Fedor Sizov

+4915141388187 | fedia.sizov@gmail.com | linkedin.com/in/fedor-sizov | github.com/sfedia

Bio

I am pursuing a Bachelor's degree in Natural Language Processing at Saarland University. I work as a Machine Learning Engineer at natif.ai, where I develop large language model applications for intelligent processing of layouts and scanned documents.

Education

Saarland University   Saarbrücken, Germany
Bachelor's degree in Natural Language Processing / Computational Linguistics   Oct. 2023 – Jun. 2026

Russian University of Humanities   Moscow, Russia
Bachelor's degree in Linguistics   Sept. 2019 – Oct. 2023

Publications

1. **Fedor Sizov**, Cristina España-Bonet, Josef van Genabith, Roy Xie, and Koel Dutta Chowdhury. Analysing Translation Artifacts: A Comparative Study of LLMs, NMTs, and Human Translations, 2024. Accepted to Ninth Conference on Machine Translation (WMT24), EMNLP 2024. (pdf)
2. Daria Zhornik and **Fedor Sizov**. Linguistic variation and minor languages corpora: a case study of mansi dialects. In *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2018"*, May 2018. (pdf)

Experience

Machine Learning Engineer   August 2025 – *Currently*
natif.ai   Saarbrücken, Germany
- Working on large language model applications for intelligent processing of layouts and scanned documents

Research Assistant   June 2024 – *Currently*
German Research Center for Artificial Intelligence   Saarbrücken, Germany
- Currently focus on multilinguality in large language models, including fairness and robustness.
- Previously worked on identifying and analyzing translation artifacts in human and machine translations.
- First author of the paper "Analysing Translation Artifacts: A Comparative Study of LLMs, NMTs, and Human Translations", accepted to EMNLP.

Software Engineer Intern   June 2022 – Sep 2022
Yandex   Moscow, Russia
- As a part of monitoring and automation development group, I automated a range of hardware lifecycle scenarios such as upgrades and migrations

Conference Talks

- Analysing Translation Artifacts: A Comparative Study of LLMs, NMTs, and Human Translations. Oral Presentation at WMT24 (EMNLP 2024).
- Revisiting nltk-drt: DRT extension for NLTK updated and improved. Conference talk at TaCoS 2024 (33rd Computational Linguistics Student conference).

Technical Skills

- Programming: Python, JavaScript, C++
- ML: PyTorch, Transformers, PEFT
- Deployment: Docker, Git, Bash, Linux
- Tools: Jupyter Notebook, LaTeX, VS Code
- Languages: English (B2), German (C1), Russian (native)