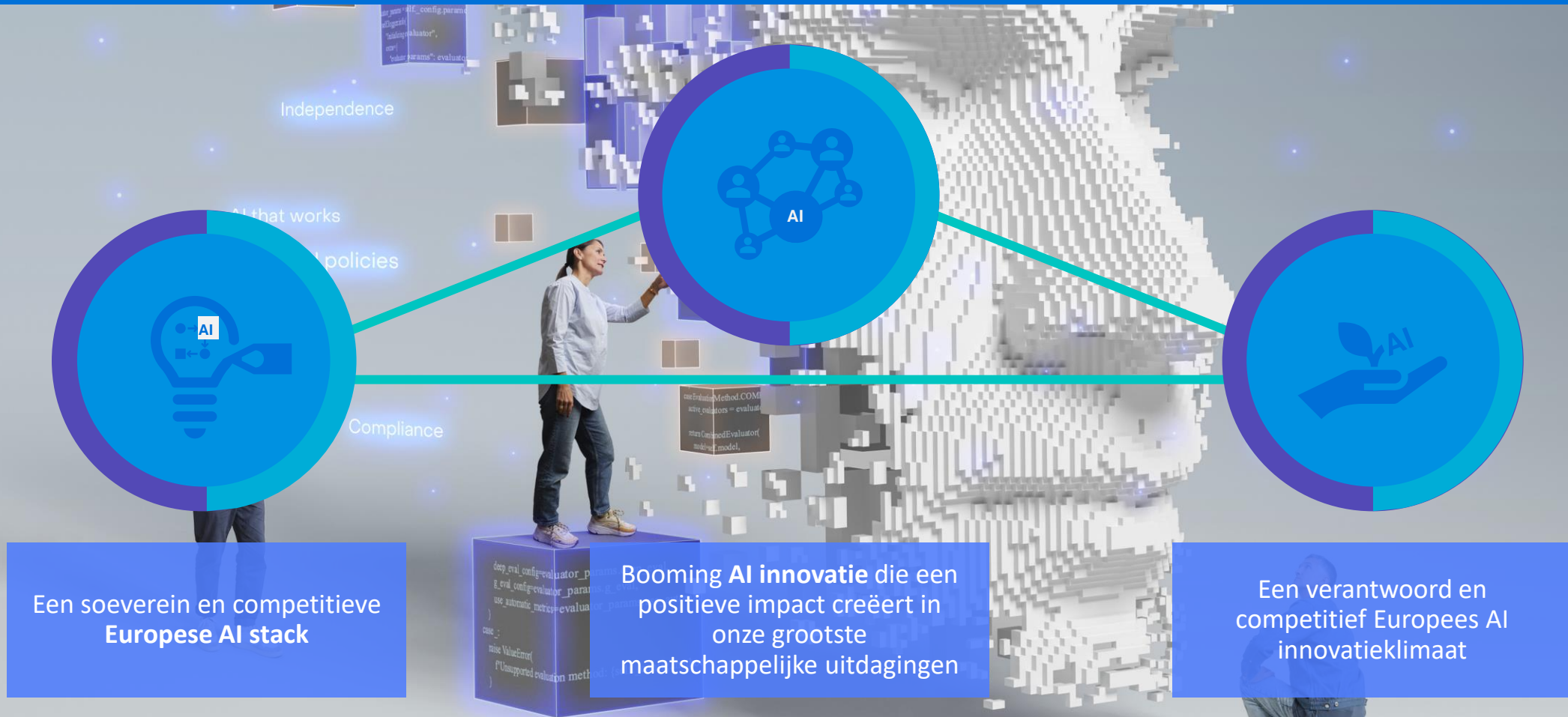


GPT-NL

TNO innovation
for life



Onze missie: de wereld verbeteren met verantwoorde AI



De Standaard

⊕ GPT-nl neemt het op tegen de Amerikaanse AI-reuzen: “Het hoeft niet zo gigantisch te zijn als big tech lijkt te beweren”

Dominique Deckmyn

19 december 2025 om 23:59

Moet publiek gefinancierde AI per se gratis zijn? GPT-NL zegt nee, en terecht

22 juli 2025



Winners Dutch Privacy Awards 2026 announced!



TNO, NFI & SURF – GPT-NL

First Dutch LLM developed fully GDPR-compliant, with transparent training data, strong privacy assurance and an innovative licensing model; a blueprint for sovereign and responsible AI in Europe.



De Nederlandse
AI-AWARD

Het eerste soevereine Nederlandse LLM

GPT-NL is afgestemd op Nederlandse contexten — en vroege trainingsresultaten laten al zien dat het beter presteert dan vergelijkbare modellen zoals OLMo2 bij taken zoals tekstsamenvatting, waarmee wordt bewezen dat *‘quality in yields quality out’*.

