

ABHISHEK KUMAR TRIPATHI

Software Engineer | AI Agentic Systems • LLM Evaluation • Backend Platform Engineering
+91 94664 92956 • abhishek27198@gmail.com • linkedin.com/in/abhishektripathi27 • github.com/Codemonk101

SUMMARY

Software engineer with 5+ years building production backend systems, now specialising in agentic AI. Core engineer on a LangGraph-based multi-agent platform automating KYC and Source-of-Wealth verification for a global private bank, where I own the evaluation platform that measures those agents — scoring completion accuracy, hallucination rate, and confusion matrices against curated ground truth. Built test-first and configuration-driven, and equally at home in the layer underneath: FastAPI services, Azure/AKS, Kubernetes, and CI/CD. I care about AI systems that are measurable, not just demoable.

EXPERIENCE

EPAM Systems

Jun 2025 – Present

Software Engineer • Client: Global Investment Bank (UBS)

Pune, India

Multi-Agent KYC & Source-of-Wealth Platform — *LangGraph, LangChain, Python, FastAPI, Azure*

- **Core engineer on a team building a multi-agent orchestration system in LangGraph** that automates KYC and Source-of-Wealth verification for high-net-worth individuals, decomposing an analyst workflow into specialised agents for document ingestion, entity extraction, cross-source corroboration, and adjudication.
- **Built the document-understanding pipeline** that extracts structured client and asset data from heterogeneous unstructured sources, reconciling conflicting evidence into a single auditable wealth provenance record.
- **Implemented stateful agent graphs with deterministic control flow** covering tool routing, retries, human-in-the-loop checkpoints, and failure isolation, so a single agent failure degrades gracefully instead of collapsing the run.
- **Delivered RAG retrieval and generation services** on the bank's internal AI platform, owning the ingestion, chunking, embedding, and retrieval layers end to end.

Agent Evaluation Platform — *LLM-as-judge, ground truth datasets, confusion matrices*

- **Built an evaluation agent and platform that scores other agents** against curated ground-truth datasets, producing completion accuracy, hallucination rate, and per-field confusion matrices — turning subjective "the agent seems better" debates into regression-tested numbers.
- **Designed the framework as a configuration-driven system** using design patterns (Strategy, Factory, Observer) and SOLID principles — new metrics, agents, and datasets are onboarded by declaring config rather than writing code, keeping the scoring core untouched.
- **Developed test-first (TDD)** with unit and integration coverage over scoring logic and dataset loaders, so a platform that certifies other systems' correctness is itself verifiable.
- **Lifted measured completion accuracy from 76% to 92% and cut hallucination rate by 45%** by feeding eval findings back into prompt, retrieval, and agent-routing changes across successive benchmark cycles.

Firmwide Agent Development SDK — *Python library, internal open source*

- **Contributed to the firm's shared AI agent SDK** — the standard library used by multiple engineering teams across the bank to build and ship agents, turning per-team bespoke implementations into one supported abstraction.
- **Designed for adoption by other engineers** — stable public interfaces, backwards-compatible releases, and documentation, treating internal developers as the customer.

Platform Engineering & Mentoring — *Azure, AKS, Docker, GitLab CI/CD*

- **Worked hands-on with deployment on Azure and AKS** — containerised services with Docker, made changes to the staging CI/CD pipeline, and built working knowledge of Kubernetes workloads and Git-based build, test, and release automation.
- **Mentored junior engineers** — reviewed their pull requests with design and code-quality feedback, and distributed work by criticality and complexity to match each engineer to the right task.

Blenheim Chalcot IT Services Pvt. Ltd.

Nov 2024 – Jan 2025

Python Developer, Specialist

Mumbai, India

- **Architected and deployed a scalable LLM-based healthcare agent** using FastAPI, AWS Step Functions, and Lambda, delivering intelligent healthcare services for the UK public sector.
- **Cut request processing time from 292s to 63s (4.6x)** by introducing multithreading, materially improving system responsiveness and reliability.
- **Reduced average issue resolution time by 40%** by integrating Sentry for real-time observability and error tracking; tuned database connection pooling to improve query latency under load.
- **Accelerated release cycles by 60%** with CI/CD pipelines on AWS CodePipeline and Docker, and lowered operational cost 20% by right-sizing Lambda memory and Step Functions orchestration.
- **Enforced UK healthcare data compliance** through encryption, role-based access control, and anonymisation, and partnered with data scientists to align ML outputs with clinical needs.

