# Suraj Sharma

Senior Data Scientist | Generative AI & Data Solutions

New Delhi, India | +91 83778 33094 | suraj.mae27@gmail.com | LinkedIn

## Professional Summary

Senior Data Scientist with 6+ years of experience in Machine Learning, Deep Learning, and Generative AI. Expert in Large Language Models (LLMs), prompt engineering, and fine-tuning transformer architectures (GPT, BERT) for NLP, semantic discovery, and enterprise automation. Proven track record in building end-to-end ML pipelines, deploying models on AWS SageMaker & Azure ML, and delivering data-driven solutions for BFSI, FMCG, Healthcare and SaaS domains.

## Core Competencies

• Generative AI & LLMs: GPT, BERT, Prompt Engineering, Fine-tuning, Domain Adaptation

• Deep Learning & NLP: TensorFlow, PyTorch, Transformer Models, NLP2SQL, NER, Semantic Search, Sentiment Analysis

• MLOps & Cloud Deployment: AWS SageMaker, Azure ML, CI/CD Pipelines, Docker, Kubernetes

• Data Engineering & Platforms: Medallion Architecture, Databricks, Snowflake, BigQuery, Kafka

• Visualisation & BI: Power BI, Executive Dashboards

• Stakeholder Engagement: Discovery Workshops, Customer Collaboration, Technical Mentorship

## Professional Experience

### Senior Data Scientist | Publicis Sapient | Aug 2025 – Present

• Leading development of Generative AI solutions for healthcare clients, including NLP-driven automation and customer analytics.

### Senior Data Scientist | DQLabs.ai (Remote) | Mar 2025 – July 2025

• Designed and deployed LLM-powered NLP2SQL solutions using fine-tuned GPT models, automating query translation and reducing handling time by 60%.

• Built a semantic data discovery engine leveraging transformers and advanced prompt engineering, enhancing enterprise data accessibility.

• Led discovery workshops with enterprise clients to co-create Generative AI strategies aligned with business objectives.

### Engineer – Data Science | Pentair Waters | Oct 2023 – Mar 2025

• Led development of scalable ML models and introduced medallion architecture for multi-layered, resilient data pipelines.

• Developed a patent-pending computer vision pipeline for geo-image perimeter extraction, exploring generative AI methods for improved analytics.

• Delivered AI-driven executive dashboards and NLP-based automation tools, decreasing enterprise response time by 40%.

• Mentored junior data scientists and worked with cross-functional teams to embed AI-driven insights into decision-making.

### Data Scientist | LTIMindtree | May 2022 – Oct 2023

• Developed and deployed a transformer-based Named Entity Recognition (NER) model for FIU-IND, facilitating automated extraction of entities from suspicious transaction reports with an accuracy rate of 75%.

• Designed sophisticated risk-scoring frameworks aimed at prioritizing high-risk entities within Anti-Money Laundering (AML) workflows.

• Constructed time-series forecasting models to identify anomalous transaction patterns, thereby enhancing compliance monitoring mechanisms.

### Data Scientist | iLink Digital | Jul 2019 – May 2022

• Designed a telemedicine platform powered by conversational AI, utilizing advanced NLP pipelines for tailored patient interactions.

• Created an RNN-based system for sentiment analysis to monitor social media in real time and gather brand insights.

• Implemented OCR-based medical enrollment solutions using Tesseract OCR and Kubernetes to enable scalable deployment.

• Used machine learning analytics to optimize revenue for FMCG clients, deploying scalable solutions on Azure ML and AWS SageMaker.

### Data Analyst | NSYUS | Jun 2017 – Aug 2018

• Conducted A/B testing and behavioral analytics, improving website conversion rates by 15%.

• Redesigned customer journeys, resulting in a 20% increase in online sales.

## Education

• M.Tech in Data Science – BITS Pilani

• PGP in Data Science – INSOFE Hyderabad (2018–2019)

• B.Tech in Mechanical & Automation – GGSIPU Delhi (2013–2017)

## Certifications

• Microsoft Azure AI Fundamentals (AI-900)

• AWS Certified Cloud Practitioner

## Technical Skills

• Programming: Python, SQL

• Libraries: TensorFlow, PyTorch, scikit-learn

• Deployment: AWS SageMaker, Azure ML, Docker, Kubernetes, Flask

• Data Processing: Hadoop, Spark, H2O.ai

• Visualization: Power BI