Large dataset news organizations for Dutch AI language model GPT-NL

Dutch news publishers contribute a vast collection of articles to train the GPT-NL AI language model

TNO and members of NDP Nieuwsmedia who represent the vast majority of Dutch news publishers are working together to further develop GPT-NL, the first large-scale Dutch AI language model trained entirely on legally obtained data. Members of NDP Nieuwsmedia are making a substantial portion of their archives—containing articles from over 30 national and regional news titles—available to train the model. The addition of the dataset is expected to double the amount of high-quality Dutch data used for training. News agency ANP is also joining the initiative. It's the first time worldwide that private news publishers are collaborating collectively with a organization developing an AI model.

The collaboration between NDP Nieuwsmedia and TNO supports the shared goal of GPT-NL and the Dutch government: to create a language model that respects copyright and sets an benchmark for how to handle copyrighted content in AI systems. Strict agreements have been made to prevent the articles from being technically extracted from the AI-model. When GPT-NL is released, publishers will receive appropriate remuneration.

GPT-NL is a non-profit initiative by TNO, NFI, and SURF, and offers a responsible alternative to existing language models. It is built for the Netherlands using high-quality Dutch data. Unlike some international models that use large parts of content scraped from the internet without permission, GPT-NL collects copyrighted data carefully and ethically, and remunerates those who contribute their content. It also complies with European laws like the AI Act. The model is being developed for specific tasks such as summarizing, simplifying, and extracting information from text.

News article data
Thanks to the collaboration with NDP Nieuwsmedia—the trade association of private news publishers such as DPG Media, Mediahuis, Erdee Mediagroep, and De Groene Amsterdammer—and the participation of ANP, the model gains access to over 20 billion tokens. The articles cover many topics, from politics and economics to healthcare and science, and form a rich source for training GPT-NL. The datasets contain billions of tokens—small pieces of text that help AI understand and process language. A token can be a word, part of a word, or even a punctuation mark.

Rien van Beemen, chair of NDP Nieuwsmedia:
"Big Tech companies have trained their models on news articles without permission or payment. This partnership between NDP Nieuwsmedia members and TNO shows there is an alternative route. We're setting a precedent that helps the advancement of AI and strengthens journalism in the Netherlands. AI innovation can be ethical and responsible—without using the work of our journalists without permission. This step gives that movement a boost."

Selmar Smit, Manager of Science & Technology at TNO and founder of GPT-NL:
"We're proud of this collaboration. NDP Nieuwsmedia members are not only providing high-

quality data, but also sending a strong message: AI can be developed responsibly, with respect for copyright and public values."

Timeline

TNO, NFI, and SURF began developing GPT-NL in 2023. Training started in June 2025. In the fourth quarter of 2025, the model will be improved and prepared for first use. Earlier Dutch contributors of data to GPT-NL include DNB (De Nederlandsche Bank), ICTRecht, and Het Utrechts Archief.

============== Attention !!!!: UNDER EMBARGO untill juli 17th 2025 12.00 PM ==============

NDP Nieuwsmedia
De brancheorganisatie voor nieuwsbedrijven

GPT-NL

TNO   SURF   Nederlands Forensisch Instituut
Ministerie van Justitie en Veiligheid

============== Attention !!!!: UNDER EMBARGO untill juli 17th 2025 12.00 PM ==============

Information for the press:

On NDP Nieuwsmedia
Stefan Heijdendael, strategic advisor, +31-(0)633428547,
S.Heijdendael@ndpnieuwsmedia.nl

On GPT-NL

Rein Croll, spokesperson +31 (0)61 171 87 84, rein.croll@tno.nl.

TNO

TNO is the largest independent research and technology organisation in the Netherlands and one of the largest in Europe. TNO innovates and researches. By innovating, but also by orchestrating (inter)national consortia and ecosystems, we create technological and methodological breakthroughs that enhance the earning capacity of the Dutch economy. This is how TNO contributes to a safe, sustainable, healthy and digital society. In addition, we inform government and society with sound scientific research and reliable measurements. In this way, we provide a scientific basis for decision-making.

More information: www.tno.nl

NDP Nieuwsmedia
NDP Nieuwsmedia is the trade association for news companies. NDP Nieuwsmedia promotes the journalistic and business interests of Dutch news companies. The vast majority of news companies in the private sector are members of NDP Nieuwsmedia. The 11 news companies jointly operate dozens of news brands: daily newspapers, news sites, news apps and news programmes on radio and television.

More information: www.ndpnieuwsmedia.nl

============== **Attention !!!!: UNDER EMBARGO untill juli 17th 2025 12.00 PM** ==============