PARAG SONAR

Gen Ai Engineer | AI & ML Engineer

# Professional Summary

Innovative Generative AI Engineer with 6 years of experience designing, developing, and deploying AI models and generative applications in NLP, computer vision, and multimodal domains. Skilled in fine-tuning large language models (LLMs), building diffusion-based image generation systems, and implementing scalable AI pipelines. Passionate about transforming business challenges into intelligent, automated solutions.

# Work History

**Senior Software Development Engineer Optimum Solutions** – Pune

**Client - MasterCard**

06/2024 to 06/2025

* Worked on end-to-end delivery of data science projects for pattern recognition, statistical modeling, and fraud detection, which involved data extraction, EDA, wrangling, and building machine learning ensemble models like Random Forest, XGBoost, CATBoost, LightGBM, SVM, DT, and hyperparameter tuning.
* Work on segmentation: used transaction-level and demographic data to build customer

profiles. An ensemble of K-Means and transaction rules was applied to build the segments using SQL and Python.

**Project**

**GEN AI Driven AML (Anti-Money Laundering) Compliance Bot**

* Developed a fine-tuned LLaMA-2 model to analyze flagged transactions for suspicious activity in line with FATF and RBI AML guidelines.
* Implemented XGBoost anomaly scoring to detect patterns such as structuring, and rapid movement of funds.
* I built a semantic search knowledge base in Weaviate to cross-reference past AML cases.
* Automated SAR (Suspicious Activity Report) drafting reduces compliance documentation time from three days to a few hours.
* Deployed a serverless architecture using AWS Lambda for cost-efficient batch processing.
* **Tech Stack: Python, PyTorch, Hugging Face Transformers, LLaMA-2, LangChain, Weaviate, XGBoost, AWS Lambda, S3, CloudWatch, and REST APIs.**

**Real-Time Transaction Fraud Detection and Explainability Platform.**

* Built a real-time fraud detection pipeline using Isolation Forest and Autoencoders to detect unusual transaction patterns at millisecond latency.
* Integrated **GPT-4 via LangChain** to generate explainable fraud alerts, providing investigators with natural-language summaries.
* Deployed Kafka streaming for continuous ingestion from payment gateways, scoring via SageMaker endpoints.
* Used Pinecone vector DB to retrieve similar past fraud cases for context-aware LLM outputs.
* Scaled the system using Kubernetes, and stored aggregated risk data in Snowflake for downstream analytics.
* **Tech Stack: Python, PyTorch, Hugging Face Transformers, GPT-4, LangChain, Pinecone, Isolation Forest, Kafka, AWS SageMaker, Docker, Kubernetes, Snowflake.**

**Cross-Border Payment Fraud Detection with LLM Risk Scoring**

* Designed a system to monitor cross-border payments for suspicious patterns, such as structuring, round-tripping, and shell accounts.
* Trained BERT-based embeddings for transaction metadata, and stored them in Neo4j for graph-based fraud pattern detection.
* Used GPT-4 with financial compliance prompt engineering to produce **risk scores and**

**narrative justifications**.

* Implemented Airflow pipelines to process daily settlement data in BigQuery, and trigger fraud alerts in real time.
* **Tech Stack: Python, TensorFlow, BERT, OpenAI GPT-4, Neo4j, LangChain, Airflow, GCP BigQuery, Vertex AI, Docker.**

**Data Scientist**

**Grihum Housing Finance** – India

03/2023 to 07/2023

* Develop unsupervised and semi-supervised NLP and computer vision models to profile textual and visual data into pre-defined or undefined classes and values, along with relationships, to achieve auto-encoding of textual data.
* Identify fraudulent transaction metrics by building a fraud detection model using

classification models.

* Build an ANN-based model for predicting the approval or rejection of loan applicants' cases.
* **Tech Stack: Python, PyTorch, SQL, AWS, Power BI.**

**Manager - Analytics & Reporting PNG Jewelers** – Pune, India

08/2022 to 02/2023

* Identify valuable data sources and automate operational processes using data analytics and machine learning techniques.
* Created a Power BI dashboard for bridging the gap between technology and the retail business using data analytics to deliver data-driven recommendations and reports to executives and directors.
* Used SQL and Power Query for end-to-end product sales data pipeline.
* It helps with sales prediction in festive seasons. Analyze large amounts of sales information using Python and R: a program to discover trend analysis and pattern recognition.
* Build predictive models and machine-learning algorithms.
* **Tech Stack: AWS, Python, PyTorch, Power BI, and SQL**

**Product Data Analyst**

**Reality Premedia Services** – Pune, India

11/2021 to 06/2022

* Created data visualizations and a data crawling tool using Python and R programming for visualization with Power BI, and presented it to the Technology and Production Team.
* Developed and deployed advanced statistical models, predictive models, and machine learning methods (Random Forest Method, Decision Tree Method, SVM, NLP, Gradient Boosting, supervised/unsupervised learning, clustering, classification, and regression modeling).

**Project**

**AI-Powered Health Scorecard Prediction System**

* Built a predictive model to **generate a health score (0–100)** for individuals based on demographic data, medical history, lifestyle habits, and lab test results.
* Collected and preprocessed data from hospital EHR systems and wearable IoT devices,

performing feature engineering (BMI, risk indices, and chronic disease markers).

