

# Jiaxiong (Jason) Guan

(929)-453-2255 | [jasoguan10@gmail.com](mailto:jasoguan10@gmail.com) | [linkedin.com/in/jiaxiong-guan/](https://www.linkedin.com/in/jiaxiong-guan/) | [github.com/Jguan10](https://github.com/Jguan10)

## EDUCATION

**Hunter College, City University of New York**

Aug 2022 – Dec 2025

*B.A. - Computer Science, Focus in Human Biology — GPA: 3.4*

## TECHNICAL SKILLS

**Languages:** Python, C++, JavaScript, Java, SQL

**Tools:** TensorFlow, Flask, Git, WordPress, Tableau, PyTorch, HuggingFace Models, Jupyter Notebook

**Databases:** PostgreSQL, MongoDB, Neo4j, Supabase

**Libraries:** Pandas, NumPy, Matplotlib, Seaborn, Scikit-Learn, Selenium

## EXPERIENCE

**Data Science Intern**

Jun 2025 – Present

*Memorial Sloan Kettering Cancer Center*

- Fine tuning Named Entity Recognition (**NER**) models to de-identify hospital records and ensure patient privacy
- Applying **weak supervision** to develop test datasets and strengthen training data quality
- Exploring and evaluating **image segmentation** models on radiology scans and pathology slides
- Collaborating with Database Administration team to integrate new models into current systems

**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*

- Managed customer relations database, ensuring accurate input of partner and donor information to track the distribution of grants and funding totaling hundreds of thousands of dollars
- 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
- Built survey application with **React Native** and **SQLite** to collect event attendee feedback and sentiment in offline and low-bandwidth environments

## PROJECTS

**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
- Trained multiple **CNNs** to extract pill attributes (shape, color, imprint) and combined their outputs using an **XGBoost** classifier, increasing top-1 accuracy by **30%**
- Designed a **QA RAG** system using **DeepSeek**, optimizing context retrieval to improve domain-specific reasoning
- Built a recursive agent workflow to evaluate responses, retrieve additional data, and regenerate outputs as needed, ensuring accurate responses and reducing hallucinations by **15%**

**Dishcovery** | *FAISS, Sentence Transformers, NLTK*

Sep 2024 - Dec 2024

- Collaborated with a team to design and implement a personalized recipe recommendation system utilizing user preferences such as cuisine, taste, and desired ingredients
- Leveraged an ensemble strategy combining **k-Nearest Neighbors** and **Approximate Nearest Neighbors**, trained on a dataset of 500,000 recipes
- Applied **NLP** through **Sentence Transformers** and **NLTK** to process and generate embeddings, allowing for near-instant querying with vector database