

Rishabh Sai

rishabhsaiv@gmail.com | +1 (814) 954-9156 | [linkedin.com/in/rishabh-sai](https://www.linkedin.com/in/rishabh-sai)
[rishabhsai.com](https://www.rishabhsai.com) | Newark, CA

F-1 student eligible for internships through CPT, which needs no visa sponsorship from the employer

Education & Certificates

- B.S. in Computer Science from Pennsylvania State University in PA Expected May 2028
- B.S. in Data Science from Indian Institute of Technology Madras (online) Expected May 2028

Work History

Technical Intern at Cisco, San Francisco, CA

June 2026 to September 2026

- Built software that acts like real Cisco data center hardware, using Go, Python, Linux, and containers, so engineers could test cloud networking features without physical machines.
- Extended the production simulator in Go and Python for Cisco Nexus One – Hyperfabric to report temperature, power, and hardware health data for switches, servers, and BMCs by testing each device's data format against production collectors.
- Added live port splitting to the simulator's virtual Linux network by pre-creating interfaces and writing a watcher that enables them on demand, with 0% dropped traffic and no restarts.
- Used AI coding agents to build parts of the project in parallel, checking each with automated tests, end-to-end tests, and code review, which cut the planned timeline by 60%.
- Worked with 7 engineers across software, networking, and hardware teams to design and deliver server support, writing the proposal, leading design reviews, and shipping Linux, Kubernetes (K3s), and container profiles.

AI/LLM Intern at Mashreq Bank, Dubai, UAE

June 2025 to August 2025

- Built AI tools for a large bank using Python, LLMs, and React that helped employees review code, policies, and documents faster.
- Built 3 internal AI products for code optimization, policy checks, and document quality by connecting LLMs to the bank's internal policy documents with RAG, improving review and compliance work.
- Launched an MCP server that gave the bank's internal AI tools one shared API, so teams could reach every tool from one place.
- Built an intern operations portal using React, FastAPI, and PostgreSQL so intern tasks and information lived in one web app.

Projects

Atrax – Cloud Platform for AI Agents

atrax.run

- Built a cloud platform in TypeScript where AI coding agents can put full web apps online by themselves, each with a public link and its own SQL database.
- Designed one config file that lists an app's deploys, SQL databases, and database migrations,

so agents could ship and update apps without manual setup on any cloud provider.

- Received \$20K in AWS cloud credits as funding to host and grow the platform.

NotionBench – Benchmark for AI Coding Agents

notionbench.rishabhsai.com

- Built an open-source tool in TypeScript and Docker that tests how well 8 AI coding agent setups finish 38 real developer tasks, to compare their quality and cost.
- Graded 890 agent runs automatically by running each result's code and checking it through APIs, so scores did not depend on human opinion.

ByRight PGH – Housing Lot Finder (AI Housing Hackathon)

byright-pgh.vercel.app

- Built a web app in TypeScript and Python that finds which of Pittsburgh's vacant city lots can fit small homes by checking zoning rules against public parcel data and cost estimates.
- Won a \$10K individual cash prize at the AI Strike Team PA Housing Hackathon.

Briefly – Newsletter App (Nas Company AI Residency)

- Built Briefly, a newsletter product for online creators, as 1 of 11 residents in an AI startup program, taking it from user research through launch.

AI Research – LLM Memory and Medical Imaging

hf.co/rishabhsai

- Wrote PyTorch code in Python that shrinks the memory an LLM uses while generating text (MiniCache KV-cache compression), with tests and benchmarks to check accuracy.
- Built a live demo that uses a pretrained ML model to classify retinal eye scans, highlight the area it based its answer on, and draft a short report.

Open Source Contributions

github.com/rishabhsai

- Fixed bugs in Go, Python, and C++ that were merged into Kubernetes, vLLM, Kubeflow, PyTorch, Hugging Face, Google DeepMind's MuJoCo, and the MCP Go SDK through maintainer code review.

Leadership

Campus Lead at OpenAI, State College, PA

- Explained ChatGPT and AI coding tools to technical and non-technical students by running workshops and studio hours as president of Penn State's OpenAI Student Collective.

Skills

- Languages: Python, Go, C++, SQL, TypeScript, JavaScript, HTML, CSS
- Tools: PyTorch, LLMs, RAG, AI agents, FastAPI, React, REST APIs, Docker, Kubernetes, PostgreSQL, Redis, AWS, Azure, Git, Linux, Bash
- Interests: Poker, golf, 3D printing, entrepreneurship, homelab and self-hosting, investing, Pokémon/TCG collecting, hiking