Robin Kokot

⊙ Leuven, Belgium   ✉ robin.edu.hr@gmail.com   📞 +32 47 808 70 43   ⊙ rokokot

About me

I am a graduate student in Artificial Intelligence at KU Leuven specializing in NLP, Language Engineering, and LLM interpretability. I work on design and evaluation of language and reasoning models, with a focus on natural language understanding, question answering, and non-textual semantics. I am interested in the ways humans produce and interact with noisy data and different ways of learning from these interactions in order to build universally scalable task-and-value aligned AI systems. My recent research has touched on topics in multilingual alignment (RAI) and computational modelling of non-verbal communication.

Education

KU Leuven, Faculty of Engineering Science   *Sep 2024 — ongoing*
M.Sc. in Artificial Intelligence (Speech & Language Technology)

KU Leuven, Institute of Philosophy   *Sep 2022 — Sep 2024*
M.A. in Philosophy (Cum Laude)
B.A. in Philosophy (Cum Laude)   *Sep 2019 – June 2022*

Research Projects

Type and Linguistic Complexity Signals in Multilingual Question Representations
Oct 2024 – Present (supervised by Miryam de Lhoneux)

In the scope my thesis project, I developed a framework for studying how multilingual languages models encode differences between interrogative structures across 7 languages. The project implements a set of traditional ML baselines using a variety of linear models and ensembles, and compares performance to diagnostic models trained on top of frozen glot500 embeddings in order to evaluate extraction of abstract linguistic properties. I developed a custom dataset and designed experiments to scale efficiently on HPC infrastructure using Slurm and distributed parallel execution modules.

Technologies used include PyTorch, HF Transformers, Hydra, Submitit, Scikit-learn, and Pandas.

PictoNMT: AAC Picto-to-text Machine Translation With Semantic Schema Induction
February 2025 – Present (supervised by Vincent Vandeghinste)

I am actively developing a neural MT system that converts ARASAAC pictogram sequences into fluent French text for users of AAC transcription and email services. The main contribution of the project is a custom decoding algorithm based on semantic schema induction that constructs linguistic templates from parallel sequence data. The system includes functional word prediction and structure-aware beam search to generate grammatically correct output even when source pictograms lack explicit syntactic markers. I evaluated the system using both traditional MT metrics and specialized measures for pictogram-specific challenges.

Technologies used include OpenNMT (eole), Vision Transformers, Streamlit, Fast API, and Seq2Seq Library.

Technical Skills

Scripting: Python, Bash, SQL, LaTeX

ML/NLP: I implemented systems for model interpretability and probing (PyTorch, TensorFlow), RAG pipelines and quantization techniques for efficient deployment (LlamaIndex, Rasa). My work has involved designing experiments to investigate linguistic properties in model embeddings, with a focus on extracting and visualizing semantic representations across languages

HCI and User research: I have experience with designing experiments bridging technical systems for real human needs, applying mixed-methods research, and incorporating user studies in backend design.

Languages

Native: Croatian — **Fluent:** English — **Intermediate:** French, German

Service

KU Leuven Student Council, Stura   *September 2022 – September 2023*

- Student chair of the WG Recruitment and Admissions and WG Support for International Students
- Member of the Policy Group for Internationalization
- Organized the Assembly for International Students, creating a platform for direct feedback

Una Europa Alliance   *November 2022 – September 2023*

- Represented KU Leuven students on the Una Europa Student Board
- Participated in organizing Student Council meetings at the University of Edinburgh
- Led the KU Leuven delegation at the Una Europa Student Congress at Jagiellonian University, Krakow