**Chenfeng Huang**

Los Angeles, CA | (415) 341‑2037 | [chengfengeric@gmail.com](mailto:chengfengeric@gmail.com) | chenfenghuang.info

**education**

**University of Southern California** January 2023-December 2024

**Master of Applied Data Science**

* GPA: 4.0/4.0

**Santa Clara University** September 2020-June 2022

**Bachelor of Mathematics and Computer Science**

* GPA: 3.5/4.0

**experience**

**Research Assistance** February 2023-Present

University of Southern California, Los Angels, CA

* Performed data cartography of 2.4M interaction steps on Calvin language-conditioned policy learning benchmark (CALVIN), especially language annotations associated with synthetic robotic arm trajectories dataset, using imitation learning baseline model.
* Conducted data cleaning on CALVIN in terms of results of data cartography information to achieve 10 percent better success rate with 30 percent less dataset size, 200 hours less training time, and 50 percent less GPU memory
* Leveraged GPT prompting to enhanced dataset language annotation quality, resulting in a 5% improvement in success rate
* Optimized batch size, narrowed training set to language-annotated portions, and leveraged multi-GPU parallel computing capabilities within Slurm cluster; reduced training time from 30 to 1 hour per epoch

**Research Assistance** April 2021-September 2021

Massachusetts Institute of Technology, Cambridge, MA

* Applied transfer learning on a ResNet50 model, selectively freezing initial layers. Later unfrozen and refined at a decreased learning rate, achieving specific adaptability for breast cancer diagnosis with a 94.7% accuracy
* Engineered an advanced chatbot leveraging Natural Language Toolkit (NLTK) capabilities, acclaimed by over 2,000 users for its intuitive communication interface, saving more than 10K US dollars each year for diagnosis
* Synthesized technical details and achievements in academic prose, culminating in a role as a core author for a paper at 2021 2nd International Seminar on Artificial Intelligence, Networking, and Information Technology (AINIT 2021)

**Machine Learning Engineer Intern** May 2020-October 2020

Shuzhilian Co, Sichuan, China

* Utilized a pre-trained BERT model for Named Entity Recognition (NER) and Relation Extraction tasks, specifically fine-tuning model on legal texts to precisely identify entities and determine relationships over more 3000 cases each month
* Employed BERT embeddings from extracted entities for semantic similarity, streamlining historical case indexing. The optimization reduced manual effort and saved over 50K in labor costs every month

**publication**

["An interactive prediction system of breast cancer based on ResNet50, chatbot, and PyQt"](https://ieeexplore.ieee.org/document/9725102)

* Co-first author, 2021 2nd International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT 2021)

**projects**

[**Firebase Emulation with Citi Bike Station Query Platform**](https://github.com/91MrHuang/firebase_emulation_citibike) January 2023-April 2023

* Replicated Firebase RESTful API using Flask. Established synchronization of JSON data from MongoDB database for data consistency
* Constructed a Shell command-line interface mimicking Firebase's functionalities (Get, Put, Post, Patch, and Delete) for user-friendly interactions to query and modify database
* Designed a web-based platform focused on Citi Bike station data for functionalities to search and update station information to improve user experience and data utilization

[**Aspect-based Customer Review Analysis**](https://github.com/91MrHuang/ABSA) August 2022-December 2022

* Developed an aspect-based sentiment analysis classifier with PyTorch, tailored for restaurants and laptops, targeting the SemEval 2014 Task 4 dataset. By discerning topics within customer reviews, classifier accomplished 95.6% accuracy
* Leveraged MLflow for end-to-end model management and detailed performance tracking during training to packaged code into reproducible runs and developed essential artifacts to deploy various environment

**skills**

* Programming: Python, Java, C/C++, Shell, SQL
* Data Science skills: PyTorch, TensorFlow, Keras, BERT, Hadoop, Firebase, MongoDB, Slurm, BRAT, NLTK, MLflow