NBREAL 2025

The 1st Workshop on Nordic-Baltic Responsible Evaluation and Alignment of Language Models

Proceedings of the Workshop

March 2, 2025

Published by:

University of Tartu Library, Estonia
Indexed in the ACL Anthology

ISBN 978-9908-53-116-8

Introduction

We are excited to welcome you to NB-REAL 2025 (Nordic-Baltic Responsible Evaluation and Alignment of Language Models), a half-day workshop focusing on the responsible evaluation and alignment of Large Language Models (LLMs) for Nordic and Baltic languages. The workshop was held on March 2, 2025, bringing together researchers and practitioners working on ethical benchmarks, culturally sensitive alignment datasets, and responsible LLM evaluation techniques for Nordic and Baltic languages.

The NB-REAL workshop aimed to address critical challenges in the development and evaluation of language models for Nordic and Baltic languages, with a particular focus on ethical considerations and cultural sensitivity. The program included a keynote presentation, three oral presentations, and three poster presentations, covering a diverse range of topics from cultural awareness evaluation to multilingual tweet analysis.

We received 9 submissions this year. Each submission underwent a rigorous double-blind review process, with three reviewers assigned to each paper. Our program committee, consisting of 9 dedicated reviewers, provided thorough evaluations and constructive feedback. After careful consideration of the reviews and discussions, we accepted 7 papers, while 1 paper was rejected and 1 was withdrawn, resulting in an acceptance rate of 78%. The accepted papers were presented either as oral presentations or posters, based on their content and format.

The workshop program featured three oral presentations covering important topics such as cultural awareness evaluation of Danish language models, crowd evaluation of translations, and the development of Danish idiom datasets. The poster session showcased three additional papers focusing on multilingual LLM evaluation, particularly for Baltic languages, and image-text relation prediction.

A workshop of this scale requires the dedication and support of many individuals, and we have many people to thank. We extend our gratitude to our program committee members for their thorough reviews and valuable feedback: Barbara Scalvini, Garðar Ingvarsson, Iris Edda Nowenstein, Kenneth Enevoldsen, Lars Bungum, Mathias Stenlund, Peter Ebert Christensen, and Steinunn Rut Friðriksdóttir. Their expertise and commitment were essential in ensuring the high quality of the accepted papers.

This workshop was organized as part of the TrustLLM project (the European Commission, grant agreement no. 101135671), an EU-funded initiative aimed at developing trustworthy large language models. We gratefully acknowledge this support, which made the workshop possible.

Finally, we thank all the authors who submitted their work to the workshop and all participants who contributed to making NB-REAL 2025 a success. Through their contributions, we have taken important steps toward establishing more responsible and culturally aware approaches to LLM evaluation and alignment for Nordic and Baltic languages.

Hafsteinn Einarsson, Annika Simonsen and Dan Saattrup Nielsen, Program Chairs

Organizing Committee

Program Chairs

Hafsteinn Einarsson, University of Iceland, Iceland
Annika Simonsen, University of Iceland, Iceland
Dan Nielsen, Alexandra Institute, Denmark

Program Committee

Program Committee

Lars Bungum
Peter Ebert Christensen
Kenneth Enevoldsen
Steinunn Rut Friðriksdóttir, University of Iceland
Garðar Ingvarsson Juto, Miðeind ehf.
Iris Nowenstein, University of Iceland
Barbara Scalvini, University of the Faroe Islands
Mathias Stenlund, University of Iceland

Invited Speakers

Annika Simonsen, University of Iceland, Iceland
Dan Nielsen, Alexandra Institute, Denmark

Keynote Talk

Aligning and Evaluating Language Models: Challenges for Low-Resource Languages

Dan Saattrup Nielsen and Annika Simonsen
Alexandra Institute and University of Iceland

2025-03-02 09:15:00 – Room: **Venue at Hestia Hotel Europa**

Abstract: This keynote presentation examines two crucial aspects of developing reliable language models: alignment strategies and evaluation frameworks. The first part will focus on language model alignment, particularly for Germanic languages, presenting recent work from the TrustLLM project. We discuss key challenges in ensuring reliable and ethically sound language models, especially addressing the scarcity of alignment data for low-resource languages. The second part will provide a comprehensive overview of language model evaluation approaches, from traditional benchmarks to emerging methodologies like LLM-as-a-judge. We examine evaluation frameworks with special attention to low-resource languages, highlighting both available resources and critical gaps in evaluation datasets. The presentation emphasizes the interconnected nature of evaluation and alignment in developing trustworthy language models.

Bio: Annika is a Faroese computational linguist and PhD student at the Department of Computer Science, University of Iceland. As part of the TrustLLM project, her research focuses on Germanic language model alignment, building high-quality training and evaluation data, and aligning models.

Dan is a Senior AI Specialist from the Alexandra Institute in Denmark. He has a PhD in Mathematics and has worked with AI within both academia and industry, with 5+ years of experience in low-resource NLP. He is the creator and lead maintainer of the European LLM evaluation framework ScandEval.

You can find Dan on platforms such as GitHub, Hugging Face, LinkedIn, Bluesky, etc. with the username **saattrupdan**. His website is saattrupdan.com

Table of Contents

Program

Sunday, March 2, 2025

09:00 - 09:15 *Opening Remarks*

09:15 - 10:00 *Keynote Speaker: Annika & Dan*

10:00 - 10:30 *Coffee Break*

10:30 - 10:50 *Paper Presentation 1*

DaKultur: Evaluating the Cultural Awareness of Language Models for Danish with Native Speakers
Max Müller-Eberstein, Mike Zhang, Elisa Bassignana, Peter Brunsgaard Trolle and Rob Van Der Goot

10:50 - 11:10 *Paper Presentation 2*

What's Wrong With This Translation? Simplifying Error Annotation For Crowd Evaluation
Iben Nyholm Debess, Alina Karakanta and Barbara Scalvini

11:10 - 11:30 *Paper Presentation 3*

The Danish Idiom Dataset: A collection of 1000 Danish idioms and fixed expressions
Nathalie Hau Sørensen and Sanni Nimb

11:30 - 11:40 *Closing Remarks and Future Directions*

11:40 - 13:00 *Poster Presentations*