

Anu Sihra

San Francisco, California, USA | anupreetsihra5@gmail.com | in/Anupreet-Sihra | GitHub

EXPERIENCE

Labelbox | *Forward Deployed Engineer*

May 2026 - Present

- Embedded with frontier-lab customers as engineering point of contact, coding side-by-side with their teams to turn AI agent failure modes into human-in-the-loop eval and preference-data programs that shipped 10,000+ expert-rated comparisons.
- Built and shipped a full-stack AI agent evaluation platform integrating customer systems via REST APIs, orchestrating sandboxed agent runs, LLM-as-judge and human grading, and guardrails catching hallucinated or off-policy responses.
- Designed evals and deterministic graders measuring agent resolution rate, answer correctness, and failure recovery, surfacing model trade-offs and limitations as measurable quality benchmarks across 180+ audited rollouts.
- Rapidly prototyped and shipped a computer-use agent pipeline that generated tasks, executed agents, and captured full trajectories (actions, screenshots, step-level rewards), delivering 500+ structured examples to improve agent reliability.
- Partnered with Sales, Success, and Product to drive adoption, presenting eval results to technical and non-technical stakeholders and codifying recurring customer needs into product feedback and reusable eval playbooks.

Paycom | *Software Engineer, AI* | *AI Software Engineer Intern (May 2024 - Aug 2024)*

Jan 2025 - Feb 2026

- Served as the engineering point of contact for enterprise clients, partnering with Product and Support to triage client-reported issues and ship tailored fixes, and turning recurring customer feedback into platform features across scheduling and payroll modules serving 100,000+ client accounts.
- Shipped AI-driven automation for enterprise labor-budget auditing - multi-step MCP tool calling over scheduling and payroll data to surface discrepancies - cutting manual auditing effort 30%.
- Built and scaled backend REST and GraphQL APIs for scheduling workflows - shift assignments, shift exchanges, and real-time labor cost tracking - with SQL query optimization that cut data-fetching overhead 40% under enterprise load.
- Re-architected the Scheduling Import Center with rebuilt React upload flows and server-side bulk validation, cutting import time for 50,000-row enterprise files from 3+ minutes to under 30 seconds.

Barclays | *Software Engineer (BA3)*

Aug 2022 - Aug 2023

- Built global tax applications (React, C#/.NET, SQL Server) computing charges across APAC, UK, and US regions; resolved 15 critical production defects with zero-downtime releases across year-end cycles.
- Automated the Year-End Tax Longview Reporting Pack using RESTful APIs and batch pipelines, cutting report generation from 3-4 hours to 40-50 minutes (80% faster).
- Executed multi-region tax-data migrations across transactional databases with automated validation, completing cutovers 87% faster than the prior process.

PROJECT

Avals - Local-First AI Agent Evaluation Platform | github.com/anupreet/avals

- Built a local-first AI-agent evaluation platform (FastAPI, React 19) that launches any agent x model x task combination and live-scores trajectories on reward, cost, and correctness - 10+ concurrent eval jobs streamed over one multiplexed WebSocket.
- Engineered a durable sequence-numbered JSONL event log: every run survives server restarts with zero event loss, and clients replay full trajectories from any offset.
- Generated TypeScript clients from the OpenAPI spec - 100% of the API surface type-checked end-to-end - with a trajectory dashboard rolling up reward, cost, and token metrics per task.

HyperFrames Video Studio - Local AI Video Production Pipeline | github.com/anupreet/hyperframes-video-studio

- Contributed to a fully local AI video studio (React/Vite, Express/TypeScript) that turns a script into a rendered 1080p MP4 - voiceover, captions, animated scenes - with 0 cloud API calls end-to-end.
- Implemented pipeline stages spanning Kokoro TTS narration, Whisper caption alignment, and 30 animated scene templates, with ffmpeg renders streamed live over SSE - script to final cut in under 10 minutes.

EDUCATION

Rice University | *Master of Computer Science, Computer Science* (GPA: 3.9/4)

Houston, Texas, USA Aug 2023 - Dec 2024

Banasthali University | *Bachelor of Technology, Computer Science and Engg* (GPA: 8.97/10)

Jaipur, India Jul 2018 - Jun 2022

SKILLS

Languages: Python, TypeScript, JavaScript, SQL, Java, C#/.NET

Frameworks & APIs: React.js, Node.js, FastAPI, REST APIs, GraphQL, WebSockets, OpenAPI

Cloud & Data: AWS, Docker, Kubernetes, PostgreSQL, CI/CD (GitHub Actions, Jenkins)

AI & LLM: AI Agents, Agentic Workflows, Tool/Function Calling, MCP, LLM Evals (LLM-as-Judge), AI Guardrails, RAG, Vector Databases, Prompt & Context Engineering, Conversational AI, RLHF & Preference Data, LLM APIs (OpenAI, Anthropic)