

# Yueran (Hannah) Sun

205 S State St, Ann Arbor, MI 48104 | +1 (415)-549-6332 | [hysun@umich.edu](mailto:hysun@umich.edu) | [sun770311.github.io](https://sun770311.github.io)

## EDUCATION

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

**Ann Arbor, MI**

*BS in Data Science, Minor: Mathematics, GPA: 3.69/4*

Expected Apr 2025

**Coursework:** Math (Financial Mathematics, Calculus, Matrices, Discrete Math, Probability, Differential Equations, Statistics, Regression); Data Science (Database Management, 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

**Tools:** Git, Linux, MongoDB, SQLite, IBM Db2, DBeaver, React, Flask, FastAPI, NumPy, Pandas, PyTorch, LangChain

## PROFESSIONAL EXPERIENCE

**ProQuest LLC, University of Michigan College of Engineering**

**Ann Arbor, MI**

Capstone Project Internship Student Developer

Jan 2024 – Dec 2024

- Automated textual and spatial segmentation of Detroit Free Press newspaper front pages into individual articles
- Achieved 91% average recall for title classification on a test set of 20 pages covering each decade from 1923 to 1994
- Classified majority of text through a GMM, reducing pipeline cost to 15 cents per page, half of manual segmentation
- Eliminated visual noise including photos and marks by identifying large dark regions in binarized front page images
- Prompted GPT-4o to identify indexes and advertisements in page sections, adjusting portion size threshold by decade

**Pachira Information Technology (Hengqin) Co., Ltd**

**Hengqin, CN**

NLP Algorithm Intern

May 2024 - Aug 2024

- Partnered with Toyota Motor Corporation to design and develop a smart in-car Q&A system set to launch in late 2025
- Created a function for handling warning light queries and deployed multiple system iterations in simulator testing
- Engineered LangChain prompt to manage response refusal, multi-intent recognition, and context-based query translation
- Optimized RAG pipeline by implementing batch processing, achieving a web page embedding speed of under 5 seconds

**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
- Led 1-on-1 and group exam review sessions, guiding students through 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 with a team of 3 to support the 2023 Due Diligence Reporting Project
- Extracted primary key combinations to structure SQL tables for efficient tracking of transaction records from the past year
- Optimized query to retrieve 100,000 customer relation records and 1 month of global remittance entries under 30 seconds

## PROJECT EXPERIENCE

**University of Michigan EECS 484 Course Project**

**Ann Arbor, MI**

Fakebook Database

Sep 2024 – Dec 2024

- Designed SQL database schema for a fictional social media platform with key entities (Users, Messages, Photos, Events)
- Ensured data integrity through foreign keys, triggers, and sequences, testing output accuracy in the SQL\*Plus environment
- Utilized JDBC to implement SQL queries in Java, extracting user birth month, name statistics, and mutual friends
- Executed sorting, grouping, and nested subqueries, optimizing runtime to an average of 0.075 seconds across nine tasks
- Developed MongoDB aggregation pipelines to generate Fakebook monthly engagement and activity reports

**PothoAI**

**Jersey City, NJ (Remote)**

Back End Developer

Jun 2024 – Sep 2024

- Engineered the backend of a chatbot platform with the AutoGen framework, integrating LLMs, tools, and human input
- Implemented SQLite database operations for user sessions and message history, ensuring data retrieval and management
- Tailored sequential and autonomous agent workflows to handle queries in individual and group chat settings

**MobiDrop (Zhejiang) Co., Ltd.**

**Shanghai, CN (Remote)**

Single-cell Large Language Model

Nov 2023 - May 2024

- Designed a pre-training step for scGPT, a Transformer model for predicting single-cell gene expression levels
- Sampled 300,000 human blood cells from CELLxGENE, preserving all cells from each randomly selected organism ID
- Preprocessed samples by filtering, normalizing, discretely binning, and tokenizing highly variable genes
- Monitored training progress, achieving a 54% reduction in training error and a 6% drop in validation error over 6 epochs