

SRI PHANI SUBRAMANYAM

Data Scientist | AI Engineer | Python Developer

@subbu27498@gmail.com
subbu-art

+447833692502

Englefeild Green, Egham

Surrey, UK

subbu27

EXPERIENCE

Data Scientist

AI4Less

October 2023 - Present

London, UK

- Experienced in implementing and optimizing various **state-of-the-art Large Language Models (LLMs)** including **Mixture of Experts, GPT, LLAMA, and BERT** for diverse natural language processing tasks.
- Employed **LoRA (Local Rank Adaptation)** on a **Bloom Large Language Model** to construct a fine-tuned model and then trained this model using personalized data.
- Created **RAG pipelines** using **Langchain** to make the LLM to learn the customized data effectively and fine tuned its responses via **Prompt engineering**.
- Created an **invoice data extraction pipeline** utilizing a Large Language Model (LLM) for both **structured and unstructured data**. This pipeline enables users to seamlessly **analyze invoices** and more effectively extract relevant information.
- Effectively managed **Pinecone Vector Database** to optimize **Large Language Models (LLMs)** through efficient **vector storage, retrieval, and similarity computations**, enhancing performance in natural language processing tasks.

Senior AI Engineer

Capgemini Invent

Jan 2022 - September 2022

India

- Explored **Graph Neural Networks (GNNs)** and different architectures like **Graph Convolutional Networks (GCN), Graph Attention Networks (GAT)**, and others, and have developed a solid grasp of **PyG**.
- Engineered node embeddings by training the GNN model on a **bipartite graph dataset**, enhancing representations with additional features.
- Employed **Seismic Architecture** within a **Graph Neural Network (GNN)** using **Graph Attention Networks (GAT)** and **Graph Convolutional Networks (GCN)** to develop a classifier for predicting Drug-Target Interactions.
- Contributed to advancing understanding and adoption of GNNs in recommendation systems, showcasing potential for **personalized and context-aware recommendations**.

Senior AI Engineer

Capgemini Invent

Dec 2020 - December 2021

India

- Led the end-to-end development of a cutting-edge AI application specializing in the extraction of information from diverse **structured and unstructured data sources**.
- This innovative tool significantly optimized operational efficiency by reducing manual workforce investments by **62%**. Ensured uninterrupted **client support with 24/7 availability**, showcasing a commitment to delivering reliable and impact solutions.
- Achieved a **92% accuracy** in text processing and classification, harnessing **Spacy models** for comprehensive data understanding and analysis.

SUMMARY

Bringing over 4 years of software development expertise, skilled in multiple programming languages and leveraging various cloud services, including AWS. Proficient in designing and implementing scalable architectures for improved performance on AWS cloud infrastructure. Experienced in utilizing Large Language Models (LLMs) like BERT, GPT, etc.

MOST PROUD OF



Extra Mile Award

Recipient of the prestigious Extra Mile Award at Capgemini



Out of the Box

Awarded in recognition for my outstanding performance.



Persistence & Loyalty

I persevered through challenging moments, demonstrating my commitment to remain with Capgemini.



Innovative AI Invoice Automation

Led the development of an application reducing manual invoice handling with great accuracy.

EDUCATION

M.S. in Artificial Intelligence

Royal Holloway University of London

Sept 2022 - December 2023

M.S. in Computer Science

University of York

Aug 2021 - August 2022

Bachelors in Electronics

GITAM University

May 2015 - April 2019

SKILLS

LLMs

Python

GNNs

Langchain

PyTorch

PyG

SQL

Vector DB

Knowledge Distillation

RAG

Django

- Attained a **95% accuracy** rate utilizing **BERT's** advanced contextual comprehension for intricate language patterns in text classification tasks
- Utilized **pattern-based OCR** technology with **89% accuracy** to extract text from scanned documents, enhancing data accessibility.
- Created a **RabbitMQ pipeline** that efficiently transmits data from individual modules as files undergo processing within the system.
- Designed **UiPath** workflows tailored to automate the retrieval of data from various client environments such as **CITRIX, VMware**, etc., utilizing information extracted by AI/ML modules.
- Designed and implemented AWS CI/CD pipelines using **AWS CodePipeline** and **AWS CodeBuild**, automating build, test, and deployment processes for efficient software delivery.

