

# Haitian(Enoch) Liu

<https://hliu3.com>

Email : [haitian.job@gmail.com](mailto:haitian.job@gmail.com)

Mobile : +1 (347) 925-3966

## EDUCATION

### Harvard University

MS, Data Science (Cross-registration at MIT)

Cambridge, MA

Aug. 2022 - May. 2024

### University of California, Berkeley

BA, Applied Mathematics; BA, Data Science

Berkeley, CA

Aug. 2019 - Dec. 2021

Courses: Machine Learning, Deep Learning, Computer Vision, Probability Theory, System Development, Visualization

## EXPERIENCE

### Harvard Business School [ Database | Python | MySQL | Stata | JavaScript | HTML ]

Cambridge, MA

Research Associate on Impact Investments Project

Feb. 2024 – Present

- Optimized ETL pipelines for data transformation and consolidation in PCRI database, resolving discrepancies from varied data types across 30+ data vendors over 100,000+ private capital companies, achieving 98% data integrity
- Redesigned and developed Private Capital Research Institute (PCRI) website via HTML5, JavaScript and CSS

### Nock Lab, Harvard University [ LLM | PyTorch | HuggingFace ]

Cambridge, MA

Research Assistant Supervised by Ph.D. Yaniv Yacoby and Taylor McGuire

Sep. – Dec. 2023

- Curated the first benchmark dataset for detecting LLM-generated text in mental health surveys through human subjects simulation using LLMs, text augmentation and rigorous quality validation
- Developed tree-based, transformer-based, and few-shot in-context learning models for LLM text detection, enhancing survey data quality and achieving 96.2% test accuracy
- Presented at *Women in Data Science Conference 2024* to hundreds of professionals from industry and academia

### YunExpress USA [ Python | GeoPandas | Tableau | HTML | Google Apps Script ]

Vernon, CA

Data Scientist

Apr. – Aug. 2022

- Developed a custom website for the Operations department, boosting data processing efficiency by 20%
- Built a pallet recycling tracking system and automated email solution, reducing costs by \$4 million annually
- Utilized data analysis and visualization techniques (e.g. Tableau dashboard, GeoPandas) on 24 million parcel records to improve the package delivery operations, resulting in a 1.2 days average reduction in delivery time

### Department of Mathematics, UC Berkeley [ PDE | MATLAB ]

Berkeley, CA

Research Assistant for Prof. Di Fang on Bifurcation and Clustering of Biological Networks

May. – Dec. 2021

- Designed numerical algorithms of biological networks models, especially on leaf venation
- Built an efficient semi-implicit marching algorithm that preserves gradient flow structure
- Calculated Fréchet derivative for the gradient flow of the constraint optimization problem

## PROJECTS

### Sentiment Analysis for Tinder Review using Natural Language Processing [NLP]

Feb. – May. 2023

- Utilized various NLP techniques (e.g., simple Recurrent Neural Networks, Bidirectional Long Short-Term Memory, pre-trained word embeddings (GloVe) on Bidirectional Gated Recurrent Units, simple transformer, BERT, and Roberta) to categorize the sentiments expressed in Tinder app reviews, achieving accuracy of 94%

### MRI Brain Tumor Image Diagnosis and Classification [ CV | TensorFlow ]

Feb. – May. 2023

- Implemented and built our own deep convolutional neural network (e.g., VGG Network, Residual Network, and Google Inception Network) to classify MRI images into meningioma, pituitary, glioma, and no tumor. Evaluated model performance by confusion matrix and F1 score, achieving test accuracy of 97%

### Mortality prediction in General Population [ Survival Analysis | Cox Regression | Elastic Net ]

Oct. – Dec. 2022

- Constructed predictive models leveraging Regularized Regression (e.g., Ridge, Lasso), Principal Component Regression, Random Forest, Boosting, Neural Networks, and survival prediction techniques in the National Health and Nutrition Examination Survey (NHANES) Follow-up Study, achieving an accuracy of 83%

## SKILLS

**Programming Languages:** Python, Java, SQL, R, JavaScript, HTML, CSS, MATLAB,  $\LaTeX$

**Platform & Tools:** Git, Unix/Linux, Docker, Tableau, Jupyter Lab, R Studio, VS Code, PyCharm, IntelliJ

**Libraries:** PyTorch, TensorFlow, HuggingFace, pandas, NumPy, Seaborn, Matplotlib, Scikit-learn, D3.js, Bootstrap

**Mentorship:** Teaching Fellow for *AC 207: System Development* at Harvard University

**Entrepreneurship:** Founded Triple Six Group, supplying beauty products to 3000+ global SMEs via digital marketing