

Aabid Mohamed A

Bengaluru, India | +91-70255 25039 | aabidmohamed2003@gmail.com | GitHub | LinkedIn

SUMMARY

Associate Software Engineer with experience building and deploying production AI systems on live banking infrastructure at PwC, including NLP entity extraction pipelines, ML classifiers, and agentic research workflows. Combines GenAI application development, prompt engineering, and backend engineering to take systems from concept to production. Strong track record of rapidly mastering new technologies and delivering high-impact solutions in fast-moving environments. Seeking Backend, AI Engineer, and GenAI Platform roles focused on building scalable intelligent systems.

EXPERIENCE

Associate Software Engineer

Aug 2025 – Present

PwC — Client: SMBC (Sumitomo Mitsui Banking Corporation)

- Eliminated a fully manual workflow for 25 salespeople by engineering an end-to-end agentic deal platform — async microservice pipeline processing ~50 deals/day at sub-60-second latency across ingestion, extraction, classification, and approval stages.
- Boosted downstream ML extraction quality 30%+ by building a chat ingestion and preprocessing pipeline that normalises raw deal conversations into retrieval-ready structured representations, enabling clean downstream NLP and vector-database-compatible document storage.
- Removed 100% of manual entity-tagging by developing a prompt-driven NLP extraction layer (investor name, tranche details, deal size, terms) with per-call evaluation logging to catch extraction drift and hallucination risk before downstream scoring.
- Cut irrelevant blotter entries ~40% by implementing an ML deal-likelihood classifier scoring execution probability before deals reach the sales desk.
- Sustained 25 concurrent users at zero downtime by building the blotter backend API with approval/rejection workflows, state persistence, and full audit logging.
- Saved ~2 hours/day in status-check overhead by developing a live potlist view aggregating all approved deals in real time for the sales desk.

Machine Learning Intern

Jan 2025 – Feb 2025

SchoolHack.AI

- Slashed per-query latency to under 3 seconds by building a multi-source research agent orchestrating web search (Tavily API) and professional data retrieval (Hunter API), delivering structured outputs at scale.
- Reduced manual SDR prospecting time ~60% by developing a LinkedIn and Gmail outreach agent automating contact discovery and personalised message generation.
- Tripled system throughput by exposing both agents via async FastAPI endpoints, enabling concurrent query handling without blocking.

PROJECTS

RenewAI — Insurance Policy Renewal Engine ↗ GitHub

Python · LangGraph · Gemini · ChromaDB · FastAPI · SQLite · Streamlit

- Automated renewal outreach for ~500 policies/day across 5 agent layers (strategic, execution, quality, learning, escalation) with a <2% human escalation rate and 100% IRDAI audit compliance.
- Eliminated hallucinated policy terms by engineering a RAG pipeline grounding every Gemini prompt in a 170-document ChromaDB vector database (FAQs, objection scripts, IRDAI rules) with a TF-IDF keyword fallback for hybrid retrieval.
- Blocked non-compliant outbound messages at 0% bypass rate by building a 5-node quality gate — Critique → Safety → Compliance → Sentiment → Scorer — routing sub-70-score messages to human specialists.
- Achieved full LLM cost and latency observability by instrumenting a per-call token tracker across Gemini, ElevenLabs, Twilio, and Razorpay with a SHA-256 append-only audit chain and a Rs.500/day budget alert.
- Delivered 206 unit tests passing in ~8s by implementing a mock-first agent architecture, enabling CI/CD with zero external API keys.

FinCore Banking Assistant ↗ GitHub

Python · LangGraph · FastAPI · Neo4j · FastMCP · Next.js · SSE

- Achieved P90 latency under 400ms across 8 end-to-end banking chat scenarios by building a LangGraph multi-agent system routing queries across Account, Loan, Fraud, and Compliance specialist agents.
- Reduced account and customer lookup times ~3× vs. equivalent SQL joins by implementing a Neo4j knowledge graph (50 customers, 120 accounts, 40 loans) with multi-hop Cypher traversal for relational queries.
- Eliminated 100% of runtime validation errors across inter-service boundaries by defining typed Pydantic schemas across 4 isolated FastMCP servers (core_banking, credit, fraud, compliance) over SSE transport.
- Maintained full auditability across all pipeline stages by adding LangGraph checkpointing, structured audit logs, and SSE streaming endpoints delivering live tool-call events to the Next.js dashboard.

Paint Intelligence Platform — Dealer CRM ↗ GitHub

Node.js · Express · SQLite · Docker · DeepLab · Playwright

- Shipped a full-stack multi-tenant SaaS platform from zero: JWT-authenticated REST backend (15+ endpoints), offline-first frontend with server sync, and Dockerised deployment with Prometheus monitoring and GitHub Actions CI.
- Cut wall-color decision time to under 60 seconds by engineering a real-time ML recolor engine fusing DeepLab v3 semantic segmentation with a custom HSL region-growing algorithm for accurate live wall masking.
- Enabled accurate shade recommendations across 63 SKUs from 4 brands (Asian Paints, Dulux, Berger, Nerolac) by building a catalog search and dominant-color extraction pipeline.
- Achieved zero data loss on poor connectivity by architecting offline-first lead capture with graceful server sync, localStorage fallback, and a 600-event analytics engine tracking full dealer funnel KPIs.
- Validated full user flow quality with automated Playwright E2E tests covering photo upload through lead export, catching regressions across ~15 critical paths.

TECHNICAL SKILLS

Languages	Python, JavaScript, TypeScript, C++
Backend	FastAPI, Node.js, Express, REST APIs, gRPC, AsyncIO, WebSockets, SSE Streaming, Redis, Kafka
Databases	PostgreSQL, SQLite, Neo4j, ChromaDB, Qdrant (vector databases)
Infrastructure	Docker, Kubernetes, GitHub Actions CI/CD, Prometheus, Grafana, Linux, Git
AI / ML	LangGraph, LangChain, LlamaIndex, RAG Pipelines, Hybrid Retrieval, MCP, Multi-Agent Systems, LLM Evaluation, Prompt Engineering, NLP Extraction, OpenAI API, Gemini API
Cloud	AWS, GCP
Testing	Playwright (E2E), pytest, API smoke testing

EDUCATION

National Institute of Technology Calicut
B.Tech in Computer Science

2021 – 2025

LEADERSHIP & ACTIVITIES

- Joint Secretary, Enquire (NIT Calicut):** Drove 20,000+ school students statewide, organising Samasya, Kerala's largest inter-school science competition.
- Chief Link Representative, IEEE NIT Calicut:** Reached 10,000+ participants across 40+ events, earning Best Student Branch in India.
- Public Relations Lead, Tathva & Ragam:** Managed PR for NIT Calicut's flagship technical & cultural festivals, each drawing 30,000+ attendees annually.