GPT-NL is uniek en het eerste LLM in zijn soort dat:

- een **eerlijk licentiemodel** heeft. We hanteren een kick-backstructuur voor datapartners, zodat zij een eerlijke positie krijgen in het generatieve-AI-ecosysteem.
- **volledig compliant** is. Het voldoet aan alle vereisten van de EU AI Act, de AVG (GDPR) en het auteursrecht.
- **Geen ongeoorloofd datascrapen** — alleen transparante en ethische dataverzameling.
- voortkomt uit **een non-profitorganisatie, TNO**.

Ongekende en unieke samenwerking met mediabedrijven



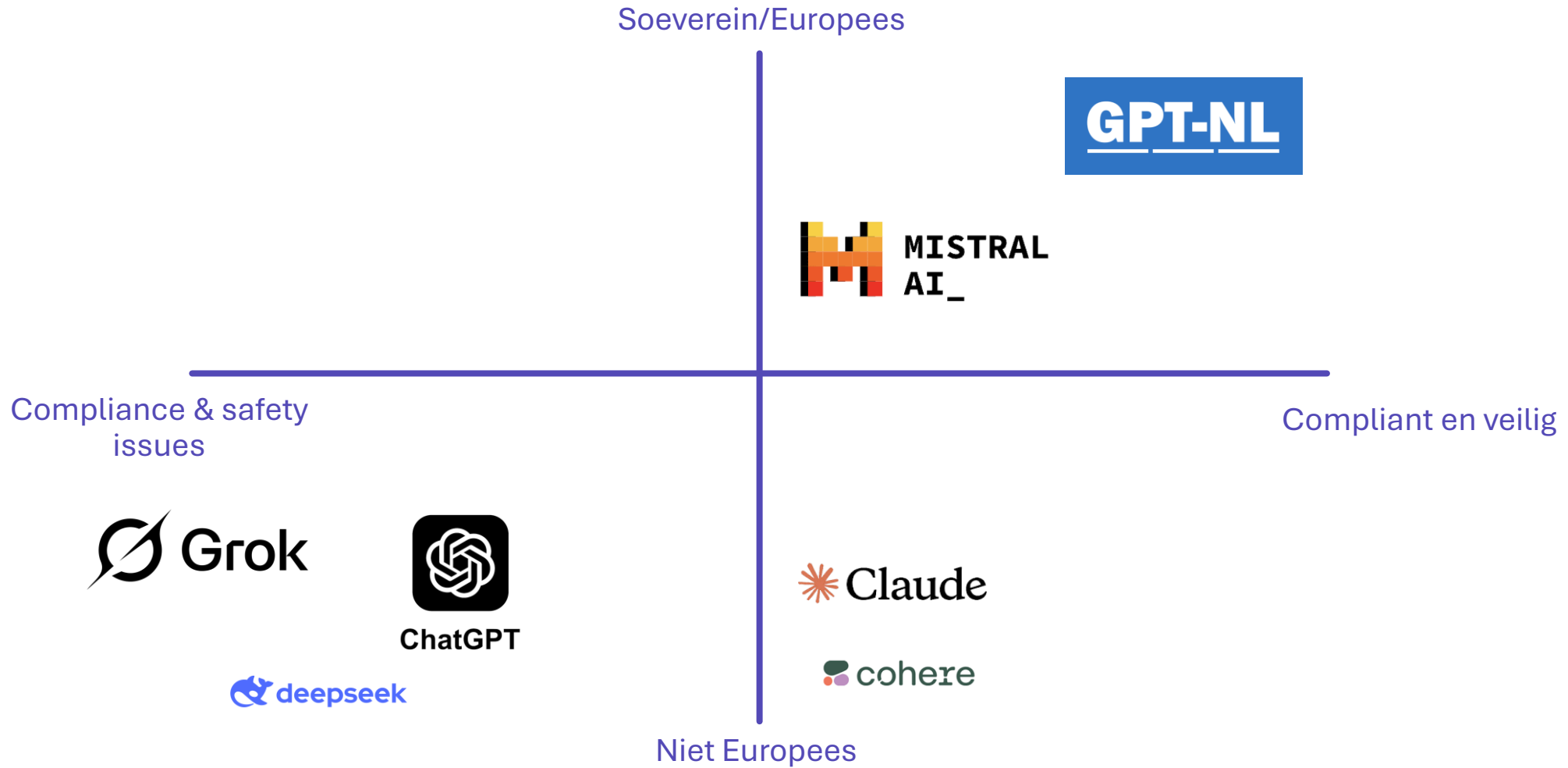
NDP Nieuwsmedia
676 volgers
1 mnd • 🌐

De leden van NDP Nieuwsmedia en het ANP gaan samenwerken met GPT-NL van TNO, het eerste grootschalige Nederlandse AI-taalmodel dat volledig wordt getraind op rechtmatig verkregen data. De nieuwsbedrijven laten daarmee zien dat hoogkwalitatieve data voor AI-training onder de juiste voorwaarden beschikbaar zijn voor licensering.
<https://lnkd.in/e-EdCbZ>

