

Shichen Zhang

Brooklyn, NY | (646) 408-5362 | sz4968@nyu.edu | [LinkedIn](#)

Education

New York University

M.S. in Computer Science

New York, NY

Sep 2025 – May 2027 (Expected)

- Coursework: Design and Analysis of Algorithms I, Software Engineering I, Principle of Database System, Open Source/Professional Software Development, Machine Learning, Engineering Application of Deep Learning

Harvard Extension School

Graduate Coursework in Computer Science

Cambridge, MA (Online)

Jan 2025 – Aug 2025

- Coursework: Web Application Development using Node.js, Computer Vision, Deep Learning

Macalester College

B.A. in Philosophy

Saint Paul, MN

Sep 2017 – May 2022

Skills

Programming Languages: C++, Python, JavaScript, TypeScript, HTML/CSS

Web Development: FastAPI, Node.js, Express.js, Angular, MongoDB

Machine Learning & Data Science: TensorFlow, PyTorch, scikit-learn, NumPy, Pandas, Matplotlib, MNE

Developer Tools: Git, GitHub, VS Code, Jupyter, Linux/Unix Shell, Postman

DevOps & Production: Docker, Terraform, CircleCI

Projects

AI Email Inbox Agent | [GitHub](#)

Sep 2025 – Dec 2025

- Built a Discord bot that lets users manage their email inbox through natural-language chat, including listing, reading, and summarizing emails.
- Designed and implemented a **FastAPI (Python)** backend that receives Discord messages, runs LLM-driven multi-step agent workflows, and returns responses in real time.
- Built an **MCP-style** tool-use framework with **JSON-Schema** tool definitions and a registry/runner, enabling the LLM to invoke email tools (list/fetch) and auth tools (login/logout/status) in multi-step agent workflows.
- Built OAuth2 authorization endpoints to securely link each user's email account; persisted per-user tokens in **SQLite** to maintain durable sessions for multiple users.
- Established **CI/CD** and deployment workflows with **Git/GitHub**, **CircleCI**, **Ruff**, **MyPy**, **Pytest** (unit/integration/end-to-end), **Docker**, and **Terraform**.

GPT-style Language Model Training

Sep 2025 – Dec 2025

- Trained a 124M-parameter GPT-2-style **Transformer** in **Python/PyTorch** on FineWeb-Edu dataset (~10B tokens) with a next-token prediction objective.
- Built a scalable data pipeline to download and preprocess a large-scale web text corpus, tokenize with **tiktoken**, and generate sharded, streaming-ready token datasets for high-throughput training I/O.
- Scaled training to **8x A100** using **Distributed Data Parallel (DDP)** and **gradient accumulation**, optimizing throughput with **BF16/FP32 mixed precision**, **FlashAttention**, and **fused AdamW**.
- Implemented checkpointing and periodic evaluation (validation loss, zero-shot HellaSwag), achieving ~30.5% accuracy.

PhotoBoard | Photo Sharing Web App

Jan 2025 – May 2025

- Built a photo sharing platform supporting photo uploads, browsing, and editing to enhance user interaction.
- Designed and implemented a backend **RESTful API** with **Express.js/Node.js** and **MongoDB** (via **Mongoose**) to handle CRUD operations for photo metadata.
- Developed a responsive **Angular (TypeScript, HTML/CSS)** frontend with reusable components and services, enabling smooth API integration and maintainable design.
- Managed version control with **Git/GitHub** and deployed the application on DigitalOcean using **Linux/Unix shell**.

Research

Visual Semantic Retrieval from EEG Brain Activity

Sep 2025 – Dec 2025

- Developed neural decoding models in **Python/PyTorch** to predict **CLIP** image-label embeddings from multi-channel EEG time-series for visual semantic retrieval, using a dataset of 20,880 stimulus events.
- Analyzed stimulus-evoked responses with ERP component measurements and hypothesis testing (**t-tests**, **ANOVA**) to validate condition effects.
- Designed a reusable preprocessing system with **MNE/NumPy/Pandas** that standardizes filtering, epoching, artifact handling, and trial alignment, with caching of intermediate results for fast reruns and reproducible experiments.
- Built a temporal EEG encoder (**1D CNN + Transformer**) and trained it with **contrastive learning** to align EEG embeddings with **CLIP** image-label embeddings, evaluated with Top-K retrieval metrics.