# Shivam Johri

+91-9108404965 | shivamjohri247@gmail.com | linkedin.com/in/shivam-johri | github.com/Shivamjohri247

## Summary

Results-driven Senior Machine Learning Engineer with 9+ years of experience, including 5+ years specializing in architecting and deploying end-to-end MLOps pipelines and delivering impactful AI/ML solutions. Expert in automating model lifecycles, leveraging deep learning, LLMs, and cloud-based technologies. Proven ability to translate complex business needs into scalable, high-performance solutions. Google Cloud Associate Cloud Engineer certified.

## Technical Expertise

**MLOps  Data Engineering:** Apache Airflow, MLflow, Kubernetes, Docker, Git/GitHub, CI/CD pipelines, Apache Spark, Apache Solr, ElasticSearch

**AI/ML Development:** Deep Learning (BERT, Transformers), LLMs, Recommendation Engines, NLP, TensorFlow, PyTorch, Keras, Scikit-learn, FastAPI. Huggingface

**Cloud  Big Data:** Google Cloud Platform (Vertex AI, BigQuery, GKE), AWS SageMaker, Hadoop Ecosystem (Spark, Hive), Distributed Computing

**Programming  Data:** Python, SQL, Go, Shell Scripting, Data Preprocessing, Visualization, Data Governance, Responsible AI

## Professional Experience

### Senior Machine Learning Engineer — EPAM Systems — 02/2024 - Present

- 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 XGB-classifier for search relevancy improvement and enhanced result ranking for search.

### ML Engineering Senior Analyst — Accenture — 09/2021 - 02/2024

- 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.

### Machine Learning Engineer — Tata Consultancy Services — 06/2019 - 08/2021

- 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

**B.Tech in Electronics and Communication Engineering**
APJ Abdul Kalam University, Lucknow (2011 - 2015) — GPA: 6.24/10

## Certifications

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