

# Bo-Jian (Eddie) Ho

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

## 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 • San Francisco, CA • 12/2024 – Present**

##### **AI Engineer, Member of Technical Staff • Full-time**

- 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.
- Led R&D initiatives to implement SOTA fine-tuning techniques, including GRPO, SFT, and DPO, enhancing LLM accuracy by 40% while improving reasoning and task execution capabilities in distilled language models.
- 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 • Los Angeles, CA • 05/2024 – 11/2024**

##### **Data Scientist • Full-time**

- 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; 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, improved data query performance by 40% and reduced storage costs by 25%; designed a normalized Snowflake schema & 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.

### SKILLS

#### **Data Management:**

Airbyte, Airflow, Cassandra, Dagster, dbt, Delta Lake, DuckDB, Neo4j, PostgreSQL, Snowflake, Spark

**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)

### EDUCATION

#### **Bachelor of Science in Data Science**

University of California, San Diego  
San Diego, CA  
09/2021 – 03/2024

Provost honors for 6 quarters, 3.8 overall GPA

**UC San Diego • San Diego, CA • 09/2023 – 03/2024**

**Student Engineer • Part-time**

- 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.

## PROJECTS

**Langflow • 01/2025 – Present**

Langflow

Contributed to Langflow ([github.com/langflow-ai/langflow](https://github.com/langflow-ai/langflow)), an app builder for retrieval-augmented generation and multi-agent AI applications, by identifying and fixing bugs and implementing additional integration features that improved user experience.