☎ *+91-9108404965*
✉ *shivamjohri247@gmail.com*
in *shivam-johri*
Shivamjohri247

Shivam Johri

Summary

Senior AI Engineer with 9.5+ years of experience delivering enterprise-grade AI and ML solutions, including 6+ years in architecting scalable MLOps pipelines and production systems. Specialized in **large language models and agentic AI**, with proven expertise in designing autonomous agents, intelligent code generation frameworks, and collaborative multi-agent workflows. Notable contributions include development of a **Designer Agent** for automated code solution generation and an **Underwriting Assistant Agent** to streamline risk evaluation and discrepancy detection in enterprise underwriting. Experienced in **leading and mentoring a team of 5 AI engineers**, driving innovation and ensuring successful delivery of complex projects. Adept at integrating deep learning, cloud-native platforms, and modern orchestration frameworks to translate complex business requirements into measurable impact. **Google Cloud Associate Cloud Engineer certified.**

Technical Expertise

MLOps / Data Eng.	Apache Airflow, MLflow, Kubernetes, Docker, Git/GitHub, CI/CD pipelines, Apache Spark, Apache Solr, ElasticSearch
AI/ML Dev.	Deep Learning (BERT, Transformers), LLMs, Recommendation Engines, NLP, TensorFlow, PyTorch, Keras, Scikit-learn, FastAPI, Huggingface
LLM / Agentic AI	LangChain, LangGraph, Semantic Kernel, Google Agent Builder (Vertex AI Agents), OpenAI Function Calling, Retrieval-Augmented Generation (RAG), Vector Databases (FAISS, Pinecone, ChromaDB)
Cloud / Big Data	Google Cloud Platform (Vertex AI, BigQuery, GKE), AWS SageMaker, Hadoop Ecosystem (Spark, Hive), Distributed Computing
Programming	Python, SQL, Go, Shell Scripting, Data Preprocessing, Visualization, Data Governance, Responsible AI

Professional Experience

Jul 2025 – Present **Senior AI Engineer**, *Suzega*, India

- ○ Leading development of agentic AI systems capable of autonomous task execution, decision making, and workflow orchestration using sdk's like Langgraph,Agent Development Kit, Semantic Kernel, smolagents etc.
- ○ Building a **Designer Agent** for code generation that produces optimized code solutions based on user prompts, enabling faster prototyping and development.
- ○ Designed and deployed a full workflow for an **Underwriting Assistant Agent**, comprising multiple collaborative agents that generate assisted underwriting reports, flagging discrepancies and potential fallouts for enterprises.

Feb 2024 – Jul 2025 **Senior Machine Learning Engineer**, *EPAM Systems*, Remote, Poland

- ○ Worked on automating end-to-end MLOps pipeline for financial **Named Entity Recognition (NER)** utilizing Transformers fine-tuning, significantly enhancing search relevancy and accuracy by 77% in search platform and deployed in production.
- ○ Developed and deployed an LLM-powered query rewriting system for indexing services, driving substantial improvements in search relevancy and query processing.
- ○ Worked on creation of Query classification pipeline using fine-tuning approach using BERT embeddings and XGBClassifier for search relevancy improvement and enhanced result ranking for search.

Sep 2021 – Feb 2024 **ML Engineering Senior Analyst**, *Accenture*, India

- ○ Led the development of an LLM-based query expansion service for financial queries using Gemini models, achieving a 25% increase in query coverage and enhancing overall search effectiveness for a finance domain client.
- ○ Developed high-accuracy signature detection models for fraud prevention in handwritten documents using advanced computer vision techniques like FasterRCNN and YoloV3.
- ○ Designed and deployed 5 personalized fashion recommendation models on GCP's Recommendations AI, boosting user engagement by 15% for a retail client.
- ○ Automated regulatory report generation processes, resulting in a 40% reduction in manual effort and improved compliance efficiency for Google partnered regulatory reporting clientele.

Mar 2016 – Aug 2021 **Machine Learning Engineer**, *Tata Consultancy Services*, Bangalore, India

- Developed an end-to-end pipeline for text detection and information extraction from ID cards using CTPN, automating critical data processing workflows using ETL pipelines.
- Built and deployed machine learning models for document classification and data extraction from complex PDF documents (using Tesseract OCR), reducing validation costs by 30%.
- Implemented a sophisticated email indexing and entity recognition system, enabling advanced data retrieval and analysis for clients.

Education

2011 – 2015 **B.Tech, Electronics and Communication Engineering**, *Shri Ramswaroop Memorial College of Engg. & Management (APJAKTU)*, Lucknow, India

GPA: 6.24/10

Certifications

Google Cloud Associate Cloud Engineer (Valid till 11/2025)

Projects

LLM Fine-tuning Pipeline Baseline pipeline for fine-tuning SLMs on downstream tasks; extensible for custom needs.

Emotion Classifier Open-source project using Huggingface Transformers for emotion detection in text. GitHub Link

Volunteering

Feb 2015 – Mar 2015 **Organizer**, *Shri Ramswaroop Memorial College Fest "XERO-2K15"*, Lucknow

Led organizing team for two consecutive years; managed cultural + technical events with high participation.