Vaisakh Prakash

Senior Architect | AI & Full Stack Engineering

Full Stack and AI Engineer with over a decade of experience in designing and building scalable, data-driven systems across healthcare, e-commerce, and NLP domains. Proven expertise in architecting AI-powered applications, integrating backend intelligence with frontend interfaces, and deploying production-ready solutions leveraging LLMs, RAG pipelines, and modern microservice architectures.

Experience

Senior Architect | IQTrek Consulting Pvt Ltd

Feb 2025 – Present

- Architecting enterprise AI systems integrating backend intelligence with frontend interfaces.
- Designed distributed RAG and orchestration pipelines using Python, Kafka, and Spring Boot.
- Collaborating across engineering, design, and product teams to deliver GenAI-driven solutions for clients.
- Led the architecture and deployment of **PostRAI**, an AI poster design assistant transforming creative automation, and **Elenora,** a trip concierge

Full Stack Developer | KeHE Distributors / The Bridge (now North Highland)

Sep 2020 – Jan 2025

- Developed multi-platform e-commerce applications using Angular (frontend) and .NET backend.
- Implemented Kafka-based asynchronous messaging between microservices for real-time order and inventory processing.
- Deployed applications on AWS serverless infrastructure for scalability and reliability.

Data Science Engineer | Auxo Technology Labs Pvt Ltd

Mar 2017 – Sep 2020

- Designed AI-driven applications combining NLP, Rasa, and Kafka for real-time conversational inference.
- Built medical assistant solutions leveraging Python, Quarkus, and Kafka for streaming data processing.
- Applied ML models using Apache Spark MLlib for large-scale data inference and analytics.

Senior Software Engineer | QBurst Technologies Pvt Ltd

Sep 2012 – Feb 2017

- Delivered enterprise-grade Java and Spring Boot applications integrating GraphDB and NLP modules.
- Engineered Kafka-based ingestion pipelines for unstructured medical and social media data.
- Implemented Dropwizard REST APIs and custom NLP pipelines using Stanford NLP and YACC.

Technical Skills

- **Languages & Frameworks:** Python, Java, JavaScript, Angular, React, NodeJS, Spring Boot, Django, Quarkus
- **AI & Data Engineering:** Apache Kafka, Spark MLlib, Hadoop, RAG pipelines, NLP (Stanford NLP, Rasa, NLTK), Langchain, Langgraph, OpenAI, Pocketflow

- **Cloud & DevOps:** AWS (Lambda, S3, CloudWatch), Docker, Jenkins, Serverless Framework
- **Databases:** PostgreSQL, MongoDB, Elasticsearch, GraphDB, Redis, HBase, VectorDB
- **Tools & Others:** RabbitMQ, Dropwizard, Hive, Flume, TensorFlow, Data Pipeline Design

---

Key Projects

PostRAI - AI-Powered Poster Design Assistant

Tech Stack: Python, React, FastAPI, AWS, LangChain, VectorDB, Generative AI, RAG Pipelines

- Led the architecture and development of **PostRAI**, an AI-driven design automation platform for businesses and creators.
- Integrated **RAG (Retrieval-Augmented Generation)** pipelines using LangChain and VectorDB to enable contextual poster generation - allowing the AI to retrieve relevant templates, brand elements, and design inspirations dynamically.
- Designed brand consistency automation through embedded logos, color palettes, and typography.
- Built scalable **multi-language** and **multi-format** support for cross-platform publishing (Instagram, Facebook, WhatsApp, print).
- Delivered a **90% reduction in design turnaround time** and ensured consistent on-brand visuals across campaigns.
- Established the foundation for future integrations with marketing and hiring platforms for automated creative workflows.

Elenora - AI Trip Planning Concierge

Tech Stack: Python, FastAPI, LangChain, VectorDB, Embeddings, RAG Pipelines

- Designed and implemented an automated crawler to extract and structure content, images, and metadata from company blog pages.
- Consolidated policies, terms, and documentation into a unified **knowledgebase** for semantic retrieval.
- Built an **embedding-powered RAG pipeline** using word embeddings and a vector database for contextual content retrieval.
- Developed a **progressive user profiling system** that dynamically tailors questions and recommendations based on responses.
- Integrated **AI-generated, context-aware travel recommendations** with inspirational imagery and verified sources.
- Created **FastAPI endpoints** for personalized recommendations, user profile updates, and conversational chat.
- Optimized the system for **scalability and production**, including caching, database optimization, and CDN strategies.

Cascade - Medical Transcription & Inference Engine

Tech Stack: Spring Boot, Kafka, Stanford NLP, GraphDB

- Built an AI-based medical transcription system converting doctor-patient audio into structured medical records.
- Developed Kafka-powered ingestion pipelines and NLP-based extraction workflows for real-time insights.

CliniBot - Real-Time Conversational AI Assistant

Tech Stack: Quarkus, Python, Rasa, RabbitMQ, Kafka

- Designed an AI assistant for real-time doctor-patient conversation analysis using NLP and Rasa.
- Streamlined asynchronous communication and inference via RabbitMQ and Kafka.

Synapse - Medical Data Inference Platform

Tech Stack: Java, Apache Spark MLlib, Elasticsearch, Spring Boot

- Developed ML-driven analytics platform for unstructured clinical data with Spark-based scalable inference