Abhishek Pandey

AI & ML Engineer

+91 9008938916

pandey86@live.com

Experienced Machine Learning Engineer (4+ years) with a strong focus in generative AI, computer vision and deep learning, utilizing advanced analytics and ML technologies to develop and optimize scalable AI solutions that drive business value and innovation.

What I Can Offer:

1. Expertise in developing and deploying applications leveraging Large Language Models (LLMs), ML and DL algorithm's enhancing business solutions and client engagements.
2. Specialized in designing custom RAG frameworks using agents to facilitate advanced data retrieval and integration.
3. Proficient in deploying open-source models in production environments, ensuring optimal performance, and maintaining high-quality standards.
4. Gained industry-standard expertise through collaborations with high-profile clients such as Airbus, Valeo, P&G and Cisco delivering tailored technological solutions

Technical skills:

Machine Learning modeling and Deep Learning algorithms | Chatbot Design and Development | Generative AI for NLU, NLG and Images | LLM model Fine-tuning ;Lora, QLora| Prompt Engineering | RAG | NLP applications | Computer Vision applications | AI applications in production | Time Series forecasting | Statistics | Data Visualization & Analysis using Tableau, PowerBi and Matplotlib

Programming Languages: Python | Vector database
Libraries and Frameworks: TensorFlow & Keras | PyTorch | Scikit-image | Spacy | OpenCV | Scikit-Learn | Transformers | Ffmpeg | YOLO | NLTK | Langchain | LangGraph | Llama Index | Huggingface | ollama | Numpy | Pandas
Cloud & API: GCP Vertex AI | OpenAI API | Huggingface | Amazon Bedrock
Deployment: Streamlit | Fast API | Git & GitHub | Metaflow | W&B (Weights & Biases) | TensorBoard | Docker

Work Experience:

Sr. Software Engineer- AI and ML

Randstad Digital Pvt. Ltd., Bengaluru **Nov'23-Present**

- Led the development of an HR chatbot for Randstad Digital, utilizing multi document Retrieval-Augmented Generation (RAG) techniques with cutting-edge LLMs like GPT-4, Google Gemini models, and open-source models from Hugging Face.
- Orchestrated a team of three engineers to develop Generative AI Solution for Legacy Code Modernization: Utilized open-source large language models (LLMs) for specific tasks like pseudocode generation, text-to-code translation, code testing and documentation generation.
- Engineered a Generative AI-powered Synthetic Data Generator leveraging large language models. Implemented advanced statistical analysis techniques, including KDE and chi-square tests, to create privacy-preserving datasets maintaining original data distributions and correlations. Developed asynchronous batch processing, real-time progress tracking, and comprehensive statistical comparison features, ensuring high-fidelity synthetic data generation at scale.
- Constructed a Natural Language Understanding (NLU) chatbot for HR management, incorporating prompt engineering and generative AI to handle policy checks, leave management, and meeting scheduling, resulting in a 20% reduction in HR administrative time.

- Conducted research on state-of-the-art generative AI models such as Llama, Mistral AI optimizing them for application use across content generation, conversational AI, and other applications.

Analyst-Advanced Analytics (R&D)

Optimal Strategix Group Inc, Bengaluru **Mar'22-Sep'23**

- Automated Video Analytics Pipeline for Customer Insights: Developed a complete pipeline for analyzing screen-recorded videos to understand customer behavior during their online shopping journeys. Implemented OCR technology to detect and extract text from video frames, generating valuable insights for feature automation. The automation reduced processing time by 50%, significantly accelerating the delivery of customer insights. The solution enhanced the efficiency of video analytics, providing quicker and more accurate business decisions.
- Time Series Forecasting: Developed an end-to-end time series forecasting tool, capable of handling both univariate and multivariate data. The tool includes pre-processing steps, automatic model selection, and hyperparameter tuning to optimize forecasts. Results are displayed in an intuitive dashboard, making it easy to interpret and analyze forecasting outcomes. This solution streamlines the time series prediction process for a wide range of data inputs.
- Integrated range of OpenAI, LlamaIndex and HuggingFace services into data driven applications. and explored a range of LLMs- LLaMA, Phi-1, Falcon, Vicuna 33B assessing their benchmark performances.

Graduate Engineer- Research and Development

IFB Automotive Pvt Ltd, Bengaluru **Mar' 19-Sept 19**

- Developed python script for data ingestion, data cleaning and preprocessing, and automated the process for large automotive datasets.
- Used statistical analysis techniques to identify trends, developed data visualizations and reports, increasing the accuracy of data- driven decision making by 20%.
- Worked with various high reputed clients such as Valeo, Renault & MG Motors and completed their projects within time and budget parameters.

Logistics Analyst

Prakash Parcel Services Ltd, Bengaluru **Oct'17- Dec'18**

- Identify, track, and report regularly on key operational performance metrics.
- Develop quantitative analysis and reports to support key operations decisions, including process optimizations and logistics management.
- Compile and create reports, graphs, charts, documents and slide presentations using SPSS, Microsoft PowerPoint, Excel.

Education

- **Masters of Technology**, Data Science and Machine Learning | 8.99 GPA | PES University | Bengaluru | 2022
- **Bachelors of Engineering**, Electronics and Communication | 62% | Sapthagiri College of Engineering | Bengaluru | 2018
- **Higher Secondary**, Physics, Chemistry, Mathematics, Computer Science | 89.16% | MES Kishore Kendra | Bengaluru | 2014
- **Senior Secondary** | 93.44% | Jindal Public School | Bengaluru | 2012

Certifications Azure Cognitive Services Specialist(*) | OpenAI-L100 | Azure OpenAI Fundamentals