

Yekaterina (Katya) Yegorova

Computer Science PhD Student

yay2@illinois.edu

<https://www.linkedin.com/in/katya-yegorova>

<https://huggingface.co/katyayego>

EDUCATION

PhD in Computer Science

Aug. 2023 - Present

University of Illinois at Urbana-Champaign, Urbana-Champaign, Illinois

Advisors: Mark Hasegawa-Johnson, Ph.D.; Julia Hockenmaier, Ph.D.

BSc in Computer Science, BA in Linguistics, Honors College

Aug. 2019 - May 2023

Purdue University, West Lafayette, Indiana

Overall GPA: 3.83/4.0

RESEARCH EXPERIENCE

Aug. 2023-Present. UIUC. Graduate Research Assistant.

Supervisors: Mark Hasegawa-Johnson, Ph.D.; Julia Hockenmaier, Ph.D.

- Conducting speech processing and NLP research for code-switched speech and text
- Developing interpretability techniques for LLMs and ASR systems to analyze deep learning models

May 2021-May 2023. [Indigenous and Endangered Languages Lab](#).

Principal Investigators: E. Benedicto, Ph.D.; R. Zúñiga-Argüello, Ph.D.; E. Opoku, Ph.D. Candidate

- Led a team of 4 in designing a virtual data collection method, enabling scalable linguistic fieldwork.
- Introduced automatic data formatting script, saving 100+ hours of researchers' time. Python
- Optimized data processing with regular expressions, accelerating data analysis. ELAN
- Documented all processes, and built a framework allowing researchers to follow developed processes
- Annotated and analyzed 400+ utterances in ELAN, contributing linguistic insights to two Ph.D. dissertations on indigenous language syntax

May 2021-May 2022. [Purdue University](#). Undergraduate Research Assistant

Principal Investigator: Eric Waltenburg, Ph.D.

- Analyzed the correlation between climate change legislation success and media attention
- Coordinated work of 3 research assistants, defined common data storage and exchange formats, and developed a scraper for legislation data. GraphQL, Pandas, Python
- Selected the machine learning model to use based on previously coded frames and modern techniques

PROJECTS

That's Deprecated! Probing and Steering Code Generation Under Knowledge Conflicts

Aug. 2024-Dec. 2024. CS 546 - Advanced Topics in NLP (Dr. Heng Ji)

- Expanded knowledge conflict research in LLMs from Q/A tasks to code-generation tasks
- Proposed a novel framework for testing the behavior of LLMs on code-generation tasks
- Demonstrated that the concept of knowledge conflicts is not prevalent in LLMs via linear probing and steering

UniLID: Towards Universal Code-Switched Language Identification

Aug. 2024-Dec. 2024. ECE 537 - Speech Processing Fundamentals (Dr. Mark Hasegawa-Johnson)

- Explored methods of creating a universal code-switched language identification for automatic speech recognition
- Utilized a speech processing model integrating CNNs, Transformers, and phonotactic embeddings for joint language identification and phoneme segmentation

Trigger Word Detection and Phone Recognition as Features for Code-Switched Speech Recognition
Aug. 2023 - Aug. 2024. CS 597 - Independent Study (Dr. Mark Hasegawa-Johnson)

- Developed the frame-level language identification system for Mandarin-English code-switched speech by leveraging a finetuned Wav2Vec2Phoneme, Mandarin-English code-switching trigger words, and an LLM for Mandarin-English code-switched automatic speech recognition (ASR)
- Improved language identification accuracy by 16% (from 82% to 95.4%)
- Contributed a new phonetically transcribed Mandarin-English code-switched speech [dataset](#)

Jan. 2024 - May 2024. RetroProbe. CS 597 - Independent Study (Dr. Julia Hockenmaier)

- Created a probabilistic way of probing for dependency trees from mBERT embeddings

PRESENTATIONS

- “The Role of Locatives in Determining Telicity in Akan,” Purdue Linguistics Symposium. West Lafayette, IN. May 2023
- “What a Little Verb Contributes to Telicity: The Case of Verbal Series in Akan,” Purdue Undergraduate Research Conference. West Lafayette, IN. April 2023.
- “The Structure of Motion Predicates in Akan,” University of Illinois at Urbana-Champaign. ECE590SIC Urbana-Champaign, IL. March 2023

TEACHING EXPERIENCE

Jan 2022-May 2023. Purdue University. Teaching Assistant for Foundations of Computer Science

- Led weekly interactive recitations for 60+ students
- Held weekly office hours for a class with 400 students
- Developed standardized grading rubrics and graded weekly homework for 400 students

INDUSTRY EXPERIENCE

May 2022-Aug. 2022. DocuSign. Software Engineering Intern

- Redesigned and implemented API for data export functionality, used daily by thousands of users
- Improved user experience and reduced server utilization by over 10% through asynchronous processing and optimized querying. Added throttling to the export endpoint to limit exports per hour
- Implemented access to exported data via a secure link. C#, .Net

June 2020-Aug. 2021. [EduSource](#). Apprentice Software Engineer

- Full-time intern for two summers and a part-time remote apprentice during the school year
- Created and unit-tested CQRS back-end for an internal tracking and billing application. EventStore, GraphQL, .Net, C#, Stripe, Plaid, CosmosDB, Azure
- Worked on multiple web-based applications (e.g. survey creation, billing, and scheduling) for external customers. C#, .Net, TypeScript, React, Material UI, PostgreSQL, and GraphQL
- Received award for most project points completed by a first-year apprentice

HONORS AND AWARDS

College of Liberal Arts Outstanding Senior, Purdue University, West Lafayette, IN
Dean's List and Semester Honors

2023
2019-2023

Yekaterina (Katya) Yegorova

Computer Science PhD Student

yay2@illinois.edu

<https://www.linkedin.com/in/katya-yegorova>

<https://huggingface.co/katyayego>

Boeing Scholarship, Purdue University, West Lafayette, IN

2021

SERVICE

Reviewer

2025 IEEE Transactions on Audio, Speech and Language Processing (TASLP)

COURSEWORK

Graduate Coursework: Natural Language Processing; Signal Processing; Machine Learning; Advanced Topics in NLP; Speech Processing Fundamentals; Deep Learning for Computer Vision; Language, Interfaces, and Communication

Undergraduate Coursework: Problem-Solving and Object-Oriented Programming, Foundations of Computer Science (Discrete Math), Data Structures and Algorithms, Systems Programming, Data Mining and Machine Learning, Introduction to Artificial Intelligence, Introduction to Analysis of Algorithms, Natural Language Processing, Operating Systems, Computer Security, Compilers, Computer Networks

SKILLS

Programming: Python, Java, C, C++, C#, R, TypeScript, JavaScript, HTML

ML/NLP: PyTorch, HuggingFace, NumPy, Pandas

Tools: React.js, GraphQL, .NET, SQL, Linux, Windows, Git, Bash/Shell,

Languages: English (native), Russian (native), Spanish (intermediate)