

# Rounak Show

📍 Howrah, West Bengal, India    ✉ rounak.1210@gmail.com    📞 +91 98755 61533    🌐 rounak.wiki  
in rounak-show    🔄 ShowRounak

## Profile

---

AI/ML Engineer specializing in real-time conversational AI, including voice and video agents, LLM fine-tuning, and multi-agent systems. At TruGen AI, fine-tuned and deployed custom classification models into production real-time voice pipelines, and built meeting-bot integrations enabling AI agents to join external multi-participant meetings. Background also spans Generative AI application development, retrieval-augmented generation, and scalable cloud deployment. Recognized multiple times as Employee of the Month for consistently delivering high-quality, fast results. Passionate about applied AI, real-time systems, and innovative problem-solving to drive business impact.

## Skills

---

### Programming

**Frameworks:** Python, FastAPI, LangChain, LangGraph, LiveKit Agents SDK

### Generative AI

**NLP:** LLMs, Fine-Tuning (LoRA/PEFT), Prompt Engineering, Chatbots, Retrieval-Augmented Generation (RAG), Knowledge Graphs, Multi-Agent Systems

### Real-Time

**Voice AI:** LiveKit, WebSocket Streaming, STT/TTS Orchestration, Meeting-Bot Integrations

### System Design

**MLOps:** Scalable Pipelines, Cloud Deployment (AWS, GCP), Docker, GitHub Actions, GPU Cloud Deployment (RunPod)

### Databases

**Storage:** SQL, NoSQL, Vector Databases (Embeddings)

## Professional Experience

---

### AI/ML Engineer

*Feb 2026 – Present*

*TruGen AI*

- Fine-tuned a hybrid GatedDeltaNet/full-attention transformer (Qwen3.5-0.8B) as a binary classifier for real-time AI agents, achieving 97.9% accuracy and 0.975 Macro F1 on held-out test data.
- Built a meeting-bot pipeline enabling AI agents to join external multi-participant meetings (Zoom, Google Meet, Microsoft Teams) in real time, including WebSocket-based transcript and event streaming.
- Shipped a vision-capable screen-share agent enabling AI agents to interpret shared screen content live during meetings.
- Integrated MCP-based tool calling into production AI agents, improving reliability of session handling and tool routing.

### AI Engineer

*March 2024 – Jan 2026*

*FutureSmart AI*

- Led end-to-end AI projects for global clients, delivering Generative AI solutions, custom agents, and NLP-driven applications.
- Rapidly prototyped and deployed AI MVPs tailored to client needs, balancing speed with production readiness.
- Designed scalable AI systems including multi-agent workflows, modular pipelines, and third-party integrations (e.g., Odoo, Databases).
- Implemented retrieval systems using SQL + VectorDBs for real-time context-aware responses.
- Deployed and scaled applications on AWS and GCP with containerization (Docker) for reproducibility.
- Mentored interns and junior engineers on best practices, system design, and prompt engineering.
- Recognized multiple times as Employee of the Month for excellence in fast, high-quality delivery.

## Education

---

**Bachelor of Science in Anthropology**  
*Narasingha Dutta College, Howrah*  
CGPA: 8.3

*Sept 2019 – Sept 2022*

## Certificates

---

AWS Certified Cloud Practitioner

## Courses

---

Full Stack Data Science Bootcamp 2.0 by iNeuron  
Data Science Mentorship Program by CampusX  
Generative AI with Large Language Models