Yu-Wei Lai

412-251-7392 | therachellai@gmail.com | Linkedin |Github

EDUCATION

Georgia Institute of Technology

Master of Science in Computer Science—Machine Learning Specialization *Aug. 2023 - Expected Dec. 2025*

Carnegie Mellon University

Bachelor of Science in Mathematical Sciences *Sep. 2018 - May 2022*

SKILLS

Domains: Machine Learning, Deep Learning, Math Modeling, Finance, NLP, CV, Web/Mobile development, UI/UX

Programming Languages: Python, Java, C++, C, R, SQL, Javascript/Typescript, HTML/CSS

Technologies: PyTorch, Tensorflow, Scikit-learn, AWS, MongoDB, Express, React, Node.js, Flask, Django, Docker, Spark

ACADEMIC PROJECTS

High Frequency Automated Trading Link | *GA Tech* | *Spring 2024*

- Implemented from scratch a reinforcement learning-based trading agent and rule-based trading strategies, integrating **SMA**, **MACD**, **BBP** indicators and leveraging **NumPy** and **Pandas** for data analysis to design an optimal trading strategy for the JPM stock over a duration of two years
- Achieved outperformance of 35% against benchmarks in out-of-sample analyses with RL model and 5% with fixed rules
- Researched and discovered positive correlation between trading frequency and both returns and volatility, negative correlation between impact value and returns, and positive correlation between impact value and volatility

Generative Pre-Trained Transformer (GPT) Clone Link | *GA Tech* | *Summer 2024*

- Developed from scratch a character-level autoregressive **transformer** language model in **PyTorch** to generate text resembling Shakespearean writing style

Twitch Clone Link | Northeastern | *Fall 2023*

- Developed and launched a full-stack streaming platform using **MongoDB**, **Express**, **React**, and **Node.js**, enabling seamless live broadcasting by integrating **OBS** and incorporating **SocketIO** for real-time chat functionality
- Implemented user authentication systems, facilitating login and registration processes for improved platform security
- Conducted comprehensive API testing utilizing **Postman**, ensuring robust functionality and performance of the streaming website, while leveraging **React** for building dynamic and interactive user interfaces

Food Nutrition Analysis & Recommendation System Link | Northeastern | *Spring 2024*

- Implemented web crawlers to collect food health scores from EWG website and nutritional information from Costco and utilized **OCR** to extract nutritional data from images, storing both data systematically in **AWS S3** for preprocessing
- Ensembled models: Linear Regression & Decision Tree model constructed with **scikit-learn** to predict rating from nutrition label & open-sourced **LLM** using few-shot prompt engineering technique to predict rating from ingredient list
- Built interactive recommendation system for healthier foods, validated through UX research and a **Streamlit** dashboard

Olympic Games Data Pipeline Northeastern | *Summer 2024*

- Engineered an ETL pipeline leveraging **Azure Data Factory** to extract Olympic Games data, stored it in **Data Lake Gen 2**, transformed the data using **Apache Spark** on **Azure Databricks**, and visualized analysis result on **Tableau** dashboard

RESEARCH EXPERIENCE

Project Manager (Research Project with CMU's Human-Computer Interaction Department) *CMU* **|** *Spring 2022*

- Engaged with disabled users and collaborated with Colombia researchers to explore augmentative & alternative communication systems to design a more usable communication device that utilizes shortcuts and user eye movements
- Led user-centered research; analyzed feedback; crafted prototype with **Figma**; delivered design process documentation

WORK EXPERIENCE

Data Analyst **Powerline (startup)** *Sunnyvale, CA* | *Mar. 2022 - Jan. 2023*

- Researched & assessed EV driver charging patterns, weather correlations, emissions, and electricity pricing to gain crucial insights for team's development of a user recommendation system for optimal EV charging locations & timings
- Analyzed & visualized time-series user benefit patterns data with **pandas**, **seaborn**, & **plotly**, demonstrating product's potential to reduce charging expenses by 32% in various usage scenarios, guiding investor decision towards our solution
- Utilized **Selenium** for web scraping of EV database, curating optimal test vehicle selection for successful product launch