

Zhengyang (Jason) Qi

+1 (404) 797-9976 | zqi2@andrew.cmu.edu | <https://github.com/Jasonqi146>

EDUCATION

Carnegie Mellon University <i>Master of Science in Intelligent Information Systems</i>	Pittsburgh, PA August 2022 - December 2023
Georgia Institute of Technology (Georgia Tech) <i>B.S., Computer Science</i>	Atlanta, GA January 2020 - May 2022
Emory University <i>B.A., Mathematics & Economics</i>	Atlanta, GA August 2017 - December 2019

RESEARCH EXPERIENCES

Georgia Tech - Visualizing Climate Change Discourses <i>Research Assistant</i> Dr. John Stasko, Dr. Diyi Yang	Atlanta, GA March 2022 - Present
<ul style="list-style-type: none">• Collected and organized data into tree-like discourse structure through climate-related Reddit channels.• Generate and fetch supportiveness measurement, political stance, and user demographics through deep sentiment analysis and web scraping tools.	
Georgia Tech - Financial Question Generation for Earnings Conference Calls <i>Research Assistant</i> Dr. Diyi Yang	Atlanta, GA August 2020 - December 2021
<ul style="list-style-type: none">• Formulated a dataset for question generating through matching questions to contexts by similarity metrics on name entities, LDA topic modeling, and BERT embeddings tuned on earning call texts.• Researched and trained classifiers for question intentions (F1 = 87%) to enable conditional question generation.• Cleaned question corpus leveraging GPT-3 to improve transformer models' performance on the dataset.	
Georgia Tech - Spatiotemporal Modeling of Covid-19 <i>Research Assistant</i> Dr. Paula Gómez	Atlanta, GA August - December 2020
<ul style="list-style-type: none">• Researched on improving agent-based simulation models for the airborne transmission of Covid-19.• Designed and programmed 3 high school egress simulations with varied egress strategies by Rhino and PedSim.• Conducted controlled experiments for the effects of space, airflow, and social distancing on the infection risk rate of Covid-19, gaining analytical and modeling skills.	
Zhejiang University - Graph Fourier Analysis for Learnability of Graphs <i>Research Assistant</i> Dr. Yang Yang	Atlanta, GA June - August 2019
<ul style="list-style-type: none">• Learned graph theory application and network embedding to the convolutional neural network (CNN).• Tested the association between CNN's convergence behavior and graph Fourier analysis on the Laplacian matrix for social network graphs.	

WORK EXPERIENCES

Intuit Inc. <i>Artificial Intelligence Engineer Intern</i>	Mountain View, CA July - August 2022
<ul style="list-style-type: none">• Researched and constructed a deep image captioning model to generate alt-texts for visual disabilities.• Trained a transformer-based model with CLIP and GPT2 on MSCOCO dataset with AWS Sage Maker.• Deployed the model to Intuit ML platform with docker prediction tools and KubeFlow training pipelines.	
EPI-USE America, Inc. <i>Software Engineer Intern</i>	Atlanta, GA June - July 2021
<ul style="list-style-type: none">• Implemented a graphical interface for interactive visualization of information networks for ADP.• Built a BERT classifier for privacy leakage detection and graph algorithms for clustering and flow optimization.• Assembled Node.js user interface, Neo4j graph database, Splunk query, and Flair backend into a docker app.	
Kiwi Smart Shopping Assistant <i>CTO & Machine Learning Engineer</i> Dr. Chu Xu	Atlanta, GA December 2020 - August 2021
<ul style="list-style-type: none">• Lead a team of six to develop Kiwi, a browser extension for cross-website price comparison on Google Chrome.	

Zhengyang (Jason) Qi continue

- Trained ensemble models through 3 iterations of data collection, feature engineering, and model tuning, attaining accuracy of 89%, acquiring industry machine learning (ML) experience.
- Delivered AWS Lambda back-end with ML algorithms, automatic web scraping, and database update.

Huawei Technologies Co., Ltd.

Shandong, China

Project Manager Intern

May - July 2018

- Coordinated manufacturers, conveyances, and contractors for base stations' installation in 17 regions.
- Collected and reflected regional progress through visualized daily reports and presentations.

PROJECTS

American Express Default Prediction - Kaggle

July - August 2022

Ranking: 152/4935 (Silver)

- Cleaned and downsized the large dataset and conducted feature engineering with aggregation and lag features.
- Trained boosting models with 10-fold cross validations including LightGBM, CatBoost, and XGBoost.
- Improve model performances by parameter tuning with Optuna and ensemble with out-of-fold predictions.

U.S. Patent Phrase to Phrase Matching - Kaggle

May - July 2022

Ranking: 313/1889

- Constructed and trained DeBERTa, PatentBERT, Electra, and RoBERTa models for classification and regression tasks to assign matching scores for concatenated phrases (PCC = 0.8593).
- Improve the model performance by 5-fold cross validation and multi-version ensemble.

Natural Language Processing - Georgia Tech

August 2021 - December 2021

Style Transfer between Standard American English (SAE) and African American Vernacular English (AAVE)

- Collected a parallel dataset of SAE and AAVE and experimented with encoder-decoder models for generation, including LSTM, GPT-2, BART, and T5 on the dataset, achieving the best BLEU of 0.351.
- Improve model performances by adding a paraphrasing state-of-the-art model as the encoder.

SKILLS

Programming Languages: Python (proficient), Java (proficient), JavaScript (medium), C++ (medium), C

Machine Learning: PyTorch (proficient), Scikit-Learn (proficient), HuggingFace Transformers (proficient), Keras (medium), Gensim, SpaCy, NLTK, OpenCV, Pandas, NumPy, SciPy

Web Data Extraction: Selenium (proficient), BeautifulSoup (proficient), Scrapy

Cloud Services: AWS Sage Maker (proficient), AWS Lambda (proficient), AWS S3 (proficient), Neo4j Graph Database (proficient), Kubeflow (medium), Splunk

Software Development: Node.js (intermediate), Docker (proficient), shell (proficient), Flair, Ubuntu

Analytics & Visualization: R (proficient), Stata (proficient), MATLAB (proficient), Matplotlib, PowerBI