UDATYA DEB

Kolkata/Bangalore/Pune, India
+91 70442 22356 | udatya.developer@gmail.com | linkedin.com/in/udatya-deb-604785200 | github.com/UDATYA-DEB

PROFESSIONAL SUMMARY

Machine Learning Engineer with 2 years of experience designing agentic AI systems and production RAG pipelines. Specializes in multi-step LLM workflows, hybrid retrieval, and context optimization for on-premise open-source models — applied to document intelligence, workflow automation, and enterprise system integration.

CORE SKILLS

Agentic AI & LLMs: Multi-Agent Workflows, LangGraph, LangChain, Prompt Engineering, Context Optimization, Human-in-the-Loop Systems, Llama / Qwen / Gemma (via Ollama), On-Prem LLM Deployment

Retrieval & Data: RAG, Hybrid Retrieval, Semantic Search, ChromaDB, Sentence-Transformers, Knowledge Graphs, Metadata Filtering, SQL

Technologies & Software: Python, FastAPI, PyTorch, Vector Databases, Langfuse / LangSmith (LLM observability & tracing), Jira / Confluence APIs, Workflow Orchestration, Computer Vision (YOLOv8, FaceNet), OpenAI Whisper

PROFESSIONAL EXPERIENCE

Machine Learning Engineer **April 2024 – Present**
Calsoft Kolkata, West Bengal, India

AI Sales Assist — Automated Order Processing & Quoting

- Built an agentic email-to-quote pipeline processing 30,000+ inbound messages per month, classifying order intent and extracting per-site product, sub-category, and quantity requirements before routing qualifying orders into the quoting flow
- Developed a feasibility and quoting engine that validates extracted requirements against internal product and pricing systems, returning a priced quote in 2–3 minutes with no analyst involvement, replacing a fully manual review-and-pricing process
- Implemented a multi-site clarification workflow that detects missing per-site specifications and auto-drafts templated follow-up emails, generating structured Excel collection sheets when gaps span multiple sites, enabling independent quoting for each site within a single order

Automated Requirements Specification (SRS) Generation

- Built an automated SRS generation system that ingests product definition, system requirement, and product requirement documents, maps cross-document dependencies, and drafts complete specifications using on-prem open-source LLMs, cutting manual authoring effort by ∼85%
- Engineered a context-optimization layer (semantic chunking, hierarchical summarization, relevance ranking) compressing source corpora [X:1] to fit a fixed context window without truncation loss, enabling full-corpus reasoning on locally hosted models
- Designed a ChromaDB retrieval layer over versioned requirement documents with Confluence integration for live sourcing, improving factual grounding and reducing irrelevant context
- Auto-generated sprint-ready Jira epics and stories from finalized SRS sections, with ∼65% accepted without major edits, closing the loop from raw requirements to actionable backlog

Automated Bug Triaging System

- Built an automated bug triaging system using the Jira API, Sentence-Transformers, a vector database, and an on-prem LLM with RAG and prompt engineering, achieving 82% classification accuracy without model fine-tuning
- Designed a weighted knowledge graph with statistical thresholding to prune irrelevant tickets pre-retrieval, combined with hybrid semantic search and metadata filtering to reduce noisy context

Additional Delivery

- Delivered a multimodal video CMS (Whisper, keyframes, VLM person ID) and an image-based product search system

EDUCATION

Bachelor of Technology (BTech), Computer Science (CGPA 9.2) **2020 – 2024**
Netaji Subhash Engineering College