je laten zien dat het wél kan.” GPT-NL is volgens haar zo’n poging. In gesprek met Lokke, schijnen we licht op de urgentie van digitale soevereiniteit.

De urgentie is volgens Moerel niet alleen technisch, maar vooral geopolitiek en economisch. Moerel: “Als we onze toekomstige welvaart en strategische autonomie veilig willen stellen, moet Europa fors investeren in sleuteltechnologieën zoals AI, digitalisering en klimaatinnovatie. Digitale infrastructuur is daarin geen bijzaak meer, maar een fundament.

Digitale soevereiniteit als urgent pijnpunt

Soevereine AI moet onderdeel zijn van een bredere digitale basis. Dat is geen ideologisch verhaal, maar een strategische noodzaak.”

Ze pleit daarom voor een doordacht tweesporenbeleid. Enerzijds investeren in eigen digitale infrastructuur en diensten die voldoen aan Europese regelgeving en publieke waarden. Anderzijds actief sturen op soevereiniteit binnen mondiale netwerken, zodat Europa niet afhankelijk wordt van single points of failure. Niet isoleren, maar zorgen dat je positie sterk genoeg is om samen te werken zonder kwetsbaar te zijn.

Niet alleen regels maken, maar technologie bouwen

Volgens Moerel is GPT-NL juist in dat licht belangrijk. Europa is sterk in reguleren, maar dat is niet genoeg. “Als je alleen regels maakt voor technologie die anderen bouwen, blijf je altijd achter de feiten aanlopen. Je moet ook zelf technologie ontwikkelen, anders blijf je aan het infuus van buitenlandse aanbieders hangen en heb je geen onderhandelingspositie.”

GPT-NL als hefboom in het digitale landschap

Wat Moerel vooral interessant vindt, is de ketenreactie die GPT-NL inmiddels in gang heeft gezet. Ze ziet dat het project steeds vaker opduikt in strategische plannen, waar het eerder nauwelijks werd genoemd. “Je merkt dat andere Europese lidstaten meekijken en vragen hoe het ecosysteem van alle publiek en private samenwerkende partijen juridisch en organisatorisch is ingericht. We zijn het eerste land ter wereld

dat erin is geslaagd een samenwerking aan te gaan met de gezamenlijke uitgevers. Dus we krijgen vaak vragen over hoe we de uitgevers zover hebben gekregen hun content ter beschikking te stellen, tegen welke voorwaarden, en welke rol hebben ze in de governance van het project.”

Voor haar bewijst GPT-NL dat samenwerking tussen publieke en private partijen op een andere manier mogelijk is. Niet vanuit een eigen belang van een specifieke partij, maar vanuit vertrouwen en een nationaal gedeeld belang. “Je ziet dat dit project iets losmaakt. Het geeft inhoud aan het idee dat we als we samen op trekken echt veel meer kunnen dan we denken. ” Dat raakt volgens haar ook aan andere grote transitie, zoals klimaat en energie. “Ook daar geldt dat technologische innovatie, investeringskracht en publiek en privaat samen moeten optrekken. Zonder stevige digitale basis blijven die domeinen kwetsbaar.”

Van experiment naar maatschappelijke infrastructuur

Tegelijk is Moerel realistisch over de fase waarin GPT-NL zich bevindt. “Kijk, dit is ergens nog een startup. De mensen die hier werken leren ongelooflijk veel. Maar de echte uitdaging begint nu: hoe maak je hiervan iets dat schaalbaar is en structureel onderdeel wordt van de maatschappij?”

Die stap vraagt volgens haar om andere verhoudingen. Meer capaciteit, meer investeringen, sterkere koppelingen met marktpartijen en publieke organisaties. “Niet tien losse use

Je kunt kritiek hebben op Big Tech en op de manier waarop large language models worden ontwikkeld, óf je bouwt een goed alternatief.

cases tegelijk willen oplossen, maar gericht opschalen. En vooral: durven erkennen dat soevereine alternatieven geld kosten. We zijn in Nederland soms goed in experimenteren, maar terughoudend in echt investeren. Het beginkapitaal is er vaak wel. De vervolgstap veel minder. Terwijl dit juist het moment is waarop je moet doorpakken.”

