

Benjamin Liu

liuben398@gmail.com | 650-441-6068 | Palo Alto, CA | linkedin.com/in/benliu8 | github.com/bliu8
benliu.vercel.app

EDUCATION

University of Pennsylvania
B.S.E., Computer Science

May 2028
GPA: 3.97/4.00

PROFESSIONAL EXPERIENCE

AI/ML Engineering Intern | SailPoint Jun 2026 – Present

- Built a **LangSmith evaluation framework** for Harbor Pilot, SailPoint's production identity-security agent.
- Supported an enterprise identity security platform used by **53% of the Fortune 500**.
- Engineered a **LangSmith → Parquet → DuckDB** pipeline for trace analytics and SQL reporting.
- Migrated production analytics workloads to **Snowflake** for scalable, reliable reporting.
- Analyzed agent traces for **correctness, regressions, hallucinations, and multi-step failure modes**.

Software Engineer (Part-Time) | Aquara.ai Apr 2026 – Present

- Built an LLM tool-calling pipeline extracting **M&A transactions** from the web for a **500-company** backfill.
- Benchmarked **7 frontier models**, achieving **83% recall** overall and **89% recall** on major deals.
- Cut projected backfill cost **74%** (\$130→\$34) via model routing, prompt caching, and async **Batch API**.
- Engineered provider-agnostic extraction with source tiering, entity deduplication, and idempotent DB upserts.

Software Engineer (Part-Time) | Penn Labs Feb 2026 – Present

- Own the **Django/DRF backend** for Penn Mobile Games in a mobile app serving **10,000+** Penn students.
- Built REST APIs for daily puzzles, score submission, server-side validation, and per-day leaderboards.
- Launched Word Hunt with one-submission-per-user constraints and production leaderboard workflows.
- Wrote **pytest/Django** coverage and collaborated with PM, design, and frontend through production code review.

PROJECTS & RESEARCH

Chief, Real-Time Overhearing Voice Agent May 2026

- Placed **3rd among 80+ teams** at **Y Combinator's Call My Agent Hackathon**.
- Built **Chief**, a real-time overhearing agent that infers intent from live calls to act as an AI chief of staff.
- Architected an **event-driven Electron/Python pipeline** streaming mic capture through STT into live belief store.
- Engineered **salience gate** that abstains on ~90% of utterances, deciding act-vs-silent at **<1-second latency**.
- Offloaded retrieval, memory writes, and action planning to **concurrent async agents** kept off the live critical path.

ML Developer | Shen Lab, Perelman School of Medicine at Penn Oct 2025 – Present

- Co-developed a **ResNet-50** model to detect spectral unmixing errors in high-parameter flow cytometry.
- **Caught 82%** of expert-annotated errors on unseen patient samples, training only on synthetic data.
- Engineered an **end-to-end pipeline** (preprocess → inference → QC), enabling automation at scale.
- Deployed via Docker and Pennsieve, enabling department-wide access; **sole undergraduate** contributor.
- Presented at CYTO 2026 (international research conference) as **first author**.

TECHNICAL SKILLS

Languages: Python, C++, Java, TypeScript/JavaScript, Swift, SQL

Backend: Django, Django REST Framework, REST APIs

Data: PostgreSQL, Snowflake, DuckDB, Parquet, pandas, NumPy, Polars, PySpark

AI/ML: PyTorch, TensorFlow, scikit-learn, LangGraph, LangSmith, RAG, AI agents

Testing/DevOps: pytest, GitHub Actions, CI/CD, Docker