Souvik Ghosh

AI Researcher — Data Scientist

📱 +91 8697282257
✉ souvik.g@research.iiit.ac.in
in /souvikg544
⬡ /souvikg544
ⓘ /souvik.g

Education

2024–2026 **MS by Research, CSE**, *IIIT Hyderabad*, Grade: 8.0/10
Advisor: Prof. C. V. Jawahar, Vinay Namboodiri
Research Areas: Lip-to-Speech, Controllable Multimodal Generation.

2019–2023 **BTech, Applied Electronics and Instrumentation**, *HITK, Kolkata*, Grade: 8.12/10

Experience

Dec 2023–Aug 2024 **Research Engineer - Internship**, *Sync Labs (YC W24)*
- Developed GAN-based models for temporal consistency in face enhancement, achieving substantial reductions in background flicker across video sequences.
- Pioneered diffusion-driven lip-sync models, to surpass SOTA baselines in both perceptual face quality and sync accuracy.
- Spearheaded research on novel methodologies to enhance lip-sync technology for the multimedia and film industry.

Aug 2023–Jan 2024 **Research Fellow**, *CVIT, IIIT Hyderabad*
- Researched Hypernetworks for personalized lipreading models.
- Conducted research on lipreading and lip-to-speech technologies, developing on-device models with 40% parameter reduction.
- Implemented and analyzed state-of-the-art research in facial dynamics.

May 2023–Dec 2023 **Data Scientist - Freelance**, *Brainbox Innovations*
- Directed the transition to AI-first systems, increasing adoption of LLM-powered solutions by over 3× within 6 months.
- Specialized in applied LLMs, building custom wrappers and integrating with large-scale vector databases for retrieval-augmented generation.
- Led end-to-end development of intelligent search, conversational agents, document retrieval, and AI-driven database query systems.

Jul 2022–May 2023 **Data Scientist - Internship**, *Whilter AI*
- Pioneered initiatives in advertising and media through deepfakes, Text to Speech for Indian languages, voice cloning, and lip sync.
- Led a large-scale Indian wedding video analysis project, developing an automated framework for unsupervised video clustering and personalized trailer generation.
- Researched GAN-based approaches for generating target faces with specific makeup styles.

Publications
---

1. **Souvik Ghosh**, C. V. Jawahar, Vinay Namboodiri. *LipAdapter: Text-to-Video Alignment is All You Need for Lip-to-Speech.* **[Under Review]**

2. **Souvik Ghosh**, C. V. Jawahar, Vinay Namboodiri. *Aavaz: An Early Attempt to Give Voice to the Voiceless.* **[Under Review]**

3. Shaon Bhattacharya[†], **Souvik Ghosh**[†], Prantik Deb[†], Ajoy Mondal, C. V. Jawahar. *Adapting Vision-Language Models for Hindi OCR.* **ICDAR**, 2025.

4. Shankhanil Ghosh, Chanda Saha, Nagamani Molakathala, **Souvik Ghosh**, Dhananjay Singh. *reSenseNet:*

Ensemble Early Fusion Deep Learning Architecture for Multimodal Sentiment Analysis. **IHCI**, 2021.

5. Shankhanil Ghosh, Chanda Saha, Nagamani Molakathala, **Souvik Ghosh**, Dhananjay Singh. *Speech@SCIS: Annotated Indian Video Dataset for Speech-Face Cross Modal Research*. **SCI**, 2021.

Skills

AI & ML	Multimodal Generative AI, Controllable Editing, Lip-sync, Lip-to-Speech, Lipreading, Face Enhancement, Text to Speech (TTS), Large Language Models (LLM), Representation Learning, Computer Vision, Machine Learning.
MLOps	AWS, Azure, Google Cloud, FastAPI, Flask, Docker.
Frameworks	PyTorch, TensorFlow

Achievements

2023	Recognized as a **Top Rated Freelancer** on Upwork (Top 10% globally), maintaining a 4.5+/5 rating and 100% job success rate across 50+ international clients.
2023	Awarded a **$3000 Global Youth Mobilization Grant** to develop women safety solutions. Successfully launched *HelpBy*—an app enabling immediate nearby assistance.