

Nursulu Sagimbayeva

Applying for an Internship in Natural Language Processing/ML

Open to relocation

✉ nurs.sagimbayeva@gmail.com

🐙 [GitHub/ Portfolio](#)

🌐 [LinkedIn Profile](#)

📝 [Medium](#)

RELATED EXPERIENCE

•Research Assistant

Interdisciplinary Institute of Societal Computing, Saarland Informatik Campus

June 2023-present

Saarbrücken, Germany

- Working as a student NLP researcher in the interdisciplinary team led by **Prof. Dr. Ingmar Weber**.
- Topic: detecting shifts in words' meanings in diachronic corpora using word embeddings
- Techniques used: Fine-Tuning LLMs, Pre-training LLMs, Clustering, Scraping, Word2Vec

•Technical Content Writer

Hasty.ai (CloudFactory)

July 2022-July 2023

Berlin, Germany (remote)

- Creating and maintaining documentation for an AI-powered image annotation platform
- Writing articles on Computer Vision/Machine Learning

PROJECTS

- **All you need is love?** Exploring most popular Spotify song topics around the world using LDA topic modeling
- **Mental health in literature:** Diagnosing literary characters with mental disorders according to ICD using cosine similarity between descriptions
- **YouNameIt!** A chatbot assisting patients with aphasia train their memory (*model: prefix-tuned FLAN-T5 trained on a reverse-dictionary task*)
- **GoTogether:** A restaurant recommendation app that considers preferences of 2+ users (*custom recommendation system*)

EDUCATION

•Universität des Saarlandes

M.Sc. in Language Science and Technology (Current GPA: 1.5/5.0, best: 1.0)

October 2022-present

Saarbrücken, Germany

- Holder of **Deutschlandstipendium** for high-achieving students (acceptance rate in 2023: 16%)
- **Relevant Coursework:** Computational Linguistics • Statistics with R • Statistical Natural Language Processing • Empirical Methods • Math (Linear Algebra, Calculus) • Python • Neural Networks • Machine Learning • Software Projects
- **Seminars:** What Do Pretrained Language Models Really Understand? • ML for NLP • Brain-Computer Interfaces • Bridging Language in Machines with Language in the Brain

•University of Warsaw

Exchange student

Winter Semester 2021 - 2022, Winter Semester 2019-2020

Warsaw, Poland

- 2 times scholarship holder from the Ministry of Education and Culture (Republic of Kazakhstan)
- Active member of a debate club (2019-2020); Circle of the Philosophy of Mind (2021-2022)

•Al-Farabi Kazakh National University

B.A. in Translation Studies, GPA: 3.75/4 (best: 4.0)

2018-2022

Almaty, Kazakhstan

- **Thesis topic:** exploring how peculiarities of user-generated content in the Kazakh language affect the subjective quality of machine translation in Kazakh-Russian language pair (grade 98/100)
- Active member of a debate community; winner and top-10 speaker of numerous international competitions

PUBLICATIONS

- **Winner of the best paper in the NetMob challenge 2023.** Unveiling Local Patterns of Child Pornography Consumption in France using Tor (Till Koebe, Zinnya del Villar, Brahmani Nutakki, Nursulu Sagimbayeva, Ingmar Weber).
- Karolina Kuligowska, Agnieszka Repka, Nursulu Sagimbayeva, **Text Mining in overcoming disinformation: human cognitive security**, [in:] Paulina Szymczyk, Joanna Kozłowska (eds.), Education and society - an interdisciplinary analysis of selected issues, Scientific Publisher TYGIEL, Poland - Lublin 2022 (ISBN: 978-83-67104-44-9)

ACTIVITIES

•I2SC Kick Off Event: Opening of the Interdisciplinary Institute for Societal Computing

Presenting a poster on our research (semantic shift detection)

September 2023

University of Saarland

•TaCoS 2023 (international student conference on Computational Linguistics)

Giving a talk about my project on topic modeling in songs

July 2023

University of Düsseldorf

•STAL-2023: Slurring Terms Across Languages

Presenting my essay on a Kazakh slur and venturing new hypotheses on slur reclamations

June 2023

Ecole Normale Supérieure, Paris

SKILLS AND INTERESTS

Languages: English (proficient) • Russian (native) • German (intermediate) • Polish (intermediate) • Kazakh (conversational)

Libraries (Python): NumPy, Pandas, Scikit-Learn, Matplotlib, PyTorch, Transformers, Streamlit, nltk, Gensim, BeautifulSoup; familiar with: HTML, CSS

Tools: Git, LaTeX, VSCode, HuggingFace