GPT-NL

Mission & Vision

GPT-NL Model

Introduction

The GPT-NL initiative, a collaboration between TNO, SURF, and NFI, aims to support the Large Language Model community through the development of a state-of-the-art research facility and the pursuit of a sovereign, lawful, trustworthy and transparent language model. We aim to create useful and yet law compliant and sovereign large language models that provide an alternative to big tech's generative AI models. We share the knowledge gained during model development to foster a collaborative and open environment for LLM research. Furthermore, we strive at a fair and responsible use of the technology that benefits contributors and society alike.

Why

Alternative

01#

The current LLM-market does not offer sufficient transparency and cannot ensure that they comply with our laws and values.

Complexity

02#

The quality of current LLMs, that are trained on translated Dutch texts, is insufficient when it comes to understanding Dutch texts in complex cases.

Believe

03#

We believe that technology should be reciprocal, trustworthy, transparent, and ensure sovereignty of our citizens and institutions.

How

01#

A competitive model

By offering a useful alternative to existing LLMs, we'll demonstrate that it is possible to create language models that comply with our laws and values.

02#

A new standard

By being open and transparent about the data sources and decisions made during the design process, we'll set a standard on how to develop LLMs in responsible way and in compliance with legislation.

03#

A strong ecosystem

By cooperating with stakeholders, we'll create a strong ecosystem that strengthens the European AI innovation landscape.

What
A RESPONSIBLE LARGE LANGUAGE MODEL BUILT FROM SCRATCH

Data

We will develop, train and fine-tune one large language model using mostly Dutch, English, German, and Code sources. This will be a combination of only:
· Opt-in data
· Data that is legally accepted for the training of LLMs
· Non-IP infringing synthetic data

Performance

The GPT-NL model will be trained on at least three hundred billion text tokens and can perform text generation, summarization, and simplification tasks at a level of performance comparable to the Llama2 7B model, GPT-3 175B models. The dataset will be finished with 1,5 hundred billion tokens of code.

Transparency

The GPT-NL Model will be supported by extensive documentation on the GPT-NL model to enable understanding and transparency, including a datasheet describing qualities of the dataset and a model card describing qualities of the model.

Core
Values

Sovereignty

We believe in freedom of choice and independence from the big tech market. GPT-NL also believes in the relevance of cultural diversity language models.

Trustworthy

The GPT-NL Model is designed to be compliant with Dutch and European laws and to have a competitive, technical quality. Our organisation is reliable and committed to building Responsible AI.

Reciprocity

We believe innovation should benefit everyone and should contribute to a fair and inclusive society. Technology should be built in cooperation with important stakeholders.

Transparency

We are accountable, verifiable, and honest. We strive to be as open as possible.

Who we are
Personality

Inspiring   Honest

Open   Committed

Our
Vision &
Mission

What is our future dream?

Our future dream envisions a Europe capable of building powerful LLMs in the way that serves citizens best, wherein the technology adheres to our laws, norms, and values.
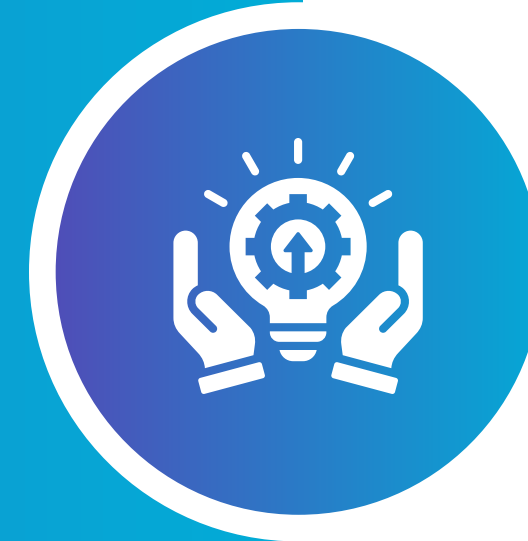
What will we achieve?

GPT-NL will strengthen the Dutch and European AI innovation landscape by supporting a fair, sovereign, and safe ecosystem of knowledge-share around performant, responsibly developed, and legally compliant LLMs.

How will we get there?

Through the development of the GPT-NL Model, comprehensive documentation, and active engagement with LLM communities, we will build a strong ecosystem where knowledge is shared.

Future
Goals

In 1 year from now

we have launched the GPT-NL Model and developed a fair business model for data suppliers and content creators for their contributions.

In 5 years from now

our key partners have integrated customized versions of GPT-NL into their core processes, while the general public can access it through various trusted providers. This in turn has inspired the creation of multiple Dutch and European LLMs.

In 10 years from now

we will have strengthened the digital sovereignty of Europe by increasing capability in AI, by developing technologies aligned with the AI Act, and earning the trust of the European people.

GPT-NL

December, 2024 | info@gpt-nl.nl