YUANRONG HAN

rex.han@berkeley.edu | Toronto, ON, CAN | linkedin.com/in/rexhanh | github.com/rexhanh | rex-han.com/

EDUCATION

University of Toronto **01/2024 - 12/2025**
Master's, Computer Engineering

University of California - Berkeley **08/2018 - 05/2021**
Bachelor's, Computer Science

SKILLS

Skills: LLM, Natural Language Processing (NLP), Machine Learning, Neural Networks, Reinforcement Learning, Git, Docker, Kubernetes, HTML/CSS, Java, Rust, TypeScript, Python, Data Structures & Algorithms, Digital Ocean, FastAPI, JavaScript, Jest, Jupyter, LangChain, Linux/Unix, NumPy, Pytorch, React.js, React Native, Redux.js, REST APIs, Scikit-learn
Languages: Mandarin, Chinese, English

PROFESSIONAL EXPERIENCE

University of Toronto **Toronto, ON, Canada**
Graduate Research Assistant *10/2024 - 07/2025*

- Engineered a service-oriented backend architecture using FastAPI, supporting low-latency data retrieval and high-concurrency request handling for lecture content delivery.
- Designed and implemented RESTful API endpoints with well-defined schemas and dependency injection patterns, improving code maintainability and integration stability.
- Integrated external data sources and university authentication systems, handling session management, access control, and secure data pipelines for protected course materials.
- Collaborated with institutional IT and Accessibility Services to validate system behavior under data security and privacy constraints, ensuring compliance with internal data governance policies.
- Successfully integrated Azure services to call LLM APIs, enhancing backend capabilities for AI features and improving system efficiency and response times.

Dream Formula Education **San Francisco, CA, USA**
Mobile Software Engineering - iOS *05/2021 - 06/2023*

- Led the development of an iOS app to assist students in organizing and managing their academic responsibilities.
- Collaborated closely with stakeholders, including educators and students, to understand requirements and gather feedback for iterative improvements.
- Implemented secure Firebase-based authentication and database, providing a user-friendly onboarding experience.
- Delivered iterative updates based on feedback from pilot users, improving overall performance and usability.

PROJECTS & OUTSIDE EXPERIENCE

Retrieval-Augmented Virtual Teaching Assistant - *Link to project* **Toronto, ON, Canada**
06/2024 - 04/2025

- Designed and implemented a retrieval-augmented chatbot (RAG) using Python (Langchain) and JavaScript, delivering real-time, contextually accurate responses.
- Enhanced the professor’s website with a responsive frontend (HTML, CSS, JS), improving UX for students seeking immediate clarifications.
- Optimized the backend to achieve sub-second response times, increasing user engagement by 20%.

Taro - Web *10/2025 - 12/2025*

- Collaborated in a 4-person team to build a fullstack coffee chat booking app that connects students and professionals for 1:1 networking conversations.
- Contributed to an eventdriven backend running on Kubernetes, using Redis Streams, PostgreSQL, and CloudEvents to implement atleastonce delivery, deadletter queues, and a remediation pipeline for resilient invite and notification workflows.
- Develop a React + Vite frontend with Firebase authentication, realtime push notifications (FCM/Expo), and a dashboard for creating, managing, and tracking coffee chat invitations.
- Worked with a modern cloudnative stack including Docker, FluxCD (GitOps), KEDA autoscaling, Cilium, Prometheus, and Grafana to implement CI/CD, autoscaling, and endtoend observability in a productionlike environment.

QuickLit - *Link to project* **Toronto, ON, Canada**
09/2025 - 12/2025

- Built a fullstack literature review assistant that retrieves, ranks, and summarizes arXiv papers using a FastAPI backend, OpenAI models, and a React/TypeScript frontend.

- Implemented a streaming chat interface that lets users ask naturallanguage questions about selected papers, backed by a Qdrant vector database for retrievalaugmented generation (RAG).
- Containerized frontend and backend with Docker and Docker Compose, including health checks and environmentbased configuration for local and productionlike deployments.
- Designed evaluation scripts to measure retrieval relevance (cosine similarity, precision metrics) and LLM summary quality (coverage, concision, faithfulness, writing quality) using embeddingbased scoring.