

Bo-Jian (Eddie) Ho

San Francisco, California • edwin.ho.bj@gmail.com • linkedin.com/in/eddie-bojian-ho • eddieho.xyz

AI & Data Engineer

AI & Data Engineer with 1+ years turning raw event streams into governed, ML-ready models at Chima and LESSO. Expert in dbt, Airflow, Snowflake, Spark and RAG pipelines. Built a multi-terabyte warehouse that trimmed ETL runtime, powering Looker dashboards used by 7 business units.

WORK EXPERIENCE

Chima
AI Engineer, Member of Technical Staff • Full-time

12/2024 – Present
San Francisco, CA

- Designed & deployed advanced AI agents, including ReAct, SQL-based, and self-improving agents, to automate end-to-end enterprise workflows across 10+ platforms, significantly enhancing operational efficiency and scalability.
- Investigated and adopted SOTA AI methodologies, such as chain-of-thought prompting, policy gradient optimization, and retrieval-augmented generation, to maintain a leading edge in AI-driven enterprise solutions.
- Led end-to-end initiatives from conceptualization to deployment, enhancing product performance and accuracy significantly through collaborative innovation and diligent execution.
- Directed process automation projects that successfully reduced operational costs and improved resource efficiency, demonstrating strong teamwork and the capability to drive complex projects to successful completion.

LESSO
Data Scientist • Full-time

05/2024 – 11/2024
Los Angeles, CA

- Orchestrated 25 Apache Airflow DAGs that ingested 5 GB/day of ERP CSVs and Shopify events into S3, staged them, and executed dbt-style ELT in Snowflake; slashed data-latency SLA from 3h to 2h and freed analysts from manual uploads.
- Embedded Great Expectations tests inside Airflow-orchestrated Python/SQL tasks, augmenting our dbt ELT to enforce 40+ column-level checks; sustained 99% validation pass-rate.
- Utilized optimized data architecture strategies, such as Snowflake schema and virtual warehouse within Snowflake, improve data query performance by 40% and reduce storage costs by 25%.
- Designed a normalized Snowflake schema plus dimensional marts; implemented RBAC and dynamic data-masking to meet PII & SOX controls; raised data-governance audit score from 80 to 98%.
- Championed data-contract reviews with Marketing & Supply-Chain teams; reduced ad-hoc SQL rework by 35% and accelerated dashboard rollouts in Looker.

UC San Diego
Student Engineer • Part-time

09/2023 – 03/2024
San Diego, CA

- Accelerated cloud-warehouse queries by 20% by tuning Amazon Redshift sort & distribution keys, adopting BigQuery time-partitioned clustering, and applying S3 performance-design patterns to trim I/O on 5TB fact tables.
- Built Airflow-orchestrated ELT and landed raw IoT logs into S3, auto-cataloged them with AWS Glue Data Catalog, then loaded curated layers into Redshift and BigQuery
- Embedded data-quality gates (AWS Glue Data Quality & Great Expectations) in every DAG task achieved 99.9% validation pass-rate and cut manual error fixes by 75%.
- Adopted the Medallion lakehouse pattern: Bronze, Silver, and Gold layers on AWS S3 and BigQuery; cut cloud-storage spend ~30% by standardizing columnar formats (Parquet & ORC) in Bronze and Silver layers.

EDUCATION

Bachelor of Science in Data Science
University of California, San Diego
Provost honors for 6 quarters, 3.8 overall GPA

San Diego, CA • 09/2021 – 03/2024

PROJECTS

CSES Website

08/2023 - 02/2024

UC San Diego

Collaborated with a team of student engineers to rebuild the UCSD Computer Science & Engineering Society web application (csesucsd.com) from scratch. Utilized React, MongoDB, Express.js, Node.js, and implemented unit and integration testing using Jest and Cypress to ensure reliability.

SKILLS

Machine Learning: Deep Learning, Keras, Neural Networks, NLTK, Pandas, Reinforcement Learning, scikit-learn, Supervised Learning, TensorFlow, Unsupervised Learning

Programming & Software: Agile Methodology (Expert), Docker (Intermediate), Git (Advanced), Javascript (Advanced), JIRA (Advanced), Python (Expert), R (Advanced), SQL (Expert), Typescript (Advanced)

Data Management: Cassandra, DynamoDB, MongoDB, MySQL, Neo4j, PostgreSQL, Spark

Web Application: Alembic, Angular, Django, Flask, NextJS, Node.js, Prisma, React, Redis, Restful API

Computer Vision: Image Segmentation, Object Detection, OpenCV, Pattern Recognition