Yueran (Hannah) Sun

205 S State St, Ann Arbor, MI 48104| (415) 549-6332| hannahxsun0504@gmail.com| https://github.com/sun770311

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 Dec 2025
Relevant Coursework: Math (Calculus, Matrices, Discrete Math, Probability, Differential Equations, Statistics, Regression Modeling); Data Science (Machine Learning, Data Mining, Simulation in C++); Finance (Market Microstructures, Options)

SKILLS, LANGUAGES & INTERESTS

Programming and Markup Languages: C++/C, Python, R, SQL, Bash, HTML, CSS, JavaScript, LaTeX, XML
Frameworks and Tools: Git, Linux, AWS, DBeaver, Excel, Flask, Scikit-learn, SpaCy, PyTorch, LangChain
Languages: English (Native), Mandarin (Native), Cantonese (Native)
Interests: Visual Art (Painting, Digital illustration, Calligraphy)

PROFESSIONAL EXPERIENCE

Pachira Information Technology (Hengqin) Co., Ltd   **Hengqin, CN**
NLP Algorithm Intern   May 2024 - Present
- Partnered with the Tsinghua-Toyota Joint Research Institute to design a smart in-car Q&A system set to launch in 2025
- Trained the model on owner's manual data to answer dashboard symbol queries using word and image explanations, and updated the deployment server to support these queries in the vehicle test simulator
- Optimized web page encoding speed, embedding model accuracy, and query translation consistency in the RAG process

Bank of China Macau Branch   **Macao, CN**
Information Technology Academic Exchange   May 2023 - July 2023
- Updated customer relation and remittance transaction data to support 2023 due diligence reporting
- Identified primary keys from 20+ tables to extract and structure intermediate SQL tables
- Optimized query to retrieve 100,000 customer relation and 1 month of remittance transaction entries in under 30 seconds

PROJECT EXPERIENCE

PothoAI   **Jersey City, NJ (Remote)**
AI Agent Developer   Jun 2024 - Present
- Built FinSense, a financial recommendation system that dynamically selects and executes functions based on user query
- Implemented stock forecasting using SARIMAX, automating data collection, parameter selection, and plot visualization
- Integrated the Finnhub API to retrieve real-time stock price, company news, earnings surprises, and basic financial metrics
- Designed a user interface with interactive features including real-time chat with the LLM and modal windows

ProQuest LLC, University of Michigan College of Engineering   **Ann Arbor, MI**
Automated Newspaper Article Segmentation   Jan 2024 - Present
- Developed automatic newspaper article segmentation to enhance online news database query
- Clustered text by font size to differentiate headline from body text, fixed headline spelling using Tesseract OCR
- Used BERTopic to filter out advertisement and index artifacts, improving artifact identification recall by 42% on average
- Ran segmentation on 550 front pages of the Detroit Free Press in the ProQuest TDM Studio

MobiDrop (Zhejiang) Co., Ltd.   **Shanghai, CN (Remote)**
Single-cell Large Language Model   Nov 2023 - May 2024
- Constructed a pre-training step in Transformer-based model scGPT for downstream gene expression prediction tasks
- Sampled 300,000 human blood cells from CELLxGENE, retaining and shuffling all cells from selected donor IDs
- Preprocessed sample by filtering, normalizing, binning, tokenizing, and padding highly variable genes with Scanpy
- Logged training progress, achieving a 54% reduction in training and 6% decrease in validation error after 6 epochs

University of Michigan EECS 445 Class Project   **Ann Arbor, MI**
Tiffany's Tensor Tourism   Feb 2024 - Mar 2024
- Trained a CNN to classify images of the Pantheon and the Hofburg Imperial Palace
- Applied transfer learning, image augmentation, and Grad-CAM to improve model robustness and test performance
- Achieved test AUROC of 0.84 after enhancing using dropout, LASSO regularization, and grayscale data augmentation

University of Michigan EECS 281 Class Project   **Ann Arbor, MI**
Puzzle Solver   Aug 2023 - Sep 2023
- Developed a multi-layered maze puzzle solver in C++
- Stored and managed maze data using 2-D vectors, 3-D vectors, and deques to optimize memory and runtime efficiency
- Implemented both BFS and DFS to efficiently solve puzzles of up to 400 million cells with interactive doors and traps