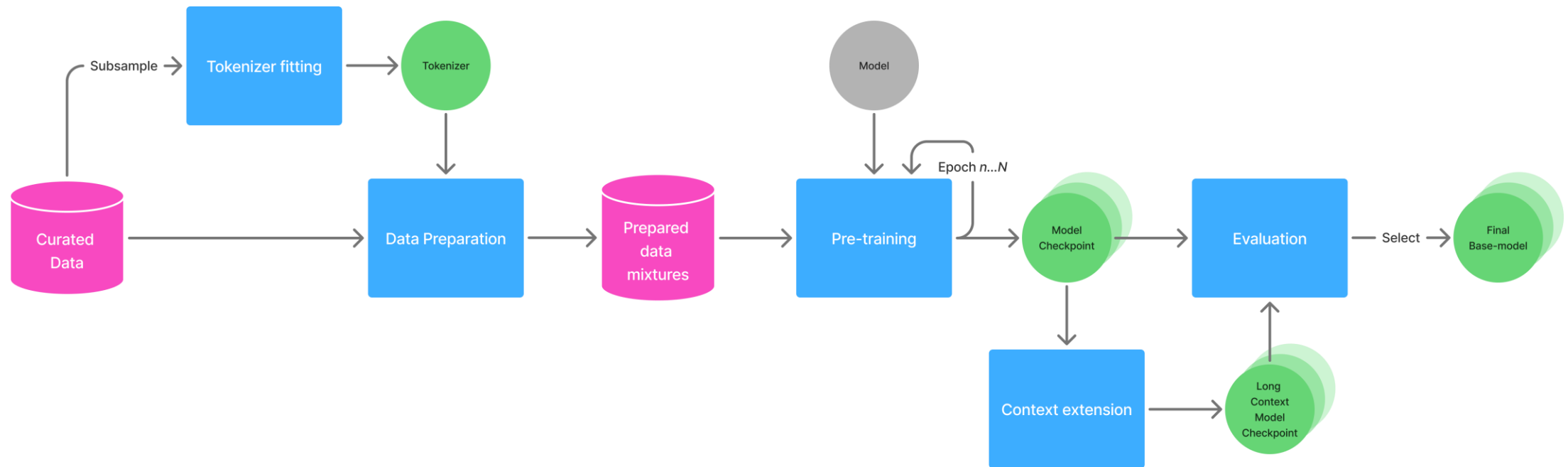
P Nieuws Omvangrijke dataset nieuwsbedrijven voor Nederlands AI-taalmodel GPT-NL
ndpniewsmedia.nl

Kernfunctionaliteiten: samenvatten, versimpelen, RAG

GPT-NL: soeverein, compliant & security by design



Pre-training



GPT-NL Model

Model Size: 26 billion parameters

- Balanced for performance and practical use
- Large enough for complex tasks, small enough to run efficiently

Architecture: Based on Llama 3

- Proven, modern design
- Optimized for speed and quality

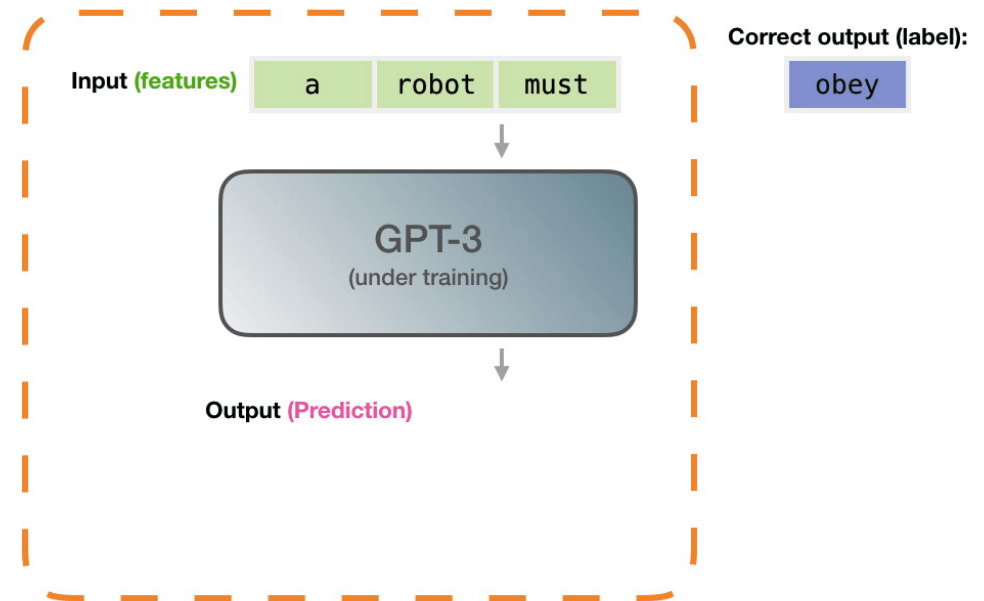
Training Infrastructure:

