

Guoxuan Xu

No Sponsorship Required | (+1) 909-760-9522 | jason2025xu@gmail.com | [Linkedin](#) | [Github](#)

EDUCATION

University of California San Diego — *B.S. Data Science, GPA 4.00 (Provost Honors)*

Sep 2022 – Jun 2026 (Expected)

INDUSTRY EXPERIENCE

Cadre AI (AI Consulting) - Part-Time AI Engineer

July 2025 – Present

Project: Serverless AI Agent Automation Platform

- Developed a **cloud-native, serverless AI agent framework** using Azure Functions, Python, and Blob Storage, enabling document automation with zero infrastructure overhead.
- Designed and deployed 5 modular **AI agents** (Sales Proposal, Preflight, SOP, Compliance, Q&A), reducing doc processing time from **2h → 10m** for **100+** employees with **0%** formatting/compliance errors.

Project: MCP Proxy for Enterprise API Integrations

- Built a **FastAPI-based MCP proxy** enabling ChatGPT to securely call BigQuery/Zoom via **OAuth 2.0**, improving dataset exploration for 20+ internal users.
- Achieved **>99.9% reliability** for long-running ChatGPT sessions through proactive token refresh and request forwarding.
- Streamlined analytics workflows, reducing manual **BigQuery** console navigation by **~70%**.
- Designed an **OpenAPI→MCP** auto-generator (in progress) expected to cut new API onboarding time by **80–90%**.

Project: RAG Admin Portal & User Management System

- Designed and implemented a full-stack **admin portal** for a production RAG system, providing **user management, role-based access control (RBAC)**, and **secure operational governance**.
- Built **end-to-end workflows** for user invitation, token-validated signup, role assignment, and account deactivation, integrating frontend state management with backend APIs.
- Identified and remediated security gaps in exposed authentication endpoints, and improved performance through **client-side caching** and **controlled refresh mechanisms**.

Project: Intelligent Email Automation Platform (Azure)

- Architected a cloud-based, event-driven email automation system using **FastAPI** and **Azure App Service**.
- Implemented **Microsoft Inbox webhook subscriptions** with automatic validation, renewal, and fault-tolerant startup handling.
- Integrated **OCR** and **LLM** pipelines to extract structured data from heterogeneous document formats and automate downstream validation and response drafting.

LLM Strategies (AI startup) - Data Engineering Intern

Dec 2024 – Jun 2025

- Built distributed ingestion services (**Celery + Redis + Docker**) to process **10M+** patent & trademark records across APIs, scrapers, and XML feeds, improving throughput and system reliability.
- Developed **Flask-based** microservices exposing REST APIs for data/task management and built a **high-accuracy document extraction pipeline** (USPTO API + OCR + Gemini), achieving **99.2%** accuracy on 500+ legal documents.
- Designed a **SigLIP + Qdrant vector search engine** enabling sub-second image similarity retrieval, and reduced redundant metadata by **85%** through algorithmic normalization.

The Wu lab, Scripps Research - Software Engineering Intern

Nov 2024 – Feb 2025

- Built **FastAPI** services and memory-efficient Python parsers to normalize large-scale genomic datasets (**PostgreSQL + S3**).

PROJECT EXPERIENCE

Trading Capture System – Execution Module - Group Project

Jul 2025 – Present

- Developed core modules of a **distributed trading system** using microservices (Java/Spring Boot) and React, enabling secure execution management and real-time communication across services.
- Implemented **FIX-driven** REST APIs and **Kafka-based** asynchronous reporting, ensuring reliable event-driven message flow and improving system responsiveness in distributed environments.
- Designed and optimized **MySQL** schemas and integrated **Redis** caching for token/session management, improving query latency and authentication efficiency.
- Built **React + Redux** interfaces for creating, updating, and tracking execution records, and implemented validation logic with field-level constraints to ensure data integrity and smooth user experience.

Climate Emulation with Deep Learning - Kaggle Competition | Top 10 (9/85)

May 2025 – Jun 2025

RESEARCH EXPERIENCE

Student Researcher - Data Science & Learning Lab @ UCSD

Mar 2024 – Jul 2024

- Enhanced the **Pandas Tutor visualization tool** by improving DataFrame introspection for selective rendering of large datasets—reducing memory usage and improving in-browser interactivity. Published this work as **first co-author** at ACM SIGCSE (paper: <https://dl.acm.org/doi/10.1145/3641554.3701933>).

SKILLS & TOOLS

Languages & Frameworks: Java, Python, SQL/PostgreSQL, JavaScript, R, Shell, PyTorch, scikit-learn, React

Tools & Platforms: Git, Docker, Kubernetes, Redis, MySQL, Kafka, RESTful APIs, Azure, Flask, FastAPI