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, NLTK
Tools: Git, WordPress, Tableau, Jupyter Notebook, BigQuery, AWS, Hugging Face, Docker
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
- Built recipe recommender system, leveraging **K-Nearest Neighbors** and **Approximate Nearest Neighbors** algorithms to search and match within **FAISS** vector database based on user-inputted description and ingredients
- Cleaned dataset of over 500,000 recipes pulled from Food.com through NLP techniques, such as **tokenization** and **lemmatization**, standardizing the data for cosine similarity matches and increasing database lookup speeds
- Engineered full-stack computer vision AI workout tracker, utilizing **TensorFlow** and **MediaPipe** model to detect repetitions and a **ChatGPT**-based coach to provide real-time feedback on workout form

Technical Operations Intern   May 2024 – Aug 2024
The Bee Conservancy
- Automated SQL database updates with Python **ETL** pipeline, increasing efficiency and reducing manual workload
- Launched survey application with **React Native** and **SQLite** to collect event attendee feedback and sentiment in offline and low-bandwidth environments
- Designed community-focused educational pages based on detailed analysis of search & user data from HotJar and Google Analytics, resulting in **20%** increased site traffic

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, LangChain, BGE Embeddings, Computer Vision, Supabase, REST API*   Feb 2025 - Jun 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**
- Designed modular **ETL** pipelines to preprocess pill images and metadata for ML training and structured storage in **Supabase** to support scalable querying and downstream applications
- Trained multiple **ResNet-18** models to extract pill attributes (shape, color, imprint) and combined their outputs using an **XGBoost** classifier, resulting in a top-5 accuracy of **93%**
- Built a recursive agent workflow to evaluate responses, retrieve additional data, and regenerate outputs as needed, increasing accurate responses by **15%**