

HAORAN CHENG

📞 971-900-6780 ✉ haorancheng@brandeis.edu 💻 [Haoran Cheng](#) 🌐 [Rsirp0c](#)

Education

Brandeis University, Waltham, MA

Sep 2021 – May 2025

Bachelor of Science in Computer Science / Bachelor of Art in Anthropology

CS GPA: 3.8

Awards: Dean's List (Fall 2022 - present)

Coursework: Data Structure, Operating System, Natural Language Processing, Deep Learning, Distributed System, Software Engineering, Information Visualization, Object-oriented programming in Java, Database Management, Computer Networking

Experience

Chewy

June 2024 – Aug 2024

Software Engineer Intern

Boston, MA

- Developed the infrastructure automation for a corporate-level internal tool built on Backstage, utilized by over 500 developers; automate the EKS cluster deployment through **Helm charts**, **Terraform**, and **CI/CD** pipeline.
- Coded the backend script using **Typescript** and **Cookiecutter**, later simulate and verify with 10+ Jest unit test.
- Implemented **GitHub Actions** and **Jenkins** CI/CD pipelines to streamline testing and deployment workflows.
- Utilized **Jenkins**, **Docker**, **Poetry**, and **Python** script to auto deploy custom Docker image to **AWS ECR**.

BayMeters | IoT Stealth Startup

Aug 2024 – Present

Founding Engineer

San Jose, CA

- Developed the company website with **TypeScript**, **React**, **Tailwind CSS**, **Node.js**, **RESTful API**, and **Elasticsearch**, delivering a responsive design and robust product search functionality.
- Implemented a **CI/CD pipeline** using **GitHub Actions** and deployed the service with **Docker**, **Nginx**, **AWS ECR**, **ECS**, and an **Application Load Balancer**, integrating with GoDaddy DNS service.

Zbyte Technology | Data Warehouse Startup

May 2023 – Sep 2023

Technical Product Manager Intern

Hangzhou, China

- Involved in the design, development and product launch of a **RAG** (LLM-dataset) web application, improving domain-specific response accuracy; collaborated with a cross-functional team of 10+ members.
- Directed **design reviews** bwtween UI/UX and engineer teams, resolving over 150+ frontend **React** component errors.
- Drafted **documentation** for the data warehouse, including, 50+ SQL Commands and 10+ Build-in Functions.

Branda | Brandeis Campus App

Jan 2023 – Jan 2024

Software Engineer

Waltham, MA

- Responsible for developing mobile application using **React Native**. Deployed a interactive calendar daily used by 1.6K student to keep track of school events. Used **Redis** to cache data, reducing the workload on main database by 20%.
- Coded **REST API** in Node.js and Express.js to fetch data from **MongoDB** as backend.

BitbyBit Coding | a Google for Non-Profit K-12 after-school coding program

Sep 2023 – present

President

Waltham, MA

- Initiated and managed the in-person program. Designed the **Python** curriculum and communicated with 5 + Waltham local clubs and school executives to establish partnership.

Projects

Job-Search Chatbot | *Next.js, Typescript, FastAPI, Vercel, Cohere, Pinecone* | [link](#)

- Developed an **LLM-powered RAG** chatbot to search for matching jobs based on user queries, deployed using **Vercel**.
- Implemented **chat streaming** and **citation support** using **Next.js** and **FastAPI**, along with an **agentic logic** to decide whether to perform a RAG search and to dynamically adjust the frontend reply format.
- Built a **web scraper** to extract job from sources such as Google and LinkedIn, and upserted the data into **Pinecone**.

RAG LLM-dataset chatbot | *Pinecone, Langchain, Cohere, OpenAI api, Streamlit* | [link](#)

- Developed a **retrieval-augmented generation** chatbot that support difference choices over Embedding model, chunking strategy, top k retrieval, LLM model, prompt engineering, and meta-data retrieval search.
- Implemented the file uploading workflow; Utilized **Langchain splitter module** and **Python** script to clean and chunk the uploaded file and use **Huggingface sentence transformers** and **Pinecone** to vectorize and store data.
- Used **SpaCy NER model** to retrieve meta data from prompt and ran a **hybrid search** using meta-data and vector.

Brandeis Course Evaluation Web Application | *React, MongoDB, Express, Node.js, AWS* | [link](#)

- Launched a course evaluation web app for Brandeis students to review and share courses, exceeding 200+ active users.

Technical Skills

Programing Languages: Python, Java, C++, Golang, JavaScript, Typescript, HTML/CSS, Groovy, HCL

Frameworks: Pandas, PySpark, D3.JS, FastAPI, Flask, React, Node.JS, Express.JS, Docker, Kubernetes, Kafka

Database and DevOps: Jenkins, Terraform, PostgreSQL, DynamoDB, Pinecone, AWS, Github Actions

Other: Figma, OpenAI API, Excalidraw, Trello, REST