Vooruitkijken: soevereine cloud en nieuwe modellen

Voor het komende jaar hoopt Moerel dat GPT-NL kan worden aangeboden als een SaaS oplossing vanuit een soevereine cloudomgeving. Niet alleen als technische keuze, maar ook als nieuw economisch model. Een infrastructuur waarin Nederlandse of Europese providers niet alleen hosten, maar ook zelf technologie ontwikkelen, beheren en doorbouwen. “Als je ziet hoeveel geld er nu naar buitenlandse AI- en cloudpartijen gaat, dan is het logisch om te vragen: waarom zou dat geld niet deels hier kunnen landen, bij partijen die in Nederland en Europa innoveren?” Ze ziet daar kansen voor nieuwe allianties, nieuwe diensten en een steviger digitale positie. “In Nederland zie ik vaak een Calimero-houding, terwijl er zulke fantastische innovaties worden gedaan. We kunnen veel meer dan velen denken, als we maar samenwerken.”

Dit moment telt

Voor Moerel staat vast dat GPT-NL zich op een kantelpunt bevindt. Het project is verder dan een idee, maar nog niet groot genoeg om vanzelfsprekend te zijn. Precies daarom is dit het moment om keuzes te maken over hoe Nederland en Europa zich willen verhouden tot een wereld waarin AI, data en digitale



infrastructuur steeds meer de basis vormen voor economie, overheid en samenleving.

“Juist als niemand zich meer kan voorstellen dat het ook anders kan, moet je laten zien dat het wél kan. GPT-NL is daar het bewijs van. Nu is het zaak om op te gaan schalen.”

Over Lokke Moerel | Lokke Moerel is hoogleraar Global ICT Law aan Tilburg University en een toonaangevend expert op het snijvlak van AI, technologie en governance. Ze adviseert overheden en organisaties over digitale soevereiniteit, kunstmatige intelligentie en platformstrategie, en is partner bij advocatenkantoor Morrison Foerster. Vanuit die rol is ze nauw betrokken bij het publieke debat over verantwoord AI-gebruik en de positionering van initiatieven als GPT-NL binnen Nederland en Europa.

Jullie zijn gestart met de Feasibility Studies. Kan ik nog meedoen?

We zijn met de eerste partijen die willen investeren in de doorontwikkeling van GPT-NL gestart met de Feasibility Studies. Er is ruimte om dit kwartaal nog in te stappen, afhankelijk van de use case en capaciteit. Neem hiervoor contact op via het contactformulier op onze website.

Wanneer kan ik GPT-NL inzetten in mijn organisatie?

We verwachten de bredere uitrol te doen in de tweede helft van 2026. Dan kan GPT-NL worden afgenomen middels de professionele licentie. We werken ook toe naar een hosted versie van GPT-NL, hierover volgt later meer info.

Wanneer kan ik GPT-NL gebruiken als onderzoeker?

We verwachten de bredere uitrol te doen in de tweede helft van dit jaar, dat geldt ook voor het gebruik van GPT-NL voor onderzoeksdoeleinden.

Verzamelen jullie momenteel content?

Momenteel ligt de focus bij het uitvoeren van de Feasibility Studies en het voorbereiden van de bredere uitrol in de tweede helft van dit jaar. We starten daarom later met actieve acquisitie van content en datasets voor GPT-NL v2.0. Mocht u een grote dataset beschikbaar hebben, kunt u ons hierover contacteren via de website.

Wanneer komt de dataset van de v1.0 beschikbaar?

Onze collega's leggen de laatste hand aan het opmaken van de publieke dataset zodat deze in Q1-2026 gepubliceerd kan worden op HuggingFace. Over de auteursrechtelijk beschermde content die we hebben gebruikt publiceren we metadata. Zo kunnen we transparantie bieden over de samenstelling van de gehele dataset.

De data acquisitie voor GPT-NL 1.0

Van pionieren naar partnerschap

De start van eerste pilots met de beta-versie van GPT-NL markeert een goed moment om terug te blikken op de dataverzameling en de dataset voor het model v1.0. In deze Deep Dive lees je welke aanpak we hebben gehanteerd en welke uitdagingen kwamen kijken bij het bouwen from scratch.

