

HAOHUI LIU

<https://liuhh02.github.io> | +1 4122095194 | haohuil@andrew.cmu.edu

EDUCATION

Carnegie Mellon University | B.S. in Computer Science; Concentration in Machine Learning Dec 2024

- Activities: Quant Club (Executive Committee), Blockchain Group, Business Technology Group, Gym
- Courses: Principles of Imperative Computation (15122), Principles of Functional Programming (15150), Introduction to Mathematical Finance (21270), Calculus in 3D (21259), Matrices and Linear Transformations (21240)

INTERNSHIP EXPERIENCE

Machine Learning Intern, PayPal May 2022 – Aug 2022

- Performed statistical analysis, topic modelling with Latent Dirichlet Allocation (LDA) and Named Entity Recognition in Python to uncover trends across imbalanced, longform documents and short transcripts
- Trained state-of-the-art NLP Transformer models using PyTorch to improve text classification through iterative self-learning

Data Science Intern, Amili (Bioinformatics Start-up) Apr 2021 – Jul 2021

- Performed statistical analysis, data visualization, feature selection, dimensionality reduction (t-SNE, PCA, etc.) in R & Python to extract useful insights from multi-dimensional rRNA data containing over 1 million features
- Published research abstract as 1st author in *Gut* (23.1 Impact Factor)

Deep Learning Intern, National University of Singapore Jun 2019 – Nov 2020

- Proposed use of and developed conditional GANs & ResNets in Python on 3D MRI scans with PyTorch, Keras, Tensorflow
- Improved early detection of Alzheimer's Disease by 67%
- Published paper as 1st author in *European Journal of Nuclear Medicine & Molecular Imaging* (9.2 Impact Factor; 10 citations)

Machine Learning Intern, DSTA Singapore Oct 2018 – Jan 2019

- Trained 1D CNNs, bidirectional LSTMs & stacked ensembles using Keras and scikit-learn to detect fake news using NLP
- Improved generalizability and increased accuracy of classification from baseline of 44.3% to 84.9%
- Won 4th award in category (Robotics and Intelligent Machines) at International Science and Engineering Fair (ISEF)
- Published paper as 1st author in *2019 IEEE Big Data Conference* (18.7% Acceptance Rate; 7 citations)

HACKATHONS AND COMPETITIONS

UChicago Trading Competition – Selected from over 20 teams to represent CMU 2022

- Coded a SARIMAX algorithm for time series analysis to predict the fair value of lumber prices based on rainfall predictions
- Built a market making bot to place orders and execute trades, realizing profit of over \$400K in simulation

FinTech Nations Hackathon Winner and Best Use of SQL at cmd-f 2021

- Developed interactive full-stack website displaying technical analysis & sentimental analysis of stocks
- Scraped financial data from Yahoo Finance and used CockroachDB hosted with Google Cloud to store data in SQL database

Champion at Superposition V 2021

- Fine-tuned state-of-the-art T5 NLP model on code-natural language pairs using PyTorch & Huggingface
- Coded VS Code extension in Typescript & Javascript and linked the NLP backend to the extension using axios

First Place in Tech Takes on Unemployment Hackathon 2020

- Trained bidirectional-LSTM and Universal Sentence Encoder to calculate similarity between resume and job postings
- Served models with Flask and deployed website using Heroku

SKILLS

- **Languages:** Python, C++, C, Java, JavaScript, R, SQL, SML, Bash, Mandarin Chinese (Bilingual), German (Intermediate)
- **Deep Learning & Machine Learning:** Natural Language Processing, Computer Vision, Time Series, Interpretable ML
- **ML Libraries:** PyTorch, Fastai, Keras, Tensorflow, HuggingFace, OpenCV, XGBoost, Scikit-learn, SHAP, Numpy, Pandas, spaCy
- **Software Engineering & Full-Stack Web Development:** React, Nodejs, Flask, HTML, CSS, Streamlit, SQL, Firebase, Google Cloud, Heroku, PythonAnywhere, Linux

AWARDS

- **Other Hackathons:** Quantathon 2022 First Runner's Up (Sponsored by Goldman Sachs), Best Application of Data Hack at HackCMU 2021 by Hudson River Trading, Best Use of MongoDB Atlas & Best COVID-19 Hack at hths.hacks() 2020
- International Science and Engineering Fair (ISEF) 2019: Fourth Award in Robotics and Intelligent Machines
- **Certifications:** Stanford Game Theory, Yale Financial Markets, Stanford Algorithms & Data Structures Specialization

PUBLICATIONS

1. **H. Liu**, et al. Improved amyloid burden quantification with nonspecific estimates using deep learning. *Eur J Nucl Med Mol Imaging* (2021).
2. **H. Liu**, et al. IDDF2021-ABS-0140 Gut microbiota significantly correlate with body constitution in traditional chinese medicine. *Gut* (2021).
3. **H. Liu**, et al, "Deep Learning-Based Estimation of Non-Specific Uptake in Amyloid- PET Images from Structural MRI for Improved Quantification of Amyloid Load in Alzheimer's Disease," 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems, Rochester, MN, USA.
4. **H. Liu**, "A Location Independent Machine Learning Approach for Early Fake News Detection," 2019 IEEE Big Data, Los Angeles, CA, USA.
5. Y. Nai, **H. Liu**, A. Reilhac, Alzheimer's Disease Neuroimaging Initiative. Validation of deep learning-based nonspecific estimates for amyloid burden quantification with longitudinal data. *Physica Medica* (2022).