

Weifan(Fred) Yu

📍 San Diego, California ✉ weifanyufred@gmail.com ☎ +1 8583665562 in <https://linkedin.com/in/weifanyu>

Education

University of California, San Diego (UCSD)

09/2024 – 06/2026(Expected)

Master of Science in Computer Science GPA: 4.0/4.0

Shenzhen University (SZU)

09/2020 – 06/2024

Bachelor of Electronic Information Engineering GPA: 4.17/4.5 Rank: 1/204

Skills

Languages: Python, Java, JavaScript, SQL, MATLAB

DataBases & Cloud: MySQL, PostgreSQL, AWS EC2, Google Cloud, Microsoft Azure

Web/Mobile Development: HTML/CSS, React, Node.js, Android

Tools & Framework: PyTorch, TensorFlow, Linux, Jina, Git, Docker, LangChain

Work Experience

Software Engineer Intern

07/2023 - 11/2023

Microsoft, STCA

An AIGC-based E-commerce Product Advertising Backdrop Generation Web Application

- Developed a full-stack AIGC-based E-commerce Advertising Web Application from 0-1 for generating visually appealing product backdrops, reducing reliance on costly photoshoots and speeding up advertising workflows.
- Improved the quality of base model (**SD/SDXL**) outputs through **prompt engineering**, achieving a 25% increase in customer satisfaction by aligning generated results more closely with user expectations.
- Designed and deployed a robust two-stage image processing pipeline using **Meta SAM API** for binary mask generation and **SD-WebUI** RESTful API, delivering a boost in processing efficiency and seamless deployment on **Microsoft Azure**.
- Applied **ControlNet** to address edge artifacts, ensure style consistency, and maintain adherence to physical rules, significantly improving generation accuracy and realism by 35%.
- Trained **LoRA** to improve **SD/SDXL**'s performance under in-house scenarios, enhancing models' ability to understand complex 3D scenes.

Selected Projects

Staybooking: An online stay rental application based on React and Spring Boot

11/2024 - 01/2025

- Designed and built a web application using **React**. Bootstrap the development with mature component library **Ant Design**.
- Developed the backend services based on **Spring Boot** to support stay upload, delete, search, and reserve functionality.
- Utilized **PostgreSQL** for data storage, managing user-generated content such as stay information and reservation history, while integrating **Google Cloud Storage** for media file management, ensuring seamless and secure data handling.
- Created geo index by **PostGIS** to support geo-based stay search based on user's selected locations.
- Deployed backend services to **Google Cloud Run**, improving scalability and reliability, and reducing downtime by 15%.

MiniSpotify: A Modern Kotlin-Based Music Favorites Mobile App

10/2024 - 11/2024

- Designed a **Kotlin** Spotify favorite app by fully leveraging **Android Jetpack Library** and **Hilt Dependency Injection**.
- Implemented the BottomBar & App Navigation using **Jetpack Navigation component**.
- Created a mock RESTful API json-server and used the **Retrofit** to handle requests, reducing response times by 25%.
- Built the feed/album/favorite UI in **Jetpack Compose** following **MVVM** architecture.
- Enabled the local cache ability for favorite feature by using **Room** Database, increasing user retention by 10%.
- Integrated the **Google Exoplayer** to handle the global music playback, boosting overall app performance and user satisfaction.

Twitch+: A Personalized Twitch Resources Recommendation Engine

09/2024 - 10/2024

- Designed and developed a full-stack **Spring Boot** application, enabling users to search Twitch resources (streams/videos/clips) and receive personalized recommendations, enhancing user engagement by 25%.
- Created a rich, user-friendly web interface using **React** and **Ant Design**, improving user experience and navigation.
- Implemented RESTful APIs with **OpenFeign** HTTP client to fetch real-time Twitch resources.
- Hosted **MySQL** database on **AWS RDS**, performing efficient **CRUD** operations with **Spring Data JDBC** to manage user and content data.
- Integrated **Spring Security** for user authentication, supporting secure register, login, and logout functionalities.
- Utilized Spring Boot Caching with **Caffeine**, optimizing response times and reducing server load.
- Developed **unit tests** for backend APIs using **JUnit**, ensuring robust error handling and achieving 80% code coverage.

Next AI: Full Stack Web-Based Q&A AI Agent for PDF Document Queries

07/2024 - 09/2024

- Developed an interactive conversational UI using **React** and **Ant Design**, enabling seamless real-time interaction with PDF documents, enhancing user engagement by 30%.
- Implemented high-performance RESTful APIs with **Express** and **Node.js**, optimizing request handling and improving response times by 40%.
- Integrated **OpenAI's GPT-3.5 Turbo** API and **LangChain** technologies, creating a sophisticated AI Agent capable of document loading, splitting, storage, and efficient output, improving information retrieval accuracy by 25%.