

JUN ZE HE

(626)-592-2842 | junzehe977@gmail.com | <https://www.linkedin.com/in/junze-he-9146a3267/> | <https://github.com/JunJul>

EDUCATION

University of California, Los Angeles (UCLA), Los Angeles, CA Expected June 2025
Bachelor of Science, Data Theory

- GPA: 3.324
- Relevant Coursework: Python, R programming, Introduction to Data Analysis, Introduction to Computational Statistics, Mathematics for Data Theory, Statistical Optimization in Machine Learning, and Data Mining

Break Through Tech at Cornell University, New York, NY May 2024 – Aug 2024
Certificate In Machine Learning Foundations (eCornell)

- Relevant coursework: Machine Learning, Computer Vision, Natural Language Processing, Language Models

CAMPUS INVOLVEMENT / EXTRACURRICULAR ACTIVITIES

National Student Data Corps @ UCLA, Los Angeles, CA Jan 2024 – June 2024
Project Lead

- Led 5 teammates to complete the credit score and the heart disease classification project by **Python**
- Managed workflow of **data understanding / visualization, feature engineering, modeling, and model evaluation** in both projects by **Trello**
- Accomplished **90%** true positive rate by **XGB** in the Credit Score classification project and **95%** true positive rate by **Random Forest** in the heart disease classification project

PROJECTS

Honkai Star Rail's Player Reviews Analysis Aug 2024 - Sep 2024

- Scraped Honkai Star Rail's player reviews data on Google Store by **serpapi**
- Analyzed *Honkai Impact Star Rail* player reviews using **LatentDirichletAllocation, Nonnegative Matrix Factorization, Top2Vec**, and **BERTopic** by **Python** to identify key themes in complaints and wishes
- Summarized player concerns with game difficulty, login issues, and requests for controller support

WORK EXPERIENCE

Accenture, Los Angeles, CA Aug 2024 – Dec 2024
AI Studio Program (Machine Learning Engineer Intern)

- Developed a **Technology News Insight Engine** to identify Technology and Tech industry trends to **support** Tech startup clients' business development
- Designed a **text cleaning pipeline** to **tokenize** sentences, **remove** stopwords, **lemmatize** words, and **summarize** long text articles by frequency word-based Summarization and TextRank Summarization
- **Organized** Tech News articles into each group by **BERTopic** and **created topic names** for each group of the News articles by **Groq API**
- Utilized **retrieve and re-rank** methods from **sentence transformers, RAGs, sentiment scores**, and article publish date to build an interactive **recommendation system**
- Empowered Accenture consultants to gather targeted technology insights aligned with the specific technologies utilized in clients' businesses.

Mt San Antonio College, Walnut, CA June 2022 - Aug 2022
Research Assistant

- Assisted a computer science lab in University of California, Irvine in **integrating** a health watch with their Institute of Future Health platform
- Collected blood pressure from over 20+ patients by the health watch and a medical blood pressure device
- **Cleaned** data and investigated the watch's data collection accuracy through **data visualization** and **statistical measurements** in Python
- Demonstrated the difference between data from the watch and the medical device was small based on **mean squared error, variance**, and **p-values** by **Scipy**.

SKILLS

- Programming Languages: Python, R, SQL, HTML, CSS, JavaScript, C++, Java
- Programming Frameworks: TensorFlow, pyTorch
- Computer: Google Suite, Microsoft Office, Trello, Github, Git
- Language: Mandarin (Fluent), Cantonese (Fluent)