

Sasi Boyapati

+18064512735 | sasiboyapati.12@gmail.com | <https://www.linkedin.com/in/boyapatiasasi/> | : Virginia, USA

AI DEVELOPER | MACHINE LEARNING

- **Experienced AI/ML Developer with 3+ years of hands-on expertise in designing, developing, and deploying machine learning and deep learning models using Python, TensorFlow, PyTorch, and Scikit-learn.** Proficient in NLP and LLM-based applications, leveraging tools such as **Hugging Face, OpenCV, RAG, and LangChain**. Skilled in building scalable data pipelines using **SQL, Pandas, and Spark**, and deploying models via **Docker, AWS, and RESTful APIs**. Proven ability to deliver production-ready AI solutions that enhance performance, automate workflows, and drive measurable business outcomes and expertise in GenAI
- **Technical Skills:** -
 - Programming Languages:** Java, Python, SQL
 - Frameworks & Libraries:** TensorFlow, PyTorch, Keras, OpenCV, Pandas, NumPy, Spring Boot, Flask, LangChain, HIPAA, RAG.
 - Web Development:** HTML, CSS, React, Django, Node.js
 - DevOps Tools:** Git, Docker, Kubernetes, Jenkins, JIRA, CI/CD, MedCoder
 - AI Tools:** OpenAI API, Microsoft Azure, Amazon Bedrock, Hugging Face Transformers, Rasa, AWS (EC2, S3).

WORK EXPERIENCE

Aegis Technology

Jan 2025 – Present

AI Developer

- Engineered and implemented **Retrieval-Augmented Generation (RAG) pipelines using LLMs, ChromaDB, and LangChain**, These mainly used for retrieving clinical, and safety-related information, and they've helped reduce retrieval latency by around 30%.
- Built scalable **multi-agent AI** workflows for NLP and computer vision using **OpenAI APIs**, custom retrievers.
- Developed full-stack applications using **Streamlit** and **FastAPI** for real-time model inference and seamless user interaction.
- Deployed scalable solutions on **AWS (Lambda, EC2, S3)** with **Docker** and **Git-based CI/CD**, boosting deployment speed by **40%** and ensuring **99.9% uptime** in production and aligning with MLOps best practices for reliability and scalability.

Texas Tech University

Aug 2023 - Jan 2025

AI Developer

- Formulated and implemented advanced **LLM-based applications** using **LangChain**, including **multi-agent architectures** and **RAG pipelines**, improving contextual retrieval accuracy by **15%** and reducing response time by **30%** through task automation.
- Led end-to-end development of **NLP systems**, including data preprocessing, feature engineering, and augmentation for **large-scale datasets (1M+ records)**; improved model performance by **18%** through **hyperparameter tuning**.
- Deployed scalable AI solutions using **PyTorch, TensorFlow, FastAPI**, and **AWS**, collaborating with cross-functional teams to integrate transformer models, resulting in **20% faster deployment cycles** and enhancing production system accuracy by **12%**.

Norm Software Solutions

May 2022-May 2023

Data Analyst Intern

- Selected appropriate datasets and data representation methods. Analyzed EHR system data to uncover usage patterns, improving patient data access **efficiency by 40%**.
- Engineered data-driven solutions to optimize **RESTful API integration**, enabling real-time sync with external databases and ensuring **100% data accuracy**.
- Ensured secure handling of patient data and reporting workflows in compliance with **HIPAA**, improving audit-readiness.

PROJECTS

Sign Language Conversion Application:

- Developed a machine learning application using **CNN and OpenCV**, achieving **98% accuracy** in translating hand signs to text.
- Built bi-directional system using **TensorFlow** and **PyTorch** for **text-to-sign** and **sign-to-voice conversion**, enhancing accessibility for **500+ users**.

Mathematical Reasoning in LLMs:

- Built and Tested Transformer-based model on a dataset of **50K+ arithmetic problems** improving mathematical reasoning accuracy by **7%** using rule-based vs. retrieval-based evaluation frameworks..

EDUCATION

- **Master of Science in Computer Science, Texas Tech University** | GPA: 3.66/4.0 | 2023-2025
- **Bachelor of Engineering in Computer Science and Engineering, Parul University, Gujarat** | GPA: 3.25/4.0 | 2019-2023

CERTIFICATIONS

- AWS Certified AI Practitioner, **AWS**, 2025 & Amazon Bedrock Getting Started with Generative AI, Coursera, 2025
- Master Generative AI: Professional level LLM application development, Euron, 2025
- AI Extern , **AT&T**, 2024