Alex Chen

(814)380-8601 | aclheexn1346@g.ucla.edu | aclheexn1346/aclheexn1346.github.io | U.S. Citizen

Education

University of California Los Angeles (UCLA)

Los Angeles, California
Expected Graduation Date: December 2025

PhD Candidate in Statistics

GPA: 4.0/4.0

Pennsylvania State University

University Park, Pennsylvania
Graduation Date: May 2020

Double Major in Data Science and Statistics

GPA: 3.83/4.0

Skills

Programming R, Python, SQL, SAS, Docker, MongoDB, Tableau, Excel, Data Structures and Algorithms
High-Performance Cluster parallel computing, RShiny, AWS, C, C++

Notable Graduate Classes Taken Causal Inference, Graphical Models, Monte-Carlo Optimization, Machine Learning

Experience

Machine Learning Research Intern — *Seattle, Washington*
Informatics and Predictive Sciences (IPS) Internship at Bristol Myers Squibb — *June 2024 - Present*

- **Developed innovative computational methods:** Identified over 10 potential gene regulators for distinguishing T-cell states and designed a method for integrating multiple single-cell RNA-seq datasets into one consensus causal graph, supporting target identification from gene networks. Developed R package and work to be submitted and published in selected journal.
- **Advanced CAR T cell therapy development:** Collaborated with IPS colleagues to analyze causal graphs, identifying new drivers of T-cell exhaustion. Directly contributed to improving therapeutic strategies by leading discussions and presenting findings at team and division-wide meetings with stakeholders.
- **Continuing work to enhance causal graph estimation:** Post-internship collaboration to incorporate context-specific data such as eQTL information, refining methods for greater biological accuracy and therapeutic relevance.

PhD Research — *Los Angeles, California*
UCLA Causality Gaussian DAGs from Network Data with Dr. Qing Zhou — *June 2021 - Present*

- **Causal Inference on Network Data:** Led NSF-funded research to develop an algorithm for causal inference in dependent data, improving accuracy by 20-50% compared to standard methods. This work has direct applications for industry in fields with data dependencies, such as marketing analytics for customer segmentation, financial data modeling and behavioral prediction. *Submitted to AISTATS 2025
- **RNA-seq Clustering and Causal Network Detection:** Implemented unsupervised cell clustering techniques on single-cell RNA-seq data, leading to significant improvements in identifying gene regulatory network structures.

Data Science Internship — *Redmond, Washington*
Database and Machine Learning Intern at Genie Aerial Work Platforms — *June 2021 - Sept. 2021*

- **Outlier Detection and Usage Modeling:** Continuously pulled and merged IoT sensor data from AWS Redshift using SQL and Python, modeling fault detection and analyzing user usage time to create live dashboards for real-time monitoring in manufacturing that reduced downtime and operational costs. A/B testing revealed 22% increase in dashboard engagement after including fault detection feature.
- **Recommendation System for Client Purchases:** Implemented a collaborative filtering recommendation system in R, enabling the sales team to optimize sales strategies for major clients and increasing business performance.

Projects/Awards

2024 **NLP and Knowledge Graph Extraction,** Developed a pipeline to segment and analyze Yelp reviews. Stored over 300,000 reviews in MongoDB, finetuning LLM(GPT-4o) for semantic understanding and Neo4j for analysis

2018 **1st Place Overall & Best Insight Award – PSU DataFest Competition,** Analyzed Indeed job market UX/UI data to uncover temporal trends in job postings and salary fluctuations across regions. Proposed new features for Indeed's platform, including a metric comparing job salaries to local cost of living, enhancing user decision-making for job seekers. Delivered data-driven insights that led to strategic recommendations, recognized for innovation and practical impact on improving the user experience.