- 88 high-performance GPUs (H100) on Snellius HPC
- 22 compute nodes (4 GPUs per node), with fast network

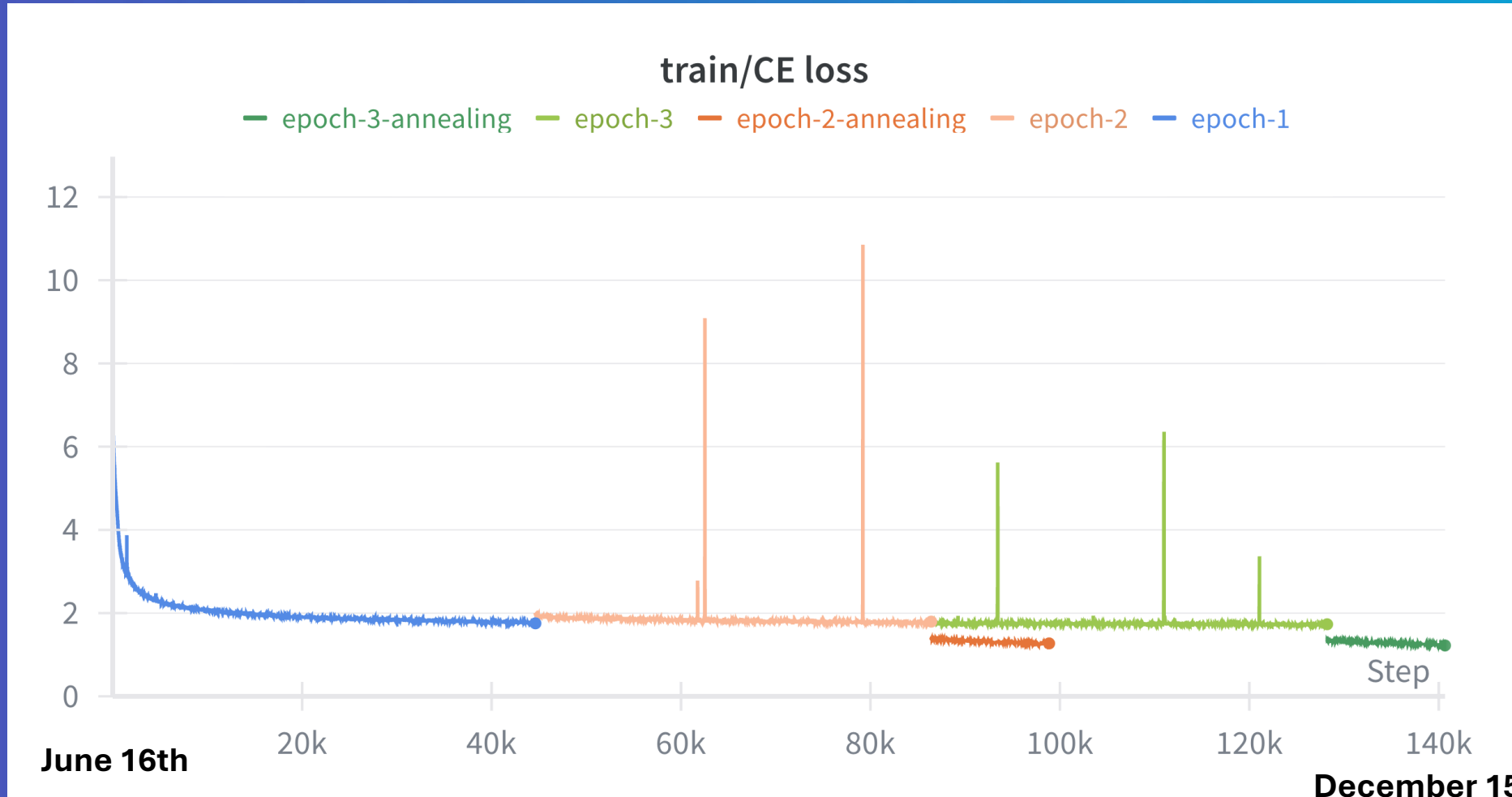
Context window:

- Handles 4,096 tokens at once (~3,000 words)

Unsupervised Pre-training



The loss curve over the course of training



December 15th

~ 1.8T tokens

Post-training: Why pre-training itself is not enough

Pre-trained model is a raw language generator that is not useful yet: it does not follow instructions.

