Jason (Guoxuan) Xu

No Sponsorship Required | (+1) 909-760-9522 | guoxuan.xu8@gmail.com | Linkedin | Github

EDUCATION

University of California San Diego   **La Jolla, CA**
BS. Data Science *(GPA: 4.00, Provost Honors)*   *Sep 2022 – Jun 2026 (Expected)*

INDUSTRY EXPERIENCE

Cadre AI (AI Consulting) - *AI Engineering Intern*   *July 2025 – Aug 2025*
- **Developed a cloud-native, serverless AI agent framework** leveraging **Azure Functions, Python, and Blob Storage**, enabling traditional firms to automate document handling and storage without maintaining infrastructure.
- **Designed and deployed 5 production-ready AI agents** as modular pipelines, cutting document processing time from **2 hours to 10 minutes** for 100+ employees while ensuring **0% error rate** and compliance with formatting standards.
- **Engineered reusable pipelines** with **HTTP triggers & serverless orchestration**, reducing AI agent development and deployment cycles from weeks to hours across **4 enterprise clients**.
- **Led end-to-end system design and implementation**, collaborating with the AI manager to align requirements, enforce compliance, and ensure seamless adoption across client workflows.

LLM Strategies (AI startup) - *Data Engineering Intern*   *Dec 2024 – Jun 2025*
- **Designed and implemented distributed pipelines** with **Celery** + **Docker**, enabling scalable ingestion of millions of patent and trademark records from **REST APIs**, **HTML scrapers**, and **XML feeds**.
- **Built production-ready services** for high-accuracy document image extraction (USPTO API + OCR fallback + Gemini API fallback), achieving **99.2% accuracy** across 500+ multi-format court documents.
- **Optimized system performance** by implementing streaming XML parsing (*lxml.iterparse*) for weekly 10k+ records and reducing redundant metadata by **85%** (2.5M → 362k) through algorithmic cleaning and normalization.
- **Developed a full-stack data platfor**m by integrating ETL pipelines with a **PostgreSQL → Data Warehouse backend**, exposing analytics through **Flask APIs**, and building **React dashboards** for real-time trend visualization.

The Wu lab, Scripps Research - *Software Engineering Intern*   *Nov 2024 – Feb 2025*
- **Developed a FastAPI service** to expose genomic datasets as RESTful APIs, integrating **PostgreSQL** for structured metadata and **S3 storage** for raw files, enabling global researchers to query standardized biological data.
- **Engineered high-performance Python parsers** with **lxml** and generator-based pipelines to normalize raw genomic datasets, applying rule-based validation to ensure data quality while preventing memory bottlenecks on large files.

PROJECT EXPERIENCE

Trading Capture System – Execution Module - *Group Project*   Jul 2025 – Present
- **Developed the execution module** of a distributed trading system with a **microservices backend** and **React frontend**, enabling secure execution management and real-time communication.
- **Implemented RESTful APIs** to process **FIX messages**, ensuring reliable event-driven communication across distributed services.
- **Designed and optimized MySQL schemas**, and integrated **Redis caching** for token management (logout & refresh), improving query performance and authentication efficiency.
- **Built frontend features** with **React + Redux** to create, update, and delete execution records, providing real-time execution tracking and improved user experience.
- **Implemented execution validation logic** by enforcing field constraints and dependency checks, ensuring data integrity.
- **Engineered Kafka-based asynchronous reporting**, enhancing system responsiveness and reliability in a distributed environment.

Climate Emulation with Deep Learning - *Kaggle Competition*   May 2025 – Jun 2025
- Engineered a **U-Net model with CoordConv2d and Residual Layers** in PyTorch, achieving **Top 10 (9/85)** in forecasting surface temperature and precipitation.

RESEARCH EXPERIENCE

Student Researcher - *Data Science & Learning Lab @ UCSD*   *Mar 2024 – Jul 2024*
- Enhanced the **Pandas Tutor visualization tool** by improving DataFrame introspection logic, enabling selective rendering of large datasets to reduce memory usage and speed up interactivity in-browser.
- Published as **first co-author** at **ACM SIGCSE**, analyzing data analysis behaviors while contributing to the design and optimization of large-scale visualization features.

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