Anirudh Herady

602-793-8590 | aherady@asu.edu | linkedin.com/in/anirudhherady | github.com/avh17 | anirudhvhportfolio.vercel.app

Professional Summary

Full-stack engineer specializing in scalable backend systems and AI applications. Delivered production microservices for video streaming and intelligent chatbot solutions. Excel in fast-paced, high-impact technical environments.

Education

Arizona State University Tempe, AZ
M.S. Computer Science – GPA: 3.78/4.0 *Aug. 2024 – May 2026*
Manipal Institute of Technology Manipal, India
B.Tech. Computer and Communications Engineering *Jul. 2018 – Jul. 2022*

Technical Skills

Languages: Python, JavaScript, TypeScript, SQL, C/C++, Bash
Backend Engineering: FastAPI, Node.js, Flask, RESTful APIs, JWT, OAuth2, GraphQL, Microservices Architecture
Cloud & DevOps: AWS (Lambda, EC2, S3, SQS, ECR), Docker, CI/CD (Github Actions), Linux
AI & Machine Learning: RAG (Retrieval-Augmented Generation), LangChain, LangGraph, LLMs (GPT, Gemini), Vector Embeddings
Databases: PostgreSQL, MongoDB, Redis, SQLite, SQLAlchemy ORM, PostGIS
Frontend: React.js, Next.js 14, Tailwind CSS, Streamlit

Professional Experience

Software Intern Jul. 2025 – Aug. 2025
Potters Tech *Remote*

- Solely designed and implemented the backend for a location-aware LLM chatbot, supporting 10+ natural-language query types (proximity search, rating-based ranking, category filtering).
- Re-architected a hardcoded chatbot into a Retrieval-Augmented Generation (RAG) system, reducing hallucinated responses by 60% by grounding outputs in OpenStreetMap data.
- Ingested and indexed 100k+ OpenStreetMap POIs into PostgreSQL + PostGIS, enabling sub-200 ms radius-based spatial queries.
- Built multi-step LLM workflows using LangChain and LangGraph, enabling conversational state and multi-turn query refinement across sessions.

Associate Software Developer Aug. 2022 – Sept. 2023
Valtech India *Bengaluru, India*

- Built and owned the backend for a multi-tenant video streaming CMS over 7 months, supporting 3–5 client platforms (e.g., news and sports) from a single shared codebase.
- Designed a configuration-driven architecture using JSON-based tenant settings, reducing new client onboarding effort by 60–70% by eliminating client-specific code changes.
- Developed 15+ RESTful APIs using FastAPI, leveraging async request handling, dependency injection, and Pydantic validation to deliver low-latency, well-documented endpoints.
- Implemented JWT-based authentication with OAuth2, securing 100% of CMS and administrative endpoints and enabling role-based access control across tenant environments.
- Designed and managed data storage using PostgreSQL and MongoDB, modeling and querying 10+ relational entities while supporting flexible, schema-less tenant configurations.
- Containerized backend services using Docker, improving local setup time by 50% and ensuring consistent environments across development and staging.

Projects

Distributed Face Recognition System | *AWS (Lambda, EC2, SQS, S3), Python, Docker*

- Designed two serverless architectures (Lambda vs EC2) for video frame processing that achieved 99%+ recognition accuracy at <3s latency under 100+ concurrent requests, providing cost-performance analysis for production deployment.
- Developed a Python autoscaler that dynamically provisions 0-15 EC2 instances based on SQS queue depth, reducing infrastructure costs by 45% during off-peak hours while maintaining performance during traffic spikes.
- Ensured 99.5% system reliability during burst traffic by decoupling ingestion and processing layers with SQS, successfully handling 10x traffic surges without service degradation.

StoreIt: Cloud Storage Solution | *Next.js 14, TypeScript, Appwrite*

- Built a full-stack file management platform with secure authentication and storage using Appwrite, supporting 50+ concurrent users with real-time file upload/download capabilities.
- Developed 20+ reusable React components with dynamic storage analytics charts, reducing UI development time by 40% for future features and improving user engagement by 25%.