

Makilesh M

Bangalore • makilesh24225@gmail.com • +919894419452 • [linkedin.com/in/makilesh/](https://www.linkedin.com/in/makilesh/) • github.com/Makilesh/

PROFESSIONAL SUMMARY

AI/ML Engineer | Data Scientist | LLM Specialist

Oracle-certified AI professional with production expertise in building and deploying LLM, RAG, multi-agent frameworks and real-time voice AI with proven ability to lead technical teams and deliver AI solutions with measurable business impact.

SKILLS

- **Programming:** Python, C++, C, Java
- **AI & Machine Learning:** Deep Learning, Neural Networks, NLP, Transformers, Generative AI, Large Language Models (LLMs), MLOps, Model Fine-tuning, Multi-Agent Systems, RAG, Prompt Engineering, Vector Embeddings
- **Frameworks & Libraries:** PyTorch, TensorFlow, Scikit-learn, OpenCV, LangChain, LangGraph, CrewAI, AutoGen, Hugging Face, spaCy, n8n, Pandas, NumPy, Matplotlib, Seaborn, NLTK, Whisper
- **Cloud & DevOps:** AWS (EC2, S3, Lambda), Azure OpenAI, Docker, FastAPI, Flask
- **Databases & Tools:** SQL, MySQL, Vector Databases (Pinecone, FAISS, ChromaDB, Weaviate), Git, GitHub
- **Web Development:** HTML, CSS, JavaScript, React.js, Spring Boot, RESTful APIs

PROJECTS

AI-Powered Legal Engine (June 2025) - github.com/Makilesh/AI_Legal_Engine

- Built AI legal assistant for Indian criminal law using RAG architecture with ChromaDB and Gemini (OpenRouter/Ollama fallback), enabling semantic search across 673 legal documents via Sentence Transformers embeddings
- Implemented multi-agent LangChain system with JWT authentication, Redis semantic caching (92% similarity), WebSocket voice I/O, and automated document generation (FIR, bail, affidavit) deployed via FastAPI backend
- **Technologies Used:** LangChain, ChromaDB, Sentence Transformers, FastAPI, Redis, WebSocket, SQL

AI-Powered Voice Agent (Sep 2025 - Oct 2025) - github.com/Makilesh/voice_engine_MVP

- Designed a production-grade real-time bidirectional voice agent from scratch achieving <200ms latency with sentiment-aware conversation, multi-threaded barge-in detection (50ms polling), and adaptive tone modulation, using open-source tools (Whisper, Llama 3) with producer-consumer threading architecture
- Deployed as multi-purpose **SME** (Subject Matter Expert) assistant across **meetings, voice calls, and AI consultation** with wake-word activation, providing on-demand expert responses with context retention and seamless interruption handling for customer support automation
- **Technologies Used:** Python, Llama 3, RealtimeSTT(Whisper), RealtimeTTS (Orpheus), Asyncio, Threading

EXPERIENCE

AI-Developer Intern | Shamla Tech Solutions | Coimbatore, India (Sep 2025 - Present)

- Designed and deployed an **AI Voice Calling Agent** with Twilio integration storing conversation history in Pinecone vector database, enabling context-aware automation of 100+ weekly outbound calls with dynamic response generation based on previous interactions and reducing manual call handling time by 80%
- Built a **Cross-platform Meeting Assistant** with real-time speaker diarization processing 50+ hours of meeting transcriptions weekly across Zoom, Google Meet, and in-person sessions, featuring wake-word activation for on-demand AI expert consultation
- Engineered end-to-end **Multi-source Lead Generation Pipeline** with 3-stage AI qualification (pre-filter + dual LLM: GPT-4 → Gemini fallback), achieving 94.6% cost savings via batch processing (20 leads/API call) and concurrent async scraping across Reddit, Discord, Slack, LinkedIn with competitor detection
- **Technologies:** Python, Twilio API, RealtimeSTT/TTS, Whisper, Pinecone, LangChain, pyannote.audio, OpenAI/Gemini APIs, Selenium, Playwright, PRAW, Discord.py, Slack SDK, AsyncIO, Flask

CERTIFICATIONS

- ORACLE Cloud Infrastructure 2025 Certified Generative AI Professional - 2025
- ORACLE Cloud Infrastructure 2025 Certified Data Science Professional - 2025
- UDEMY Complete Machine Learning, NLP Bootcamp MLOPS & Deployment - 2024
- NPTEL Python for Data Science – 2024

HACKATHONS

IMPACT X - Finalist (2024) | Team Lead, Full-Stack Developer, AI Engineer

Led a team of 4 to develop an AI-Powered Crime Detection System using YOLOv8 and CNNs.

THIRAN - Finalist (2025) | Data Scientist, Data Engineer

Developed AI Legal Engine with RAG architecture achieving semantic search across 673+ legal documents

EDUCATIONAL BACKGROUND

Sri Krishna College of Engineering and Technology

B.E. in Computer Science Engineering | CGPA: 8.0 | Expected: June 2026 | Coimbatore, India