

YASHAS VADDI

(LLMs | RAG | AI Agents | GCP | Docker | FastAPI | MLOps)

Contact

 yashas.vaddi12@gmail.com

 +91 9702120202

 <https://www.linkedin.com/in/yashasvaddi/>

 <https://github.com/Yashasvaddi>

Academic Details

B.E. Computer Engineering - Expected May 2027

Thadomal Shahani Engineering College, Bandra

AGG. CGPA: **9.01** (Top 5% of class) (avg. till 5th Semester)

Profile Summary

- Strong CS fundamentals in AI platform engineering, backend systems, REST APIs, and cloud-native pipelines.
- Interning at Eka.Care (Agentic AI); previously at TRREV Technology, shipped production ML services with 30–50% precision improvement.
- Hands-on with RAG, AI Agents, LLMs, FastAPI, Docker, GCP (production) & AWS (EC2, S3, Lambda) with end-to-end ML deployment experience.
- 7x Hackathon Winner · 2x Published Researcher (JETIR, IJTE) · CGPA 9.01 · TSEC Mumbai

Research Publications:

- Published a peer-reviewed AI/ML paper in JETIR (2024) on “Developing smart education environments”. [\(Certificate\)](#) [\(Link to Paper\)](#)
- Published IJTE research on a low-cost multimodal AI system for touch-free interaction using gestures and AI voice/chat assistance. [\(Link to Paper\)](#)

Talks & Speaking Engagements

Invited AI/ML Speaker — SIES Graduate School of Technology (IETE Students’ Forum)

- Delivered a technical seminar on deploying applications from local development to production, covering server setup, CI/CD pipelines, hosting, configuration, and monitoring. [\(Event Flyer\)](#)

Leadership & Community Engagement

Google Developers Group (GDG), Thadomal Shahani Engineering College

Joint Tech Head - AI/ML innovations & Development

- Led AI/ML workshops and technical initiatives within the GDG community.
- Coordinated hands-on development sessions for student developers.

Technical Skills

Programming: Python, C++, JavaScript.

Backend & Systems: FastAPI, REST APIs, Asynchronous Services, TCP/IP Fundamentals, Linux/OS Basics.

ML & Deep Learning: PyTorch, TensorFlow, Scikit-learn

NLP: RAG, Embeddings, Keyword Extraction, Sentiment Analysis.

MLOps & Cloud: Docker, GCP (Compute Engine, Cloud Build, Monitoring), AWS (EC2, S3, Lambda, IAM), CI/CD Pipelines, Model & API Deployment.

Data & Analytics: Pandas, NumPy, Feature Engineering

Developer Tools & Practices: Git, Unit Testing, Agile/SDLC,

Internship Experience

Agentic AI Intern

Eka.Care — Bangalore, June’ 2026 – till date

- Architecting production-grade **multi-agent AI** pipelines for clinical workflows, defining **agent roles**, tool-calling interfaces, and **orchestration** logic using LLM-based frameworks.
- Designing **inter-agent** communication protocols and memory management for reliable, **stateful agent** execution across multi-step medical tasks; targeting **GCP/AWS** deployment with fault tolerance and compliance.

AI/ML Engineering Intern

TRREV TECHNOLOGY - Mumbai, Aug’ 2025 – June’2026

- Deployed a production **FastAPI** keyword extraction service on **GCP** with **Docker**; improved precision by **30–50%** via noise filtering, text normalization, and **prompt tuning**.
- Built a real-time medical transcription pipeline (**VAD, ASR, medical NER, text-stitching**); achieved production-grade accuracy on a lightweight server.

Awards

- **1st Place** - Need for Code 4.0 (2025) · DOC. [\(Certificate\)](#) [\(Demo Video\)](#)
- **1st Place** - Codessiance (2025) · SerenMind. [\(Certificate\)](#) [\(Demo Video\)](#)
- **1st Place** - Techathon (2025) · AI Council. [\(Certificate\)](#)
- **1st Runner Up** - TSEC Hacks (2026) · ScriptChat. [\(Certificate\)](#) [\(Demo Video\)](#)
- **1st Runner Up** - HackVision (2026) · Agentic LLM Council. [\(Certificate\)](#)
- **2nd Runner Up** - M-INDICATOR National AI Hackathon - [\(Certificate\)](#)
- **Best AIML Project** - Code Craft 2. [\(Certificate\)](#)

Academic Projects

1. DOC.AI – AI-Powered Document Assistant | Python, LangChain, OpenAI, FastAPI, AssemblyAI, Vector Database, TTS

- Built production RAG system handling PDF/DOCX with SQL-backed retrieval supports multi-document comparison, source-referenced Q&A, and voice interface via AssemblyAI + TTS
- Includes voice-based chatbot, audio summaries, and ethical AI filters. [\(Project Link\)](#) [\(Demo Video\)](#)

2. MindVoice | Python, Librosa, BiLSTM, DistilRoBERTa, GCP

- Built multimodal depression detection system fusing acoustic features (MFCCs, pitch, energy via Librosa) with linguistic NLP (DistilRoBERTa) — produces explainable risk score with 3-tier severity classification
- End-to-end pipeline: voice input → prosody extraction → BiLSTM + transformer fusion → GCP deployment with ethical consent safeguards. [\(Project Link\)](#) [\(Demo Video\)](#)

3. ScriptChat | Local LLMs, Multi-Agent AI, Knowledge Graphs

- AI-powered writing IDE with context-aware assistance for story development and continuity.
- Multi-agent AI feedback system (Director, Editor, Audience) with structured evaluation.
- Local model execution with knowledge graph integration for privacy and workflow support. [\(Project Link\)](#) [\(Demo Video\)](#)