Jason Y. Yang

Product-oriented Software Engineer specializing in applied AI and scalable backend systems. Led high-impact LLM and diagnostic tools that grew ARR from \$1.3M to \$10M. Combines ML infrastructure and microservices expertise with fast-paced product delivery.

EDUCATION

Massachusetts Institute of Technology

B.S. in Computer Science and Engineering

Relevant Coursework: Machine Learning, Software Construction, Autonomous Decision Making, Complexity Theory, Software Studio

EXPERIENCE

Regard

Software Engineer

New York, NY Feb 2023 – Present

Palo Alto, CA

Jun 2022 - Aug 2022

Cambridge, MA

Aug 2018 - Feb 2023

- **Diagnostic Algorithm Builder**: Led backend design of a physician-facing no-code tool to author condition-based algorithms. Enabled a 92% reduction in dev effort per diagnosis and expanded coverage from 30 to 300+ live/tested conditions.
- LLM-Assisted Note Editor: Technical lead for backend infrastructure integrating LangChain-based AI agents across 5 stages (labs, meds, scribe text, etc.). Enabled 100+ generated notes across 8 alpha users in first two weeks of testing.
- **Extensibility & Compliance**: Designed a versioned note model with text attribution and rollback support for auditability. Supported custom physician styles using historical notes.
- **Ecosystem Upgrade**: Pushed successful migration from a Flask monolith to modular FastAPI/Postgres/Alembic stack now used by 8+ services. Authored scaffolding used org-wide.
- **Developer Tooling**: Built internal DB mirroring tool using Doppler CLI and containerized workflows, widely adopted by engineering teams to test production-like states.
- **Culture & Leadership**: Drove key initiatives proactively, mentored engineers, and frequently led high-priority cross-functional launches. Repeatedly recognized for balancing velocity and quality.

Impact: Diagnoses scaled from 2M to 7.5M, ARR grew from \$1.3M to \$10M, and team supported transition from \$20M Series A to \$61M Series B.

J.P. Morgan Chase & Co.

ML Engineering Intern

- Entity Deduplication: Developed NLP-based matching pipeline for 200M+ business records, identifying 8M duplicates in internal intelligence systems.
- **Heuristic Matching**: Proposed hierarchical address-distance heuristics to enable fast approximate comparisons across skewed datasets.
- Scaling + Tooling: Used Transformers, Scala, and multiprocessing to handle 200M+ API requests; placed 3rd in intra-firm ML hackathon.

Projects

- Explainable AI in Insurance: Co-authored capstone paper on XAI policy for underwriting/claims, presented to NAIC. Advised by Prof. Hal Abelson.
- **Production Data Tooling (Internal)**: Created make-based tooling to safely mirror production DB states locally for testing migrations. Adopted org-wide across backend teams.

TECHNICAL SKILLS

Languages: Python, Typescript, Java, SQL

Frameworks: FastAPI, LangChain, PyTorch, React, RTK Toolkit

Tools: Docker, Alembic, SQLAlchemy, AWS, OpenTelemetry, LangFuse, Doppler

Concepts: LLMOps, agentic workflows, RAG pipelines, auditability, extensibility, prompt engineering