Rushikesh Aher

Pune, India | 9518545743 | rushiaherpatil@gmail.com | LinkedIn | GitHub

Data Scientist with **3.3 years** of professional experience excelling in the development of comprehensive **Natural Language Processing, Generative AI, Azure OpenAI, Computer Vision, Deep Learning, Machine Learning.** Hands-on experience with frameworks like **Langchain** Pytorch, FastAPI, MLFLOW, and libraries like transformers, Huggingface OpenCV, OpenAI. Adept at solving complex business challenges Currently spearheading. the development of for **Large Language Models - LLMs** at **Cognizant India Pvt Ltd.**
Top-notch communicator with excellent learning skills, along with time management, administrative and delivering exceptional customer service.

Skills

Python | Databases | Vector Database
Natural Language Processing - Embedding, Classification, NER, BERT, RNN, LSTM, GPT, **Transformers**, Transfer Learning
Generative AI - Langchain, Azure OpenAI, LoRA, PEFT, LLMOPS, Langsmith, TruLens, **RAG** -Retrieval Augmented Generation, GPT
Agentic Framework, llamaindex, **Prompt Engineering**, Chain Prompt, Tree of thought Prompt, ReAct Prompt, Llama-3, QLoRA
MLOps - MLflow, Azure ML, Azure DevOps, Sagemaker, Docker, Github Action | **Linux** Ubuntu
Rest APIs- FastAPI, Flask, Streamlit as UI | **Cloud** -AWS Ec2, Lambda, API Gateway, Azure ML, Azure OpenAI, Azure AI

Professional Experience

Data Scientist, *Cognizant India pvt Ltd* 09/2021 – present | Pune, India
Role & Responsibilities: -

- Analyzed business needs and delivered data-driven solutions for multiple clients, collaborated with senior managers and clients to prepare **Statement of Work (SOW)** and design **end-to-end workflow** for each project.Worked on the latest **Gen AI** techniques, **RAG Pipelines, Prompt Engineering, LLM,** Azure OpenAI, Llama3, GPT 4, Langchain, Azure AI Search and vector database
- Expert in creating a quick NLP, GenAI and LLM concept prototype as well as converting prototype into product.

Projects

Intelligent Automation with GenAI Agent, *Cognizant India Pvt Ltd.*
***Tech**: Python, Azure OpenAI, Langchain output parser, Postgres, Prompt engineering, AWS API Gateway, AWS ECR*
Developed an advanced GenAI agent capable of comprehending customer incident tickets, extracting essential information, and efficiently tracking ticket resolution progress.

- Utilize the Resolve action express workflow tool to retrieve tickets from ServiceNow and seamlessly transfer them to the GenAI Agent for analysis.
- Employ the Langchain output parser to discern the request type embedded within the ticket and extract all pertinent details required for resolution.
- Perform post-processing of the extracted data and update predefined templates if all necessary values are present in the ticket description.
- In instances where essential information is absent from the ticket description, Agent notify the user via email and comment on the ticket, requesting the addition of specific fields.
- Efficiently track the ticket's progress and engage in communication with the user to resolve the ticket effectively

GenAI Driven Ticket Management, *Cognizant India Pvt Ltd.*
***Tech:** Python, Azure OpenAI, Langchain, Service Now, Azure AI Search, Azure Document Intelligence, Azure Vision, Vector Database*
By using GenAI and a knowledge base to efficiently manage and resolve customer issues by assessing incoming tickets and assigning
them for immediate resolution or to respective teams.

- Extract all information from tickets using prompt entity extraction logic.
- Seamless integration with a centralized knowledge base for instant solutions
- Automated ticket updates, keeping customers informed and engaged.
- Efficient team assignment for complex issues, ensuring the right experts handle them.
- Real-time monitoring of incoming tickets for swift response and resolution.

Multilingual Article Retrieval & GenAI Question Answering Chat Bot, Cognizant *India Pvt Ltd.*
***Tech**: Python, Huggingface, sentence transformers, Azure Machine learning, Azure OpenAI, Faiss index, GenAI, Langchain.*
Developed an advanced system for precise multilingual (French, English, Dutch, Spanish, Polish, German) article retrieval by using sentence similarity and providing answers to user queries within articles by using GenAI question answering system.

- Enabled semantic search for multilingual articles retrieval.
- Here I have worked with French, English, Dutch, Spanish, Polish, German languages
- In memory storage for all article against a unique ID.
- Provided answers to user queries within Saved in memory index.
- If user Query is not solved, then transferring it to agent.

Advance RAG- Confidential Document Interaction with GenAI, *Cognizant India Pvt Ltd*.
***Tech**: Python, Azure Search, Azure Document intelligence, Azure index, Agent, Azure OpenAI, Azure*
Developed a secure document interaction system, combining GenAI, LLM, RAG, Conversation Memory, and Azure Vector Store. Enables private conversations around uploaded documents, featuring secure upload, encrypted storage, AI-driven chat, context retention, and RAG for improved document privacy, reduced data exposure risk, and enhanced user experience.

- Creating Azure Workspace and then resource group
- Creating all required module (ex: azure OpenAI, azure backend deployment)
- Uploading Our own private document.
- Updating prompt w.r.t particular domain and data set
- Creating New GenAI agent if its required
- Adding different types of prompt engineering technique (ex: one shot, chain of thought, ReAct, Tree Of Thought)
- Then Deploying using azure backend service

AIOPS Solution for CPU Prediction and Root Cause Analysis, *Cognizant India Pvt Ltd.*
***Tech**: Python, Grafana, Prometheus, Open telemetry, NLP, LSTM, RNN*

- Focuses on predicting CPU performance and performing root cause analysis in a Web Application 500K real time CPU metrics per day. Utilized Recurrent Neural Networks (RNN) to predict CPU consumption patterns. Distinguishing critical alerts from non-critical ones. Extended beyond CPU performance prediction, incorporating a cascaded model to identify the specific application module responsible for CPU spikes, leveraging downstream application traces collected using Open Telemetry.
- This AIOPS solution enhances system reliability and performance by predicting CPU behavior, prioritizing alerts, and pinpointing the exact application module causing CPU spikes, contributing to efficient system management and troubleshooting.

Education

10th from HSC Board *2014 - 2015*
12th from CBSE Board *2016 - 2017*
B.E. Information Technology | Pune University *2017 - 2021*

Certification

Microsoft Certified: Azure Data Scientist Associate (DP-100)
Microsoft Certified: Azure Data Fundamental (AI-900)
Generative AI With Large Language Model - Deeplearning.Ai
Pytorch for computer vison - Udemy
Python programming - Udemy
LLMs Mastery: Complete Guide to Transformers & Generative AI