

# Carleen Li

(+1) 510-364-8555 | 2000cli@gmail.com | linkedin.com/in/carleen-li | Mountain View, CA

## Experience

LinkedIn

2021 - Present

### Senior Software Engineer - Data Management

- Accelerated DMA compliance release and updates to AI models and data pipelines through a 30% reduction in average processing time of member privacy computation data pipelines
- Executed end-to-end migration of next gen orchestration platform to **Airflow**, achieving key leadership objectives
- Advanced internal **Airflow** adoption and platform maturity through development of centralized best practice resources and detailing operational risks and feature gaps that directly informed platform improvements
- Improved operational excellence through strengthened observability of system failures, batch process resource optimization, and enhanced platform resilience
- Mentored team intern, driving project deliverables to completion while fostering learning and professional growth

### Software Engineer - Data & AI Platforms

- Collaborated with internal Spark & customer teams to drive end-to-end **Spark 2** to **3** migration to 100% completion
- Improved release confidence and engineering velocity through +17% integration testing coverage enhancement and -85% average runtime reduction
- Built framework to derive privacy compliance policies to metric datasets based on data lineage and ownership
- Streamlined development pipeline by automating routine platform maintenance, cleaning up build artifacts for +40% accelerated release cycles, and deprecating unsupported features

### Software Engineer Intern - Big Data

- Improved data discoverability and lineage completeness by linking base table metadata to views in data catalog UI
- Enhanced metadata crawler workflow to derive field-level documentation for **Hive** views from underlying tables
- Evolved metadata **Java**-based API to fetch derived documentation to update 60% of views in data ecosystem

## Languages & Technologies

Java, Python, C++, SQL, Trino, Spark, Hadoop, Airflow, Hive, Openhouse(Iceberg), Git

## Education

### University of California, San Diego

B.S. IN BIOENGINEERING: BIOINFORMATICS

## Outreach/Organizations

UCSD Dept of Bioengineering, Neural Interaction Lab

2019 - 2021

### Research Assistant

- Characterized EMG signal data of laryngeal muscles during food consumption with **Python**
- Developed flexible and stretchable sensor prototypes for noninvasive EGG testing with microfabrication

Bluebonnet Data

2020

### Campaign Data Fellow

- Analyzed census data to provide insight into best use of campaign resources to secure funding and grow audience
- Improved outreach efficiency by analyzing VoteBuilder data to identify high-propensity donors and supporters

## Honors & Distinctions

Data Guard: A Fine-grained Purpose-based Access Control System for Large Data Warehouses 

2025

accepted to 2026 IEEE International Conference on Data Engineering (ICDE), contributor

Nexus Fellow - The D.E. Shaw Group undergraduate fellowship

2020