

# Mohd Zahid Faiz

Data Scientist - Applied ML

Phone: +91-9425231257 Email: mohdzahid.faiz@gmail.com

LinkedIn: [www.linkedin.com/in/mohd-zahid-faiz/](https://www.linkedin.com/in/mohd-zahid-faiz/) GitHub: [github.com/ZahidFaiz](https://github.com/ZahidFaiz)

## Professional Summary

---

Data Science professional with hands-on experience designing and scaling machine learning systems—ranging from **predictive modeling** and **anomaly detection** to **generative frameworks**—from problem formulation through production deployment. Regularly navigate modeling choices, trade-offs, and operational constraints to build solutions that perform reliably at scale and inform critical business decisions.

## Technical Skills

---

**Languages:** Python (Expert), SQL, PySpark

**Machine Learning/Math:** PU Learning, Graph ML (PageRank/Centrality), Market-Invariant Modeling, Time Series, Bayesian Optimization, A/B Testing

**NLP & LLM Systems:** LLM Fine-Tuning (PEFT, QLoRA, LoRA), Retrieval-Augmented Generation (RAG), Embeddings & Semantic Search, Prompt Engineering, LLM Evaluation Frameworks, Tool Calling, Structured Outputs, Multi-turn Conversations, Vector Databases

**Agentic Frameworks:** LangChain, AutoGPT, Agentic Workflows, LLM Orchestration

**Deep learning:** GANs (ClinicalGAN, CycleGAN), PyTorch, TensorFlow **Data Engineering:** Distributed Computing, Spark Optimization, ETL Pipelines, Airflow

**Cloud & MLOps:** AWS (SageMaker, S3, Redshift), Docker, CI/CD

**Databases & Tools:** PostgreSQL, MySQL, Git, Linux, Jupyter

## Professional Experience

---

**Associate Consultant / Senior Data Scientist**

ZS, India

*Jan 2024 – Present*

- **Developed a production-grade agent-driven content authoring and evaluation system** to automatically generate training material and assessment Q&A from multi-document enterprise knowledge sources.
  - Architected a modular LLM orchestration workflow (LangChain-style graph orchestration) with: Hybrid RAG (vector + Azure AI search)
  - Tool/function calling for structured Q&A generation, and Multi-turn conversational validation agents.
  - Designed evaluation framework to benchmark hallucination rates, answer relevance, and retrieval precision.
- **Orchestrated a 1.6-year technical roadmap** as the lead architect bridging Data Science and Engineering for a demand forecasting in revenue management .
  - Developed a **market-invariant generalization framework** that enabled a ML model to outperform localized price elasticity curves across heterogeneous markets and solving for demand forecasting.
- **Engineered a reusable Generative Digital Twin framework** using **ClinicalGAN** to synthesize high-fidelity patient data.
  - Enabled **scalable clinical simulations** and cross-therapeutic data augmentation across multiple therapy areas.
  - Mitigated local optima convergence via a **multi-stage transfer learning strategy**, significantly improving **AID and FID** distribution metrics.
- **Deployed a PoC for automated ICD-10 code extraction** using **GPT-4o-mini** and an **OpenAI Deep Research agent** for patient cohort identification.
  - Orchestrated a multi-stage pipeline on **Databricks**: automated web-fetching from ICD-10 CM, fuzzy matching for disease aliases, and LLM-based filtering to eliminate false positives. Integrated a **Deep Research agent** to autonomously resolve missing mappings for rare diseases, increasing matching accuracy from **30% to 60%**.

**Data Scientist / Data Science Associate**

ZS, India

July 2021 – Dec 2023

- Developed a high-precision HCP Lookalike Modeling framework using **Positive-Unlabeled (PU) Learning** to solve the **Cold-Start problem** for newly launched drugs with sparse adoption data. By identifying high-potential HCPs based on prescribing patterns, automated the prioritization of sales-force call plans, significantly increasing targeting precision and enabling data-driven market entry strategies.
- Engineered a **Graph-based KOL Identification system** by synthesizing disparate data sources (referral networks, claims, and speaker programs) into a **heterogeneous network** to uncover latent influence patterns among healthcare practitioners.
- Developed a **ranking methodology** leveraging **PageRank and Centrality measures** to automate dynamic HCP prioritization; maintained the production pipeline that drove a market share increase from **2.5% to 6%** within 12 months.

**Data Science Intern**

ZS, India

Dec 2020 – June 2021

- Built and automated scalable EDA pipelines integrated into firm-wide analytical tools, utilizing PySpark to accelerate data discovery for billion-row healthcare datasets.

## Education

---

**M.Tech in ECE (Signal Processing & Pattern Recognition)**

IIIT Bangalore, India

2019 – 2021

CGPA: 3.64 / 4.00

- **Graduate Teaching Assistant — Machine Learning:** Mentored a cohort of 120+ students; conducted technical office hours and evaluated advanced examinations for the Graduate ML curriculum.
- **Hackathon Lead:** Designed and orchestrated a department-wide ML hackathon; developed the problem statement and automated evaluation metrics for 30+ competing teams.

**B.Tech in Electrical and Electronics Engineering**

Bhilai Institute of Technology, India

2012 – 2016

Percentage: 71.18%

## Achievements

---

**Won three times Kaggle Bronze Medal**

Top 1.7% nationwide (Ranked 3,331 out of 201,737 candidates).

**Game Changer Award — ZS**

Recognized for exemplary performance and measurable business impact.

## Certifications

---

Generative AI with Large Language Models by Deeplearning.AI