

Melinda Fang

Cell: (+1) 607 262 5689

<https://github.com/MelindaFang-code>

Email: cf348@cornell.edu

EDUCATION

Cornell University, College of Engineering, Ithaca, NY

Expected May 2024

Bachelor of Science, Computer Science

GPA: 4.07, Dean's List (all semesters)

Relevant Courses: OOP & Data Structures; Probability; Statistics; Reinforcement Learning; Natural Language Processing; Deep Learning; Operating Systems; Cloud Computing; Database; Operation Research Tools for Financial Engineering

SKILLS

Programs: Python, Java, SQL, JavaScript, C, PyTorch, NodeJS, React, .Net, C#, Azure, AWS, Flask, docker, OCAML

WORK EXPERIENCE

Citadel Securities, *Quant Developer Intern*

Jun 2023-Aug 2023

- Re-architecture the pricing tool to apply volatility assertion, reduced 70% latency and improve production stability
- Implemented a lightweight volatility curve library, enable users to instantiate curves and tweak surface more flexibly
- Redesign message protocol to enable backwards compatibility and fasten deploy of new pricing models for trader
- Implemented WebSocket library, support WebSocket & HTTP connection, JSON & bond binary encoding format

Stripe, *Software Development Engineer Intern*

Jan 2023-Apr 2023

- Built workflows and dashboards to migrate petabytes of data from 1k+ replica sets to upgrade & rebalance shards
- Orchestrated data migration using Temporal Workflow; crafted MongoDB schema, Protobuf, and gRPC API
- Set up SignalFx detectors in Terraform to trigger Puppet alerts and Splunk Dashboard for infrastructure monitoring

Meta, *Data Engineer Intern*

Jun 2022-Aug 2022

- Built a centralized scalable data framework for FAIR research, enabling fast query, automated testing, and debugging
- Built an end-to-end data validation CLI to optimize data ingestion of Ego4d, largest egocentric video dataset
- Developed a 5-stage data pipeline that automates incorporation of 1 million data points from 8 work streams

Microsoft, *Software Development Engineer Intern*

Jun 2021-Aug 2021

- Invented a scheduling algorithm to arrange engineers' on-call rotations with perfect fairness if attainable
- Developed a web application for engineers to track roles and tasks and auto generate weekly summary emails
- Built REST microservices using C# and .Net Core for backend, React for frontend, and Azure pipeline for CI/CD

PROJECT EXPERIENCE

MyCourseIndex, Cornell University, *Lead Machine Learning Engineer*

Sept 2020-July 2022

- Developed Question-Answering Restful APIs for a search platform to query course materials using **Flask** and **PyTorch**
- Deployed an end-to-end NLP model doing passage retrieval, answerability detection and answer selection
- Achieved overall 0.7 precision by experimenting with Transformers vs Elasticsearch and their ensemble
- Implemented backend on AWS RDS database; Improved memory usage and query speed using FAISS document store

Third Place of Citadel DataOpen (global datathon with 2k+ applicants)

July 2022-July 2022

- Led 3 people to analyze social disclosure's effect on loan funding and repayment using LendingClub loan descriptions
- Applied sentence-encoder to extract text features, PCA to reduce feature dimension, K-Means to cluster similar descriptions, class-based TF-IDF to identify keywords and phrase patterns and infer loan applicants' characteristics
- **Panel regress** word patterns on interest rate to analyze impacts; built default prediction model by fine-tuning BERT

Winner of ProjectX 2021 (global ML research competition with 1k+ participants)

Sept 2021-Feb 2022

- Researched the spread of false covid tweets and developed a warning system, accepted by the UoT AI Conference
- Classified misinformed tweets by Bi-LSTM; invented virality metrics and produced virality estimation based on BERT
- Improved results by text preprocessing using NLTK, data augmentation, and bagging, achieving 0.91 F1 score

International Mathematics Contest in Modeling (top 5 %), *Research lead*

Feb 2021-Feb. 2021

- Led team of 3 to predict distribution of hornets and provide quantitative guidelines for pest control
- Extracted image feature by pretrained VGG16 and classified hornets by training an RBF-SVM with 0.97 precision
- Applied logistic regression on SVM output to produce a probabilistic estimate of misclassification with precision of 1
- Fit an SVM on geolocation and temporal data combined with sliding window to predict distribution with 0.88 precision