

KOUSHIK NAG

AI / GenAI Engineer | Agentic AI • RAG • LLM Systems

Kolkata, India | +91 86372 00185 | koushiknag596@gmail.com | linkedin.com/in/koushik-nag-929315200

SUMMARY

AI/GenAI Engineer with 5+ years building production LLM applications, multi-agent systems and RAG pipelines for banking, healthcare, automotive and government clients. Hands-on with Claude (AWS Bedrock), GPT-4, Gemini and Google ADK, plus LangChain, CrewAI, AutoGen and MCP tool calling. Strong in Python, AWS, Azure and GCP, with a record of shipping secure, observable, CI/CD-driven systems in regulated environments.

CORE SKILLS

Languages: Python, SQL

GenAI / LLMs: Claude, GPT-4, Gemini, Llama, BERT, prompt engineering, AWS Bedrock, Vertex AI

Agentic AI: Google ADK, LangChain, LangGraph, CrewAI, AutoGen, MCP, tool calling, human-in-the-loop workflows

RAG / Retrieval: text-embedding-3-large, Titan, Azure AI Search, FAISS, Elasticsearch, BM25 hybrid search, document chunking

Document AI / OCR: Azure Document Intelligence, OCR, AWS Textract, PyMuPDF, OpenPyXL

Backend / APIs: FastAPI, Flask, Django, REST, JSON, Swagger/OpenAPI, Postman, enterprise API integration

AWS: Bedrock, Lambda, S3, DynamoDB, SQS, API Gateway, IAM, Secrets Manager, CloudWatch

Azure: Databricks, AKS, Key Vault, Document Intelligence, AI Search

GCP: Vertex AI, Google ADK, Gemini, BigQuery, Cloud Storage (GCS)

DevOps: Docker, Podman, Kubernetes, Git, GitHub/GitLab CI/CD

ML / Data: XGBoost, Prophet, ARIMA, K-Means, PyTorch Geometric, Neo4j, SHAP, Pandas, NumPy, Hugging Face

PROFESSIONAL EXPERIENCE

HashedIn by Deloitte — AI Engineer

Bangalore | 02/2026 – Present

Agentic AI Financial Fraud Detection | Global Banking Client

- Led design of a multi-agent fraud-investigation workflow on Google ADK covering transaction analysis, customer profiling, risk assessment and fraud decision support.
- Built reusable Python agent tools and enterprise API integrations for evidence retrieval from Bitbucket, internal knowledge bases and enterprise APIs.
- Integrated Gemini / Gemini Enterprise; added human-in-the-loop review and explainability for regulated workflows; ran GitHub branching, PRs and CI/CD.

Agentic BPMN Process Automation Platform | Enterprise Client

- Built a platform that ingests BPMN files, decomposes them into atomic steps and uses Claude + LangChain orchestration to generate candidate agentic workflows for SME selection.
- Matched generated process graphs against existing SAP records to recommend reusable patterns, speeding up new-project onboarding.
- Implemented RBAC, SSO, admin delegation and Azure Key Vault for multi-tenant governance; deployed on Databricks and Kubernetes with Alembic migrations.

KPMG — Senior AI Engineer

Delhi | 09/2025 – 01/2026

Demographic Deduplication | Jharkhand Government

- Led entity resolution across multi-million records with no unique ID, reaching ~92% precision and 88% recall.
- Combined blocking, clustering, TF-IDF and hybrid BM25 + embedding retrieval in scalable Python pipelines on FAISS and Elasticsearch.

Financial Fraud Platform | Jharkhand Government

- Deployed real-time fraud detection with an XGBoost + Autoencoder ensemble, cutting false positives by 28%.
- Built GCN fraud-network models (Neo4j, PyTorch Geometric) and a GPT-4 + Azure AI Search RAG claim summarizer with SHAP explainability.

TCS — AI Engineer

Kolkata | 05/2021 – 08/2025

Enterprise RAG Chatbot | UHG / Optum

- Architected a GPT-4 RAG chatbot using text-embedding-3-large, chunking and Azure AI Search top-k retrieval, with Databricks as the vector layer.
- Parsed ServiceNow, CSP, SOPHIA and iCES data with Azure Document Intelligence; delivered grounded citations and evaluation via GitLab CI/CD.

Spare Parts Inventory Intelligence | Jaguar Land Rover

- Built Prophet + XGBoost forecasting across 5,000+ dealer-part combinations with K-Means vehicle segmentation.
- Developed a Claude on AWS Bedrock assistant with multi-source RAG over parts catalogs, service manuals, dealer inventory and warranty data.
- Implemented MCP-based agents and serverless backends with Lambda, S3, DynamoDB, API Gateway, Secrets Manager and CloudWatch.

EDUCATION

B.Tech, Electronics & Telecommunication — KIIT, Bhubaneswar

2017 – 2021