

Shreyash Hamal

San Francisco, CA | +1 (415) 361-7070 | shreyash.hamal@gmail.com | [Linkedin](#) | [Website](#) | [Github](#) | [Devpost](#)

Education

University of San Francisco

San Francisco, CA | Aug 2021 - May 2025

Bachelor of Science in Computer Science, Graduating May 2025

Skills

Languages: Python, TypeScript, JavaScript, SQL, Bash, Go, C, C++

Frameworks & APIs: FastAPI, Next.js, React, Node.js, Express.js, REST APIs, GraphQL

AI & Data: LLM integrations (Gemini / Vertex AI, OpenAI), Agno multi-agent workflows, embeddings, vector search, Weights & Biases Weave, reinforcement-learning-style optimization

Cloud & Security: AWS (Cognito, EC2, S3, Bedrock), OAuth 2.0, JWT / JWKS validation, RBAC, Semgrep / security scoring in ML loops

Data & State: Convex (schema, realtime sync, jobs), MongoDB, Redis, SQLite, ETL-oriented pipelines

Tools & Practices: Docker, Kubernetes, Git, CI/CD, pytest, unit and integration testing, observability, technical documentation

Highlighted Experience and Projects

Head of technology | PersistOS

San Francisco, CA | Aug 2025 - Feb 2026

- Built production-style AI memory backend in FastAPI with async lifecycle management, graceful shutdown, and validated embedding APIs feeding vector similarity search.
- Shipped modular API surface with SSE streaming chat, authenticated streaming proxy on Next.js App Router, and background job orchestration for multi-entity intelligence processing
- Integrated Convex for schema-driven app state, embeddings, conversations, and job persistence; delivered automated tests across API, concurrency, failure recovery, CORS, and orchestration.

Chief Technology Officer | DivertissementAI

San Francisco, CA | Mar 2025 - Aug 2025

- Developed full-stack AI memory platform with React/Next.js and FastAPI backend.
- Designed a modular API framework that integrates evolutionary algorithms and LLM-driven optimization logic for automated system self-improvement.
- Implemented robust data persistence layers and vector store integrations, enabling low-latency state management for real-time AI context processing.

Backend Developer | Consilient Labs

San Francisco, CA | Aug 2024 - Dec 2024

- Built and deployed secure REST APIs in Node.js utilizing AWS Cognito for enterprise-grade role-based access control (RBAC).
- Designed scalable system architecture and a reverse proxy solution to ensure resilient dynamic routing and high availability.
- Managed CI/CD pipelines and automated testing workflows, reducing deployment errors and improving platform reliability.

HACKATHON WINS

Winner, Reinforcement Learning Track | Weavehacks-2 (Google Cloud)

- Self-Improving Agents: Developed an API optimization framework using Google Cloud to automate system tuning.
- Automated the generation of training data through self-improving policy networks to ensure system convergence toward optimal behavior.

Winner Best Use of Semgrep | AWS AI Agents Hackathon

- Built a modular agent framework on AWS, focusing on infrastructure scalability and seamless API integrations.
- Utilized AWS serverless components to manage real-time data ingestion and processing for autonomous agent societies.