Amit Kumar Mahapatra

 Bhubaneswar, IN   mahapatraarhas200@gmail.com   +91-8114-362199

 amitmahapatrav5   amitmahapatrav5

Generative AI • RAG • LangChain • Agentic AI • LangGraph • MCP • Traditional AI •
Deep Learning • PyTorch • Backend • Python • FastAPI • Frontend • React •
TypeScript

Experience

IT Analyst   *2020-10-08—Present*

Tata Consultancy Services

Generative AI Engineer

- Pluggable Multimodal RAG for Secure Knowledge Retrieval
- EHS&S Incident Reporting Gen AI Chatbot
- API Development and Integration using FastAPI, SQLAlchemy, and OAuth 2.0

Projects

Pluggable Multimodal RAG for Secure Knowledge Retrieval   *2025-02-01—Present*

Developed a pluggable Retrieval-Augmented Generation (RAG) framework that avoids costly
fine-tuning by leveraging Hugging Face pre-trained models with ChromaDB vector search,
laying the foundation for multi-hop reasoning and grounded answers in later phases.

- Implemented document ingestion for structured (.csv, .xls, .json) and unstructured (.txt,
  .md, .pdf, .docx, .pptx) formats using LangChain loaders, ensuring modular support for
  future plug-ins.
- Built preprocessing pipeline with LangChain Text Splitters (Recursive, Character,
  Markdown, NLTK, spaCy) to dynamically generate high-quality chunks tailored to file
  types.
- Integrated Hugging Face Embeddings for vectorization and persisted results in ChromaDB,
  enabling semantic similarity search as the backbone of the RAG loop.
- Designed modular ingestion architecture with caching, loaders, splitters, enrichers, and
  vector store executors, improving maintainability and scalability.
- Exposed ingestion and retrieval workflows via FastAPI endpoints, routed through an API
  Gateway with service-specific APIs (SQL metadata API, content API, and GenAI query
  API).
- Implemented query handling pipeline: Retriever $\rightarrow$ PromptTemplate $\rightarrow$ Hugging Face LLM
  (ChatHuggingFace) $\rightarrow$ StrOutputParser, enforcing concise, grounded answers strictly from
  retrieved context.
- Improved retrieval reliability and reduced implementation risk by adopting pre-trained
  Hugging Face models with vector search instead of costly fine-tuning, enabling incremental
  upgrades in later phases.
- Established secure, team-specific retrieval boundaries by embedding ownership metadata,
  preventing cross-team data exposure while improving knowledge accessibility.

EHS&S Incident Reporting Gen AI Chatbot - MVP 2 - POC *2025-02-01—Present*

Prototyped the GenAI chatbot enhancement featuring vector similarity search against validated Golden Records database, implementing dynamic clarifying question generation to improve data collection accuracy, and expanding incident classification capabilities to support multiple EHS&S incident types including Good Save and Near Miss reporting scenarios.

- Implemented vector embeddings using ChromaDB vector database to index and retrieve Golden Records, enabling semantic similarity search with sub-second query performance for contextual reference matching during incident reporting.
- Developed intelligent clarifying question generation system leveraging Langchain's retrieval-augmented generation (RAG) architecture, dynamically formulating context-specific questions based on similar Golden Records to capture comprehensive incident details and reduce ambiguity in user submissions.
- Built automated incident classification engine utilizing OpenAI-4o model to accurately categorize incidents across multiple taxonomy levels, distinguishing between Good Save, Near Miss, and other EHS&S event types with high precision.
- Optimized prompt engineering strategies implementing few-shot learning techniques and structured output templates, enabling context-aware response generation with improved factual accuracy and consistent formatting for downstream system integration.

EHS&S Incident Reporting Gen AI Chatbot - MVP 1 *2024-10-01—2025-10-01*

Developed a GenAI-powered chatbot system to replace traditional form-based data entry for Environmental Health, Safety and Sustainability(EHS&S) incident reporting, reducing data collection rework and licensing costs while improving user engagement through conversational AI interface.

- Engineered RESTful API to process free-text user inputs and deliver summarized, structured responses, enhancing data quality and consistency for downstream processing, achieving <2-second response time per API interaction and ensuring real-time user experience.
- Leveraged Langchain framework to orchestrate OpenAI-4o LLM integration, enabling intelligent context summarization and natural language understanding from unstructured user inputs, supporting >5,000 Good Saves per month adoption rate through improved user accessibility.
- Implemented Guardrail AI validation pipeline to ensure AI safety through multi-layer content filtering, mitigating risks of hallucinations, bias, and toxic content generation in LLM responses, maintaining <5% data error rate through chatbot-driven validation and quality checks.
- Developed robust audit system using FastAPI and SQLAlchemy ORM, implementing Pydantic models for type-safe serialization/deserialization of conversation histories, ensuring compliance and traceability requirements while contributing to ~70% reduction in license-related costs by eliminating dependency on traditional form-based systems.
- Integrated submission API with Internal System to automate Good Save record Draft creation, streamlining SME review and reducing manual data entry overhead, achieving 90% successful API submission rate for seamless data flow to backend systems.

API Development and Integration using FastAPI, SQLAlchemy, and OAuth 2.0 *2023-03-01* *—2022-03-01*

Designed and implemented a secure, high-performance REST API using FastAPI for document retrieval and delivery, leveraging SQLAlchemy ORM for database operations, Pydantic for data validation, and OAuth 2.0 for authentication. Deployed the solution on IIS with robust logging and auditing mechanisms.

- Developed REST API endpoints using FastAPI to enable secure and efficient document retrieval and delivery.
- Implemented OAuth 2.0 authentication with 60-minute token validity for secure API access.
- Utilized SQLAlchemy ORM for database interactions, improving maintainability, reducing boilerplate SQL, and enabling database-agnostic operations.
- Applied Pydantic models for request and response validation, ensuring data integrity and type safety.
- Configured API hosting on IIS for enterprise-grade deployment and scalability.
- Designed and implemented a structured logging system using Python's logging module with `TimedRotatingFileHandler`, capturing request/response details, status codes, and error traces for audit and troubleshooting.
- Created YAML-based logging configuration for modular and maintainable log management, including log rotation and multi-level loggers for API, DB, and utility modules.
- Implemented error handling and response standardization with appropriate HTTP status codes for robust client-server communication.

Languages

English: Professional Working Proficiency
Odia: Native speaker

Skills

Generative AI: RAG, LangChain, ChromaDB, FAISS
Agentic AI: LangGraph, MCP
Backend Engineering: RESTful API, gRPC, FastAPI, SQLAlchemy, PostgreSQL, Docker