

# Chenyu (Fred) Zhang

(667) 213-8216 | zhangchenyu555@gmail.com | <https://fredzcy.github.io> | New York, NY

## EDUCATION

---

Johns Hopkins University | Baltimore, MD

**M.S.E. in Data Science** | GPA: 3.73/4.0 | Dec 2022

Xiamen University | Sepang, Malaysia

**B.E. in Computer Science and Technology** (Honors) | GPA: 3.89/4.0 Rank: 1/61 | Aug 2021

## WORK EXPERIENCE

---

**Data Scientist** | Morgan Stanley

New York, USA | Feb 2023 – Present

- To be filled...
- To be filled...

**Data Scientist Summer Intern** | Morgan Stanley

New York, USA | Jun 2022 – Aug 2022

- Established data extraction pipeline for market risk dashboard using Tableau and helped maintain intranet Appstore page for 100+ existing dashboards.
- Increased company issue classification accuracy to 90% by proposing multimodal transformer trained on 50+ features and deployed the classifier to the audit department of 500+ employees.
- Collaborated with 4 analysts to conduct research project on electronic trading and presented to department.

**Data Scientist Intern** | Xiamen Small Bee Big Data Technology Co., Ltd.

Xiamen, China | Sep 2020 – Jan 2021

- Created human action dataset of 1000+ videos and developed measurement of the concentration of students.
- Improved accuracy of concentration analysis by 20% by applying pre-trained deep-learning based pose estimation models (OpenPose, AlphaPose & FSA-Net) on preprocessed data.
- Proposed strategies to improve students' concentration and demonstrated to clients (lecturers) based on results.

## RESEARCH EXPERIENCE

---

**Affiliation Name Disambiguation** | IEEE - Johns Hopkins University

Baltimore, USA | May 2022 – Aug 2022

- Extracted affiliation names and corresponding locations from AWS Redshift Database of over 700,000,000 IEEE publication records using keywords and 4 APIs (MediaWiki API, Google Places API, Geopy API & ROR API).
- Achieved disambiguation accuracy of 89% by performing the agglomerative hierarchical clustering based on cosine similarity of BERT embeddings and geographical distance.
- Improved the pipeline's efficiency by 90% using multiprocessing on AWS EC2 instance with 48 CPUs.

**Image Segmentation and Volume Estimation of Food Object** | Xiamen University Malaysia

Selangor, Malaysia | Apr 2019 – Jan 2020

- Applied industrial robot arm to mount object, procured dataset of 600+ images and performed data preprocessing.
- Implemented instance segmentation to extract object using U-Net & Mask R-CNN and obtained physical features (e.g. diameter of cross-section) using statistical formulas.
- Achieved model accuracy of 95% on predicting volume of 16 classes of ham with replaced outliers.

**Chatbot for Stock & COVID-19 Information Queries** | Xiamen University Malaysia

Selangor, Malaysia | Dec 2018 – Jul 2019

- Generated dataset of 500+ examples on stock and COVID-19 inquiries based on regular expression.
- Constructed intent classifier and entity extractor with Rasa NLU tools.
- Developed multi-turn conversation with Finite-state Machine by retrieving real-time data from two external data sources (IEX Cloud & JHU COVID-19 Dashboard) and deployed chatbot on WeChat with wxpy.

## SKILLS

---

**Programming Languages:** Python, R, SQL, C/C++, HTML, MATLAB

**Frameworks & Tools:** NumPy, Pandas, Matplotlib, Sklearn, Flask, Tensorflow, Keras, PyTorch, Anaconda, CUDA, MySQL, SQLite, MongoDB, Spark, Hadoop, Tableau, Linux, GCP, AWS, RESTful API

**Languages:** English (Fluent), Mandarin (Native)