Rohan Saxena

rohaan.saxena14@gmail.com | (+91) 9760496704 | LinkedIn: rohan-saxena-2846a5163 | Github: RohanSaxena14

Career Synopsis

Philomath and **AI specialist** with over **three years of experience** in developing production-grade AI systems, including Large Language Models **(LLMs)**, **intelligent agents**, Text-to-speech **(TTS)**, and Speech-to-Text **(ASR)** modules.

Technical Skills

- **Certifications** : *Deep Learning Specialization*, *Machine Learning*, *Data Structures and Algorithms*, *Python*, *Natural Language Processing*
- **Skills**: Python, C, SQL, PyTorch, Tensorflow, Next Gen Kaldi, Coqui-TTS, SpeechBrain, NVIDIA NeMo, DeepSpeech, Transformers, Crew-AI, Langchain, torchserve, Flask, Django

Work Experience

Navana Tech

AI Scientist *(July 2022 to March 2025) & (May 2025 to Present)*

- Developed a **Context Biasing LM algorithm** for ASR, enabling faster text-based improvement cycles with negligible added latency.
- Built a real-time **TTS system** with TorchServe, optimized for GPU concurrency and delivering latency under **200 ms**.
- Led migration to production-grade **ASR systems** for **10 Indian languages**, achieving streaming inference latency of **200 ms on CPUs**.
- Developed a production-grade **Speaker Diarization Pipeline** with Nemo and Next-Gen Kaldi, leveraging experience from the **BMGF project**.
- Contributed to the prestigious **BMGF project**, designing a voice parameter system to detect fraudulent speaker metadata with over **95% accuracy** and creating high-quality datasets covering **38 dialects**.
- Spearheaded development of **LLM-powered RAG-based WhatsApp chat agents** for top clients (e.g., **Asian Paints, Boat, Fireboltt**), significantly enhancing customer engagement and support.
- Assisted in building **LLM-based product search algorithms** using Contrastive Learning and fine-tuning.

BharatGen at TIH, IIT Bombay

Applied-MLE *March 2025 to April 2025*

- Contributed to **core LLM development** as part of the **Supervised Fine-Tuning team**, focusing on improving model performance through systematic experimentation.
- Designed and executed **scalable hyperparameter search workflows** using small models, enabling efficient tuning strategies transferable to larger LLMs.
- Developed in-house **evaluation protocols tailored** for the Indian context to identify and analyze failure modes in language models.

Publications & Achievements

RESPIN-S1.0 Corpus: A read speech corpus of 10000+ hours in dialects of nine Indian Languages *June 2025*
*Research accepted in **NeurIPS 2025 Datasets and Benchmarks Track**.* *IISc, Bangalore.*

Model Adaptation for ASR in low-resource Indian Languages *July 2023*
*Research accepted in **ASRU** arxiv.* *IISC, Bangalore*

CQFaRAD: Collaborative Query-Answering Framework for Research Article Dataspace *September 2023*
*Research accepted in **International Journal of Information Technology** arxiv.* *BIET, Jhansi*

An Intelligent Recommendation-cum-Reminder System *August 2021*
*Research accepted in **Cods-COMAD: ACM, 2021** arXiv.* *IIT, Indore*

ASEAN-India Hackathon

- Grand Finalist of **ASEAN-India Hackathon**, held between **11 countries**. Developed a **image segmentation** model to identify ships from satellite images to solve **Marine Traffic Management** under the Blue Economy.

Winner of Smart India Hackathon

- Developed a **Face Identification model** to spot criminals from low quality CCTV footage, **old images, Sketches**.

Education

Bundelkhand Institute of Engineering and Technology, Jhansi, India

ranked under 2% in AKTU *August 2018 - July 2022*