Pranav Sharma

pranavsharma.org | +91 9354166121 | reach.sharmapranav@gmail.com | LinkedIn/Pranav | GitHub/Pranav

Speech AI Research Engineer working on full-duplex speech-to-speech models, multilingual ASR, and the evaluation of conversational speech systems. Co-author of Human-1, with depth in benchmark and annotation design, failure-mode analysis, large-scale data curation, and distributed training.

RESEARCH

Human-1: A Full-Duplex Conversational Modeling Framework in Hindi using Real-World Conversations *arXiv*

- **Co-developed an open-source full-duplex speech-to-speech dialogue system** for Hindi, adapting the **Moshi/Mimi** architecture for real-time turn-taking, overlaps, and backchannels.
- **Designed a 32k-token Hindi SentencePiece tokenizer** and built a distributed multi-GPU training pipeline on a **26,000 hour stereo corpus** across an **8x NVIDIA H100** cluster.
- **Achieved a 4.10/5.00 human naturalness score**, with **66.9%** of generations tying directly with genuine human speech in perceptual quality and turn-taking baseline tests.

Hear to Correct: Audio-Informed Multilingual Disfluency Correction via Cross-Modal Knowledge Distillation *In revision*

- **Architected a 3-stage cross-modal knowledge distillation framework** for multilingual speech disfluency correction across Hindi, Bengali, Marathi, and English.
- **Applied parameter-efficient fine-tuning (PEFT)** using LoRA adapters on Qwen2.5-3B and Llama-3.2-3B LLMs, improving BLEU and chrF2 over zero-shot baselines.
- **Constructed a multilingual Indian speech disfluency dataset** consisting of 2,166 spontaneous recordings.

EXPERIENCE

AI Researcher - Josh Talks Dec 2025 - Present

- **Fine-tuned an English full-duplex speech model and benchmarked it using Full-Duplex-Bench**, evaluating overlap handling and conversational turn-taking against the base Moshi model.
- **Built checkpoint-level evaluation pipelines** surfacing degradation patterns - volume fluctuation, noise buildup, hallucination loops, unstable backchanneling - and traced each failure mode back to training data and fine-tuning strategy, turning eval signal into targeted fixes.
- **Built a speaker diarization evaluation dataset end to end** for in-person and online multi-party calls with **up to 10 speakers**, owning the pipeline from boundary generation to human annotation protocol and tooling to **model evaluation**.
- **Designed a human evaluation framework benchmarking 15+ commercial and open TTS systems** on Indic and Indian-accented English, covering protocol design, rater workflow, and reporting.
- **Engineered targeted speech model optimizations** through rigorous error analysis, strategic data curation, and multi-stage fine-tuning, achieving a **25-50% relative WER reduction** across multilingual holdout evaluation sets.
- **Fine-tuned and evaluated Whisper, Wav2Vec2 and other SOTA ASR models** on large multilingual datasets, optimizing distributed PyTorch multi-GPU training and establishing competitive baselines across **low-resource Indic languages**.
- **Curated multilingual speech datasets at scale (100k+ hours)**, covering cleaning, quality filtering, metadata generation, and assembly of task-specific training sets; built high-throughput **vLLM** inference pipelines on distributed multi-GPU clusters.

AI/ML Engineer - Intern at Rely Technologies Apr 2025 - Jun 2025

- Built a **Redis-backed RAG chatbot** (multi-turn context, hybrid LLM routing with deterministic guardrails) and a Django Celery hospital CRM for backend workflow automation.

AI Backend Engineer - Intern at Persist Ventures Nov 2024 - Feb 2025

- Cut transcription **latency 70%** (**Deepgram Nova + OpenAI Whisper**) and **50% fewer wake-word false detections** in a conversational-AI voice pipeline.

EDUCATION

Amity University Jan 2025 - Dec 2026
Master of Computer Applications (AI/ML Specialization)

Delhi Skill & Entrepreneurship University Oct 2021 - Jun 2024
Bachelor of Computer Applications **CGPA: 8.93 / 10**

TECHNICAL SKILLS

Programming Languages: Python, JavaScript
Core Domains & Research: Full-Duplex Speech Models, ASR, Conversational AI, Multilingual Speech Processing
Deep Learning & Speech: PyTorch, Transformers, Torchaudio, DDP, FSDP
Data & Datasets: Hugging Face Datasets, Pandas, Parquet, Mozilla Common Voice, OpenSLR, IndicVoices, FLEURS, LibriSpeech
GenAI & LLM Stack: Hugging Face Transformers, vLLM, PEFT (LoRA), RAG, LangChain, LangGraph
Speech Evaluation: Full-Duplex-Bench, jiwer, WER/CER, BLEU, chrF2
MLOps & Infrastructure: GCP, AWS, Runpod, Docker, MLflow, Git, GitHub
Backend: FastAPI, Django, PostgreSQL, Redis