



Qijia He

+86 17752877602 | [✉ heqj2021@mail.sustech.edu.cn](mailto:heqj2021@mail.sustech.edu.cn) | [🌐 Personal Website](#)

EDUCATION

Southern University of Science and Technology (SUSTech) **Shenzhen, China**

B.Sc. in Data Science and Big Data Technology, GPA: 3.88/4.0, Major GPA: 3.95/4.0 Aug. 2021 - Jun. 2025 (Expected)

- **Awards:** 2nd Class Scholarship of Zhicheng College (Top 10%, 2022 & 2023), Elite Student in Dept. of STAT & DS
- **Relevant Courses:** Computer Vision, Data Structures & Algorithm Analysis, Data Science Practice, Big Data Analysis Software & Application, Computer Programming, Multivariate Statistical Analysis, Mathematical Analysis

Stony Brook University **Stony Brook, NY, USA**

Exchange Student, GPA: 4.0/4.0 Aug. 2023 - Dec. 2023

- **Relevant Courses:** Data Analysis, Data Mining, Foundations of Machine Learning, Introduction to Visualization

RESEARCH & INTERNSHIP

TAPALL **Shenzhen, China**

Computer Vision & Cloud Engineer Intern, R&D Department Jun. 2024 - Aug. 2024

- Scraped 20,000+ uncopyrighted videos from several websites (Pikwizard, Vecteezy, Mazwai, etc.) through Beautiful Soup using Python; designed a video filtering algorithm to select the videos with flat surface, camera in motion from the dataset for product demonstration in various environments.
- Employed OpenCV to identify basic optical flow and video characteristics and applied SegFormer (a semantic segmentation framework) to detect the specific objects in the video. The detection algorithm achieves an F1 score of 0.7.
- Utilized VideoLLaMA2 to generate captions for each video and classify them into several predetermined categories.
- Developed a cloud scheduling system to implement the company's 3D Modelling and Digital Human solutions in Huawei Cloud. Reduced the back-end operational expenses by deploying the GPU instead of engaging in long-term leasing.

Exploiting Offline Dataset to Improve Online Bandit Learning Efficiency with Relative Feedback **Shenzhen, China**

Research Assistant, PI: Prof. Fang Kong, SUSTech Sept. 2024 - Present

- Conducted a comprehensive literature review on the state-of-the-art bandit algorithms, focusing on estimating model parameters using offline stochastic data for online preference feedback by accommodating the LinUCB algorithm.
- Proposed several algorithmic approaches and conducted regret analysis, mathematically demonstrating that the offline dataset can effectively reduce regret within the proposed algorithm framework.
- Delivered a seminar presentation on stochastic linear bandits based on *Bandit Algorithms* [Tor Latimore et al, chapter 19].

Nature Index Periodicals: A Comparative Analysis of Impact and Innovation in Publications **Shenzhen, China**

Research Assistant, PI: Prof. Yifang Ma, SUSTech Jun. 2022 - Aug. 2023

- Collected 21,848 journal information from over 10 million papers from the Web of Science and a private database, featuring paper indexes like conventionality, hit5, hotspot, and novelty.
- Visualized KDE plots for over 1.5 million papers in NI and non-NI journals with similar journal impact factors in the same field, finding that NI journals do not exhibit an advantage in novelty.
- Analyzed the affiliations and publications in NI and non-NI journals, concluding that NI journals tend to favor works from institutions in the U.S. and Europe. Suggesting either a limitation in the Nature Index's ability to differentiate highly influential journals or an inherent bias in the selection practices of publishers.

COMPETITIONS

Optimization of the CLIP Model (ViT-B-32) with Few-shot Learning and Zero-shot Learning **Shenzhen, China**

Core Member, The 2024 Jittor Artificial Intelligence Challenge Apr. 2024 - Aug. 2024

- Based on hierarchical clustering, selected four photos from each dataset (Animal, Caltech-101, Food-101, and Tsinghua Dogs) for few-shot training, with Hamming distance as the metric and image vectors encoded by the original model.
- Fine-tuned the prompt from “a photo of xxx” to various forms by merging self-written and GPT-4-written captions, and selected the final prompt based on increased accuracy in training datasets.

