Samuel Lee

(408) 650-2618 | sl5806@columbia.edu | https://www.linkedin.com/in/samuel-lee-12142a198/

EDUCATION

Columbia University New York, NY
MS in Computer Science -- Thesis Expected Dec 2026
Focused Research Areas: Machine Learning and LLMs; Causal Inferences

Northwestern University Evanston, IL
BA in Computer Science, 3.973/4.0 Dec 2024
Focused Research Areas: Causal Inferences – Adjusting for Measurement Bias from Unobserved Confounders; Deep Learning Models for Large Scale EHR

EXPERIENCE

Northwestern Dept. of Computer Science Evanston, IL
Machine Learning and Causal Inference Research Assistant Jan 2023 - Present

- Led the development process of advanced Machine Learning models (e.g. Transformers -- BERT, Large Language Models, Deep Learning) using PyTorch and Postgresql to impute unknown confounders for causal relationships
- Applied causal methodologies (e.g. measurement error + noisy proxy predictions) to examine causality between TTE, mortality, and smoking proxy (Python) in MIMIC-III, a large-scale (40000+ patients) EHR database

Intuit Inc Mountain View, CA
Software Engineer Intern Jun 2024 - Aug 2024

- Proposed and implemented persistence enhancements to current API functionalities (Java, SpringBoot) to more efficiently handle large-scaled (200K) transactions daily
- Designed and optimized resiliency in the backend (Java) to catch + retry 100% of unpublished transactions (Kafka), improving platform robustness for every-day use
- Ensured automation and 100% adherence to schedule by deploying persistence and resiliency onto schedulers (AWS, Kubernetes)
- Initiated collaboration with sister teams to contribute to new features and fix past bugs in the codebase
- Tools: Java, PostgreSQL, OracleDB, Apache Kafka, SpringBoot, AWS, Kubernetes

Ford Motor Company Dearborn, MI
Software Engineer Intern Jun 2023 - Aug 2023

- Developed an end-to-end ML features pipeline (Express, Javascript, MongoDB) to collect 1000+ KPI data generated from infotainment system in vehicles and visualize the data through a monitoring dashboard (React.js)
- Proposed the system design, collaborated with senior engineers, and received approval from tech leadership

PUBLICATIONS

Contro. for Unobserved Confounding with Large Language Model Classification of Patient Smoking Status (Under Review), Samuel Lee and Zach Wood-Doughty Dec 2024

- Selected for Oral Presentation at Advancements in Medical Foundation Models (AIM‑FM) Workshop @ NeurIPS 2024 (7 out of 81)
- Github Repo: https://github.com/controllingunobservedpaper/NEURIPS_submission_2024
- Paper link (arxiv): https://arxiv.org/abs/2411.03004
- Shoutout from NU Engineering on the paper: https://www.mccormick.northwestern.edu/computer-science/news-events/news/articles/2024/strong-northwestern-presence-at-the-2024-neurips-conference.html!

Exploring to Learn Winning Strategy, Samuel Lee, July 2019, IADIS International Conference Game and Entertainment Technologies 2019 (GET'19) Jul 2019

- Developed a novel algorithm UCB* to make more intelligent decisions, enhancing the winning chances within games (ex: Tic-Tac-Toe) by 200% through the utilization of a well-known tree search algorithm, Monte Carlo Tree Search (MCTS)
- Received the **Encouragement Award** from Get'19 -- given to a young researcher whose paper was of highest standard based on a blind review process
- Github Repo: https://github.com/penguinman6678/AI4Fun

SKILLS

Programming Languages: Python, Java, Javascript, R, SQL, C
Tech Tools + Frameworks: PyTorch / Tensorflow, Jupyter, Shell Scripting, Flask + FastAPI, React.js, Node.js, SpringBoot, Git / Github
Systems + Cloud: Apache Kafka, AWS -- EC2, Lambda, RDS; DBMS -- NoSQL, MySQL, PostgreSQL, OracleDB, PineconeDB