

Adarsh Sahu

+91-9555212727 | adarshsahu1077@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

EDUCATION

National Institute of Technology Karnataka, Surathkal
Bachelor of Technology in Information Technology

July 2020 – June 2024
Mangalore, India

EXPERIENCE

HCLTech

Senior Software Engineer

July 2024 – Present
Bangalore, India

- Cut analyst regulatory review from **~45 min to under 2 min** as **sole engineer** on a LangGraph multi-agent pipeline over Azure OpenAI GPT-4.1, with Pydantic-enforced structured extraction and drug-mention detection triggering automated FDA approval lookups and ClinicalTrials competitor searches persisted to Azure SQL.
- Scaled regulatory content ingestion to **1,000+ articles/day** at **sub-90s latency** by engineering a 3-tier LangGraph pipeline (config-match → Selenium → BeautifulSoup) across **50+ agencies**, extended with PyPDF2 and Azure Document Intelligence for PDF coverage and URL delta-detection.
- Eliminated **silent pipeline failures** across all monitored sources by deploying an APScheduler cron system with automated per-job metric tracking in Azure Cosmos DB, atomic SQL cleanup on both success and error paths, and per-URL plus job-summary email alerts via Azure Communication Services.
- Achieved **89.9% F1** on AML/PEP risk classification and **91.4% straight-through approval rate** by designing and deploying an agentic KYC platform using CrewAI, LangChain, and FastAPI to orchestrate multi-agent document parsing, risk scoring, and case routing with full audit observability.

PROJECTS

Intervyu

AI voice interview platform with live code execution, CV analysis, and performance reports

Oct 2025 – Present

[Live](#) | [GitHub](#)

- Built a real-time voice pipeline with **Deepgram STT**, **Claude Haiku 4.5** via **AWS Bedrock**, and clause-split **Azure TTS** streaming MP3 over WebSocket mid-generation → cut first-audio latency from **~6s to ~1.7s**.
- Parallelized STT and session fetch via `asyncio.gather`, replacing HTTP clients with a persistent `httpx` singleton, switching to `converse_stream` → reduced Deepgram latency from **3726ms to 620ms**.
- Deployed a cloud-native backend on **AWS EC2 + Lambda** (code execution, CV analysis, scoring) with **Supabase** PostgreSQL and JWT-authenticated WebSockets → **\$0.76 vs \$3.15/session** on Retell.

Epistlo

Custom RFC-compliant mail server stack with real @epistlo.com mailboxes

Aug 2025 – Feb 2026

[Live](#) | [GitHub](#)

- Built **asyncio SMTP/IMAP servers** implementing RFC 5321/3501 handshakes, MIME parsing, and dot-stuffing, backed by distributed three-service **FastAPI** architecture and **Supabase** for @epistlo.com delivery.
- Engineered fire-and-forget async email dispatch with automatic **DB rollback on failure** and parallel attachment processing via `asyncio.gather`, cutting API response latency from **5–10s to <200ms**.
- Implemented **three-tier graceful degradation** across **Elasticsearch** → Supabase ILIKE for search, **Redis** → DB for caching, and **S3** → local disk for attachments, ensuring full functionality with zero hard service dependencies.

TECHNICAL SKILLS

Languages: Python, Java, JavaScript, TypeScript, SQL

Backend & Web: FastAPI, Node.js, Express, Next.js, React

Databases & Messaging: PostgreSQL, Azure SQL, DynamoDB, Redis, RabbitMQ, Vector Databases

GenAI & ML: LangChain, LangGraph, CrewAI, OpenAI APIs, Ollama, PyTorch, OpenCV

Cloud & DevOps: AWS (Lambda, S3, API Gateway), Azure (AI Search, Blob Storage), Docker, Kubernetes

CERTIFICATIONS & ACHIEVEMENTS

AWS Certified Data Engineer – Associate

[Link](#)

AWS Certified Machine Learning Engineer – Associate

[Link](#)

Google Cloud Professional Data Engineer

[Link](#)

LeetCode **Knight** – Rating **2035**, 1000+ problems solved

[Link](#)

Brazilian Jiu-Jitsu **National Gold Medalist**; State-level **Basketball Team Captain**