

HITESH ATKAR

Generative AI Developer

Pune, Maharashtra | +91 8317261096 | atkarhitesh@gmail.com | github.com/HiteshAtkar | [LinkedIn](#)

SUMMARY

Generative AI & MLOps Engineer skilled in building production-grade Agentic AI platforms, multi-agent systems, LLM-powered document intelligence pipelines, and hybrid RAG architectures. Experienced in NLP, OCR, and secure AI solutions with hands-on expertise in end-to-end MLOps workflows and cloud deployment across AWS, Azure, and GCP consistently delivering scalable, real-world AI systems that replace legacy infrastructure with intelligent, automated solutions.

EDUCATION

B. Tech in Information Technology | CGPA: 8.56

Aug 2022 – Jun 2025

D. Y. Patil School of Engineering and Technology

Diploma in Computer Engineering | Percentage: 83.43%

Jun 2019 – Jun 2022

Samarth Rural Educational Institute Polytechnic College

WORK EXPERIENCE

Generative AI Developer

Sep 2025 - Present

Combat Solution AI

- Developed and deployed Agentic AI and multi-agent systems using LangChain, LangGraph (ReAct, Plan-and-Execute, Reflex, Deep Agents), along with a custom MCP (Model Context Protocol) Server-Client architecture for autonomous task planning, agent orchestration, and inter-agent communication.
- Built scalable AI-powered applications and APIs using FastAPI, Flask, and Python by integrating LLMs, external tools, Twilio (WhatsApp & Voice AI), and HeyGen avatars to deliver intelligent automation and multimodal conversational AI experiences
- Built high-throughput OCR document processing pipeline with Celery worker replicas, Redis task queue, PostgreSQL, and Google Document AI, processing concurrent documents with zero downtime on Hostinger VPS using Docker containerized infrastructure.
- Designed and implemented Retrieval-Augmented Generation (RAG) pipelines with custom knowledge bases, semantic search, and vector databases using FAISS, Pinecone, Elasticsearch, and web-scraped data sources, fused with Qdrant vector search + Neo4j Knowledge Graph, LLM-based guardrails, and session management.
- Engineered hybrid AI search system combining Elasticsearch vector search + Neo4j Knowledge Graph re-ranking, replacing legacy keyword search and improving retrieval accuracy via intent detection and industry-code pre-filtering.
- Built production ML inference service on Kubernetes (HPA, probes, rolling updates, PSI drift detection CronJobs, auto-retraining) with microservices architecture (API Gateway + FastAPI services), Prometheus/Grafana observability, and distributed trace IDs via Docker Compose.
- Built face recognition & fraud/anomaly detection microservices with Elasticsearch biometric indexing and Docker health checks, secured via OAuth 2.0, CSRF protection, JWT authentication, and Google OAuth integration.

Data Analyst

Apr 2024 – Jun 2024

Altair (Remote)

- Built and evaluated ML models (Random Forest, XGBoost) on real-world datasets using scikit-learn, applying feature engineering, hyperparameter tuning, and experiment tracking with MLflow and DVC for model versioning and pipeline reproducibility.
- Conducted EDA on real-world datasets and developed interactive dashboards using Matplotlib, Plotly, and Pandas, surfacing model performance metrics and feature importance insights that supported data-driven decisions for cross-functional stakeholders.
- Automated ML preprocessing pipelines with Pandas/NumPy, integrated DVC data versioning and MLflow model registry to streamline cross-functional collaboration and reduce model development cycle time.

PROJECTS

Bot Builder (No-Code Agentic AI SaaS Platform)

- Built a no-code SaaS platform where non-technical users visually design, configure, and deploy production-ready AI agent workflows and chatbots using a drag-and-drop interface powered by React Flow entire agent logic, tool connections, and knowledge base setup through UI without a single line of code.

- Under the hood powered by LangGraph state machines, MCP Server-Client architecture, Qdrant RAG, third-party tool integrations (Gmail, HubSpot, Google Maps), Twilio voice/WhatsApp, and HeyGen avatars monitored via Prometheus/Grafana with GitHub Actions CI/CD.

