Jiaxiong (Jason) Guan

(929)-453-2255 | jasoguan10@gmail.com | linkedin.com/in/jiaxiong-guan/ | github.com/Jguan10

Education

Hunter College, City University of New York Aug 2022 – Dec 2025
B.A. *- Computer Science, Focus in Human Biology —* ***GPA****: 3.46*

Technical Skills

Languages: Python, C++, JavaScript, Java, SQL
Libraries: TensorFlow, PyTorch, Pandas, Scikit-Learn, Selenium, Seaborn, Flask, React, Next.js
Tools: Git, WordPress, Tableau, Jupyter Notebook, BigQuery, AWS, Hugging Face
Databases: PostgreSQL, MongoDB, Neo4j, Supabase

Experience

Data Science Intern Jun 2025 – Aug 2025
Memorial Sloan Kettering Cancer Center

- Fine-tuned Named Entity Recognition (**NER**) models to identify and classify clinical documents, achieving **95%** F1-score and outperforming an industry-ready benchmark model with 91% F1-score
- Applied **weak supervision** through broad LLM-generated labels to construct test datasets and enhance training data quality, resulting in an overall **22%** more accurate dataset
- Collaborated with Database Admin team to integrate ML models with new Electronic Health Record system

Data Science Fellow Jul 2024 – May 2025
CUNY Tech Prep

- Selected for a competitive data science fellowship with students from across the 11 CUNY senior colleges where Fellows create technical projects using tools such as **PyTorch**, **TensorFlow**, **SQL**, **Sci-kit learn**, and **Pandas**
- Participated in weekly courses and learned industry best practices for exploratory data analysis (EDA), feature engineering, data collection and processing, statistical modeling, data visualization, machine learning techniques, data science process, and big data

Technical Operations Intern May 2024 – Aug 2024
The Bee Conservancy

- Leveraged Google Analytics and Hotjar to monitor website performance and user flow, implementing insights that boosted web pages to the **top 2 SEO** rankings for targeted keywords
- Designed and developed community-focused educational web pages, resulting in **20%** increased site traffic
- Automated processes within CRM SQL database with Python scripts to reduce manual workload

Sustainability Market Analyst Intern Mar 2024 – Apr 2024
New York City Economic Development Corporation

- Spearheaded a comprehensive market research initiative focusing on the life sciences sector in NYC, analyzing industry trends and projections to guide strategic investment decisions
- Developed and presented strategic programs aimed at enhancing growth and competitiveness of NYC's life sciences ecosystem, with a focus in promoting sustainability and equity

Microbiology Research Aug 2021 – May 2022
Binghamton University

- Created research proposal and experimental design for decreasing antibiotic resistance in Vancomycin-resistant E. Faecalis that was approved by Binghamton University
- Performed literary research on a vaccine-preventative approach to microbial biofilms and presented results at a university-wide poster session

Projects

The Lounge - MSKCC Hackathon 1st Place | *Claude, AWS, LangChain* August 2025

- Developed an **AWS-based Agentic AI** social media app with an efficient RAG retrieval pipeline using **OpenSearch**, **Amazon S3**, and **AWS Lambda** to create a scalable and secure solution for young adults
- Features two key agents through **Bedrock Claude**, directing users to requested resources and helping schedule appointments along with an SMS confirmation through **Pinpoint**
- Leveraged **SageMaker** to perform safety checks, redacting sensitive information, response evaluation, and merge final answer with source tags and action chips

EZ-RX-ID | *PyTorch, DeepSeek, LangChain, BGE Embeddings, Computer Vision, Supabase* Feb 2025 - June 2025

- Full stack AI application that identifies prescription pills from images and generates medical summaries from queries using an **Agentic Retrieval-Augmented Generation** (**RAG**) System with **DeepSeek**
- Trained multiple **CNNs** to extract pill attributes (shape, color, imprint) and combined their outputs using an **XGBoost** classifier, increasing top-1 accuracy by **30%**
- Built a recursive agent workflow to evaluate responses, retrieve additional data, and regenerate outputs as needed, ensuring accurate responses and reducing hallucinations by **15%**