Raju Gopi

+91 9353973699 rajugraj29@gmail.com www.linkedin.com/in/raju-gopi-contact

SUMMARY

ML/Gen AI Engineer| 2+ Years of Experience

A results-driven ML/Gen AI Engineer with 2+ years of experience specializing in LangChain, OpenAI, and Gemini. Proficient in building solutions for Retrieval Augmented Search (RAG), code modernization, and unit testing automation. Experienced with VectorDB integration, and skilled in using models like Llama-7b. Strong foundation in machine learning, with a focus on delivering scalable, high-performance solutions. Proven track record in driving successful outcomes through effective problem-solving and collaboration.

EDUCATION

RNS Institute of Technology **08/2018 – 07/2022**
*Bachelor of Engineering – **8.1** CGPA* ***Bengaluru, India***

TECHNICAL SKILLS

Programming Languages: Python (Primary), Java
Generative AI & LLMs: Paid and Open-Source LLMs (OpenAI, Google Gemini Pro, Llama2, Hugging Face Transformers, and GitHub Copilot)
AI/ML Technologies: Fine-tuning with custom data, Vector Embedding, NLP, Prompt Engineering, QLoRA (Quantized Low-Rank Adapters), LoRA (Low-Rank Adaptation), RAG (Retrieval-Augmented Generation)
Frameworks & Tools: LangChain, Streamlit, GitHub, Docker, FastAPI
Deployment & Platforms: Hugging Face Spaces, MySQL, ChromaDB, GitHub Action
Visualization & Documentation: PowerBI, Microsoft Office
Data Science & Miscellaneous: Pandas, NumPy, Scikit-learn, SQL Alchemy, Traditional ML Algorithms, MLFlow
Frontend Development: HTML, CSS, Bootstrap

EXPERIENCE

Infosys – Client: 1

08/2022 – Present
Bengaluru, India

- **AI & NLP Model Integration**: Developed and integrated **AI models** like **Google GenAI**, **OpenAI**, and **Hugging Face Transformers** for automating business processes, including natural language understanding, task generation, and data categorization.
- **Python Backend Development:** Developed Python backend using **FastAPI** for web applications, creating scalable and high-performance APIs to support business automation and data processing workflows.
- **Data Processing & Transformation**: Built **data pipelines** for **OCR-based text extraction** using **OCRSpace API** and image processing to convert unstructured data (images, documents) into structured formats (e.g., **JSON**, **CSV**), enhancing data accessibility and usability.
- **Model Optimization & Evaluation**: Fine-tuned **machine learning models** like **BART** and **GPT**, using domain-specific datasets to improve model accuracy, and applied performance metrics such as **sacrebleu** to evaluate and optimize outputs for task generation and summarization.
- **User Interface (UI) Design**: Developed interactive **Streamlit dashboards** to provide intuitive interfaces for non-technical users, enabling easy data input, result visualization, and export capabilities (CSV, PDF reports).
- **Technical Documentation Automation**: Automated **documentation generation**, including component summaries, flowcharts, and sequence diagrams using **Mermaid.js**, enabling dynamic updates and easy sharing of technical insights.
- **Offline & Scalability Solutions**: Implemented **offline-first** solutions that allow models to run without internet connectivity, ensuring **data privacy**, **security**, and scalability for large-scale operations in varied environments.
- **API & External Tool Integration**: Integrated third-party **APIs** like **OCRSpace API** and **Google GenAI** to enrich functionality and enhance the performance of automation systems.
- **Version Control & Collaboration**: Utilized **Git** and **GitHub** for version control and collaboration across teams, ensuring smooth development workflows and effective codebase management.
- **Performance Tuning & Optimization**: Focused on optimizing model inference speed and reducing computational overhead, ensuring automated systems are highly efficient and capable of processing large datasets.

Client: 2

- **Predictive Model Development**: Designed and deployed a predictive model to forecast **Time on Wing (ToW)** for engines, managing the full lifecycle from data collection to visualization, providing actionable maintenance insights.
- **FMECA Analysis**: Led **Failure Modes and Effects Criticality Analysis (FMECA)** for gas turbines using **Excel Power Query** and **Power BI**. Automated data extraction with **Python**, reducing processing time by over 90%. Developed dashboards to highlight critical failure modes.
- **Design Iteration Management**: Oversaw **Design Definition Change Document (DDCD)** assessments, coordinating with internal teams and managing communications with **Airbus** on design changes.
- **Process Optimization & Stakeholder Communication**: Streamlined data workflows and optimized processes, improving efficiency and ensuring effective communication with internal teams and external clients.

ACCOMPLISHMENTS

- **Insta-Award- Client Appreciation:** Designed and developed intuitive and visually compelling Power BI dashboards to present complex, large-scale data in a clear, accessible, and actionable format, enabling stakeholders to easily interpret and make informed decisions.