

Ellie Kuznetsova

San Diego, CA | ye.s.kuznetsova@gmail.com | [LinkedIn](#) | [Website](#) | [Github](#)

SUMMARY

Early-career AI Language Engineer with experience building multilingual NLP, LLM evaluation, and speech AI pipelines for code-switched and low-resource language data. Skilled in Python, Transformers, dataset creation, and experimental evaluation.

RESEARCH PROJECTS

Graduate Student Researcher Sep 2025 — Present
Learning, Meaning, and Natural Language Lab (PI: Alex Warstadt), UC San Diego La Jolla, CA

- Build end-to-end pipeline evaluating XLM-RoBERTa and Llama 3 on Spanish-English minimal pairs.
- Direct dataset creation for Spanish-English code-switched grammaticality evaluation.
- Compute surprisal scores with MINICONS to measure grammaticality and acceptability preferences across models.
- Create Python scripts for dataset preprocessing, statistical testing, and visualization of model behavior.

Graduate Student Researcher Sep 2025 — Present
Bilingualism in Context Lab (PI: Anne Beatty-Martinez), UC San Diego La Jolla, CA

- Build end-to-end R pipelines for cleaning, restructuring, and analyzing repeated-reference bilingual interaction datasets.
- Compare word embeddings using fastText multilingual vectors, cosine similarity, and t-SNE visualizations.
- Implement mixed-effects regression models to quantify convention formation over repeated interactions.

Graduate Student Researcher Sep 2025 — Present
Phonetics Lab (PI: Will Styler), UC San Diego La Jolla, CA

- Fine-tune and evaluate a transformer-based ZIPA model for keyword search in unseen low-resource languages.
- Develop Python pipelines for model training, inference, and evaluation.
- Analyze the performance of the multilingual speech model in low-resource language settings for keyword search.

Independent Research Apr 2022 — Oct 2025
Northern Arizona University Flagstaff, AZ

- Led the end-to-end research process, including study design, project planning, timeline coordination, task delegation, and team leadership.
- Built a Writing Feedback Corpus (WFC) for open-access public release.
- Managed annotation workflows and co-authored two manuscripts under review.

SKILLS

Programming: Python, R, SQL, $\LaTeX$

NLP: Hugging Face Transformers, LLM evaluation, prompting, multilingual NLP, embeddings, fastText, keyword spotting

ML: fine-tuning, inference, classification, regression, dimensionality reduction (t-SNE), cross-validation

Libraries: pandas, NumPy, scikit-learn, PyTorch, matplotlib, seaborn, tidyverse, ggplot2, lme4, text2vec, Rtsne

Tools: Git, GitHub, Jupyter, RStudio

Languages: Russian, English, French, American Sign Language

EDUCATION

M.S. Computational Social Science, University of California San Diego Expected June 2026

M.A. Applied Linguistics and Teaching English as a Second Language, Northern Arizona University May 2023

B.A. Linguistics, Pushkin Leningrad State University Jul 2019

**All sections contain selected information; a complete list is available upon request.*