– WNNLP 2024 –
The 6th IN5550 Teaching Workshop on Neural Natural Language Processing

University of Oslo

June 3rd 2024



WNNLP 2024 at a glance

- 21 accepted papers submitted by 31 students (26 MSc + 5 PhDs)
- 5 different thematic tracks:
  1. Scandinavian language identification (Area Chairs: David Samuel and Petter Mæhlum)
  2. Pretraining Tiny Language Models (Area Chair: Lucas Georges Gabriel Charpentier)
  3. Fact-checking with graph evidence (Area Chairs: Roxana Pop and Sondre Wold)
  4. Artificial text detection (Area Chair: Vladislav Mikhailov)
  5. Predicting a prompt from a generated text (Area Chairs: Louis Give, and Andrey Kutuzov)
- All papers peer-reviewed by two students + one instructor.
- Workshop proceedings: to appear soon

Program

- 12:15 - 12:20 Welcome and awards
- 12:20 - 12:21 Scandinavian language identification (SLI) task overview
- 12:21 - 12:43 6 SLI lightning talks
- 12:43 - 12:44 Pretraining Tiny Language Models (TLM) task overview
- 12:44 - 13:00 5 TLM lightning talks
- 13:00 - 13:15 Break
- 13:15 - 13:16 Fact-checking with graph evidence (FCG) task overview
- 13:16 - 13:27 3 FCG lightning talks
- 13:27 - 13:28 Artificial text detection (ATD) task overview
- 13:28 - 13:45 5 ATD lightning talks
- 13:45 - 13:46 Predicting a prompt from a generated text (PPG) task overview
- 13:46 - 13:55 2 PPG lightning talks
- 13:55 - 14:00 Closing

Awards

- Only MSc students considered, not PhD fellows.

Awards

- Only MSc students considered, not PhD fellows.

Best paper

Awards

- Only MSc students considered, not PhD fellows.

Best paper

- Author: Tobias Opsahl
- Title: Fact or Fiction? Exploring Diverse Approaches to Fact Verification with Language Models

Awards

- Only MSc students considered, not PhD fellows.

Best paper

- Author: Tobias Opsahl
- Title: Fact or Fiction? Exploring Diverse Approaches to Fact Verification with Language Models

Highest scores across assignment leader-boards

Awards

- Only MSc students considered, not PhD fellows.

Best paper

- Author: Tobias Opsahl
- Title: Fact or Fiction? Exploring Diverse Approaches to Fact Verification with Language Models

Highest scores across assignment leader-boards

- Vebjørn Haug Kåsene
- Was among top-2 in all the three obligatory assignments
- Great job!

Task Overview: Scandinavian language identification

Language	Gold train	Gold valid	Silver train	Evaluation set
Bokmål	15 435	2 392	22 860	1 939
Danish	4 376	563	34 173	565
Nynorsk	14 029	1 859	24 382	1 511
Swedish	4 278	503	34 252	1 219
Other	8 805	1 165	—	1 742

Table: Number of sentences across the datasets and languages. Every sentence in the training and validation splits is annotated with a single label, while the sentence in the evaluation set can be multi-labeled.

