

# Akula Leelavathi

## Machine Learning Engineer

leelavathiakula1122@gmail.com

8919207046

Hyderabad

linkedin.com/in/leelavathi-akula  
portfolio

### PROFILE

Proven expertise in building and deploying real-time systems powered by large language models. Delivered sub-200ms response times with 99.99% uptime in production. Skilled in automating model lifecycles, optimizing inference, and ensuring scalability and reliability. Focused on creating robust, efficient solutions that reduce manual effort through smart automation.

### SKILLS

python • azure • aws • MLflow •  
langgraph • langchain •  
prompt engineering • llama

### PROJECTS

#### ParseFile

##### AI-Powered Text Extraction Engine

Developed ParseFile, an advanced text extraction engine leveraging Google Gemini 2.0 Flash-Exp for high-speed, high-accuracy recognition of both printed and handwritten text.

Key Features:

- AI-Powered OCR - Extracts handwritten and printed text with precision.
- 97% Accuracy - Outperforms traditional OCR solutions.
- Optimized for Speed & Scalability - Minimal latency, high efficiency.
- Versatile Applications - Supports document digitization, automated data entry, and intelligent processing

### EDUCATION

IT, Vishnu Institute of Technology

2020 – 2024

### PROFESSIONAL EXPERIENCE

#### Machine Learning Engineer

Zotok

04/2024 – Present | Hyderabad

##### 1.LangGraph-Based Seller-Buyer Chatbot Platform

Built and deployed a conversational AI system using LangGraph, OpenAI APIs, and Prompt Engineering to handle over 100K seller-buyer queries daily.

- Designed modular LangGraph workflows for scalable, context-aware conversations
- Created efficient prompts and embeddings for real-time query matching and product suggestions
- Connected a Streamlit frontend with an LLM backend for smooth, interactive chat experiences
- Ensured 99.99% uptime and fast 200ms response times under heavy traffic
- Cut down manual support by 80% with smart multi-turn dialogue and fallback handling

##### 2. Real-Time Data Application Development

Developed real-time applications for data retrieval and processing across product, image, invoice, and payment parsing. Utilized PromptFlow and LLM models for optimized performance and seamless integration.

- Achieved 98% prompt efficiency with 100% response accuracy.
- Reduced manual effort, enhancing operational efficiency.
- Delivered scalable, high-performance solutions aligned with business needs.

##### 3. Real-Time Speech-to-Text Application

Built a multilingual speech-to-text application using Google and AWS models, enhancing product search and customer interactions.

- Improved user experience with seamless multilingual navigation.
- Reduced customer support resolution times.
- Delivered a scalable, real-time solution for global accessibility.

##### 4. Advanced Search with Cognitive Search & ML

Implemented Cognitive Search and FAISS indexing for efficient vector-based retrieval, boosting accuracy and performance.

- Enhanced search precision for better user satisfaction.
- Enabled fast, scalable, ML-driven search.
- Improved system efficiency with intelligent retrieval mechanisms.

#### Software Development Intern

Zotok

11/2023 – 03/2024 | Hyderabad

##### 1.API Performance Monitoring (Python)

Developed a Python-based framework for real-time API monitoring with automated analytics, error detection, and Azure storage.

- Enhanced debugging and system stability with detailed insights.
- Reduced testing time by 70% through streamlined issue identification.
- Eliminated manual monitoring with automated reporting.

##### 2.Automation Framework (Python + Pytest)

Built a Pytest-based automation framework for efficient testing.

- Accelerated software delivery by automating tests.
- Improved test coverage and validation.
- Minimized human error, speeding up feedback loops.