- Developed an adapter layer for an image encoder by using the boosting method with cross-entropy loss; Trained three adapters with different structures and parameters and integrated predicted results via ensemble learning. The classification model has an improved accuracy from 62% to 69.15%, ranking top 0.6% among all participants.

Momentum: Is it Truly Exist in Tennis Game?

Shenzhen, China

Team Leader, 2024 COMAP Mathematical Contest in Modeling (MCM)

Feb. 2024

- Analyzed the shifting momentum between players and their impact on game outcomes based on the 2023 Wimbledon Championships - Men's doubles dataset.
- Quantified momentum through a self-designed time series model and developed a logistic regression model with adjusted inputs, applied Non-Maximum Suppression to figure out the momentum change, and implemented the Anderson-Darling test to compare the observed and simulated momentum effects under random conditions. The model achieved 79% accuracy in predicting tennis game winners, supporting the theory that momentum exists.

Predicting Wordle Game Result Based on Random Forest

Shenzhen, China

Core Member, 2023 COMAP Mathematical Contest in Modeling (MCM)

Feb. 2023

- Based on the dataset provided by Twitter in 2022, predicted the daily number of players of Wordle Game using an ARIMA model, determined the distribution of the number of attempts for any given word using random forest regression and assessed the difficulty of guessing a particular word using K-means clustering the accuracy of which reaches 73.6%
- Constructed a word-guessing machine using the Monte Carlo algorithm to maximize the probability of correct guesses, indicating that the real number of guesses is at least 2.75 times more than the number shared on Twitter.

ACADEMIC PROJECTS

Shenzhen Metro's Operation Schedule Optimization

May 2024

- Customized the train stop schedule during morning peak based on passenger flow data of Metro Line 5 on Aug 31, 2018.
- Optimized three different greedy algorithms with local search algorithms, and found that during peak congestion periods, adopting a flexible skipping-stop strategy can save passengers 10% to 15% of their average commuting time.
- Visualized the dynamic passenger flow of each station and the operation of the whole subway system using HTML.

Stock Price Trend Prediction Based on Dimension Reduction Techniques and Cluster Analysis

Apr. 2024

- Analyzed the operational conditions of 1,200 listed companies from three industries (pharmaceutical, chemical, and machinery) for the fourth quarter of 2023 using the financial statements scraped from BaoStock.
- Developed a company evaluation & classification system through factor analysis & K-means clustering and created a stock recommendation system for investors to make predictions about stock price trends based on Ridge regression. Buying stocks advised by the system may profit investors 2.36 RMB/stock compared to random stock purchases.

Klotski: Number Puzzle Game Solver Development

Dec. 2022

- Developed a solver for Klotski and determined if there was a solution in any given initial configuration. The solver is based on a self-designed greedy algorithm and Min Priority Queue, which gave higher priority to the move of blocks and structures that will lead the board closer to its final structure.
- The solver can quickly find solutions for a 6x6 Klotski using 1x1, 2x1, and 2x2 blocks within five minutes, whereas pure BFS and DFS algorithms are limited to solving puzzles of size 4x4 or smaller.

Bilibili User Study Data Analysis

Nov. 2022

- Based on a given dataset with 9,000+ vloggers in 2019, scraped the related data in 2022 through Google's web scraper.
- Utilized Python for exploratory data analysis (EDA), analyzing vloggers' characteristics and videos from 2019 to 2022 to identify changes in patterns and platform used.
- Provided suggestions for both new and experienced vloggers on creating videos and attracting fans accordingly.

SKILLS & MISCELLANEOUS

Programming: Python, R, Java, HTML/CSS/JavaScript, Hadoop & Spark, MATLAB

Language: Mandarin (native), English (proficient), Spanish (Basic)

Social Work:

- Assistant Instructor for SUSTech Youth Rock Climbing Course Jan. 2024
- Student Assistant in Student Affairs Center, SUSTech Jun. 2022 - Aug. 2022
- Contributed Over 80 Hours of Volunteer Service Across Diverse Activities and Organizations

Interest: Rock Climbing (SUSTech's Rock Climbing Team), Badminton (Zhicheng College's Badminton Team), Hiking.