

pronouns: they/them

speakingofdata@gmail.com

speakingofdata.com

## Research Interests

- Speech Processing
- Audio Data and Sound Synthesis
- Ethical Data Science
- Accessible Technology

## Degrees

**Master of Advanced Studies**, University of California San Diego, June 2025

- Data Science and Engineering
- Topics include: Computational Speech Processing, Bioacoustic Classification Models, Machine Learning, Neural Networks, Probability and Statistics, Python for Data Analysis, Data Management Systems, Scalable Data Analysis, Data Visualization, Data Integration and ETL
- Engineering Honor Societies: IEEE Eta Kappa Nu (HKN), Tau Beta Pi

**Master of Business Administration**, University of Phoenix

- Topics include: Statistics and Research Methods, Information Systems, Project Management

**Bachelor of Arts** with highest distinction, University of Virginia

- Majors: Cognitive Science and Music, Minor: Computer Science
- Topics include: Linguistics, Symbolic Logic, Computer Sound Generation, Human Computer Interface, Invention and Design, Cognitive Psychology, Discrete Math, Digital Logic Design, Database Systems

## Additional Education

**Neurodiversity Affirming AAC**, Northwest Augmentative Communication Society, September 2025

- Conference featuring AAC research, first-hand perspectives, and prioritizing user-centered objectives

**North American Summer School for Logic, Language, and Information**, June 2025

- Topics include: Semantics, Generalized Quantifiers, Computational Social Science on Linguistic Data

## Work Experience

**Industrial Engineering Supervisor**, UPS, 6 years

- Planned logistics and resource needs for transition to first automated facility in Colorado
- Recommended and implemented process improvements to achieve production goals
- Maintained district service area map data and on-road work measurements for 4 states
- Supervised and mentored Industrial Engineering interns and specialists

**Seasonal Scorer**, Pearson, 3 years

- Graded science responses at multiple grade levels for seasonal standardized tests

**Community Development Consultant**, U.S. Peace Corps, Ukraine, 2 years

- Opened a Community Education Center, hired and trained director and staff, developed a sustainable business plan, and created programs serving 270 local participants of all ages
- Conducted professional training in multilingual and cross-cultural environments

**Teaching Assistant**, University of Virginia, 2 years

- Introductory Computer Science C++ programming course, 50 students per semester
- Facilitated laboratory participation and assessed performance on tests, quizzes, and assignments

## Datasets

**LJ2 Corpus:** 1 voice, 26,200 transcribed sound files, 48 hours of speech audio

Tripled the size of the popular LJ Speech Dataset, with over 8 times more non-fiction source texts, employing downsampling to increase alignment with modern word usage frequency. It can be used alone in 12-, 24-, or 48-hour segments, and with any model architecture designed to work with LJ Speech.

Documentation and Repository: [https://github.com/speakingofdata/LJ2\\_Corpus](https://github.com/speakingofdata/LJ2_Corpus)

**80 Excerpts:** 4 voices, 80 transcriptions, 320 sound files

60 text selections from LJ Speech Dataset or LJ2 Corpus (non-fiction), and 20 from other sources (fiction). Recordings of four adult Americans reading each excerpt, 80 wav files for each voice.

Documentation and Repository: [https://github.com/speakingofdata/80\\_Excerpts](https://github.com/speakingofdata/80_Excerpts)

## Research

### **AUTALIC: A Dataset for Anti-AUTistic Ableist Language In Context.**

First annotated dataset for detecting anti-autistic ableist language, with machine learning analysis. Naba Rizvi, **Harper Strickland**, ..., Imani Munyaka, Nedjma Ousidhoum. Proceedings of the Association for Computational Linguistics (ACL), July 2025. <https://aclanthology.org/2025.acl-long.1022>

### **'I Hadn't Thought About That': Creators of Human-like AI Weigh in on Ethics & Neurodivergence.**

Analysis of interviews with 16 developers about accessibility, neurodiversity, and ethical implications. Naba Rizvi, Taggert Smith, Tanvi Vidyala, Mya Bolds, **Harper Strickland**, ..., Imani Munyaka. Association for Computing Machinery (ACM) Conference on Fairness, Accountability, and Transparency (FAccT), June 2025. <https://doi.org/10.1145/3715275.3732218>

### **Avian Language Models: Rethinking Bioacoustic Pipelines with Self-Supervised Learning.**

Ablation studies of bioacoustic model performance: encoder, Q-Former, large language model (LLM). Hari Prakash, **Harper Strickland**, Xinyu Bao, Jiawei Li, Alex Lehman, Sean Perry, Ryan Kastner, Data Science & Engineering Master of Advanced Study (DSE MAS) Capstone Projects, UC San Diego Library Digital Collections, June 2025. <https://doi.org/10.6075/J0NV9JM2>

## Skills

- Spoken languages: English (native), Ukrainian (proficient), Russian (basic)
- Speech processing: pitch and formant extraction, automatic speech recognition (ASR), text-to-speech (TTS), corpus development, voice transmission codecs, Praat, Kaldi, ESPnet, SpeechBrain, Unsloth
- Data management tools, exploratory data analysis: Python, Pandas, SciPy
- Machine learning, LLMs, deep learning, neural networks: Scikit-learn, TensorFlow, PyTorch
- Relational database systems: SQL, Postgres, Oracle
- Graph database systems: XML, XQuery, JSON, PartiQL, Neo4j, Cypher, SparQL
- Large data sets, cloud-based and parallel computing: Hadoop, Spark, AWS, Data Lake
- Mathematical basis for data models: probability, statistics, principle component analysis (PCA)
- Data visualization: Tableau, Matplotlib, Seaborn, Plotly, Dash
- Additional programming language experience: C++, R, Perl, RTcmix
- Natural language processing (NLP): NLTK, SpaCy
- Time series analysis, regression models, supervised and unsupervised learning, ETL, knowledge graph