



# Culture-Sensitive Assessment and Adjustment of Large Language Models – Adaptation to the Nordic-Baltic Societies

Inguna Skadiņa

Institute of Mathematics and Computer Science, University of Latvia



# Background

- Current language-centric AI takes large language models (LLMs) as its most prominent inner-wheel component
- It is, however, widely acknowledged that today's most successful and utilized LLMs – even if multilingual in their set-up – are, in essence, **English-dominated, monocultural** and **US-centric**, and that they therefore tend to **reinforce and homogenize languages and cultural norms and values**

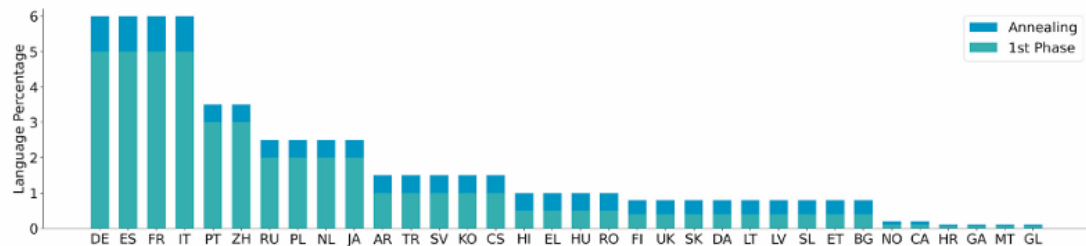


Fig. 4. Percentage of the training corpus attributed to each language, excluding English which accounts to 50% in the first phase and 32.5% during annealing. 5% of the corpus is left for datasets composed of code and math in the first phase and 7% during annealing.

# CAALLM: background

- LLMs are seen to trivialize and homogenize our languages and impact the deeply rooted and highly profiled and appreciated societal values of our region, such as **equality, democracy, trust, welfare, and inclusion.**
- CAALLM addresses these challenges from a **Nordic and Baltic perspective**



# CAALLM: background

- LLMs are seen to trivialize and homogenize our languages and impact the deeply rooted and highly profiled and appreciated societal values of our region, such as **equality, democracy, trust, welfare, and inclusion.**
- CAALLM addresses these challenges from a **Nordic and Baltic perspective**



# Culture-Sensitive Assessment and Adjustment of Large Language Models – Adaptation to the Nordic-Baltic Region

## Some requirements and solutions:

- High-quality, multilingual and monolingually derived datasets for our language communities
- Improved understanding of the models' inner workings wrt. linguistic and cultural bias
- Improved evaluation and alignment methods

## Challenges:

- Low-resourced languages, of which there are several in our regions
- **Cultural NLP** is a new and quite unexplored field of AI

# Culture-Sensitive Assessment and Adjustment of Large Language Models – Adaptation to the Nordic-Baltic Region

## Aim of CALLM:

- To facilitate the adaption of large language models towards a more **responsible** coverage and **functionality** that encompass the linguistic, cultural, and societal **diversity** in the Nordic and Baltic regions.

## The project

- Brings together language institutions and NLP research groups from the Nordic-Baltic Region

# Consortium

- Centre for Language Technology University of **Copenhagen**  
(coordinator)
- Centre for Language Technology, University of the **Faroe Islands**
- NLP Section, University of **Oslo**
- Språkbanken, University of **Gothenburg**
- The Society for **Danish** Language and Literature
- Institute of Mathematics and Computer Science, University of **Latvia**

# Three 'legs' in CAALLM

- **Diagnose** linguistic biases and cultural bias through datasets (use translation as proxy)
- **Extend** explainability of language models wrt. linguistic and cultural bias (investigate the internal behavior of LLMs)
- **Adjust** linguistic and cultural biases in LLMs (including norms, values, aboutness) through techniques such as finetuning and reinforcement learning from human feedback, direct preference optimization

# Culture-Sensitive Assessment and Adjustment of Large Language Models – Adaptation to the Nordic-Baltic Region

## We will:

- Compile a number of open-source **linguistic and cultural multi-parallel datasets** for **Danish, Swedish, Bokmål, Nynorsk, Faroese, and Latvian**, which will systematically draw on and make explicit the central aspects of the linguistic and cultural diversity of our regions.
- Based on these data, we hope to advance state-of-the-art **methods for explaining, assessing, and aligning LLMs across languages and cultures**, with particular focus on the linguistic idiosyncrasy and cultural heritage of our regions.

# There is no single recipe



Excellent project idea



Strong consortium



Alignment with call requirements

Call: Responsible use of Artificial Intelligence



Good writing skills

# However, not always good idea and partners are sufficient

- Out of 200 proposals 17 research projects have been awarded, acceptance rate **below 10%**
- CAALLM, overall grade: **6**
- Empowered North: Responsible AI for local journalism in the Nordic and Baltic countries, overall grade: **6**

*The purpose of Empowered North is to bolster local democracy in the Nordic-Baltic region by providing new AI-enabled tools for local journalists*

# Thank you!

Latvian Council of Science / NordForsk  
Support for Latvia's participation in international research and innovation programmes  
1.1.1.5/1/24/I/001 | NordForsk/2026/1



Latvian Council of Science