Conversational Engineer

Capgemini Invent

📅 Sep 2019 – December 2020 📍 India

- Demonstrated proficiency in working with **AWS, Azure**, and **GCP** cloud services, showcasing the ability to navigate and utilize various cloud services for building and deploying applications.
- Acquired in-depth knowledge and practical experience in **NLP techniques**, including **text analysis**, **sentiment analysis**, and **language generation**, enabling effective communication between machines and humans.
- Developed proficiency in building conversational agents using platforms like **Kore.ai, DialogFlow, RASA, IBM Watson** and **AWS Lex**. Successfully created and deployed a live skill on **Amazon Alexa**, IVR based AI bot on Amazon Connect, showcasing the ability to design and implement interactive voice-based applications.
- Created a Leave Management bot utilizing **Kore.ai**, streamlining the entire process as an alternative to manual submission through the website.
- Created a voice-based chatbot in a regional Indian language, leveraging Microsoft's **STT (Speech to Text)** and **TTS (Text to Speech)** services, alongside **LUIS** and the **Microsoft Bot Framework**. Achieved exceptional accuracy, enhancing conversational experiences.
- Utilized **Botium Box** for crafting **automated chatbot test scripts**, rigorously covering all scenarios to **eliminate loopholes** and ensure **comprehensive test coverage**.

PL/SQL Developer

LiquidHub

📅 Dec 2018 – Aug 2019 📍 India

- Crafting **ETL data flows** involves designing mappings or workflows to efficiently extract data from **SQL**.
- Experience in designing and creating **RDBMS Tables, Views, User Created Data Types, Indexes, Stored Procedures, Cursors, Triggers and Transactions**.
- Established data quality benchmarks by implementing tailored queries in **Python** to assess and ensure the integrity and accuracy of the source data.

CERTIFICATION

- AWS Solution Architect
- AWS Cloud Practitioner
- Team Creation - Harvard Business Publishing
- Rasa Certified Developer
- AWS Academy Graduate

ACHIEVEMENTS

- Designed and implemented the **AIHD project**, streamlining invoice processing through automation. Achieved a remarkable **62% reduction** in investment while ensuring enhanced efficiency and high availability.
- Created a **volunteer chatbot** catering to rural communities in India, delivering assistance in a regional language to support a meaningful social cause.
- Engineered a robust health check framework, slashing server downtime and elevating application availability by an **impressive 92%**.
- Created an Intelligent Invoice Extraction Assistant allowing seamless interaction with invoice data through natural language, providing users with rich insights and analytics capabilities.

PERSONAL PROJECTS

- **Tabular Data Extraction Project:** Automated Tabular Data Extraction and Excel Integration using **YOLO** and **Tesseract**.
- **Intent Routing Classifier:** Optimizing Chatbot fallbacks between Gen AI and Predefined Intents.
- **Interactive Housing Price Prediction Project:** OML Implementation with Extensive Datasets from UK and USA
- **Image Summarizing Chatbot Architect:** Leveraging LLMs, Langchain, and YOLO for Advanced Object Detection
- **Implemented Django-Based Health Check Framework:** Implemented Monitoring System with Django for CPU Performance and Module Availability Across Multiple Ports.
- **Bridging Transformers and GAT for Deeper Token Relationships:** Investigating the fusion of Graph Attention Networks (GAT) attention scoring into Transformers to harness GNNs' semantic depth and token relationships. This integration seeks to boost contextual understanding in Transformer models, potentially elevating natural language processing precision and sophistication.