GPT-NL

DEFINITION OF SUCCES 2025

TNO innovation for life

Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

SURF

Definition of Success

Auteurs: Duuk Baten (SURF)
Lieke Dom (TNO)
Freek Bomhof (TNO)

Document ID: GPTNL-RAP-25-008
Versie: 1
Datum: 4 november 2025

This publication is licensed under a Creative Commons Attribution 4.0 International.

GPT-NL is een samenwerking van TNO, SURF, en het NFI
Financiering: Faciliteiten Toegepast Onderzoek (FTO) vanuit RVO/EZK

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

TNO innovation for life
Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid
SURF

INTRODUCTION

In the Definition of Success, we specify what our deliverables will look like and illustrate their success criteria. The Definition of Success clarifies the expected project deliverables for stakeholders, making clear what is in scope and what is not. Our success criteria are ambitious, though realistic and measurable.

If you want to read about our ambitions beyond success criteria, please see our commitments to developing GPT-NL responsibly and our Mission & Vision where we explain how GPT-NL contributes to a fair and sovereign future for LLMs in Europe.

About the GPT-NL Project

GPT-NL is an ambitious project with strong ideals. The mission of GPT-NL is to offer a responsible alternative to the development of Large Language Models (LLMs). The emergence of GenAI technologies has resulted in various problems, such as unlawful use of training data, lack of alignment with European legislation, a lack of transparency, and concerning power concentration at big tech companies.

Our vision is that LLMs are only trained on data that is legally obtained, these content providers get a fair place in the development of LLMs, technology is built in accordance with AI and privacy legislation by design, and that LLMs should be developed transparently, with its developers providing as much insights as possible into how these models are built. With the GPT-NL project, we want to offer an alternative that is reciprocal, trustworthy, transparent, and ensures sovereignty of our citizens and institutions.

With the LLM and the knowledge we obtain of its development, we can achieve three objectives.

THE GPT-NL OBJECTIVES

1. To demonstrate that it is possible to create useful LLMs that comply with European laws and values.
2. To share the knowledge on how such LLMs can be developed in a transparent and responsible way.
3. To strengthen the European AI innovation landscape based on a fair exploitation of the technology and benefiting data contributors and society alike.

Deliverables

To achieve these objectives, the GPT-NL project will have four deliverables. With these we support our mission towards a fair and sovereign ecosystem for LLMs in Europe.

1. **Instruction-tuned Large Language Model (LLM)**
   A performant and compliant instruction-tuned LLM which is responsibly developed in alignment with our values. We will deliver a model with model weights, publish the training data that we are allowed to share, and source code for building the model.

2. **Documentation**
   A technical report that explains the architecture, training process, datasets, evaluations, experimental results and licensing model. This will include lessons learned and transparency reporting such as a Model Card, Datasheets, Decision Sheets, and the source code used to develop the model and its documentation. We aim for these materials to function as reporting for the project as well as provide possible good practices for future similar initiatives.

3. **Broad eco-system**
   A broad ecosystem of data providers, AI developers, researchers, and users of LLM solutions. A resilient ecosystem can be a basis for subsequent initiatives and ensure broad support for the Netherlands future digital sovereignty.

4. **Research facility**
   A research facility for TNO[1] research projects in the field of LLMs. This environment brings together computing resources and tools to strengthen TNO's ability to build knowledge on the functioning of LLMs.

Funding
In October 2023, we received funding from the Ministry of Economic Affairs and Climate (Economische Zaken & Klimaat) for a facility for applied research.[2] The four deliverables together form such a platform for further research and development of LLMs.

[1] TNO (Netherlands Organisation for Applied Scientific Research) is an independent statutory research organisation focused on applied research. An LLM research facility improves TNO's ability to perform LLM related research projects for both government and private sector.

[2] Funding Faciliteiten Toegepast Onderzoek' (EN: Facilities for Applied Research):https://www.rijksoverheid.nl/actueel/nieuws/2023/11/02/14-nederlandse-onderzoeksfaciliteiten-krijgen-184-miljoen-euro-ondersteuning

DELIVERABLES

1.
Instruction-tuned LLM

2.
Documentation

3.
Ecosystem

4.
Research facility

1. Instruct-tuned LLM

The GPT-NL project will deliver a 26B parameter large language model that:

- **is trained on data that is lawfully obtained**
- **is developed in accordance with European legislation**
- **is developed in a transparent manner to enable research and auditability**

Training data

The model will be pretrained on at least 300 billion text tokens (mostly English, Dutch and German texts) and will be further enriched with 150 billion tokens of code. This dataset only contains data that has been lawfully acquired. This means we only use data that is:

- Provided by data-owners who have opted-in to share their content;
- Licensed in a way that permits its use for training LLMs (e.g. CC-0 & CC-BY);
- Created synthetically in a way that does not infringe on intellectual property rights.

The data is then curated whereby PII (personally identifiable information) is replaced by synthetic alternatives, and toxic language and other inadmissible content is filtered out. These choices will be documented and publicly shared in a decision sheet.

Functionality

The model will perform text generation, summarization, and simplification tasks at a level of performance comparable to the Llama2 7B model and GPT-3 175B (or equivalent) models. Performance will be measured in Dutch, on benchmarks that are publicly available (i.e. in EuroEval). Where we see gaps in the quality of availably of Dutch benchmarks, we will contribute to fill these gaps with benchmarks that we will release publicly. We do not optimize the model for specific benchmarks.

Documentation

The model will be supported by extensive documentation, gathered in the accompanying technical report. This documentation enables understanding and transparency of the development process by providing insight into design choices.

