Ian Wu

ianwu99@outlook.com | ianwu13.github.io | linkedin.com/in/ianwu13 | github.com/ianwu13

EDUCATION

University of Southern California **Aug. 2022-May 2024**

Master of Science in Applied Data Science GPA: 4.0/4.0

- Coursework: Advanced NLP, Data Mining and Recommendation Systems, Foundations of Data Management, Machine Learning for Data Scientists, Machine Learning for Medical Data, Probability & Statistics for Data Scientists

Cleveland State University **Aug. 2018-May 2022**

Bachelor of Science in Computer Science, Minor in Mathematics GPA: 3.84/4.0

- Jack, Joseph, & Morton Mandel Honors College graduate with scholarships totalling 100% of tuition
- Coursework: Artificial Intelligence, Big Data, Data Structures & Algorithms, Database Concepts, Deep Learning, Discrete Mathematics, Linear Algebra, Internet Programming, Multivariate Calculus

EXPERIENCE

Course Producer (TA) **Jan. 2023-Present**

University of Southern California *Los Angeles, CA*

- Assisting in course production for over 400 students enrolled in DSCI 552: "Machine Learning for Data Scientists"
- Holding weekly office hours to provide additional instruction on topics such as linear & logistic regression, SVMs, KNN, decision trees, ensemble methods, hidden markov models, etc.

Student Worker (Machine Learning Research Intern) **May 2023-Aug. 2023**

USC Institute for Creative Technologies (ICT) *Los Angeles, CA*

- Researched methods for augmentation of LLMs for multi-issue negotiation; Created a custom code base to evaluate the negotiation capabilities of custom agents in both agent-agent and human-agent interactions
- Utilized PyTorch to train GRU/LSTM models with reinforcement learning for strategy planning in negotiations and to parallelize LLaMA across multiple nodes/GPUs on the USC High-Performance Computing Cluster

Data Science Intern **May 2022-Aug. 2022**

AI Camp *Palo Alto, CA*

- Mentored students in data science and machine learning, overseeing groups of 4-6 students in creating original projects
- Aided students in applying data analysis/visualization techniques and machine learning models such as linear & logistic regression, SVMs, neural nets/MLPs, etc, to a diverse set of datasets

Research Student **Jan. 2022-May 2022**

Cleveland Clinic Lerner Research Institute *Cleveland, OH*

- Developed an algorithmic CT scan augmentation process using Python and OpenCV to improve the performance of a domain-adapted MRI segmentation model (DenseNet)
- Tested algorithm using a dataset of 176, 3D DESS (double-echo steady-state) images collected from 16 patients at multiple hospitals; Improved Jaccard/IoU (intersection over union) score from 0.27 to 0.78

Engineering Peer Teacher (TA) **Jan. 2021-May 2022**

Cleveland State University *Cleveland, OH*

- Organized and led weekly study and review sessions and developing additional course materials to aid students
- Led lectures for "Introduction to Programming" and "Data Structures & Algorithms" at the professors' requests
- Awarded "EPT of the Month" for excellent student feedback and going above and beyond normal responsibilities

Software Engineering Intern (Machine Learning) **Jun. 2021-Aug. 2021**

Tailwind *Oklahoma City, OK*

- Created an image background removal tool which was used to add a custom background replacement feature to the company's flagship product
- Implemented the U2-Net model to perform foreground segmentation using TensorFlow
- Implemented an image background removal API using Amazon Web Services (AWS) Lambda functions in both TypeScript and Python

IT Intern **Jun. 2020-Jul. 2020**

Dominion Energy *Cayce, SC*

PROJECTS

Whole Slide Image Cancer Classifier **Jan. 2023-May 2023**

- Developed a novel method for cancer classification in WSIs using multiple instance learning and graph convolutional networks
- Achieved an improvement of 10% in recall (0.86-0.96) and 4% in accuracy (0.91-0.95) for cancer detection compared to state-of-the-art methods

Restaurant Recommender System **Apr. 2023-May 2023**

- Developed a hybrid (feature combination) recommendation model using the Yelp Dataset to predict a user's rating for a given resturaunt
- Trained XGBoost, SVD, Co-clustering, KNN, and item-based collaborative filtering models; Implemented the item-based collaborative filtering model from scratch in Apache Spark (PySpark)

Negotiation Agent with Custom Transformer Models **Aug. 2022-May 2023**

- Led research project with 3 grad students, designed and trained transformer models for negotiation prediction
- Developed custom transformer-based model using Hugging Face library; achieved an F1-BERTScore of 0.88 and a BLEU-2 Score of 0.21 in dialogue response prediction for the CaSiNo negotiation dataset

Reci-Pic Kitchen Assistant **May 2022-Aug. 2022**

- Collaborated on creation of AI-driven kitchen assistant app for ingredient recognition
- Trained YOLOv5 model with 73% accuracy to recognize 20 common kitchen ingredients
- Coded a cross platform (IOS and Android) application using React Native

Cartoons vs. Real Life Image Classifier **Jan. 2022-May 2022**

- Designed and implemented a novel deep learning model based on Feature Pyramid Networks with TensorFlow
- Collected an original dataset of over 16k cartoon and photographic images; Achieved 91% classification accuracy

Congressional Bill Sentiment Analysis System **Aug. 2021–May 2022**

- Created an automated system to gauge public sentiment on congressional bills using Twitter data
- Collected and stored over 14 million relevant tweets from the Twitter API using Python and MariaDB
- Utilized the Hugging Face library to train BERT model for sentiment analysis (classify tweet as positive, negative, or neutral), achieving 62% accuracy; Leveraged distant supervision to improve results to 78%
- Aggregated sentiment scores to generate quantitative metrics for public opinion on different congressional bills by population segment and other criteria (verified/unverified users, like-to-comment ratio, etc.)

OU Well-MB **Apr. 2020-Aug. 2020**

- Co-wrote a grant proposal for an emotional wellness support application, OU Well-MB, to help healthcare workers and educators at The University of Oklahoma (OU) during the Covid-19 pandemic.
- Developed a prototype for the application using the React.JS library and Node.JS runtime environment

TECHNICAL SKILLS

Languages: Python, JavaScript/TypeScript, Java
Libraries/Packages: Deep Graph Library (DGL), Hugging Face (Accelerate, Transformers, Datasets), Keras, Matplotlib, NLTK, Node.JS, NumPy, OpenAI, OpenCV, Pandas, Pickle, PyTorch, React, Scikit-Learn, SciPy, TensorFlow, XGBoost
Web Development Flask, HTML/CSS, Node.js, React.js
Databases Firebase, MongoDB, SQL (MySQL, PostgreSQL, SSMS)
Other: AWS (EC2, Lambda, S3), Azure, Conda, Docker (exposure), Git/GitHub, Hadoop, Jupyter Notebooks/Google Colab, LaTeX, Slurm, Spark

LEADERSHIP AND INVOLVEMENT

Society of Asian Scientists and Engineers (SASE) **Jan. 2021-May 2022**

- Established a SASE chapter at Cleveland State University; Elected as Vice President
- Coordinated with fellow officers and representatives from sponsor company, FirstEnergy, to organize meetings and networking opportunities for members

NCAA Division 1 Fencing Team **Aug. 2018-May 2022**

- Managed a full-time course load while training 20 hours per week with the Cleveland State Varsity Fencing Team
- Qualified for and competed in NCAA Regional Championships in all 4 years