Xintong (Jessica) Cai

Cell Phone: +1 (401)523-1752   Email: caixintong@ucla.edu

EDUCATION

University of California, **Los Angeles**, *Master of Science in Applied Statistics and Data Science*   **Sep.2024-Jun.2026**
- GPA: 3.95/4.0
- Relevant Courses: Mathematics Probability, Statistical Computing and Programming, Data Management

University of Chicago, *Master of Art in Social Science (Quantitative Methods and Social Analysis)*   **Sep.2023-Jun.2024**
- GPA: 3.85/4.0

Fudan University, *Bachelor of Art in Economics and Communications*   **Sep.2019-Jun.2023**
- GPA: 3.56/4.0, Ranking: 6/30

University of Helsinki, *Exchange Study (Faculty of Social Science)*   **Jan.2022-Jun.2022**
- GPA: 4.8/5.0

SKILLS

Programming: Python (Numpy, Pandas, Pytorch, Scikit-learn), R (Proficient),SQL(Proficient), AWS/BigQuery(Proficient)
Language: Mandarin(Native); English(Fluent, TOEFL:110); Wu(Fluent); Japanese (Basic, N2);German (Basic)
Tools: Latex, Markdown

PROFESSIONAL EXPERIENCES

Demandbase | San Francisco, United States   **Jun.2025 – Sep.2025**
Data Scientist Intern, ML/DS team
- Conducted exploratory data analysis (EDA) and **research on the evolution of trending hashtag**s over time to **identify ad targeting opportunities**
- Built scalable pipelines leveraging **ARIMA models for time series anomaly detection** on 800M+ daily intent records to surface emerging user interests
- Developed **attention-based algorithms** using **temporary fusion transformer** (TFT) models to predict trending keywords
- Used **AWS and Google BigQuery** to parallelize time series modeling on 300M+ daily intent records
- Deployed **Large Language Models (LLMs)** to automatically annotate firmographic attributes of B2B companies, enhancing data quality and downstream analytics

Knowledge Lab, University of Chicago | Remote, United States   **Jun.2024 – Dec.2024**
Data Analyst, Knowledge Lab
- Extracted data of dissertation of scientists since 2012 from Openalex and **conducted EDA**
- Conducted **fuzzy-match** on a dataset regarding Ph.D dissertation with over 20,000 lines of data
- Summarized over 100 research about the driving force of the **disruptive technology diffusion** and wrote an analysis

Siemens Smart Infrastructure Group China | **Shanghai, China**   **Sep.2022-Nov.2022**
Strategy Analyst
- Monitored Chinese **macroeconomic trends by analyzing National Bureau of Statistics (NBS) data using Excel**, with a focus on real estate, industry, and transportation sectors
- **Authored a strategic report on the future of new energy development** in China, drawing insights from energy policy analysis based on **World Bank and IMF reports**, providing an **outlook for the next fiscal year**
- Prepared **competitive analysis** by comparing Siemens with competitors as Schneider and ABB by financial report

PROJECT EXPERIENCES

Predicting ADHD Risk through Brain Scanning Images Data (2025 Women in Data Science)   **Jan.2025 – May.2025**
- Conducted EDA on the dataset of the functional MRI by **clustering and visualization**
- Used **ensemble methods like boosting and random forest** to analyze multilevel categorical data
- Implemented SVM and MLP(multiple layer perceptron) by PyTorch to train functional MRI data classification and optimize
- Ranked **23rd** worldwide up to Feb.23rd 2025

Predicting the CTR (Click-Through-Rate) of adversetisement recommendation system   **Jan.2025 -April.2025**
Instructed by Professor Cheng Guang, UCLA AI Trustworthy Lab
- Applied logistic regression and neural networks to explore and model user behavior from the dataset
- **Called API and used LLM to generate synthetic data** to solve the imbalance of data and tested fidelity
- Used **cat-boosting and random forest** to predict the CTR, reaching 0.89 AUC

Fading Digital Empathy: The narrative transformation of Online Support groups across Covid-19 **Jan.2024 – Aug.2024**
Instructed by James Evans, University of Chicago
- **Extracted and analyzed 76,000 posts from Reddit using Python**, conducted time series analysis on social support patterns
- Utilized **large language models for sentiment analysis** to assess shifts in online community support, discovering a 24% decline in social support intensity within mental health communities post-Covid-19
- Paper published in **PloS One 2024**