Nursulu Sagimbayeva

Munich/ Saarbrücken, Germany
Open to relocation

✉ nurs.sagimbayeva@gmail.com
⚙ nursaltyn.github.io
in LinkedIn Profile
☎ +4917674754101
Ⓜ Medium

Profile: I am a Master's student in my last year, with background in the domains of NLP, Data Science, Societal Computing, and Mechanistic Interpretability. I am looking for an internship to gain insight in the industry while working on challenging NLP/AI topics.

Libraries: NumPy, Pandas, Scikit-Learn, Matplotlib, PyTorch, Transformers, Streamlit, nltk, Spacy, Gensim, plotly, BeautifulSoup, GeoPandas, HTML, CSS; **Programming languages:** Python, R

Tools: Git, LaTeX, VSCode, HuggingFace, Docker, PowerBI

Experience: Neural Networks • Machine Learning • Computational Linguistics • Statistics • Statistical Natural Language Processing • Machine Translation • Text Mining and Social Media Mining • Trustworthy Machine Learning

Languages: English (C1) • Russian (native speaker) • German (B2) • Polish (B2) • Kazakh (B1)

WORK EXPERIENCE

•**Internship Financial Assets & Solutions Data Analytics**   *November 2024-May 2025*
Munich RE   Munich, Germany

- Developed a system for **automatic news monitoring and financial advisory** for alternative asset classes using LLMs and PowerBI visualizations
- **Automated existing tools** for a structured analysis of alternative investments
- Data intake from in-house databases and integration of third party data sources including retrieval of market data from **unstructured data sources** via LLMs
- Participated in projects on investment topics
- **Skills**: LLMs, RAG, PowerBI, Data Analysis Automation, Machine Learning, Capital Markets

•**Research Assistant**   *June 2023-present*
Interdisciplinary Institute of Societal Computing, Saarland Informatik Campus   Saarbrücken, Germany

- Initiated a project on detecting inconsistencies and contradictions in politicians' speeches in real time using LLMs;
  supervised by **Prof. Dr. Ingmar Weber.** Data sources: websites, social media, YouTube
- Collected an **original dataset** on political contradictions (work under review for ACL-2025
- Proposed a pipeline to **store** party's data and **compare** new texts with the database to detect contradictions
- **Skills:** Fine-Tuning, Pre-training LLMs, Clustering, Scraping, Word2Vec, GPT4 API, Data annotation, Crowdsourcing, Academic writing, Presentation

•**Technical Content Writer**   *July 2022-July 2023*
Hasty.ai (CloudFactory)   Berlin, Germany (remote)

- Created and maintained documentation for an AI-powered image annotation platform
- Wrote articles on **Computer Vision/Machine Learning/Neural Networks**
- Pushed ML-Wiki articles on ML/DS metrics to **top-3** results in Google search

EDUCATION

•**M.Sc. in Natural Language Processing, current GPA: 1.5 (best: 1.0)**   *October 2022-currently*
Saarland University   Saarbrücken, Germany

•**B.A. in Translation Studies, GPA: 3.75/4.0, German: 1.38**   *2018-2022*
Al-Farabi Kazakh National University   Almaty, Kazakhstan

- **Thesis topic:** exploring the effect of features of user-generated content in the Kazakh language on the perceived quality of machine translation in Kazakh-Russian language pair. Grade: 98/100

•**Exchange student**   *Oct 21 - Feb 22, Oct 19 - Feb 20*
University of Warsaw   Warsaw, Poland

PUBLICATIONS

- Koebe, T., del Villar, Z., Nutakki, B., **Sagimbayeva, N.**, Weber, I. *Unveiling local patterns of child pornography consumption in France using Tor.* Published in **Nature**: Humanities and social sciences communications 11, 807 **(2024).** https://doi.org/10.1057/s41599-024-03343-4

- 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)

PROJECTS

- **YouNameIt!** A chatbot assisting patients with aphasia in training their memory. Increased accuracy from **39%** to **92%** on a custom dataset
Skills: Fine-tuning • Prompt-tuning LLMs • Data collection • UX (Streamlit) • Chatbots • User Interaction planning

- **Brain-LLM alignment:** Estimating variance in GPT model explained by different brain regions
Skills: Ridge regression • Bootstrapping • Noise ceiling estimation • Work with fMRI data (reading and listening) • LLMs

- **All you need is love?** Exploring the most popular Spotify song topics around the world
Skills: LDA topic modeling • Data processing • Scraping • Working with APIs • Data visualization

- **Mental health in literature:** Diagnosing literary characters with mental disorders according to ICD-10
Skills: Prompt optimization • Data collection • Data visualization • Streamlit

- **GoTogether**: A restaurant recommendation app that considers preferences of 2+ users
Skills: Recommendation systems • User interviewing • Data Collection • Streamlit

- **Brain-Controlled Kitchen:** a Brain-Computer Interface to navigate through the recipe using Imagined Speech
Skills: EEG reading • EEG data processing • UI • User Interaction planning

ACTIVITIES

University Open day — *June 2024*
Presenting a poster on our research (automatic inconsistency detection) — University of Saarland

Opening of the Interdisciplinary Institute for Societal Computing — *September 2023*
Presenting a poster on our research (semantic shift detection) — University of Saarland

TaCoS 2023 (international student conference on Computational Linguistics) — *July 2023*
Giving a talk about my project on topic modeling in songs — University of Düsseldorf

STAL-2023: Slurring Terms Across Languages — *June 2023*
Presenting my essay and venturing new hypotheses on slur reclamations — Ecole Normale Supérieure, Paris

AWARDS

- **2023-2024, 2024-2025: Deutschlandstipendium** for high-achieving students (acceptance rate in 2023-2025: 16%)

- **2023**: Best paper award at the NetMob data Challenge 2023.

- **2021-22, 2019-20**: Academic exchange scholarship from the Ministry of Education and Culture (Republic of Kazakhstan)

- **2018-2022**: Winner, finalist, and semi-finalist of 20+ international and local debate tournaments (English and Russian leagues); Chief Adjudicator/judge at local and national tournaments

- **2018**: Winner of SDU nation-wide Olympiad for high school students in the section of English and World History