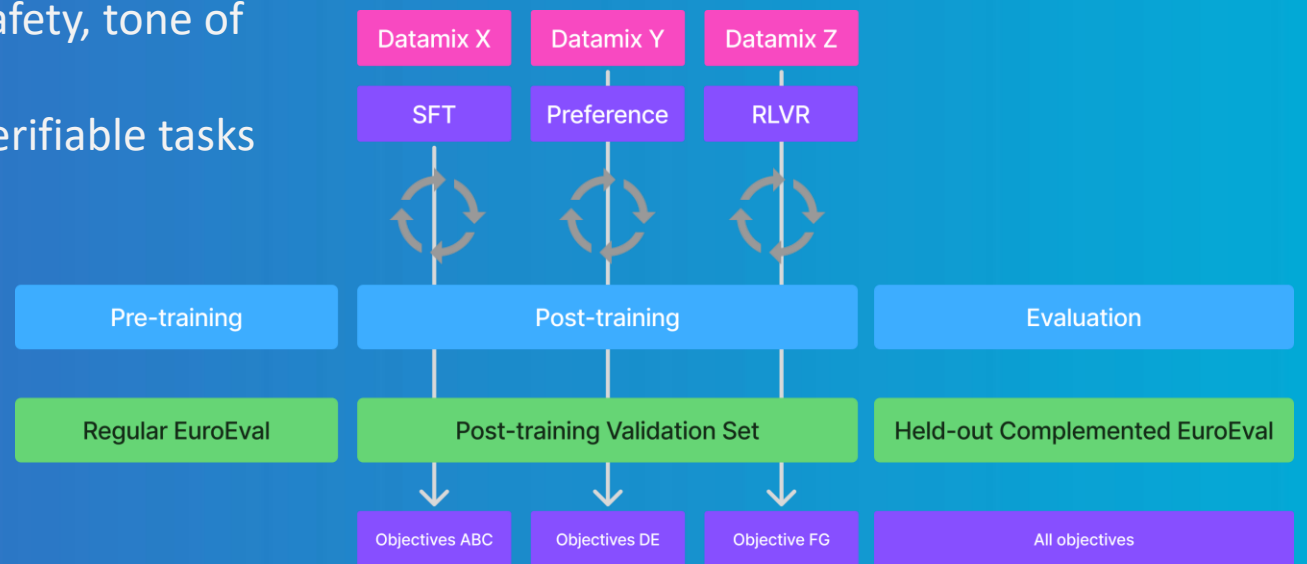
Modern post-training is a series of fine-tunes that build upon each other and require careful design:

- 1. Supervised fine-tuning:** teach formatting and base of instruction following behaviour (e.g., chat interactions, answering questions)
- 2. Preference tuning:** align to human preferences (safety, tone of voice)
- 3. Reinforcement learning:** boost performance on verifiable tasks (e.g., math, precise formatting, reasoning)

In 2025, we focused on 1!

