YiXiao Zhang

Email:q18202429280@outlook.com |Tel:+86 13558882230 |LinkedIn:linkedin.com/in/yixiao-zhang-2042ab2a3 |Web:kiko022.github.io

EDUCATION EXPERIENCE

Northeastern University Shenyang, China

Major: Information Management & Information Systems Sep.2022 - Jul.2026(expected)

GPA:3.4976/5(85/100)

Awards & Honors:**Business Elite Challenge Accounting and Business Case Competition — National First Prize, Jun.2024**

Northeastern University 11th Overseas Economic Management Scholars Seminar — Outstanding Camper, Jul.2024

Relevant Courses: An introduction to Database System(92), Probability and Mathematical Statistics (92) , Data Structures, Data analysis and data mining, Statistics, Linear Algebra, Advanced Mathematics, Business Big Data Analysis and Application.

RESEARCH EXPERIENCE

Northeastern University **Research Assistant** Jul.2024 - Mar.2025

Project:**Balanced Minimum Sum-of-Squares Clustering** Supervisor: **Assoc. Prof. Qing Zhou** **NSFC Youth Fund**

- Overview: Balanced MSSC (capacity/balance); Java + reproducible eval; beats K-means++.
- Designed and developed **balanced MSSC** algorithms under capacity/balance constraints to address poor initialization and local optima issues, delivering **~1.8×** faster per iteration neighbor scoring and **9–13%** lower objectives **(up to ~17%)** than **K means++ (50 restarts)** across six synthetic/real datasets **(n≤100k, k=10–50)**.
- Engineered a Java codebase with pluggable initializers, amortized **O(1) exchange cost updates** via statistics, and **tabu search operators (one move/swap)**, augmented by a **population based diversification** mechanism to avoid premature convergence.
- Established a **reproducible evaluation protocol** (fixed balance tolerance and restart budgets, **5 seeds, wall clock timing**) and ran **ablations** to isolate gains from the **fast cost updater, swap moves, and diversification.**

Project: **GreenCircle — Campus Second-Hand Sharing Platform** **Team Leader** Sep.2025 - Nov.2025

- Overview: Campus second-hand platform for sustainability; transactions + points + community + search/reco.
- Built a **Flask + SQLAlchemy** backend with **Flask-Login**; scaled to **1,000+** listings and **¥10,000+** monthly volume.
- **Shipped** sell/donate/rent plus **green-points/community**, increasing engagement by **35%**.
- **Delivered search/recommendation** and performance tuning (query tuning, pagination/lazy loading, Pillow) to support **1,000+** concurrent with **<200** ms latency.

INTERNSHIP EXPERIENCE

Chengdu Xiaoduo Technology Co., Ltd. **Prompt Engineer (R&D Intern)** Jul.2024 - Sep.2024

- Overview: Owned the customer-service RAG initiative for online product support, focusing on query routing, retrieval relevance, and prompt orchestration from prototype to demo.
- **Built a RAG routing agent** using GPT 4o and prompt orchestration to rewrite/classify queries into four workflows (KB Q&A, product Q&A, recommendation, comparison) with confidence based fallback, resulting in **+24%** routing precision and **−31%** misroutes on internal eval.
- Implemented **hybrid retrieval in Elasticsearch** using BM25 + dense vector HNSW with bge-m3 embeddings, with a cross encoder re-ranker and rule-based filters, achieving **+18%** top 5 recall and **−27%** hallucination rate in offline QA tests.
- **Shipped a Streamlit demo** with streaming responses to validate end-to-end RAG, delivering **p50 latency ~6 s** and **−42%** feedback cycle time through faster stakeholder reviews.

Shenzhen Bairen Biotechnology Co., Ltd. **Full-Stack Development Engineer(Remote)** Oct.2024 - Dec.2024

- Overview: Bioinformatics platform; pipeline automation & containerization for Linux workloads.
- **Automated Python analysis of pipeline outputs** using Dockerized services on Linux, resulting in **−35%** runtime, **−50%** setup time, and **+15%** success rate.
- **Designed GPT-4o prompts for JSON extraction** with schema validation and retry, achieving **90%** valid parses and **−30%** manual review.

