Chenfeng Huang

Los Angeles, CA | 415-341-2037 | chenfenghuang.info | chengfengeric@gmail.com |

RESEARCH INTEREST

Machine Learning, **Natural Language Processing**

EDUCATION

University of Southern California | *Master of Science, Applied Data Science*　　*January 2023 – May 2025*
  – GPA: 4.0/4.0
  – Coursework: Machine Learning, Database Management, Trusted System, User Studies & Design

Santa Clara University | *Bachelor of Science, Computer Science*　　*September 2020 – June 2022*
  – GPA: 3.5/4.0
  – Coursework: Theory of Algorithm, Probability & Statics, Logic Design, Theory of Automata & Language, Operating System, Data Mining

Diablo Valley College　　*January 2019 – May 2020*
  – GPA: 4.0/4.0
  – Coursework: Object Oriented Programming, Discrete Mathematics, Assembly language

Sierra College　　*September 2017 – December 2018*
  – GPA: 3.8/4.0
  – Coursework: Analytical Geometry & Calculus, Differential Equation, Linear Algebra, System Programming

PUBLICATION

- Z. Ma and **C. Huang** (2024). "Efficient Series Decomposition Learning for Time Series Generation with Diffusion Probabilistic Model". In: *submission*

RESEARCH EXPERIENCE

Research Assistance | *University of Southern California, Los Angeles, CA*　　*February 2023 – June 2023*
  – 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 SOTA success rate with 30 percent less dataset size, 200 hours less training time, and 50 percent less GPU memory.

Research Assistance | *Massachusetts Institute of Technology, Cambridge, MA*　　*April 2021 – September 2021*
  – 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.

WORK EXPERIENCE

Machine Learning Engineer Intern | *Union Big Data Co., Sichuan, China*　　*June 2023 – September 2023*
  – Utilized a pre-trained BERT model for Named Entity Recognition 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.

Computer Graphics Group intern | *Huawei Technologies Co., Sichuan, China*　　*June 2019 – July 2019*
  – Familiarized with the mechanism of networking and business by fully participating in both technical and managerial aspects of the computer graphics design project.
  – Participated in the Programming Design and developed my own project in computer graphics design.
  – Responsible for facilitating interdepartmental collaboration and assisting with digital product updates.

Teaching Experience

Mathematics Tutor | *Sierra College, Rocklin, CA* *June 2018 – December 2018*
- Tutoring MATH 30 and MATH 31 (Analytical Geometry and Calculus I & II

College Reading and Learning Association (CRLA) Certified Tutor
- Fullfilled CRLA's international Tutor Training Program and gained a certification for Certified Tutor, Level 1.

Project

A Java-Based System for Analyzing Online Healthcare Discussions *Fall 2023*
- Developed a Java application to analyze healthcare-related discussions, using N-tier architecture and modular design.
- Implemented JSON and tab-separated parsers, leveraging the Singleton pattern for logging.
- Executed geolocation matching with Cartesian distance for determining user location, generating state-level health trend reports from online discussions.

Secure Multi-Layered File System with Comprehensive Access Control *Fall 2023*
- Designed and implemented a secure multi-layered file system with robust access control, incorporating both Discretionary Access Control (DAC) and Mandatory Access Control (MAC) policies.
- Created a hierarchical structure for efficient file and directory management, with each layer handling specific tasks while maintaining strong information hiding and separation of concerns.
- Developed a comprehensive set of interfaces and internal methods for each module, ensuring secure and controlled access to system resources. The system handles various file operations (read, write, create, delete) while enforcing appropriate access controls at each level, providing a secure yet flexible foundation for file system operations.

Firebase Emulation with Citi Bike Station Query Platform *Spring 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 *Fall 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.

Technical Skill

Programming Languages : Python, C/C++, Java, SQL, Linux
Data Science Skills : PyTorch, TensorFlow, Keras, Hadoop, Firebase, MongoDB, MLflow