Ziyue Yin

+86 181 5120 0882 • ziyue.yin@dukekunshan.edu.cn • www.linkedin.com/in/ziyue-yin

EDUCATION

Duke Kunshan University (DKU) & Duke University Dual Degree Class of 2026
B.S. in Data Science, DKU Kunshan, China
B.S. in Interdisciplinary Studies: Data Science, Duke Durham, U.S.
- GPA: **3.806**/4.0; Dean's List in Fall 2022, Spring 2023, Spring 2025.

LEADERSHIP & SERVICE

TrustNet: Fake News Detection (College Student Entrepreneur Program) 06/2024 – 06/2025
Team Leader Kunshan, China
- **Developing** a **model** that integrates **text analysis** to enhance **fake news detection**.
- **Collecting and processing** a dataset of **10,000+ social media news articles**.
- **Optimizing** model accuracy using **TensorFlow & PyTorch**, refining with **precision**, **recall**, and **F1-score metrics**.

Innovation and Entrepreneurship Initiative (InE) at DKU 01/2023 – 06/2024
Senior Program Assistant, Student Worker Kunshan, China
- **Orchestrated** and **executed** high-impact entrepreneurship events, including the Entrepreneur Speaker Series (ESS) and U-Corp Lab, attracting **an average of 60 students and faculty members' participation**.
- **Developed** and **implemented** comprehensive action plans, ensuring seamless execution of DKU InE activities and program efficiency and engagement.
- **Facilitated** on-site coordination and **spearheaded** innovation-focused events in DKU Innovation Incubator (Dii), providing **over 100 incubator teams** with professional and engaging experiences for the Innovation Challenge Competition and DaChuang (College Student Innovation and Entrepreneurship) Program.

RESEARCH EXPERIENCE

Enhancement of Whisper Speech Using Audio-Visual Models 03/2025 – Now
Research Assistant to Ming Li (Professor of Electrical and Computer Engineering at DKU) Kunshan, China
- A member of the **Speech and Multimodal Intelligent Information Processing (SMIIP) Lab** at DKU.
- **Designing** a **real-time**, **low-latency speech conversion pipeline** using compressed generative models for whisper and electro-laryngeal speech enhancement.
- **Implementing** edge-optimized architectures with model pruning and quantization for on-device deployment.
- Exploring large generative models to improve zero-shot naturalness and personalization in speech output.

Multi-omics Data Analysis of Transcriptome and Metabolome in Taihu 08/2024 – Now
RA to Huansheng Cao (Assistant Professor of Environmental Science at DKU) Kunshan, China
- **Reconstructed** a **genome-scale metabolic network** for water-bloom cyanobacteria using KEGG reactions.
- **Processed** large-scale transcriptomics data with **Trinity**, **RSEM**, and **BLAST**, ensuring accurate sequence assembly and quantification, aiming at **identifying temporal metabolic shifts**.
- **Utilized** the **Kunshan Supercomputing Center (KSSC)** for computationally intensive network analysis.

Pre-trained Large Language Models for Question-Answering 06/2024 – 08/2024
Summer Research Scholar to Paul Weng (Associate Professor of ECE at DKU) Kunshan, China
- **Developed** a **Retrieval-Augmented Generation (RAG) pipeline** integrating **Llama3-8B** for question-answering, enhancing retrieval accuracy through hierarchical indexing and embedding fine-tuning.
- **Implemented** a **relevancy check mechanism** using **cosine similarity** with a **threshold of 0.6**, filtering irrelevant queries and reducing computational costs.
- **Optimized** hyperparameters by testing **75 configurations**, achieving peak retrieval performance with a **fine-tuned embedding model**, **improving *Hit Rate* by 21.4%** and *Context Recall* **by 12.9%**.
- **Evaluated** model effectiveness using *faithfulness* **(+8.3%)**, *context precision* **(+9.1%)**, and *answer relevancy* **(-6.8%)** metrics, refining retrieval for **more contextually accurate answers** while minimizing redundancy.

DCASE: Unsupervised Anomalous Sound Detection (Course Project)   08/2022 – 10/2022
Individual Contributor (Course Instructor: Prof. Ming Li & Zuchuan Li)   Kunshan, China

- **Designed** an unsupervised machine learning model using Gaussian Mixture Models (GMMs) to detect anomalous machine sounds, **achieving a peak accuracy with 64 components**.
- **Implemented Filter Bank (FBank) feature extraction**, improving sound representation and classification accuracy by effectively capturing frequency-domain characteristics.

ADDITIONAL INFORMATION

Languages:	Chinese (Native), English (Proficient), Spanish (Beginner).
Programming:	Python, Java, Shell, MATLAB.
Tools:	AGI (ChatGPT / DeepSeek), LaTeX (Overleaf), Adobe Illustrator.
Interests:	Badminton, Table Tennis, Singing, Piano.