1. Instruct-tuned LLM

We see this as being delivered successfully when:

- The model can be used in production by Dutch organisations.
- The model can summarize, simplify, and support information retrieval implementations.
- The model delivers sufficient performance for general tasks, providing a reliable base for domain-specific fine-tuning and further research.
- We have trained a model using only data that we could get permission to, either through permissive licenses (e.g. CC-0 & CC-BY) or explicit consent.
- The dataset exceeds 300 billion tokens.
- We benchmarked the model in Dutch contexts on language, performance, and bias, and shared the results publicly.
- We have clearly described licenses for the model and dataset.
- We have shared (at least) the following datasets publicly:
    - *Open training datasets*
    - *Dataset cards for proprietary training datasets*
    - *Our self-developed finetuning dataset*
- We have shared our code open-source and maintain the code-repository.
- We have published model card(s) and datasheets for the different model(s) and datasets.

2. Documentation

In line with objective 2, to share knowledge on how to develop LLMs in a responsible way, we will share extensive documentation. This will contain details on our model, project, as well as broader learnings.

Technical report
A technical report containing at least an explanation of the architecture, training process, datasets, evaluations, and experimental results of the GPT-NL model. This collection will function as reporting for the project as well as be a resource for future similar initiatives. The technical report contains:

- Datasheets: *these describe qualities of the various datasets;*
- Model card: *this describes the qualities of the model;*
- Licensing: *The documentation will clearly describe the licenses under which the model can be used;*
- Decision sheets: *these describe decisions made during the development of the model, for example choices concerning data curation or the data mix, using a self-developed framework;*
- Transparency: *we will report on our level of openness using the European Open-Source AI index[3];*
- Benchmarking results: *describing how the model scores on various benchmarks.*

Lessons learned
We have learned through experience and hope these learnings can be seen as possible good practices for responsible development of foundational models. Sharing these experiences and lessons learned through various publications is an important way to contribute to developing a Dutch and European ecosystem.

We see this as being delivered successfully when:

- We have published a technical report
- We have shared various project learnings through reports, papers, or other publicly accessible formats.

[3] *See https://osai-index.eu. Please note: We are limited in our ability to score maximal openness due to our approach to data collection as well as legal constraints, see: https://gpt-nl.nl/nieuws/open-gpt-nl/.*

3. Ecosystem

Part of the ambition of the GPT-NL project is to contribute to a robust ecosystem of data providers, AI-developers, AI-users, and researchers. This ecosystem can support future iterations of GPT-NL and similar projects.

We see this as being delivered successfully when:

- We are able to define organisations that are included in these stakeholder groups. These groups can be used for advice on choices within the GPT-NL project. As of publication (November 2025), we have defined a:
    - *Content Board for data providers*
    - *User Board for GPT-NL developers and (end-)users*
    - *Research Community*
- We have been able to give these stakeholder groups a fair and meaningful place in the GenAI ecosystem, that means that:
    - *Data providers are involved in decisions about the use of their data (for example, on the question of sharing training data for LLM research).*
    - *We cooperate with users and developers*
    - *We enable new opportunities for researchers*
- We have created a sustainable business model within this ecosystem for the future of GPT-NL.
- We have been able to bring these four stakeholder groups (Data providers, developers, users, and researchers) together, for example on a “GPT-NL community day”.
- It is possible for partners in the ecosystem to further finetune the GPT-NL model for specific purposes. There is knowledge exchange and collaboration within the EU with similar initiatives, for example with TrustLLM, OpenEuroLLM, LLMS4EU or ETH/EPFL. There are spin-off effects and initiatives within the Netherlands.

4. Research facility

We have built and exploited a research facility for LLM research within TNO. The goal of this facility is to enable research by TNO on Generative AI without being dependent on non-European infrastructures, in other words to enable sovereignty in this area.

The facility will consist of:

- Access to various training datasets
- Access to the data curation pipeline
- Access to the training software and procedures
- Platform to run automated tests (cybersecurity, functionality, benchmarks)

The facility will be available to TNO researchers and, through collaboration agreements, partners of TNO.

We see this as being delivered successfully when:

- We have made the training data available for researchers who need data for follow-up research activities that involve language data.
- We have made the data curation pipeline available in a way that it can be operated without extensive adaptations.
- We have made the training software and procedures available in a way that it can be operated without extensive adaptations.
- The automated test platform remains operational.
- We have a clear description of the requirement through which partners of TNO can gain access to the research facility.

Looking beyond success criteria

Success of the GPT-NL project can only be assessed retrospectively. The described success criteria will be used to evaluate if we have successfully concluded the project. However, GPT-NL is an ambitious project, that means we want to a step further where we can. That means we operate with the following ambitions in mind:

- We have completed three use cases with external partners as validation of the model. Evaluations will be done at least on benchmarks of language, performance, and bias; we aim to do additional (use case specific) benchmarks as well.
- Even though the model will be a ‘first version’, meaning that next training rounds to improve the model are out of scope of this project, our ambition is to release improved versions in the future.
- We maintain engagement with the community and larger audience for widespread dissemination of our learnings and good practices.
- The ambition is that also partners from TNO make use of the facility (the training data, model, curation tools, and testing platform).
- The community is strengthened by an increasing number of use case requests, increasing amount of data, and wider inclusion of underrepresented groups. An ambition is to organize follow-up events for community members (e.g., Content Board meetings, User Board meetings, Expert Sessions).

Lastly, given the impact that generative AI has had on society, we see it as our responsibility to offer transparency in a way that a general audience can understand, for example with blogs, presentations, and media outreach.

Contact us:
info@gpt-nl.nl
www.gpt-nl.nl
www.linkedin.com/company/gpt-nl/