

Scott Lee

Software Engineer

scott.lee.3898@gmail.com • scottjlee.github.io • San Francisco

Artificial Intelligence • Large-Scale Data Engineering • LLMs • Machine Learning

Education & Skills

Education

University of California, Berkeley 🐻

- **M.S. EECS** (2020)
- **B.A. Computer Science** (2019)

Skills

- **Python & Libraries:** Ray, PyTorch, TensorFlow, vLLM, HuggingFace, LangChain, Airflow
- **Other Languages:** SQL, Go, Java, R, JavaScript

Experience

[Anyscale](#) • Software Engineer (Ray Data) • San Francisco, CA • 2022 - Present

- Contributor to the **open-source Ray project** ([150+ commits](#)). Developed **core features for Ray Data General Availability**, including execution plan optimizer, data ingestion for distributed training, and observability.
- **Eng lead for new LLM workloads** at Anyscale, such as [text embeddings generation](#) and [LLM batch inference](#).
- Published [several blog posts](#) and presented a [talk at Ray Summit](#) to publicize and share our work.

[Lyft](#) • Software Engineer (Growth Platforms) • San Francisco, CA • 2020 - 2022

- **Redesigned two major components** in existing infrastructure for automated driver acquisition, efficiently scaling up marketing spend from COVID-shutdown to **\$5MM+/month marketing spend across three paid media channels**.
- **Drove multi-quarter, mission-critical initiatives** directly impacting key team OKRs, partnering with numerous other engineers and scientists in a highly cross-functional environment; leveraged and augmented team's core database of 30+ tables in actively planning the team's **short-term strategy** as well as **long-term team roadmapping**.

[RISELab](#) • Graduate Researcher • Berkeley, CA • 2018 - 2020

- Research areas: **Computer Vision** (Explainability, Few-Shot), **Medical Imaging** (EKG)
- Key work: [Neural-Backed Decision Trees](#)

[UC Berkeley](#) • Head Teaching Assistant • Berkeley, CA • 2017 - 2020

- **Managed a team of 50 TAs**, 60 tutors, and 150 lab assistants in orchestrating a **1300 student intro data science course**.
- **Led key infrastructure projects** to support scaling across data science department, including streamlining assignment development, autograding system, cheating detection, and large-scale course logistics.

Projects & Publications

More on my website: scottjlee.github.io

[NBDT: Neural-Backed Decision Trees \(ICLR 2021\)](#) • PyTorch

- Improving explainability for image classification in CNNs by generating a human-interpretable decision tree, which learns high-level concepts in images to make its decisions, rather than focusing on high-uncertainty attributes.

[BerkeleyTime](#) • JavaScript, Django, MySQL

- An augmented course catalog used by 30,000+ undergraduates at UC Berkeley.
- Provides a clean interface for serving course data, enrollment trends, grade distributions, and more.

[Fido](#) • Python, JavaScript, AWS

- A Slackbot that has a variety of features to assist teaching staff members, including roster lookup, Piazza paging, and groupshouts.