Iuliia Zaitova

✉ izaitova@lsv.uni-saarland.de　🔗 iuliiazaitova.github.io　 IuliiaZaitova　in Iuliia Zaitova　R⁶ Iuliia Zaitova

AI Researcher with a PhD in Computational Linguistics (completing early 2026) specializing in foundation models, multilingual NLP, and systematic model evaluation. 3+ years of hands-on experience developing and analyzing large-scale transformer models, with proven ability to translate research innovations into production systems. Strong engineering profile with expertise in PyTorch, distributed training, and deploying AI systems on cloud infrastructure.

EDUCATION

Universität des Saarlandes
SFB 1102, project C4
PhD in Computational Linguistics　　　　　　　　　　　　　　　　　　　　　*2022 - early 2026*

Research on foundation models with focus on multilingual capabilities, model evaluation frameworks, and interpretability. Developed novel approaches for analyzing transformer architectures across multiple modalities and languages.

Advisors: Prof. Dr. Tania Avgustinova, Prof. Dr. Dietrich Klakow, Prof. Dr. Bernd Möbius

Thesis: Non-Compositionality in Cross-Linguistic Intercomprehension

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

Relevant Coursework: Artificial Intelligence, Machine Translation, Deep Learning, Statistical NLP

Udmurt State University
Department of Translation and Applied Linguistics
BA with Honours in Linguistics (specialization: Translation)　　　　　　　　　*2014 - 2018*

WORK EXPERIENCE

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

- Developed and deployed large-scale speech and language models for production voice assistant systems, working with transformer-based architectures (10B+ parameters) and neural TTS pipelines
- Designed comprehensive evaluation frameworks for assessing model performance, fairness, and generalization across multilingual settings
- Built data preprocessing pipelines on cloud infrastructure (AWS) for training foundation models at scale
- Collaborated with cross-functional research and engineering teams to analyze and improve model behavior in production

Research Assistant
Spoken Language Systems (LSV) at Universität des Saarlandes　　　　　　　　*2020 - 2022*

- Conducted large-scale experimental studies on transformer models, developing novel methods for analyzing internal representations, attention mechanisms, and cross-lingual capabilities
- Built reproducible experimental pipelines for training, evaluating, and comparing neural models across architectures and languages
- Published findings at top-tier venues (ACL, NAACL, Interspeech), advancing understanding of multilingual model capabilities

Research Assistant
Information and Service Systems　　　　　　　　　　　　　　　　　　　　　*2019 - 2020*

- Developed and deployed neural models for sales prediction using time-series data and deep learning architectures

TECHNICAL SKILLS

Machine Learning / AI: Foundation models and LLMs (transformer architectures, BERT, GPT), self-supervised learning, multilingual and multimodal modeling, representation learning, model evaluation and interpretability, reinforcement learning (familiar), explainable AI

Deep Learning Frameworks: PyTorch (advanced - 4+ years), HuggingFace Transformers, JAX (familiar), experience implementing and modifying transformer architectures

Distributed Training & Infrastructure: Large-scale distributed training, Docker, Kubernetes, cloud platforms (AWS: SageMaker, EC2, S3), preprocessing pipelines on cloud infrastructure, experience training models with 10B+ parameters

Development & Tools: Python (advanced), C++ (intermediate), R (intermediate), Git, Linux/Unix, Jupyter, scikit-learn, NumPy, pandas, statistical analysis and visualization

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

Research & Communication: Strong track record of publishing at top-tier ML venues (ACL, NAACL, Interspeech), scientific writing and documentation, collaborative research with academic and industry partners

KEY PUBLICATIONS

Selected high-impact publications from 11+ peer-reviewed papers at ACL, NAACL, Interspeech, and related venues. Focus on foundation models, multilingual NLP, and systematic model evaluation.

"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**)

"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 (**NAACL 2025**)

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

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

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

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

Full publication list (11+ papers) available at iuliiazaitova.github.io

AWARDS & FUNDING

Interspeech 2024 Grant
Small Grant Scheme for Doctoral Researchers of the MGK – CRC 1102
Saarland Stipendium 2021

PROFESSIONAL DEVELOPMENT

Summer School on Linguistic Variability   UMLV 2023
Berlin, Germany
Summer School on Deep Learning in NLP   RANLP 2023
Varna, Bulgaria
Summer School: Virtual SemDial 24   WatchDial 2020
online