

Job Title: Founding AI Engineer

Company: Cosmium AI (backed by IIT Delhi alumni)

Location: Jaipur, Rajasthan (In-office/Remote)

Type: Full-time

The Mission

Cosmium AI is building the next generation of conversational Speech AI. We are solving the "latency problem" in voice bots—making AI sound human, instant, and intelligent.

We are looking for a hacker-engineer who loves reading papers on Monday and shipping them to production by Friday.

What You'll Build

- **The Brain:** Fine-tune and optimize LLMs for specific conversational tasks using instruction tuning, RAG and prompt engineering.
- **The Voice:** Optimize STT and TTS models to reduce latency to <500ms.
- **The Pipeline:** Build production-grade inference pipelines. You won't just train models; you will dockerize them, quantize them, and deploy them.

Must-Haves

- **Python Native:** Expert in Python and libraries like PyTorch, NumPy, and Pandas.
- **Audio/NLP Experience:** You have worked with HuggingFace Transformers, Librosa etc
- **Engineering First:** You know how to wrap a model in FastAPI/Flask and deploy it.

Bonus Points

- Experience with **WebRTC** or streaming audio data.
- Experience building with LangChain or LlamaIndex,
- Published papers or active Kaggle/HuggingFace profile.

Why Join?

- **No "Toy" Projects:** You will be building the core IP of the company.
- **Founding Team Status:** High equity, high autonomy.
- **Hard Tech:** We aren't just wrapping an API; we are optimizing the entire voice stack.
- **Mentorship:** Work directly with technical founders (IIT Delhi alumni).

How to Apply

Send your resume and **GitHub/Live Project links** to careers@cosmium.ai with the subject: *"Founding AI Engineer Application - [Your Name]"*