Farhaan Pasha

Bangalore, India

📞 +91-9108950377 ✉ fazzofficial18@yahoo.com in LinkedIn

AI professional with 2+ years of experience delivering advanced LLM and AI solutions on Microsoft Azure. Specializing in training and fine-tuning large language models (LLMs) using state-of-the-art techniques like PEFT, LoRA, and Transformer architectures. Proven expertise in designing and deploying cross-platform LLM applications and integrating inference systems, driving measurable improvements in performance and efficiency. Strong background in AI delivery, cloud solutions, and staying abreast of the latest LLM research trends.

EDUCATION

Rajiv Gandhi Institute of Technology, Bangalore   **2018 – 2022**
*B.E. - Computer Science and Engineering - **CGPA - 8.2***   *Bangalore, Karnataka*

COURSEWORK / SKILLS

Machine Learning with Python ⚭

Introduction to Deep Learning and Neural Networks ⚭

PyTorch for Deep Learning Bootcamp ⚭

EXPERIENCE

Malomatia India Technologies Private Limited   **September 2022 - Present**
Role : Junior Consultant - Data/AI   *Pune, India*

- Fine-tuned large language models such as **LLAMA2** and **LLAMA3** for NLP tasks, improving performance by 25%.
- Engineered chatbot functionalities using **Azure OpenAI** and **LangChain**, increasing real-world chatbot accuracy..
- Architected cross-platform **LLM** applications for web and mobile, reducing development overhead by 30% and ensuring seamless integration with existing infrastructure.
- Developed an AI-powered dashboard using LLMs for profile querying and real-time data visualization.
- Reduced search query response time by 15% with memory optimization techniques.

TECHNICAL SKILLS

Languages: Python, JavaScript, Node.js, SQL

Technologies: Microsoft Azure, Generative AI, Azure OpenAI, OpenCV, QLoRA, Retrieval-Augmented Generation (RAG), Redis, Azure DevOps, Deep Learning, Attention Mechanism

PROJECTS

Visual Transformer-based Image Classification | PyTorch, Transformer   **2024**

- Implemented Vision Transformer (ViT) for image classification, achieving 95% accuracy.
- Improved model robustness using **data augmentation** techniques, such as resizing and normalization.
- Monitored and refined model performance through hyperparameter optimization and TensorBoard tracking.

Customer Segmentation using LightGBM | LightGBM, Pandas, Auto-Encoder   **2024**

- Preprocessed marketing data using EDA, addressing missing values and outliers.
- Improved clustering performance by 20% through standardized feature normalization.
- Achieved 98% accuracy using LightGBM after dimensionality reduction with Auto-Encoder.
- Visualized clustering results using Matplotlib for actionable insights.

COMPETENCIES & STRENGTH

- Agile Methodologies, Data Architecture Design, Problem Solving
- Tech Community Engagement, Adaptability, Public Speaking at Tech Meetups

LANGUAGES

- English (Fluent)
- Hindi (Fluent)