Iuliia Zaitova

iuliiazaitova@gmail.com | iuliiazaitova.github.io | GitHub | LinkedIn | ResearchGate

PhD candidate in Computational Linguistics (thesis submitted, awaiting defense). I post-train and evaluate transformer-based speech and language models using SFT, RL and DPO and design the evaluation methodology that validates whether those changes actually improve behavior, not just metrics.

RESEARCH INTERESTS

Post-training and evaluation of generative language/speech models (SFT, DPO, RL); multilingual representation learning and cross-lingual generalization; evaluation methodology for generative systems (LLM-as-judge, pairwise preference, ablation-based diagnosis).

TECHNICAL SKILLS

Post-training & evaluation: SFT, DPO, LoRA and full fine-tuning; RL-based post-training; multilingual adaptation; pairwise preference evaluation with confidence intervals and paired comparison; ablation and error analysis; probing of internal representations.

Frameworks & infrastructure: PyTorch, HuggingFace Transformers, scikit-learn; experiment management and reproducible pipelines; AWS (EC2, SageMaker, S3); Docker; Linux.

Programming languages: Python (advanced), C++ (intermediate), R (intermediate)

Languages: English (C1), German (B2), Russian (native)

WORK EXPERIENCE

AI Research Engineer, Diabolocom
Paris, France – remote *06.2026 – current*

- Benchmarked open multilingual TTS systems (Qwen3-TTS, MOSS-TTS, Higgs Audio) for call-center synthesis, focusing on dialogue-turn consistency, voice cloning, and inference-time pronunciation correction for named entities.
- Extending language coverage via supervised fine-tuning, followed by RL-based post-training.

Applied Scientist Intern, Amazon (Alexa TTS)
Cambridge, UK *07.2025 – 01.2026*

- Post-trained a production multilingual TTS model via SFT followed by RL, and designed the evaluation framework used to validate the resulting behavioral changes.
- Built benchmarking datasets and automated diagnostics adopted by research and engineering teams for iterative model improvement.

Research Assistant
Spoken Language Systems (LSV), Universität des Saarlandes *2020 – 2022*

- Ran large-scale probing experiments on multilingual transformer encoders, quantifying how training regime and language typology shape cross-lingual representation transfer.

Research Assistant
Information and Service Systems, Universität des Saarlandes *2019 – 2020*

- Developed time-series neural models for sales forecasting and built the surrounding training and inference pipeline.

EDUCATION

Universität des Saarlandes **SFB 1102, project C4**
PhD in Computational Linguistics *2022 – late 2026*
Focus: evaluation of neural architectures, representation learning, multilingual modeling, and transformer interpretability.
Advisors: Prof. Dr. Tania Avgustinova, Prof. Dr. Dietrich Klakow, Prof. Dr. Bernd Möbius
Thesis: *Cross-Linguistic Analysis of Microsyntactic Units in Slavic Languages: Computational and Empirical Perspectives*

Universität des Saarlandes **Department of Language Science and Technology**
MSc in Language Science and Technology *2019 – 2022*
Advisors: Dr. Badr M. Abdullah, Prof. Dr. Dietrich Klakow
Thesis: *Modeling Mutual Spoken Word Intelligibility between Slavic languages*

Udmurt State University **Department of Translation and Applied Linguistics**
BA with Honours in Linguistics *2014 – 2018*

PUBLICATIONS

"It's Not a Walk in the Park! Challenges of Idiom Translation in Speech-to-text Systems" **Iuliia Zaitova**, Badr Abdullah, Wei Xue, Bernd Möbius, Dietrich Klakow, and Tania Avgustinova (**ACL 2025**) [code]

"Attention on Multiword Expressions: A Multilingual Study of BERT-based Models with Regard to Idiomaticity and Microsyntax" **Iuliia Zaitova**, Vitalii Hirak, Badr Abdullah, Bernd Möbius, Dietrich Klakow, and Tania Avgustinova (**Findings of NAACL 2025**) [code]

"The Effect of Word Predictability on Spoken Cross-Language Intelligibility" Wei Xue, **Iuliia Zaitova**, Bernd Möbius (**Interspeech 2025**)

"Predictability of Microsyntactic Units across Slavic Languages: A translation-based Study" Maria Kunilovskaya, **Iuliia Zaitova**, Wei Xue, Irina Stenger, and Tania Avgustinova (**NoDaLiDa Baltic-HLT 2025**)

"Cross-Linguistic Processing of Non-Compositional Expressions in Slavic Languages" **Iuliia Zaitova**, Irina Stenger, Muhammad Umer Butt and Tania Avgustinova (**CogAlex 2024** at COLING)

"Cross-Linguistic Intelligibility of Non-Compositional Expressions in Spoken Context" **Iuliia Zaitova**, Irina Stenger, Wei Xue, Tania Avgustinova, Bernd Möbius, Dietrich Klakow (**Interspeech 2024**)

"Microsyntactic Unit Detection using Word Embedding Models: Experiments on Slavic Languages" **Iuliia Zaitova**, Irina Stenger, and Tania Avgustinova (**RANLP 2023**)

"Mapping Phonology to Semantics: A Computational Model of Cross-Lingual Spoken-Word Recognition" **Iuliia Zaitova**, Badr M. Abdullah, and Dietrich Klakow (**VarDial 2022** at COLING)

"How familiar does that sound? Cross-lingual representational similarity analysis of acoustic word embeddings" Badr M. Abdullah, **Iuliia Zaitova**, Tania Avgustinova, Bernd Möbius, Dietrich Klakow (**BlackboxNLP 2021**)

"Do Acoustic Word Embeddings Capture Phonological Similarity? An Empirical Study" Badr M. Abdullah, Marius Mosbach, **Iuliia Zaitova**, Bernd Möbius and Dietrich Klakow (**Interspeech 2021**)

FUNDING

Interspeech 2024 Grant

Small Grant Scheme for Doctoral Researchers of the MGK – CRC 1102

Saarland Stipendium 2021

EXTRACURRICULAR

Summer School on Linguistic Variability **UMLV 2023**
Berlin, Germany

Summer School on Deep Learning in NLP **RANLP 2023**
Varna, Bulgaria