AJAYKUMAR NADELLA

+1 (720)840-5312 | ajaynadella369@gmail.com

AI Engineer Hattiesburg, MS, USA |LinkedIn |GitHub

ML Engineer and AI Architect with around 4 years of experience in Python Development, CI/CD Pipelines, and Cloud Technologies (GCP, AWS). Proficient in designing, deploying, and scaling machine learning solutions across various domains. Expertise in deep learning, NLP, LLM, GEN AI and automation using cloud infrastructure.

PROFESSIONAL EXPERIENCE

ML Research Assistant, The University of Southern Mississippi, Hattiesburg **Aug2023 – Dec 2024**

- Developed predictive and recommendation models in Python, improving accuracy by 20% using advanced ML algorithms.
- Led NLP system design, boosting performance by 35% using BERT, T5, and Llama for tasks like question-answering systems.
- Integrated LangChain, enhancing AI applications such as customer service and document processing.
- Implemented CI/CD pipelines for automated deployment and leveraged AWS & GCP for cloud optimization.

AI/GENAI Architect, Wipro Limited, India **Feb 2022 – Dec 2022**

- Led document automation system reducing processing time by 50%, with Python, TensorFlow, FastAPI, and Docker.
- Developed a high-performance NLP-driven chatbot using Python, increasing user satisfaction by 35%.
- Integrated ML models into CI/CD pipelines, automating deployment using Docker and Kubernetes.
- Deployed solutions across AWS and GCP, ensuring robust cloud-based architecture and scalability.

Machine Learning Engineer, HDFC Bank, India **Jan 2021 – Feb 2022**

- Led HDFC Bank's fraud detection system deployment using python, cutting fraud by 35% using ML technologies.
- Enhanced credit scoring model using Python and data analytics, improving loan decision precision by 15%.
- Collaborated with teams for seamless integration, ensuring compliance with regulatory standards.

Machine Learning Engineer, Global AI, India **May 2020 – Dec 2020**

- Designed ML models for anomaly detection and forecasting, enhancing accuracy by 20%, using Python and TensorFlow.
- Created deep learning models for image tasks, reducing analysis time by 50% with TensorFlow.
- Implemented end-to-end pipelines, integrating models into production with Docker and FastAPI.

PROJECTS

AI-Powered No-Code Chatbot Builder[GitHub]

- Developed and deployed an AI chatbot builder Platform, enabling users to build, train and deploy custom AI chatbots using LangChain, LLaMA, vector database (ChromaDB) for retrieval-augmented generation (RAG).
- Built a no-code chatbot builder with industry-specific customization, document upload support, and API/database integration for enhanced adaptability.
- Implemented secure authentication and role-based access control (RBAC) using Firebase Authentication, enhancing data security and user management
- Enabled multi-channel deployment, allowing users to integrate chatbots on websites and external platforms.
- Utilized AWS (EC2, ECS, Lambda) for scalable hosting and deployment, ensuring efficient chatbot performance and accessibility.

AI Powered Document Processor[GitHub]

- Built an end-to-end document processing system that handles both image-based and text-based documents.
- Integrated Tesseract OCR with LangChain and LLaMA to extract text and enable summarization, translation, classification, question-answering, and anomaly detection.
- Containerized the application using Docker and deployed it on AWS for scalable, real-time performance.
- Developed a Streamlit-based UI for users to upload documents and receive instant, AI-generated insights.

End-to-End Convolutional Neural Network (CNN) Pipeline for Brain Tumor Classification from X-Ray Imaging[GitHub]

- Designed an end-to-end CNN pipeline in TensorFlow for brain tumor classification with an accuracy of 90%.
- Dockerized the model and deployed it via FastAPI with a responsive web interface, enabling seamless real-time classification and user interaction.

Medical Chatbot for Women's Health and Pregnancy Inquiries[GitHub]

- Developed a chatbot for women's health, by leveraging technologies such as LangChain, Llama, ChromaDB(Vector Database)
- Implemented a Retrieval Augmented Generation (RAG) pipeline for accurate, context-aware responses in real time interactions.
- Deployed on GCP Cloud Run with a CI/CD pipeline using GitHub Actions, enabling automated deployment and scaling for seamless updates and reliability.

TECHNICAL SKILLS

Programming Languages	Python, SQL
ML Frameworks	TensorFlow, PyTorch
Libraries	Scikit-learn, Pandas, NumPy, NLTK, Transformers(Hugging Face), LangChain
Tools	Docker, GitHub, FastApi, Flask, Kubernetes, MLFlow, Jenkins, GitHub Actions
Cloud Services	Amazon Web Services (AWS), Google Cloud Platform (GCP), Azure

EDUCATION

University of Southern Mississippi | CGPA: 3.75 / 4.00 — **Jan 2023 – Dec 2024**
Masters in Computer Science — Hattiesburg, USA

Vellore Institute of Technology | CGPA: 8.00/ 10.00 — **Jun 2018 – Jul 2022**
Bachelors in Electronics and Communication Engineering — Amaravati, AP, India