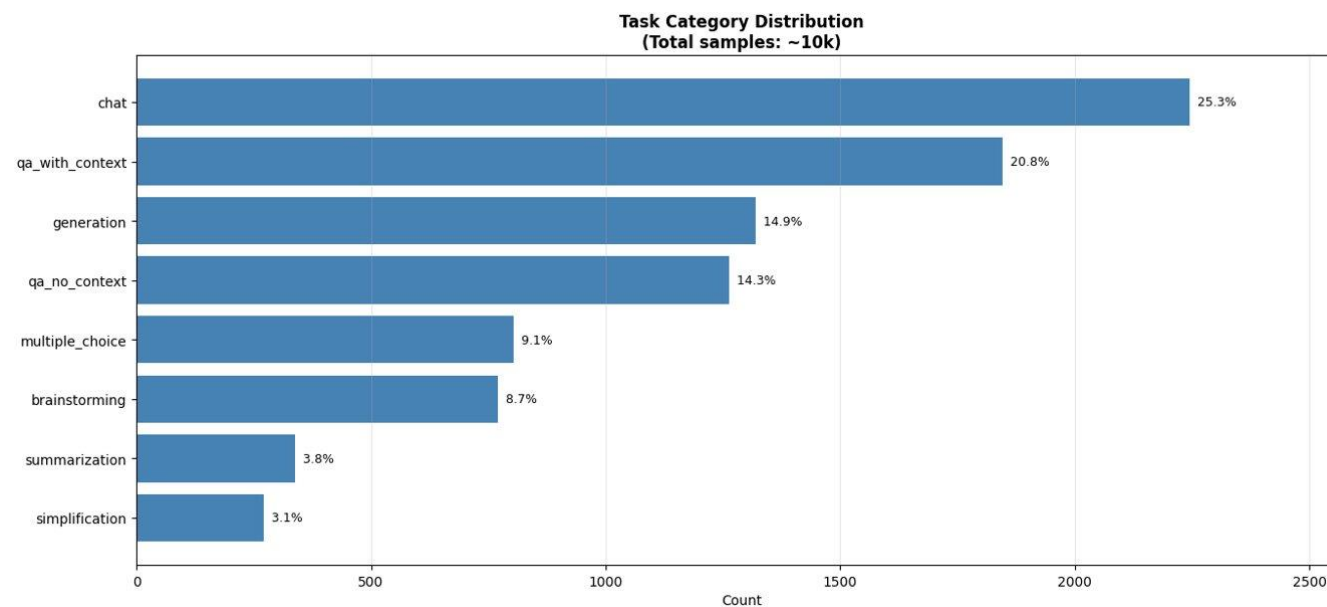
Q: In which year was The Godfather first released?

A: 1972 1974 1976 1978 1980 1982 1984 1986 1988 1990 1992...



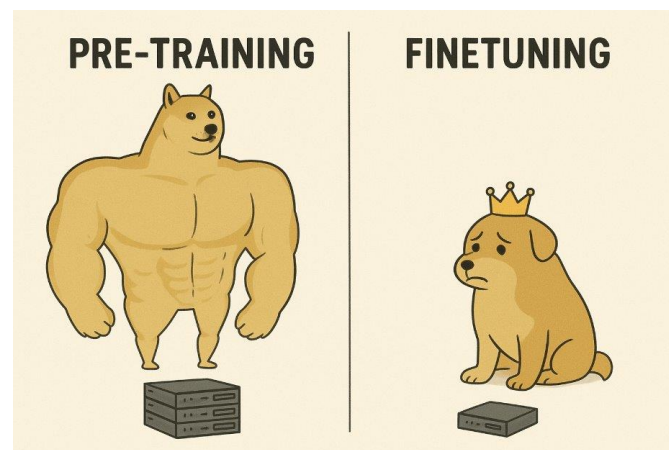
GPT-NL Instruct Dataset

- Collection guidelines
- 2 annotators: 1 for instruction, 1 for completion
- Not easy for annotators (e.g. simplification collected not good)
- ~10K samples total (3M tokens) / aiming for 15K → quite small



