

Zhao Xinyi

Columbia University, 116th and Broadway New York, NY 10027
(929) 695 7600 | xz3274@columbia.edu | linkedin.com/in/xinyizhao3274/

EDUCATION

Columbia University New York, USA
Data Science, Master of Science 2023.9 - 2024.12

Core courses: Natural Language Processing, Algorithms of Data Science, Data Analysis and Data Visualization.

Fudan University Shanghai, China
Data Science and Big Data Technology, Bachelor of Science 2019.9 – 2023.6

Core courses: Machine Learning, Artificial Intelligence, Neural Network and Deep Learning, Data Visualization, Graph Data Mining, Statistical Computing, Database Principle, Numerical Analysis, Stochastic Process.

SKILLS

Programming Languages: Proficient: Python, JavaScript, R, SQL. Intermediate: MATLAB, C.

Dev Tools: Git, Linux, Node.js, Google Colab, Tableau, Docker, Vue.js, jupyter notebooks, pandas.

Interests: Digital graphics drawing, Designing.

Language: English, Mandarin Chinese

INTERNSHIP & RESEARCH EXPERIENCES

Ping An Technology Shanghai, China
Machine Learning Engineer Intern 2023.4 – 2023.7

- Enhanced the performance of VITS, a Text-to-Speech model, by fine-tuning on a 10-hour Mandarin-English multi-speaker dataset, achieving a 0.2 increase in Mean Opinion Score through manipulation of speaker embeddings.
- Automated the process of data quality filtering and cleaning by developing Python scripts, enabling efficient examination of inconsistent pronunciations, noise, or fractured sounds using librosa and numpy.
- Optimized and experimented with over 6 existing Text-to-Speech models, including FastSpeech2 and YourTTS, on cloud servers, leading to comprehensive performance comparisons across over 4 metrics.
- Secured Python codes and models by implementing Fernet encryption and facilitated smooth deployment of the fine-tuned models to a production environment, collaborating on updates via Git.

Human Computer Interaction Institute, CMU Pittsburgh, USA
Explanatory Data Analytics Research Assistant 2022.6 – 2022.12

- Innovated an intelligent tutoring tool using Python and d3.js, analyzed over 5,000 students' problem-solving processes log data with graph analytics, leading to the development of insightful data-driven strategies. [\[paper\]](#)
- Executed a within-subject user study with hypothesis testing and uncovered over 10 innovative data insights, contributing to the enhancement of user experience and interaction.

School of Data Science, Fudan University Shanghai, China
Visual Analytics and Machine Learning Researcher 2021.7 – 2022.12

- Led a multidisciplinary team to develop a system for interactive point cloud labeling, achieving 84% accuracy by utilizing machine-aided data labeling tools and convolutional neural networks.
- Formulated a machine-learning-based approach to algorithmically analyze user interaction cost using Python and d3.js, contributing to improved user interaction and engagement.
- Proposed a framework that incorporate large language models to enhance users' visual analytics workflows, which could guide LLM in recommending insights based on system status data to facilitate mixed-initiative exploration.

School of Data Science, Fudan University Shanghai, China
Computer Vision Deep Learning Researcher 2022.5 – 2022.7

- Constructed a neural network framework with back-bone training for few-shot oracle character recognition problems and introduced an innovative non-rigid transformation approach for data augmentation, resulting in a nearly 30% increase in zero-shot classification accuracy. [\[paper\]](#)

Computer Science & Artificial Intelligence Lab, MIT Remote
Natural Language Processing Application Research Assistant 2021.4 – 2021.6

- Conducted experiments on the DistilBERT transformer architecture within the PyTorch framework, and initiated brainstorming sessions on generating relevant QA pairs to train models, achieving a remarkable 92% accuracy.