APA (Insurance Claims AI Platform)

- Built end-to-end insurance claims processing system with OCR pipeline (Google Document AI, Tesseract OCR), face recognition, and fraud/anomaly detection microservices deployed on Kubernetes with HPA autoscaling, PSI drift detection CronJobs, and automated model retraining.
- Architected distributed microservices infrastructure (API Gateway + independent fraud, anomaly, face, OCR services) with Celery, Redis, PostgreSQL, Elasticsearch biometric indexing zero downtime on Hostinger VPS via Docker.

TOT (Chemical & Tenders Extraction Pipeline)

- Built intelligent extraction pipelines that parse chemical compound names from scientific research documents and tender metadata from ODF images and mixed-format files (PDF, DOCX, scanned images) using Google Document AI, EasyOCR, and LLM-powered semantic parsing to output clean structured JSON.
- Handled multilingual documents via deep-translator and Google Translate API, processed unstructured content using Unstructured.io, and indexed extracted entities into Elasticsearch for downstream semantic retrieval and structured querying.

AI Search (Tender Search Engine)

- Replaced production Solr keyword search with a hybrid retrieval engine combining Elasticsearch BM25 + ANN vector search and Neo4j Knowledge Graph re-ranking one LLM call per unique query enriches it into 150-word semantic prose before embedding, making query and document vectors comparable for the first time.
- Designed 4-signal Knowledge Graph re-ranking system keyword overlap, co-occurrence traversal, industry alignment, and similarity propagation with dynamic weight tuning and ANN floor filtering, cutting cross-domain false positives and lifting retrieval confidence from 0.10 to 0.65.

TECHNICAL SKILLS

Languages: Python

AI & Generative AI: LangChain, LangGraph, OpenAI, RAG, Multi-Agent Systems, Agentic AI, MCP (Model Context Protocol), Prompt Engineering, Tool Calling, HITL Workflows, Multimodal RAG (text + Image), Model Fine-Tuning, GPT Models, Claude Models, Gemini Models, RLHF, LoRA, PEFT, Vector embeddings, Semantic search.

ML & Deep Learning: Scikit-learn, TensorFlow, PyTorch, Keras, XGBoost, Transformers, Sentence-Transformers, spaCy, NLTK, OpenCV, Hugging Face Transformers, NLP

Data & Visualization: NumPy, Pandas, Matplotlib, Plotly, SciPy

OCR & Document Processing: Google Document AI, EasyOCR, Tesseract, PaddleOCR

Databases & Search: PostgreSQL, MySQL, MongoDB, Redis, Elasticsearch, Qdrant, FAISS, Pinecone, Chroma DB, Neo4j

Frameworks & APIs: FastAPI, Flask, Streamlit, REST APIs, Twilio, HeyGen, HubSpot, Pipedrive

Web Scraping: Scrapy, Playwright, BeautifulSoup, Selenium

Cloud & Deployment: AWS (EC2, EKS, ECS, ECR, S3, CloudWatch, Lambda), Azure (ML Studio, AI Foundry), GCP (Document AI, Cloud Storage, Translate), Hostinger VPS, Docker, Kubernetes (HPA, CronJobs, Ingress, RBAC), A/B Testing

MLOps & DevOps: MLflow, DVC, GitHub Actions, CI/CD, Prometheus, Grafana, Celery, Alembic, PSI Drift Detection

Security & Tools: OAuth 2.0, JWT, SAML, Paramiko, Git, Postman, Jupyter

CERTIFICATIONS

- [Google Data & Advanced Analytics Certification](#) by **Google**
- [Python for Data Science, AI & Development](#) by **IBM**
- [ML Specialization](#) by **DeepLearning.AI**
- [Machine Learning Operations Certification](#) (MLOps) by **Edureka**

ACHIEVEMENTS

Recognized as Best Performer within 3 months at Combat Solutions for delivering high-impact AI solutions and consistently exceeding expectations.