

YASWANTH AMPOLU

AI Engineer — Software Development Engineer

☎ 6305151728 ✉ ampoluyaswanth2002@gmail.com [in linkedin.com/in/yaswanth-ampolu](https://www.linkedin.com/in/yaswanth-ampolu) github.com/Yaswanth-ampolu

Professional Experience

AI Engineer

Sierraedge AI Pvt Ltd

May 2025 – Present
Bengaluru, Karnataka

- Designed and deployed end-to-end AI-powered applications, integrating large language models (LLMs) with custom business logic and user interfaces
- Developed scalable APIs and backend services to serve AI inference in real-time applications
 - Fine-tuned open-source language models using task-specific datasets to improve accuracy, response quality, and contextual understanding
 - Collaborated cross-functionally with product and design teams to deliver generative AI features such as summarization, classification, and anomaly detection
 - Optimized performance, latency, and resource usage of AI pipelines across both local and cloud-based environments

Associate AI Engineer

Sierraedge AI Pvt Ltd

May 2024 – May 2025
Bengaluru, Karnataka

- Trained large language models (LLMs) and developed custom datasets to power advanced AI solutions
- Built AI applications, delivering end-to-end solutions for real-world challenges and business needs
- Developed visualization pipelines to present AI model outputs and performance metrics effectively to management

Data Science Intern

CloudTechnologies

Jun 2023 – Jul 2023
Hyderabad, Telangana

- Analyzed 500K+ records using Python, Pandas, SQL achieving 92% accuracy in churn prediction
- Created automated reporting systems reducing manual work by 75% and improving decision speed

Key Projects

Pinnacle AI Application | *React, Express.js, Postgres, Docker, Grafana*

- An AI powered and user oriented tool focused on automation and secure execution in VLSI work-load and analyze metrics increasing organization's accuracy and efficiency over 60% .
- Automated a scalable Dashboard pipeline, optimizing data visualization via log data ingestion for organizational adoption

Remote Interface MCP Server | *TypeScript, Node.js, WebSocket*

- Cross-platform terminal MCP server handling 500+ concurrent connections to handle remote server execution
- Integrated MCP server to AI application to support agentic remote execution via chatbot in remote terminal

Fullstack-App AI Assistant | *React, FastAPI, LangGraph, AI Agent*

- Integrated local model support google gemini developer's repository allowing users with local models along with API connections
- Integrated image retrieval feature support to agentic web search and developed cross platform integration fixes for different system architecture increasing cross platform deployment

Spring Test AI spec designer | *Python, RAG, TogetherAI*

- Developed a Python-based application using PyQt that automatically generates test specifications for springs. Integrated AI capabilities to streamline user input into ready-to-use test designs.
- Delivered under a tight deadline to develop client-centric application fully from scratch with very limited resources

Technical Skills

Programming: Python, JavaScript, TypeScript, Java, C++, SQL, HTML, CSS, Shell, R

AI/ML Frameworks: TensorFlow, PyTorch, Scikit-learn, Keras, LangChain, Hugging Face, OpenAI

ML Techniques: Deep Learning, Neural Networks, NLP, Computer Vision, LLM Fine-tuning

Web/Backend: React, Node.js, Express.js, FastAPI, REST APIs, GraphQL, Vite, Tailwind CSS

Data/Database: Pandas, NumPy, Matplotlib, MongoDB, PostgreSQL, Redis, MySQL, ETL

Cloud/DevOps: AWS (EC2, S3, Lambda), GCP, Azure, Docker, Kubernetes, Git, CI/CD

Education & Certifications

Bachelor of Technology in Information Technology

Aditya Institute of Technology and Management

2020 – 2024
Tekkali, Andhra Pradesh

Certifications: Coursera – AI For Everyone (Andrew Ng), University of Helsinki – Elements of AI, Google Cloud – Machine Learning Crash Course, Microsoft Learn – Career Essentials in Generative AI, IBM SkillsBuild – AI Fundamentals

Specializations: Applied AI/ML in real-world apps, Full-Stack Development, LLM RAG systems, Model Context Protocols (MCPs)

Languages: English (Fluent), Telugu (Native), Hindi (Conversational)