

Ashutosh Patra

 Pune, India  ashutoshpatra1135@gmail.com  +91 9967556873  ashutosh2211  ashutosh-2211

Summary

AI Engineer with 2 years of experience designing and deploying end-to-end scalable LLM, ML, and agentic systems. Built automation pipelines that **reduced the operational workload by more than 70% and increased the reliability of the system** in VFX, healthcare, finance, and operations. Demonstrated track record in LLM orchestration, fine-tuning, multimodal workflows, and production deployments on major cloud platforms.

Looking for long-term opportunities to architect impactful AI solutions with strong growth trajectories and meaningful scope to drive innovation.

Work Experience

AI Data Engineer

Ancrew Global Services Pvt Ltd

Pune, India

Jun 2025 – Present

- Architected an automated invoice processing pipeline using OCR (EasyOCR, Tesseract) and LLM-based validation, processing 1,000+ invoices per month and **reducing manual data entry effort by 98%**.
- Fine-tuned and deployed a domain-specific AI assistant using open-source LLMs, **cutting human review time by 30%** and improving response availability for internal users.
- Designed secure data workflows to detect and mask PII across datasets exceeding **50GB per month**, supporting regulatory compliance and enterprise data privacy requirements.

AI Developer

Blowfish Media Technology

Mumbai, India

Oct 2024 – June 2025

- Architected an AI-driven VFX diffusion pipeline using DeepFace and InsightFace to support de-aging workflows in feature films, processing **2,000+ frames** and **reducing turnaround time by 70%**.
- Designed and deployed multi-agent moderation workflows using LangGraph to detect and remediate inappropriate content across **300+ sources per hour**, flagging **10,000+ items monthly** for review.
- Orchestrated a LoRA-based AI cloning platform with built-in safety gating, watermarking, and traceability, supporting **50+ controlled model variations** for secure usage.
- Developed a PDF-to storyboard automation system that parsed **100+ page scripts** and generated **500+ storyboard frames**, **cutting manual planning effort by 93%**.

Cloud Engineer Trainee

Teleglobal International Pvt Ltd

Pune, India

Feb 2024 – Sep 2024

- Developed MongoDB-to-analytics ETL pipelines using Python and AWS Glue, transforming **millions of records per week** to support near real-time analytics workloads.
- Delivered an agentic RAG proof-of-concept using LangChain and Pinecone, indexing **50,000+ log entries** from Splunk for client UAT and achieving **92% retrieval accuracy**.

Freelance Works

AI Project Timeline Analyzer

Oct 2025

- Architected an ETL pipeline using Pandas and TimescaleDB to ingest bi-weekly XLSX reports, capturing **150+ structured data points** across **10 active projects** for timeline-based performance monitoring.
- Designed an LLM-powered summarization workflow to generate **1–6 month analytical reports**, consolidating insights from **200+ project milestones** for management review.

AI Mail Retrieval System

Aug 2025

- Designed a semantic email retrieval system using Qdrant vector database and OpenAI embeddings, indexing **5,000+ emails and attachments** and **reducing leadership review effort by 60%**.
- Deployed an LLM-driven query interface using LangChain and Gemini on Azure App Service, handling

100+ executive queries per day with consistent response quality.

AI Code Assistant Platform

May 2025

- Led development of an AI-assisted coding platform using Claude 3.5 (AWS Bedrock) and LangGraph, enabling automated generation and execution of **ETL and analytics scripts** across **6 database systems**.

Technologies

Languages: Python

Machine Learning / Deep Learning: PyTorch, TensorFlow, Scikit-learn, Hugging Face Transformers

LLMs & Agentic Systems: LangChain, LangGraph, LLaMA, GPT models, Claude, Vector Databases (Qdrant, Pinecone, OpenSearch)

Computer Vision & OCR: OpenCV, YOLO, EasyOCR, CNNs, Pillow

Cloud Services: AWS (Bedrock, Lambda, Glue, Macie), Docker, Kubernetes, GitHub Actions

Databases: PostgreSQL, MongoDB, TimescaleDB/TigerData, MySQL

Projects

ReasonForge Qwen3-4B

 GitHub

Technologies: Ray, VLLM, Qwen3, Unslloth, Redis

- Fine-Tuning and Serving a Reasoning-Optimized LLM with Ray + vLLM

TPA – The Personal Assistant

 GitHub

Technologies: LangGraph, LangChain, OpenSearch, Titan Embeddings, Docker

- Orchestrated multi-capability AI agent handling 200+ travel queries, news aggregation, and hotel searches with structured JSON results.

DailyBudget – AI Finance Manager

 GitHub

Technologies: FastAPI, SQLite, LangChain, LLMs

- Developed personal finance tracker with automatic categorization, analyzing 500+ transactions monthly using NLP.
- Generated AI-powered spending behavior summaries, identifying 15+ spending patterns per user with 88% accuracy.

Education

Master of Science in Data Science

Bhubaneswar, India

Silicon University

2022 – 2024

Bachelor of Science in Mathematics

Bhubaneswar, India

Utkal University

2018 – 2021