Shaishav Shah

✉ shaishav.shah04@gmail.com in shaishavshah04 ⌂ ShaishavShah04 🔗 Website

Education

University of Alberta   **Sep 2020 – May 2024**
Specialization in Computing Science — Business Minor; GPA: 4.0/4.0   *Edmonton, Canada*

Experience

Palantir   **May 2023 – Aug 2023**
Software Engineer Intern   *New York City, US*

- Led *AIP Assist* development by creating specialized **plugins** and selection logic to leveraging proprietary **LLMs**, supporting multi-step queries, and growing usage by **62%** .
- Developed an internal tool to answer developer queries by performing **semantic search** by leveraging sharded codebase, and passing results as context while optimizing for **token limits**.
- Devised **migration** strategies to further decompose user permissions into smaller operations, allowing support for **granular access control** and view-only modes.
- Optimized Java services with **caching, proxy objects,** and **multi-threading**, reducing the **p99 latency** for orchestration endpoints by over **41%**.
- Engineered performant, lightweight, and **modular pages** using **Typescript** and **React**, leveraging **dynamic imports** to reduce initial load times by **14%** and enhance **user state** rendering.

Niricson Software Inc.   **May 2022 – Aug 2022**
Full Stack Software Engineer Intern   *Victoria, Canada*

- Automated the machine learning workflow in **Python** to dynamically reserve, utilize, and terminate **cloud GPU instances**, saving **$17,000** yearly.
- Introduced a new **geospatial vector data** format for **AWS DocumentDB**, holding over **50%** more data points, while reducing loading time by **80%**.
- Optimized time-complexity of existing and new Javascript algorithms, allowing for users to visualize and customize **500,000+** data points dynamically.
- Designed and created REST API endpoints in **Java Spring Boot**, following **object-oriented** programming and various **software design patterns**.

Projects

Oasis (TreeHacks 2024 Winner)   **Feb 2024**

- Utilized **LangChain** and **vector databases** to build a **RAG** system allowing users to query, compare, and **chat** with large, unstructured ESG documents.
- **Fine tuned LLaMa-2 LLM** to extract metrics and structure data from ESG reports, enabling efficient data indexing, retrieval, and benchmarking across companies.
- Employed robust **text embedding models**, leveraging SentenceTransformers for semantic alignment, supporting side-by-side metric identification.

Meal Drop   **Jan 2022**

- Engineered a **full-stack** application utilizing the **MERN** stack, mitigating food wastage by allowing restaurants to schedule pickups.
- Leveraged **JWTs** to create a **client-side authentication** solution, securing the application and providing **persistent sessions** for users.

Technical Skills

Languages: Java, C, C++, Python, JavaScript / Typescript, SQL, NoSQL, HTML, CSS
Technologies: Spring, ExpressJS, MongoDB, NodeJS, React, Redux, Vue, Conjure
Infrastructure: AWS EC2, SES / SNS, S3, RDS, EFS, Google Cloud, Firebase, Git