* Implemented machine learning models (XGBoost, LightGBM, Random Forest) for baseline scoring, and deep learning models (ANN, LSTM) to capture nonlinear relationships and temporal health trends.
* Used multi-output regression to predict multiple health risk indicators (cardiac, metabolic, respiratory), and aggregate them into a composite health score.
* Applied **SHAP explainability** to show patients and doctors which lifestyle or biomarker factors most influenced their score.
* Deployed the model via AWS SageMaker, with a Flask-based API, and a Plotly/Dash

dashboard for interactive patient health monitoring.

* Achieved a 92% correlation between predicted scores and actual physician evaluations in validation trials.
* **Tech Stack: Python, Pandas, NumPy, Scikit-learn, XGBoost, LightGBM, TensorFlow, Keras, SHAP, Matplotlib, Plotly, AWS SageMaker, MySQL.**

**Data Analyst**

**Innoplexus** – Pune, India

* Saved around 70% of resource time from planning and execution.

02/2018 to 05/2021

* Saved around 80% resource time using data visualization tools like Power BI, MS Excel, Python, and R programming in the pharma domain for data forecasting.
* Worked to design an advanced neural network-based model that can predict the probability of success of clinical trials with a precision of 85%.
* Increased data scraping capabilities and data quality by 4X. Raised code quality through data reviews that help to reduce deployment problems by 60%.
* Developed an internal web-based solution database, bringing team resolution rates from

60% to 90% in three months.

* Created the Power BI dashboard for clients like Commerzbank Group, Pfizer, and Ranbaxy. Used Python and SQL for time series analysis and cluster analysis.
* Build and maintain a database pipeline of heart rate data and SQL queries to track operational productivity. It helps to monitor the Patient Health Record for the Sensor Detection Model.

**Project**

**Clinical Trial Outcome Prediction Model**

* Developed a machine learning pipeline to **predict the likelihood of clinical trial success** based on historical trial data, patient demographics, drug molecular properties, and protocol details.
* Collected and preprocessed data from **ClinicalTrials.gov** and PubChem APIs, handling

missing values, encoding categorical variables, and engineering domain-specific features.

* Trained ensemble models (XGBoost, Random Forest), and neural networks to achieve approximately 82% accuracy in predicting trial success or failure.
* Incorporated **NLP with BERT** to extract meaningful features from trial descriptions and adverse event reports.
* Used SHAP (Shapley Additive Explanations) for interpretability, identifying key factors influencing trial success (e.g., sample size, trial phase, and patient selection criteria).
* Deployed the model on AWS SageMaker with a REST API for real-time predictions, and

integrated a dashboard for researchers to visualize risk factors and probability scores.

* **Tech Stack: Python, Pandas, NumPy, Scikit-learn, XGBoost, TensorFlow, Hugging Face Transformers, SHAP, Matplotlib, AWS SageMaker, PostgreSQL.**

# Education

**Master of Science**: Economics

**Gokhale Institute of Politics And Economics** - Pune

* 6/10 GPA/CGPA

# Skills

Retrieval Augmented Generation (RAG) Model Optimization

08/2017

Prompt Engineering Multimodal Models

Model Architectures LLM Training & Fine Tuning

Generative Adversarial Networks (GAN's) Synthetic Data Generation

Cloud Platforms Containerization

Model Serving Scaling

Loss Function

# Core Competencies

* **Database Management :** SQL, Azure Cosmos DB
* **Data Visualization** : Microsoft Power - Bi, Tableau
* **Programming Language** : Python, R - Programme, STATA
* **Machine Learning and Deep Learning** : Linear Regression, Logistics Regression, Supervised and Un-Supersized Learning
* **Cloud Computing** : GCP, AWS, AZURE
* **Generative AI**: LLMs (GPT, LLaMA, Falcon), Diffusion Models (Stable Diffusion, DALL·E), GANs, VAEs
* **NLP**: Prompt engineering, text summarization, translation, conversational AI
* **Computer Vision**: Image synthesis, object detection, style transfer
* **Frameworks & Tools**: PyTorch, TensorFlow, Hugging Face, LangChain, OpenAI API, Ray
* **MLOps**: Docker, Kubernetes, MLflow, Weights & Biases, AWS/GCP/Azure AI services
* **Data Engineering**: ETL pipelines, data preprocessing, synthetic data generation

# My Time

Design and Maintain Data System

Intermediate

Fixes - Bugs/Coding issue in Real Time

Data Modelling, Data Pipeline Flow

Upper intermediate

Design and Developed of Statistical and

|  |  |  |  |
| --- | --- | --- | --- |
| Upper intermediate |  | Predictive Models |  |
|  |  | Advanced |  |
| Languages |  |  |  |
| English |  | Marathi |  |
| Advanced (C1)  Hindi |  | Bilingual or Proficient (C2) |  |
| Bilingual or Proficient (C2) |  |  |  |
| Certifications |  |  |  |
| Financial Engineering and Risk Management |  |  | 05/2020 |
| Machine Learning |  |  | 04/2019 |
| Deep Learning Specialization |  |  | 07/2019 |
| Probabilistic Graphical Models |  |  | 08/2019 |