Beginnen zonder blauwdruk

Wie nu terugkijkt op de dataverzameling van GPT-NL, ziet een proces dat grotendeels al doende is ontstaan. We begonnen echt met iets totaal nieuws. We hadden geen vast draaiboek en ook geen bestaand ecosysteem dat we konden kopiëren. Dit soort dataverzameling, samen met commerciële partijen én publieke instellingen, bestond simpelweg nog niet. Binnen TNO was er wel ervaring met data-ecosystemen, maar met GPT-NL vroegen we om iets anders. Je kon dit niet behandelen als een klassiek onderzoeksproject. En al gebeurt er wereldwijd wel veel rondom LLM's en datagebruik, best practices hiervoor zijn er nog niet. Dus op een gegeven moment zijn we gewoon begonnen.

Dat begin kreeg vorm in gezamenlijke sessies met potentiële dataleveranciers: van bibliotheken en universiteiten tot mediapartijen en erfgoedinstellingen. We hebben veel tijd gestoken in uitleg: Waarom bouwen we een alternatief voor bestaande taalmodellen? Waarom zou je je aansluiten samen met andere dataleveranciers? En wat gebeurt er eigenlijk met je data als we die gebruiken voor een trainingsdataset? Het was pionieren, maar ook waardevol dat partijen elkaars verhalen hoorden.

Geen transacties, maar een ecosysteem

Een bewuste keuze vanaf het begin was om de relatie met datapartijen niet puur transactioneel te maken. We wilden deelnemende partijen niet neerzetten als 'verkopers' van data. Het moest geen kwestie worden van: wij kopen eenmalig je data

en daar eindigt je betrokkenheid. In plaats daarvan kozen we voor de Content Board, een soort consortium met een gezamenlijk doel: een succesvol taalmodel. Iedereen kiest zelf of ze aan de Content Board willen deelnemen. Dat vereiste meer voorbereiding, zoals het inrichten van een governance voor de Content Board, maar is veel sterker voor de lange termijn.

Bouwen aan een ecosysteem

Die aanpak vroeg ook om realisme richting datapartijen. Wij bouwen een ecosysteem waarin contentleveranciers en GPT-NL elkaar versterken, geen winstmachine. We hebben niet een enorm budget en een ander verdienmodel zoals commerciële LLM-aanbieders. Dat vraagt om eerlijke gesprekken, zeker met partijen die gewend zijn hun data als economisch kapitaal te zien. Toch zien we dat dat beeld is gekanteld. Data van een partij is waardevol, maar zelden op zichzelf. De waarde ontstaat in samenhang met de data van andere partijen, binnen een model dat betrouwbaar, compliant en bruikbaar is.

Hotel California

Een vraagstuk dat daarbij speelde, was de mogelijkheid om in en uit te stappen. Het plan was aanvankelijk dat partijen altijd kunnen in- en uitstappen, maar het is simpelweg niet realistisch om dat voor individuele verzoeken uit te voeren. Je kan niet continue het model opnieuw trainen. Niet alleen vanwege de rekenkracht die daarvoor gereserveerd en gebruikt moet worden, maar ook omdat je ervoor moet zorgen dat elke afnemers van GPT-NL de laatste versie gebruikt. Daarom is gekozen voor een

andere oplossing. Als een partij zich wil terugtrekken uit GPT-NL zijn er twee opties: of we nemen maatregelen om te voorkomen dat hun content in de output van het model terug blijft komen, of ze blijven een vergoeding ontvangen zolang hun content in het model zit totdat een nieuwe versie van het model is gelanceerd. Daarbij willen we geen ‘Hotel California’-gevoel creëren waar je nooit meer uit kunt vertrekken, maar ook eerlijk te zijn over wat technisch wel en niet kan. Ook hierover zijn we transparant: voor iedere licentieovereenkomst met dataproviders gelden uniforme regels en deze afspraken zijn openbaar te vinden op onze website.

De overeenkomst met NDP

Een belangrijk mijlpaal in het traject was de licentieovereenkomst met de commerciële nieuwsmedia, via brancheorganisatie NDP. Wereldwijd zijn we het eerste initiatief waarbij het is gelukt om met alle uitgevers afspraken te maken over gebruik van hun content en een eerlijke vergoeding voor hun bijdrage. Van landelijke kranten tot platforms als NU.nl, RTL Nieuws en BNR Nieuwsradio. De NDP leverde een grote, kwalitatief sterke dataset en gaf daarmee ook een signaal af richting andere partijen. Wat maakt deze overeenkomst zo bijzonder?