Task Overview: Pretraining Tiny Language Models

Goal

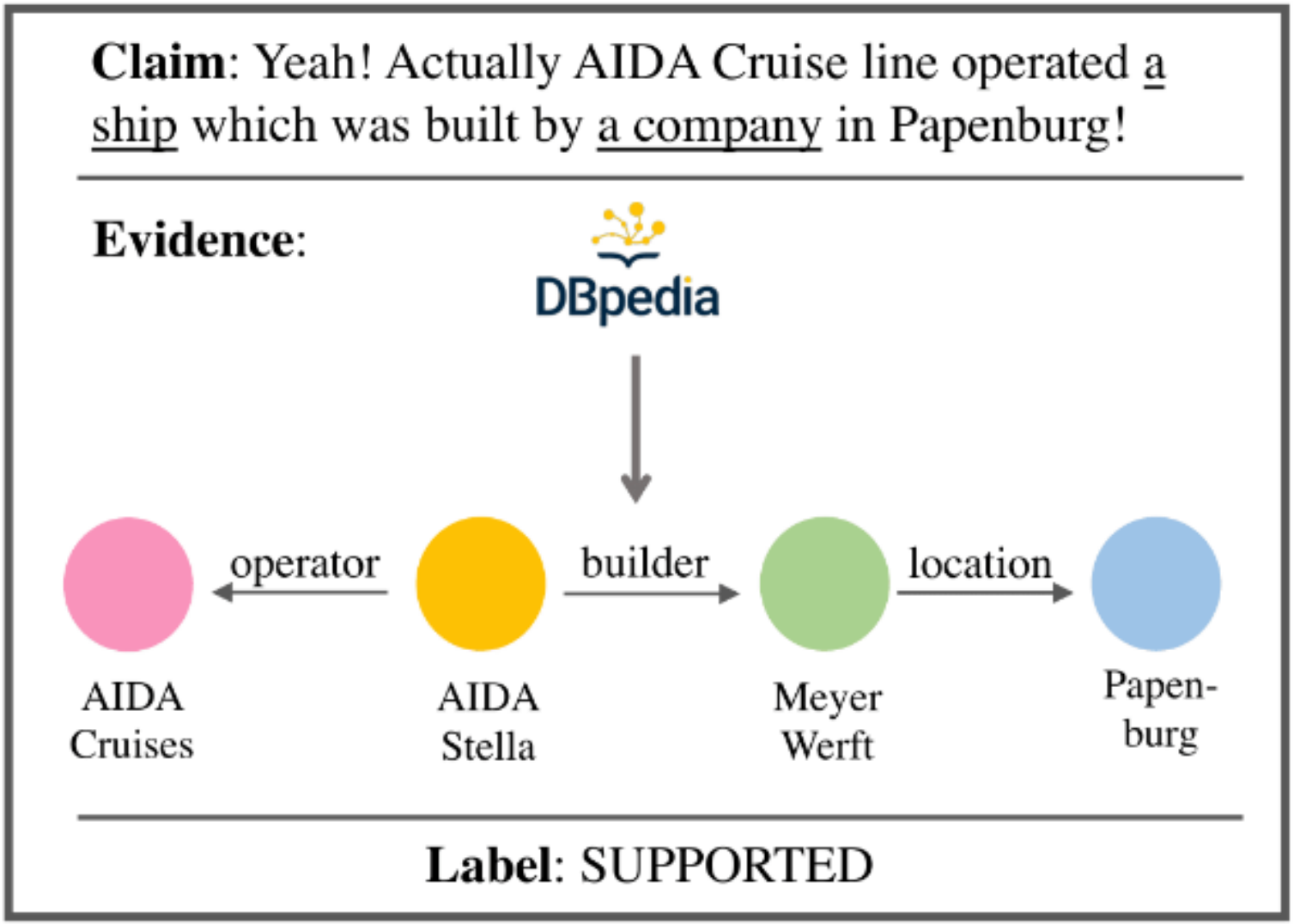
- Pretrain a language with the architecture of your choice.
- Evaluate on the two evaluation tasks provided.

Restrictions

- Your models have to have 7M parameters or less.
- The training time should not exceed 5 GPU hours (i.e. 5 hours on a single GPU).

Task Overview: Fact-checking with graph evidence

Will be using FactKG[Kim et al., 2023]



Task Overview: Artificial text detection

Goal

- Develop your own system for a task formulation of your choice.

Task Overview: Artificial text detection

Goal

- Develop your own system for a task formulation of your choice.

Data

- SemEval-2024 Task 8: Multidomain, Multimodel and Multilingual Machine-Generated Text Detection [Wang et al., 2024]
- Robustness towards language, domain, and text generator.

Task Overview: Artificial text detection

Goal

- Develop your own system for a task formulation of your choice.

Data

- SemEval-2024 Task 8: Multidomain, Multimodel and Multilingual Machine-Generated Text Detection [Wang et al., 2024]
- Robustness towards language, domain, and text generator.

Task Formulations

- Subtask A: is the text created by a human or a model?
  - Monolingual setup (English).
  - Multilingual setup (7 diverse languages).
- Subtask B: which model generated the text, or was it human who created it?
- Subtask C: when is the text continued by a model?

Task Overview: Predicting a prompt from a generated text

The main goal

To predict the prompt used to generate a given text with a given LLM

Input (LLM generated text)	Output (Prompt)
Dear [Hiring Manager's Name], I am writing to express my keen interest in the entry-level Financial Analyst position at [Company Name], as advertised. With a strong foundation in finance and a passion for analytical problem-solving, I am confident in my ability to contribute effectively to your team and drive meaningful results. [...]	Write a cutting edge cover letter for an entry-level role as a financial analyst. Ensure it makes my application stand out.

References I



Kim, J., Park, S., Kwon, Y., Jo, Y., Thorne, J., and Choi, E. (2023). FactKG: Fact verification via reasoning on knowledge graphs. In Rogers, A., Boyd-Graber, J., and Okazaki, N., editors, Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 16190–16206, Toronto, Canada. Association for Computational Linguistics.



Wang, Y., Mansurov, J., Ivanov, P., Su, J., Shelmanov, A., Tsvigun, A., Afzal, O. M., Mahmoud, T., Puccetti, G., Arnold, T., Whitehouse, C., Aji, A. F., Habash, N., Gurevych, I., and Nakov, P. (2024). Semeval-2024 task 8: Multidomain, multimodel and multilingual machine-generated text detection.