

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 in Saarbrücken, where I develop large language model applications for intelligent processing of layouts and scanned documents.

EDUCATION

Saarland University Bachelor's degree in Natural Language Processing / Computational Linguistics	Saarbrücken, Germany Oct. 2023 – Jun. 2026
Russian University of Humanities Bachelor's degree in Linguistics	Moscow, Russia 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 natif.ai	August 2025 – <i>Currently</i> Saarbrücken, Germany
<ul style="list-style-type: none">• Working on large language model applications for intelligent processing of layouts and scanned documents	
Research Assistant German Research Center for Artificial Intelligence	June 2024 – <i>Currently</i> Saarbrücken, Germany
<ul style="list-style-type: none">• 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 Yandex	June 2022 – Sep 2022 Moscow, Russia
<ul style="list-style-type: none">• 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)