

Samuel Lee

CURRENT UNIVERSITY STUDENT

☎ (408) 650-2618 | ✉ samlee0813@gmail.com | in Samuel Lee

Education

Columbia University

MS IN COMPUTER SCIENCE – THESIS

New York City, NY

Graduation: Dec 2026

Focused Research Areas: Machine Learning and LLMs; Causal Inferences

Northwestern University

BA IN COMPUTER SCIENCE – GPA: 3.974 /4.0

Evanston, IL

Graduation: Dec 2024

Focused Research Areas: Causal Inferences – Adjusting for Measurement Bias from Unobserved Confounders; Deep Learning Models for Large Scale EHR

Experience

Intuit Inc

Mountain View, CA

SOFTWARE ENGINEER INTERN

June 2024 - Aug 2024

- Proposed and implemented persistence enhancements to current API functionalities (Java) to more efficiently handle large-scaled (200K) transactions daily
- Designed and optimized resiliency in the backend (Java) to catch + retry unpublished transactions (Kafka), improving platform robustness for every-day use
- Ensured automation and 100% adherence to schedule by deploying persistence and resiliency onto schedulers using AWS and Kubernetes
- Understanding the codebase of financial products, contributing new features, and fixing other contributors' bugs
- Tools: Java, PostgreSQL, OracleDB, Apache Kafka, SpringBoot, AWS, Kubernetes

Northwestern Computer Science

Evanston, IL

MACHINE LEARNING AND CAUSAL INFERENCE RESEARCH ASSISTANT

Jan 2023 - Present

- Building advanced Machine Learning models (e.g. Deep Learning, Large Language Models) to impute unknown confounders for causal relationships
- Applying causal methodologies (e.g. measurement error + noisy proxy predictions) to examine causality between TTE, mortality, and smoking proxy
- Technologies and Tools: Transformers (i.e. BERT), Python, PostgreSQL(DB), PyTorch

Ford Motor Company

Dearborn, MI

SOFTWARE ENGINEER INTERN

June 2022 - Aug 2022

- Machine Learning features pipeline: developed an end-to-end pipeline that collects 1000+ KPI data generated from infotainment system in vehicles and visualize them through a monitoring dashboard.
- Collected requirements from senior engineers, proposed the design of the system, and received approval from the tech leadership.
- Selected for presentations for my work to senior leadership.
- Tools: Python, JavaScript, NodeJs / Express, ReactJS, MongoDB, Microsoft Azure CosmosDB, NoSQL, MySQL

Publications

Contro. for Unobserved Confounding with Large Language Model Classification of Patient Smoking Status

Dec 2024

(Under Review), SAMUEL LEE AND ZACH WOOD-DOUGHTY

- Github Repo:** https://github.com/controllingunobservedpaper/NEURIPS_submission_2024
- Selected for Oral Presentation at Advancements in Medical Foundation Models (AIM-FM) Workshop @ NeurIPS 2024 (7 out of 81)
- Paper link:** <https://arxiv.org/abs/2411.03004>
- Check out this **shoutout** from NU Engineering for the paper!

Exploring to Learn Winning Strategy, SAMUEL LEE, JULY 2019, IADIS INTERNATIONAL CONFERENCE GAME AND

July 2019

ENTERTAINMENT TECHNOLOGIES 2019 (GET'19)

- Award – Encouragement Award** from GET'19.
- Github Repo:** <https://github.com/penguinman6678/AI4Fun>

Skills

Programming Languages

Python, Java, Javascript, SQL, C

Tech Tools

PyTorch / Tensorflow, Shell Scripting, Flask, React.js, Node.js, Express, SpringBoot, Git, GitHub

Systems + Cloud

Apache Kafka, AWS – EC2, Lambda, RDS; DBMS – NoSQL, MySQL, PostgreSQL, MongoDB, CosmosDB, PineconeDB

Languages

English, French, Korean