

Hannah Sun

205 S State St, Ann Arbor, MI 48104 | (415)-549-6332 | hysun@umich.edu | sun770311.github.io

EDUCATION

University of Michigan, College of Literature, Science and the Arts

Ann Arbor, MI

BS in Data Science, Minor: Mathematics, GPA: 3.74/4

Expected Apr 2025

Coursework: Math (Financial Assets, Continuous Time Models, Calculus, Matrices, Discrete Math, Probability, Differential Equations, Statistics, Regression); Data Science (Computation Theory, Machine Learning, Data Mining, Simulation in C++)

SKILLS

Programming and Markup Languages: C++/C, Python, Java, R, SQL, Bash, HTML, CSS, JavaScript, TypeScript, XML

Frameworks: Git, Linux, Docker, PostgreSQL, SQLite, Excel, React, Flask, NumPy, Pandas, PyTorch, LangChain, AutoGen

PROFESSIONAL EXPERIENCE

ProQuest LLC, University of Michigan College of Engineering

Ann Arbor, MI

Capstone Project Internship Student Developer

Jan 2024 - Present

- Automated newspaper article segmentation to enhance database query results in a team of six students and three mentors
- Prompted GPT-4o with base64-encoded newspaper images and OCR XML data to classify article titles and body text
- Optimized LLM instruction clarity to achieve 87.6% precision and 85.5% recall across 41 front pages in the past century
- Cut LLM token costs to 4.5 cents per post-1980 front page, reducing expenses to one-sixth of manual segmentation
- Set pixel coverage threshold on binarized images to detect artifacts with strong qualitative results across entire training set

Pachira Information Technology (Hengqin) Co., Ltd

Hengqin, CN

NLP Algorithm Intern

May 2024 - Aug 2024

- Worked with Toyota Motor Corporation to develop a smart in-car Q&A system set to launch in late 2025
- Developed new function to answer dashboard warning light queries and deployed updated system for simulator testing
- Engineered prompt to handle simultaneous response refusal, multi-intent recognition, and context-based query translation
- Used batch processing to optimize web page embedding speed below 5 seconds during the RAG process

University of Michigan Math Department

Ann Arbor, MI

Undergraduate Math Lab Tutor

Sep 2023 - Dec 2023

- Tutored over 100 students in multivariable calculus, differential equations, matrix algebra, geometry, and probability
- Conducted 1-on-1 and group exam review sessions with a concentration in practice problems and challenging concepts

Bank of China Macau Branch

Macao, CN

Information Technology Academic Exchange

May 2023 - Jul 2023

- Organized customer relation and remittance data in a team of 3 to support the 2023 Due Diligence Reporting Project
- Extracted primary key combinations to structure SQL tables detailing transaction records within the past year
- Optimized query to retrieve 100,000 customer relation and 1 month of global remittance entries in under 30 seconds

PROJECT EXPERIENCE

University of Michigan EECS 484 Course Project

Ann Arbor, MI

Fakebook Database

Sep 2024 - Present

- Developed SQL database schema for fictional social media platform with key entities Users, Messages, Photos, and Events
- Created relational constraints using foreign keys, triggers, and sequences to maintain data integrity and prevent duplication
- Tested database in the SQL*Plus Linux environment to ensure schema and functionality compliance
- Utilized JDBC to implement SQL queries in Java, extracting user birth month, name statistics, and mutual friends
- Executed sorting, grouping, and nested subqueries, optimizing average runtime to 0.075 seconds across nine tasks

PothoAI

Jersey City, NJ (Remote)

Back End Developer

Jun 2024 - Sep 2024

- Built the backend of a chatbot platform leveraging the AutoGen framework to integrate LLMs, tools, and human input
- Implemented SQLite database operations to manage user sessions and message history for efficient data retrieval
- Customized sequential and autonomous agent workflow to answer input queries individually or in a group chat setting
- Used React and Nango to store Slack workspace contacts and build a mass messaging bot, deployed testing in Docker

MobiDrop (Zhejiang) Co., Ltd.

Shanghai, CN (Remote)

Single-cell Large Language Model

Nov 2023 - May 2024

- Constructed a pre-training step in scGPT, a Transformer-based model used for single-cell gene expression level prediction
- Sampled 300,000 human blood cells from CELLxGENE, preserving all cells from each randomly chosen organism ID
- Preprocessed the sample by filtering, normalizing, discretely binning, and tokenizing highly variable genes
- Logged training progress, achieving a 54% reduction in training and 6% decrease in validation error after 6 epochs