

Aditya Dey

Backend Developer — AI

Kolkata, India | +91 79085 31553 | adey9217x@gmail.com | linkedin.com/in/aditya-dey | github.com/AditHash | adithash.in

PROFESSIONAL SUMMARY

Backend engineer with **2.5 years** of experience building and shipping production systems on AWS, REST APIs, agentic and RAG pipelines, and LLM-integrated services designed for reliability at scale. Comfortable across the full delivery cycle: system design, backend implementation, evaluation against ground truth, and containerised deployment. AWS Certified Machine Learning Engineer.

PROFESSIONAL EXPERIENCE

Backend Developer — AI

Dec 2023 — Present

Workmates Core2Cloud Solutions Limited · Kolkata, India · AWS Premier Tier Consulting Partner

- Medical Insurance Claims Extraction & RAG Platform:** Took a claims-document platform from R&D to production over **2M+ documents** and **420K+ claims**: benchmarked five extraction approaches (Textract, vision-LLM, multimodal Mistral, agentic) against hand-validated ground truth before selecting the pipeline, then scaled it on ECS. Built a hallucination-resistant custom pipeline that truncates fabricated observations and enforces tool-verified answers for natural-language claims queries.
- Intelligent Document Processing Platform:** Built multimodal extraction for complex financial PDFs each page processed as rendered image + text layer through Claude on Bedrock, removing per-layout template configuration. Shipped per-page results, per-page confidence scoring with Textract OCR fallback, async jobs, and ChromaDB-backed RAG chat over extracted documents; backend covered by automated pytest.
- WhatsApp Microfinance Banking Chatbot:** Shipped a production text + voice banking bot for a microfinance lender in **11 Indian languages** (Gupshup API, AWS Transcribe, Sarvam TTS), pairing a deterministic state machine for loan/branch flows with a Strands agent for free-form RAG Q&A. Closed **all UAT-reported bugs**, including redaction of borrower PII from all LLM context and enforcing exact database values to prevent hallucinated financial data.
- Multilingual Video Intelligence Chatbot:** Built a video-RAG web app for a media company where an LLM intent router splits queries between a low-cost pgvector transcript pipeline (HNSW, Titan embeddings) and multimodal video analysis via TwelveLabs Pegasus on Bedrock avoiding the expensive model on every message. End-to-end support for 11 Indian languages with gemma models with translation fallback chains.
- Insurance Policy Recommendation Engine:** Engineered hybrid retrieval (pgvector semantic search + re-ranking) over **11K+ policy records** at low latency, secured with authenticated, rate-limited APIs and shipped fully containerised via Docker Compose.

SKILLS

Languages: Python, SQL

Backend: FastAPI, Flask, REST APIs, Microservices, Pydantic, pytest

AI/ML: RAG systems, vector search & embeddings, agentic systems (Strands Agents, LangChain), LLM routing & guardrails

Databases: PostgreSQL, pgvector, MongoDB, Qdrant, ChromaDB, MySQL

Cloud & DevOps: AWS, Docker, PM2, LangSmith

EDUCATION

Techno India University, Kolkata

GPA: 8.46 / 10

B.Tech — Computer Science & Engineering, 2024

CERTIFICATIONS & AWARDS

AWS Certified Machine Learning Engineer – Associate	Dec 2024
AWS Certified AI Practitioner	Sep 2024
MongoDB SI Associate	Sep 2025
AWS Certified Cloud Practitioner	Feb 2024
Rising Star — AI & AI Innovation Spark — Workmates Annual Business Summit	2026