

Tianyi Zhou

tzhou25@seas.upenn.edu | +1 (312) 622-3218 | Philadelphia, PA | [Portfolio: goldenexpn.github.io](https://github.com/goldenexpn)

Education

University of Pennsylvania

M.S.E. in Systems Engineering

Philadelphia, PA

Expected May 2027

University of Rochester

B.S. in Computer Science & Data Science

Rochester, NY

Aug 2021 – May 2024

Research Experience

Ale Lab, University of Pennsylvania

Philadelphia, PA

Graduate Research Assistant (Advisor: Prof. Osbert Bastani)

Oct 2025 – Present

- Designed and developed **Code Forest**, a hierarchical external memory framework organizing programming problems, AST-based atomic modules, and function implementations for LLM code reasoning.
- Built an automated distillation and retrieval pipeline utilizing AST dependency analysis, semantic embeddings, and LLM-driven module merging to enable context-efficient code synthesis.
- Evaluated function-equivalence classifiers for memory deduplication, achieving **92.1% accuracy** and **0.932 F1** score, establishing retrieval-augmented infrastructure for reinforcement-learning-guided code generation.

Huang Lab, University of Pennsylvania

Philadelphia, PA

Graduate Research Assistant

Feb 2026 – Present

- Built a supervised continual-learning benchmark across 41 pathology datasets (259 classes) evaluating foundation models (LiteFM, BioMedCLIP, CONCH) under LoRA, EWC, and distillation strategies.
- Demonstrated that replay with knowledge distillation achieves **0.911 mean balanced accuracy** on base tasks while mitigating catastrophic forgetting across 4 sequentially adapted downstream tasks (0.910 retained BA).
- Analyzed LoRA weight-space drift and adapter interference, motivating **PathCAR**, an automated prompt-conditioned adapter-routing mechanism.

Industry Experience

ECCOM Network System Co.

Shanghai, China

Machine Learning Engineer Intern

May 2024 – Aug 2024

- Engineered end-to-end classification pipelines on tabular business metrics using XGBoost, CatBoost, and TabNet.
- Conducted automated feature selection and SHAP-based feature importance interpretation aligned with client KPIs.
- Deployed an employee turnover prediction pipeline yielding **90%+ accuracy** on blind validation sets.

Red Note (Xiaohongshu)

Shanghai, China

Business Intelligence Intern

May 2023 – Aug 2023

- Implemented distributed web scrapers to track competitor sales performance and category pricing dynamics.
- Applied TF-IDF and keyword extraction on user engagement text to uncover emerging platform content trends.
- Built an LSTM sequence model predicting purchase intent from historical user activity signals.

Projects

Benchmarking Parameter-Efficient Fine-Tuning (PEFT) on Llama 3

2024

- Evaluated parameter-efficient fine-tuning architectures (LoRA-FA, AdaLoRA, VeRA) against standard LoRA on Llama-3-8B.
- Engineered an automated LLM-as-a-judge evaluation suite analyzing latency, memory footprint, and generation quality trade-offs on high-performance compute clusters (Delta HPC).

Academic Service

Teaching Assistant, University of Rochester

Aug 2023 – May 2025

- TA for *Computer Models and Limitations* (Spring 2025), *Intro to Programming* (Fall 2024), and *Computational Statistics* (2023–2024). Led weekly programming recitations and 1-on-1 debug labs in Python and R.

Technical Skills

Languages: Python, C/C++, R, SQL, Bash

Frameworks & Libraries: PyTorch, Hugging Face (Transformers, PEFT, TRL), Scikit-Learn, XGBoost, Pandas, NumPy

ML/AI Specializations: LLM Reasoning & Memory, PEFT/LoRA, Continual Learning, RAG, AST Analysis, Vision Transformers

Infrastructure & Tools: Linux/Bash, Git, HPC (SLURM), Docker, Weights & Biases