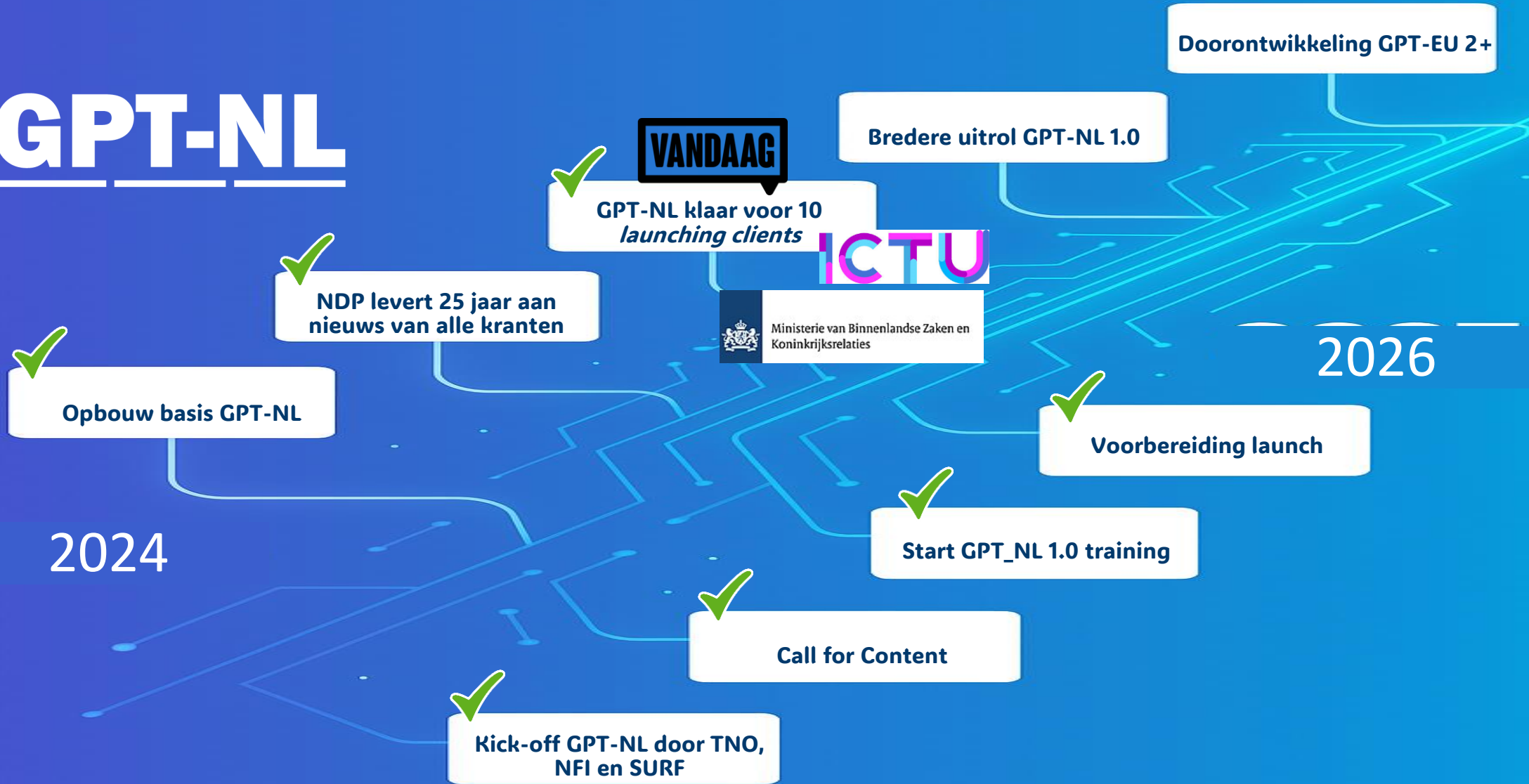
SFT Resources

	Pre-training	Finetuning
Nodes	22	8
GPUs (H100)	88	32
Duration (1 epoch)	~1.5 months	~12h
# epochs	3	1
Dataset size (tokens)	~540B	~170M