Migu Music Co., Ltd. **AI Data Intern** Dec.2024 - Mar.2025

- Overview: Spring Festival content agent (LLM + Stable Diffusion); prompts/templates and QC datasets.
- **Developed and tuned prompts for LLMs and Stable Diffusion** (text to image/video) using batch templates and negative prompts for Spring Festival assets, resulting in **~35–40%** reduction in post editing workload
- **Performed structured labeling for audio/video/images and built datasets with QC checks** using Python scripts to validate metadata and file integrity, achieving **~25%** increase in annotation throughput.

Henan Shuqing Data Technology Co., Ltd. **AI Data Training Intern(Remote)** Jun.2024 - Nov.2025

- Overview: Tianchi (Bank Customer Product Subscription Prediction); end-to-end modelling pipeline.
- **Built a data preparation and feature engineering pipeline** using Jupyter/Pandas with data integrity validation, LabelEncoder for categorical features, and Matplotlib correlation heatmaps, resulting in cleaner training data and clearer insights for targeted marketing.
- **Trained and compared logistic regression, random forest, XGBoost, and LightGBM** with scikit-learn using 5 fold cross validation and targeted hyperparameter tuning (n_estimators, learning_rate), achieving **AUC 0.87, F1 0.51, and accuracy 90%** (LightGBM best).

Midea Group (Fortune Global 500) **AI Algorithm Intern** Jul.2025 - Present

- Overview: Teachable smart-home AI across Chat & Memory Agents, focusing on memory architecture and evaluation.
- **Built user profiles and device-preference memory** using custom prompt-based dialog summarization on reset in a Chat Agent pre-demo with Streamlit UI, Redis (Docker) for chat history, and GPT 4o streaming, to enable cross-session personalisation.
- **Evaluated memory store options and selected Mem0 on-premises** using a Redis vector store with GPT 4o and bge-m3, optimising labelled memory extraction and update prompts to stabilize persistence and reduce integration issues.
- Curated an evaluation dataset from online logs using time-ordering and showroom filtering, resulting in processing of **~27,698 sessions** and yielding **2,309** scenario samples across **167** families to assess memory and dialog behaviors.
- **Delivered the Chat Agent MVP** with multi-turn dialog, correction/clarification/rejection, and memory using a single GPT-4o pipeline with legacy fallback and few-shot datasets integrated into the control screen, reducing **p50 latency to ~6 s**.
- **Improved speed and cost with two-stage SFT** using LoRA on Qwen 8B and token reduction, deploying an optimized on-premises model for inference and improving the internal benchmark from **10/22 → 12.5/22**.

SKILLS

Programming & Frameworks: Python (Pandas, NumPy, Matplotlib, scikit-learn), R (statistical analysis, data modeling), Java, SQL,
Generative AI & Conversational Agents: prompt engineering (few-shot, system/user prompts); RAG pipelines; memory systems (Mem0); embeddings (bge-m3); cross-encoder re-ranking; distillation/SFT (LoRA); Stable Diffusion (text-to-image/video)
Vector Search & Infrastructure: Elasticsearch hybrid retrieval (BM25 + HNSW); Redis (vector store); MySQL; Linux; Docker; Git; vector indices; re-ranking
Web & Application Development: Flask (backend services); Angular; Streamlit apps, Tableau

HONORS AND CERTIFICATES

Estonian National Summer School **Full Scholarship** (17/500) €1,050 scholarship 2025
Bucharest Summer University **Partial Scholarship** Selected for partial funding at 19th BSU (750+ participants) 2025
Zhejiang University**|SDG Summer School:Intelligent Technology Promotes Sustainable Development** (3 credits) Jul.2025
Shanghai Jiao Tong University**|SDG Summer Camp**:Ocean Sustainability in a Changing Climate (2 credits) Jul.2025
University of Maryland**| AI and Career Empowerment** Jun.2025
University of Chinese Academy of Sciences**| AI & Humanistic Economics Summer School** Jul.2025
Hertie School (Berlin)**| Data Science Summer School** Aug.2025
State Key Laboratory of AI Safety| **LLM Safety Assessment & Enhancement Summer School** Aug.2025
SustainMV| **The Sustainability Summer School** (online, Mecklenburg-Vorpommern) Aug.2025
Sun Yat-sen University| **Air-Space-Ground-Sea Intelligent Sensing Technology Summer School** Aug.2025
University of Oxford| **Digital Humanities @ Oxford Online Summer School: AI in Research Libraries** Sep.2025