Tingyu Zhang

tingyu.z@northeastern.edu · 206-741-8937 · LinkedIn · GitHub · Seattle · Open to Relocation

EDUCATION

Northeastern University — Seattle, Washington
Master of Computer Science, (GPA: 4/4) Core Courses: ML Sys, DL network. — *September 2025 – May 2027*

Southern University of Science and Technology — Shenzhen, Guangdong
Bachelor of Applied Mathematics(GPA:3.53/4) Core Courses: Linear Algebra, Statistics. — *August 2021 – July 2025*

SKILLS

Programming Languages: Python (proficient in data analysis and GenAI training), Java (proficient in system design), MATLAB, C/C++, JavaScript/TypeScript, PostgreSQL, MongoDB, R, LaTeX.
Developer Tools: VS Code, PyCharm, IntelliJ, DataGrip, Docker, Kubernetes, Maven.
Libraries&Framework: Pandas, NumPy, Matplotlib, Pytorch, Tensorflow, Cupy, scikit-learn, seaborn, LangChain, React, Node.js, Spring Boot, Swift.

PROFESSIONAL EXPERIENCE

Research Assistant — January 2025 – September 2025
Southern University of Science and Technology — *Shenzhen, Guangdong*

- Designed and trained a **continual learning Multi-Armed Bandit (MAB)** model with **parallel optimization**, formulating a **preference-theoretic reward function** analogous to **RLHF** feedback modeling; implemented training loop in **PyTorch,Tensorflow** with vectorized batch updates to improve convergence efficiency.
- Modeled the **influence maximization** problem as an **online combinatorial MAB** over a **stochastic graph**.

Machine Learning Engineering Intern — July 2022 – August 2022
HAOCHENG TECH HOLDING — *Shenzhen, Guangdong*

- Engineered end-to-end **ML training pipeline** integrating **Random Forest** and **Logistic Regression** for material classification on large-scale tabular data, achieving **87.3% accuracy**; applied **TF-IDF** vectorization and **NLP** preprocessing via **NLTK** and **scikit-learn** for feature extraction and downstream search optimization.
- Designed **PostgreSQL** schema with **B-tree** and composite indexes on high-cardinality video metadata fields to accelerate feature retrieval for model serving, achieving a **15× latency reduction** (30 min → 2 min).

Astronomical Image Classifier: ViT Fine-Tuning & LLM Agent — February 2025 – May 2025

- Fine-tuned a **Vision Transformer (ViT)** for multi-class astronomical image classification using **LoRA** (parameter-efficient fine-tuning) on a remote **GPU** server via **SSH**, achieving 90.4% accuracy while reducing trainable parameters by over 90%.
- Built a **supervised few-shot learning** corpus of 500+ prompt-completion pairs with interpolation curve ground-truth labels via a **prompt engineering pipeline** (cleaning, normalization, **JSON** serialization); benchmarked against **LeNet-5** trained on **MNIST** (60K images, 80.3% accuracy) as baseline.

PROJECTS

AI-Powered Job Search Agent with Ollama — *January 2026 – Present*

- Developed **autonomous job search agent** using **Ollama** with **Llama 3.2** model running locally, implementing **RAG** architecture with **ChromaDB** vector store for personalized job matching.
- Engineered **multi-agent system** for resume tailoring, cover letter generation, and application tracking using **LangChain** and **Python**, implementing **prompt chaining** and **few-shot learning** patterns that reduced application preparation time from 2 hours to 15 minutes per job.
- Built **web scraping pipeline** using **Selenium** and **BeautifulSoup** to aggregate postings from job-searching webs, integrated with **SQLite** database and **Celery** task queue for asynchronous processing, collecting 2K+ daily job listings.

Cloud-Native Microservices for Distributed System — *September 2025 – December 2025*

- Architected **cloud-native microservices** on **AWS ECS** with **Docker containerization**, **Application Load Balancer (ALB)** for traffic routing, and **auto-scaling policies**, reduced the develop time to 5-10 minutes
- Integrated **RabbitMQ** message queues with **Apache Commons Pool** for connection pooling, implementing manual acknowledgements and retry logic to ensure reliable **asynchronous processing**.

SpringBoot FullStack Book Management Web Application — *November 2025 – December 2025*

- Architected full-stack application using **React 19** with **TypeScript** and Tailwind CSS for the frontend, **Spring Boot** with **PostgreSQL** for the backend, achieving type-safe development and responsive design of the webpage.
- Developed **RESTful APIs** with **Spring MVC** supporting full CRUD operations, including multipart file uploads for book covers with client-side image preview using the FileReader API, achieving **instantaneous add and delete operations** with optimized response times.

PUBLICATIONS & PREPRINTS

[1] T. Zhang, Z. Hu, and F. Kong, "Best-of-Three-Worlds Analysis for Dueling Bandits with Borda Winner," in *the International Conference on Learning Representations (ICLR)*, 2026.

[2] K. Zhou, T. Zhang, W. Chen, and F. Kong, "Hybrid Combinatorial Multi-armed Bandits with Probabilistically Triggered Arms," arXiv preprint arXiv:2512.21925, 2025.