

Yifan (Holly) Cui

(805)-280-8850 | yifan.cui@duke.edu | Durham, NC

EDUCATION

Duke University

Durham, NC

M.S., Statistics

Aug 2023 – May 2025

B.S., Statistical Science (Data Science Concentration); Minor in Computer Science

Aug 2021 – May 2023

GPA: 3.958 / 4.0, Cumulative GPA: 3.98 / 4.0 (Graduated Magna Cum Laude)

- **Honors:** Dean's List with Distinction (Fall 2021 & Spring 2023), Chief Student Marshal (2021–2022), American Statistical Association Member, Harvard Datathon@LISH Top10, Teaching Assistant of the Year (2022–2023) for Advanced Statistics
- **Coursework:** Machine Learning & Data Mining, Natural Language Processing, Predictive Modeling, Bayesian Analysis, Study Design, Causal Inference, Advanced Data Visualization, Advanced Statistical Computing, Cloud Computing, Data Structures and Algorithms, Computer Systems, Software Quality Management, Statistical Consulting

University of California, Santa Barbara

Santa Barbara, CA

Major in Statistics and Data Science, GPA: 4.0 / 4.0

Sep 2019 – June 2021

- **Honor:** Honors Program with Distinction (2019–2021), International Student Advisory Board Member (2021), Best Undergraduate Learning Assistant of Statistics Department (2020–2021)

EXPERIENCE

Duke University

Durham, NC

Student Research Affiliate - Machine Learning in Healthcare: Lab Test Harmonization

May 2022 – Dec 2022

- Selected as the sole undergraduate amidst a competitive pool of professional candidates for Duke AI Health 2022 cohort, and earned the prestigious opportunity to present research findings at Duke AI Health Poster Showcase 2022
- Optimized lab test deduplication of grouper labels by adopting and fine-tuning Bio-BERT NLP structure pre-trained on biomedical corpora; created a new method of cross-comparison similarity evaluation based on ground-truth text embeddings, and uncovered 95% performance boost in the application to Duke lab analyte database

Hiya Inc.

Seattle, WA

Data Science Intern - Hiya Shield Project: Robocall Identification & Screening

May 2022 – Aug 2022

- Spearheaded a robocall screening process using NLP text embeddings to determine if an audio sample (or its transcript) is from a known robocall database
- Quantified the relationship between audio duration and performance of robocall classification; identified the preferred audio truncation length and optimal similarity threshold, and achieved a 67% acceleration in user experience with the introduction of a customizable screening accuracy feature for Hiya mobile App

Tsinghua University

Beijing, China

Lead Author & Research Assistant - Cross-Media Retrieval Based on Big Data Technology

Jun 2020 – Mar 2021

- Improved performance of permutation invariant training with mean squared error loss through BLSTM/LSTM and CNN in a key media separation technique; proved the improvement in two separation methods – the FIX strategy and the masking-based data augmentation strategy – and subsequently developed independent research project
- Paper Publication: *Audio-Visual Single-Channel Signal Separation based on Big Data Augmentation* published by IEEE during International Conference on Computer Networks and Electronic Communications (ICCNec 2020)

Additional Experiences:

- **UCSB - Data Science Research Assistant** (Mar-Sep 2021): automation function, NLP training, software prototype.
- **University of Chicago - Preprint Reviewer** (Aug 2021): DNA sequence pulling, automatic influenza-recognition

LEADERSHIP ACTIVITIES

Duke Impact Investment Group *Data Division - Project Manager & Data Analyst*

Aug 2021 – Dec 2023

- Managed comprehensive data analysis for start-up clients: Balanced and Joro; coordinated teams of five with project outlining, weekly meetings, and communication between student analysts and clients
- Scraped social media/website survey data and user behavior statistics; harnessed A/B testing and creative data visualizations; delivered professional business suggestions for company launch and algorithm optimization

Duke Statistical Science Majors Union *Chief Technology Officer (CTO)*

May 2022 – May 2023

- Orchestrated 15+ events such as general board meetings, pro-seminars, and faculty mixers for 280+ members
- Facilitated annual Duke Datathon, peer advising hours, and faculty-student communication; oversaw website maintenance, member enrollment, and graphic design

SKILLS & INTERESTS

Technical: R, Python, SQL, SAS, Java, C; Git, LaTeX, AWS SageMaker, Microsoft Office Suite

Languages: English, Mandarin; **Interests:** Boxing, Cycling, Piano (Level 9), Music Producing, Baking