Fynd (a Reliance Retail company)

Engineer I • Payments & Reconciliation

May 2022 – Sept 2024

Mumbai, India

- **Engineered a payments reconciliation platform** handling financial services, accounting, and seller settlements for India's largest omnichannel retail platform, operating at Reliance Retail scale.
- **Raised seller settlement accuracy from 84.6% to over 99%** by validating settlement rules and automating configuration updates to guarantee financial correctness.
- **Increased COD payout success rate by 85%** with a cron-based retry mechanism for failed RazorpayX transactions; saved ₹1L/year by replacing a third-party accounting tool with in-house automation.
- **Reduced MTTR** with end-to-end monitoring across Prometheus, Grafana, and Sentry wired to real-time Slack alerting.
- **Built Kafka producers and consumers** driving automation workflows and financial reliability reporting, within a fault-tolerant microservices architecture.

Infosys Ltd.

Systems Engineer

Jan 2021 – May 2022

Bengaluru, India

- **Built backend services for a supply-chain platform** serving a US-based, Amazon-backed EV startup, focused on logistics and vendor operations.
- **Developed end-to-end GraphQL APIs with AWS AppSync** and designed DynamoDB ingestion, transformation, and storage infrastructure for low-latency access to operational datasets.
- **Built scalable data pipelines** aggregating data warehouses and SAP, unifying data access for analytics and business users.

TECHNICAL SKILLS

- **AI & Agentic Systems:** LangGraph, LangChain, multi-agent orchestration, tool calling & function calling, stateful agent graphs, human-in-the-loop workflows, RAG (chunking, embeddings, hybrid retrieval, re-ranking), prompt engineering, structured output extraction, agent SDK / framework development
- **LLM Evaluation:** ground-truth dataset design, LLM-as-judge, completion accuracy, hallucination detection, confusion matrices, precision/recall, regression benchmarking, eval-in-CI, observability & tracing
- **Models & AI Tooling:** OpenAI GPT-4/4.5, Anthropic Claude, vector databases, embedding models, model serving & inference APIs
- **Machine Learning:** ML model integration & deployment, feature engineering, embeddings & vector search, model evaluation (precision/recall, confusion matrices, accuracy), supervised learning fundamentals, PyTorch, pandas, NumPy, PySpark, SQL analytics
- **Backend:** Python 3, FastAPI, Flask, Django, Sanic, REST, GraphQL, Kafka, async & multithreading, microservices, JavaScript, C/C++
- **Cloud & Platform:** Azure (AKS), AWS (Lambda, Step Functions, S3, RDS, ECS, DynamoDB, Athena, Textract, CloudWatch, AppSync), GCP, Docker, Kubernetes, GitLab CI, GitHub Actions, AWS CodePipeline
- **Data & Reliability:** PostgreSQL, MySQL, DynamoDB, BigQuery, Prometheus, Grafana, Sentry, caching, scaling, performance testing
- **Engineering Practice:** System design, low-level design, configuration-driven architecture, test-driven development (TDD), unit & integration testing, design patterns (Strategy, Factory, Observer), SOLID, ORM, DRY / KISS / YAGNI, designing data-intensive applications

CERTIFICATIONS

- **Anthropic Claude Certified Architect — Foundations** • IBM Data Science Professional Certificate • Machine Learning, Stanford University • CISCO Networking Workshop, IIT Delhi

SELECTED PROJECTS

AI-Based Vacation Planner

Jul 2024

Python, SQL, GPT-4, prompt engineering, RPCs

- Built the prompt-orchestration layer generating model calls and assembling responses into structured, personalised itineraries from user inputs on destinations, dates, and activity preferences.

Order Reliability System

Aug 2024

Data warehousing, SQL, APIs, alerting

- Built an hourly job monitoring unfulfilled orders and automating next-state determination, average state-change time per company, and time-series comparison — surfacing breached orders proactively, before any seller or company escalated them.

EDUCATION

Guru Gobind Singh Indraprastha University

B.Tech, Computer Science • GPA 7.50 / 10.00

Jul 2016 – Sept 2020

Delhi, India