Yueran (Hannah) Sun

205 S State St, Ann Arbor, MI 48104 | +1 (415)-549-6332 | hannahxsun0504@gmail.com | sun770311.github.io

EDUCATION

University of Washington Graduate School — **Seattle, WA**
MS in Data Science — Starting Sep 2025

University of Michigan, College of Literature, Science and the Arts — **Ann Arbor, MI**
BS in Data Science, Minor: Mathematics, GPA: 3.69/4 — Expected May 2025
Courses: NLP, Computational Linguistics, Web Systems, UI Development, Database Management, Computation Theory, Machine Learning, Data Mining, Discrete Math, Financial Math, Statistics, Regression, Probability, Differential Equations, Multivariable Calculus, Matrix Algebra

SKILLS

Languages: C++/C, Python, Java, R, SQL, PL/SQL, Scala, Bash, JavaScript, TypeScript, HTML, CSS
Tools: Git, Linux, MongoDB, SQLite, IBM Db2, DBeaver, AWS EC2, Figma, Cypress, FAISS, BeautifulSoup
Frameworks: React, Flask, FastAPI, NumPy, Pandas, PyTorch, LangChain, Jinja, BERT, SpaCy, NLTK, Gensim

PROFESSIONAL EXPERIENCE

University of Michigan Math Department — **Ann Arbor, MI**
Mathematics Course Grader — Jan 2025 – Present
- Evaluated homework assignments for Math 417: Matrix Algebra I, ensuring accuracy and adherence to deadlines
- Collaborated with Professor Ruiyuan Chen to develop grading rubrics and provide constructive student feedback

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 from 1923 to 1994
- Classified majority of text through GMM, reducing pipeline cost to 15 cents per page, half of manual segmentation
- Denoised photos and marks using computer vision by identifying large dark regions in binarized front page images

Pachira Information Technology (Hengqin) Co., Ltd — **Hengqin, CN**
NLP Algorithm Intern — May 2024 - Aug 2024
- Developed an in-car AI voice assistant set to launch in Chinese Toyota vehicles in 2025
- Created function to handle warning light queries and deployed multiple system iterations in simulator testing
- Engineered RAG LangChain prompt to manage response refusal, multi-intent recognition, and context query translation

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

PROJECT EXPERIENCE

University of Michigan EECS 485 Course Project — **Ann Arbor, MI**
Instagram Clone — Jan 2025 – Present
- Developed an Instagram clone with client-side dynamic pages, integrating a REST API by refactoring server-side code
- Built frontend using React and AJAX, supporting liking, commenting, and infinite scrolling without page reloads
- Tested API calls and UI functionality using Cypress with the Electron browser, deployed application on AWS EC2

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 sorting, grouping, and nested subqueries, optimizing runtime to an average of 0.075 seconds
- Developed MongoDB aggregation pipelines to generate Fakebook monthly engagement and activity reports

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 predicting single-cell gene expression levels
- Sampled 300,000 human blood cells from CELLxGENE, normalizing, binning, and tokenizing highly variable genes
- Wrote bash script to allocate GPU resources for training, achieving 54% and 6% reduction in training and validation error