Daarom zijn we van GPT-NL een realiteit aan het maken

GPT-NL



De 5 redenen waarom partijen voor GPT-NL kiezen

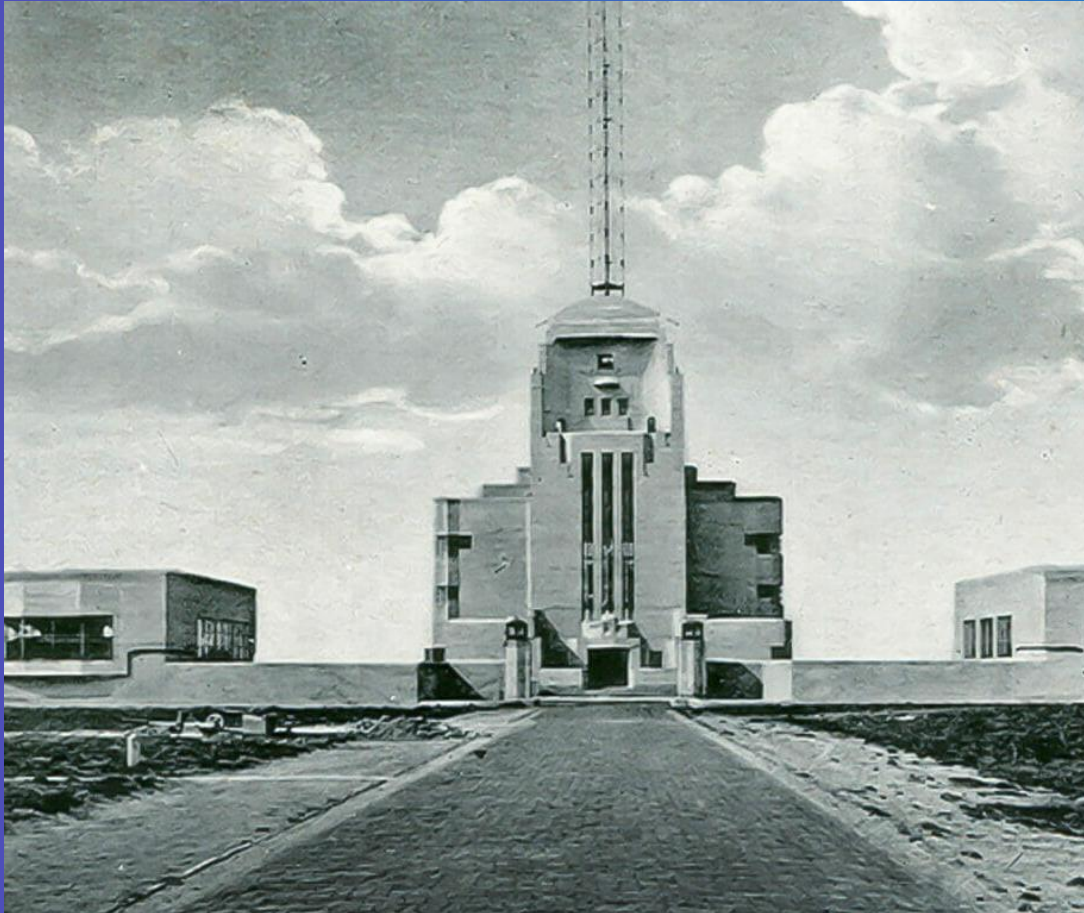


GPT-NL launching customers

- Onderzoeksgeoriënteerde samenwerking met TNO en partij met een operationele LLM-toepassing
- Doelen:
 - Benchmarken en evalueren van GPT-NL voor specifieke toepassing of context
 - Fine-tunen van GPT-NL voor specifieke toepassing of context

<p>Financier  Ministerie van Binnenlandse Zaken en Koninkrijksrelaties</p> <p>In uitvoering met</p> 	<p>Financier  Ministerie van Binnenlandse Zaken en Koninkrijksrelaties</p> <p>In uitvoering met</p> 	<p>Financier  Ministerie van Binnenlandse Zaken en Koninkrijksrelaties</p> <p>In uitvoering met</p> 
<p>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.</p>	<p>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.</p>	<p>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.</p>

Soeverein communiceren, soeverein vliegen...



... en nu op weg naar soevereine AI!

Let's keep in touch!



Saskia Lensink
saskia.lensink@tno.nl



Check onze website www.gpt-nl.nl en
volg ons op LinkedIn!



- **GEMA v. OpenAI (Munich)** — first EU court showdown over generative-AI music licensing; German collecting society alleges ChatGPT used lyrics without rights. *Signals rising EU copyright exposure for US foundation-model providers.* (29 Sep 2025). ([Juve Patent](#))
- **Austrian criminal complaint against Clearview AI** — noyb seeks criminal liability over mass biometric scraping of EU residents; highlights GDPR reach against US facial-recognition vendors. *Escalates European enforcement beyond fines.* (28 Oct 2025). ([Reuters](#))
- **EU public bodies fear dependence on US clouds/AI** (Commission briefing coverage). *Official worry about lock-in and strategic dependence on American stacks used to run AI workloads.* (6 Oct 2025). ([Euractiv](#))
- **“EU cloud and AI must be sovereign”** — interview with Denmark’s digital minister. *Political push to develop EU-controlled infrastructure rather than rely on US hyperscalers behind AI models.* (13 Oct 2025). ([Euractiv](#))
- **Europe can’t agree what “sovereign AI” means** — Bloomberg analysis newsletter. *Competing definitions leave room for US vendors to shape the agenda via investments and chips.* (20 Oct 2025). ([Bloomberg](#))
- **EU AI Act standards revolt** — CEN-CENELEC insiders warn the fast-track process risks weak/opaque norms. *Implementation turbulence that especially affects large US general-purpose models entering the EU market.* (~23 Oct 2025). ([Euractiv](#))
- **Commission unveils voluntary code for the most advanced (foundation) models** — a pre-AI-Act compliance bridge. *Targets transparency/safety duties that bite for US model providers operating in Europe.* (10 Jul 2025). ([subscriber.politicopro.com](#))
- **Meta declines to join EU’s voluntary AI code** — frictions over scope and legal risk. *Illustrates divergence between US platforms and EU compliance expectations ahead of AI Act enforcement.* (Jul 2025). ([The Verge](#))
- **EU caught in an AI copyright bind** — Bruegel policy piece. *Explains how EU copyright rules and global competition put regulators—and foreign model providers—into a tough spot on training-data disclosure and licensing.* (10 Sep 2025). ([Bruegel](#))
- **Mistral demand rises amid fears of US AI dominance** — WSJ report. *Market reaction: buyers seek non-US alternatives for sovereignty/compliance reasons.* (Jun/Jul 2025). ([The Wall Street Journal](#))