Nieuwsmedia zijn misschien wel de partijen die het meeste te verliezen hebben in de hele LLM-revolutie. Niet alleen is de content van nieuwsmedia massaal van het internet geschraapt voor training van LLMs zonder toestemming en vergoeding, deze wordt vervolgens gebruikt om nieuwsberichten te genereren die

(bijna) gratis worden aangeboden. Het is dus logisch dat een dergelijke deal tussen nieuwsmedia en LLM-ontwikkelaars niet vanzelf tot stand kwam. Er waren terechte zorgen.

Door goed te luisteren naar deze zorgen kwamen we tot extra maatregelen gekomen die zijn ingebouwd om te voorkomen dat licentiemateriaal via technische weg uit het model kan worden gehaald. Hiermee ondersteunen Nederlandse nieuwsuitgevers de ambitie om een model te realiseren dat een benchmark neerzet hoe om te gaan met nieuwsartikelen in AI-systemen. In de woorden van Rien van Beemen, voorzitter van NDP Nieuwsmedia: “We scheppen met deze beweging een precedent waarmee we de positie van de journalistiek in Nederland op termijn verstevigen. AI-innovatie kan dus ook op een ethische, verantwoorde manier tot stand komen, zónder grootschalig onrechtmatig gebruik van het werk van onze journalisten. Met deze stap geven we die ontwikkeling een impuls.”



“We scheppen met deze beweging een precedent waarmee we de positie van de journalistiek in Nederland op termijn verstevigen.”

Rien van Beemen, voorzitter
NDP Nieuwsmedia

Samenwerking met publieke organisaties

Buiten de licentieovereenkomsten die gesloten zijn binnen GPT-NL is er ook significant werk geleverd om zo veel mogelijke Nederlandse publieke data (zoals dat van de overheid of de rechtspraak) op één plek te krijgen, in de juiste vorm die geschikt is voor LLM-training. Voor deze klus hebben we veel hulp gekregen van publieke instanties en de Open State Foundation om deze data te ontsluiten. Al deze data wordt in Q1 van 2026 gepubliceerd op HuggingFace.

Met vertrouwen richting livegang in de tweede helft van '26

Een veel gehoorde opmerking aan het begin van GPT-NL was: met dit budget is dit onmogelijk. Inmiddels horen we dat niet meer: met lijkt in te zien dat het niet ging om concurreren op schaal, maar om het bouwen van iets dat nog niet bestond. Met een aanpak die eveneens vernieuwend is. Die feedback krijgen we ook van dataleveranciers. In de woorden van Marijn Storm, jurist en nauw betrokken bij de dataverzameling: “Iedereen die bij de GPT-NL sessies zat, voelde: dit is echt iets anders dan je data aan een big tech partij verkopen. De documentatie, de openheid, het opschonen van data vóór training, dat zie je elders niet.”

Met de inzet van de beta-versie van GPT-NL en de start met launching customers krijgt dat werk nu een vervolg in de praktijk.

**Namens Team
Data Acquisitie
GPT-NL**

o.l.v. Jesse van Oort

Duuk Baten, Hubert van Beusekom, Frank Brinkkemper, Lieke Dom, Karim El Assal, Leah Griffioen-van Putten, Thomas van Himbergen, Matthieu Laneuville, Saskia Lensink, Annette Langedijk, Madelon Molhoek, Irene Nooren, Peter van Schagen, Daan Vos

**Met speciale dank aan de
betrokken experts**

Open State Foundation *Rob van Dijk, Sicco van Sas, Tim Vos-Goedhart*

**Instituut voor de Nederlandse Taal / KU
Leuven** *Bram Vanroy*

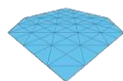
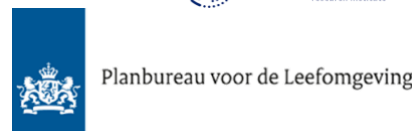
TNO *Selmar Smit, Berry Vetjens, Marc Zegveld*

Morrison Foerster *Lokke Moerel, Marijn Storm*

De dataset voor GPT-NL 1.0 is mede mogelijk gemaakt door



/instituut voor de Nederlandse taal/



Driestar



Tweede Kamer
DER STATEN-GENERAAL



GPT-NL

TNO innovation
for life



Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

