Sreyan Ghosh

+1 240 899 7685 | sreyang@umd.edu | LinkedIn | GitHub | Website | Google Scholar | College Park, USA

Education

University of Maryland College Park — College Park, USA
Advised by Dr. Dinesh Manocha and Dr. Ramani Duraiswami
Ph.D. and M.S. in Computer Science; ***GPA: 3.9/4.0*** — *Aug 2022 – May 2027*
Ph.D. supported by the 2025 NVIDIA Graduate Fellowship

Christ University — Bangalore, India
B.Tech in Computer Science and Engineering; ***GPA: 8.7/10.0*** — *June 2016 – Jun 2020*

Publications (Speech & Audio Processing)

1. Audio Flamingo 2: An Audio-Language Model with Long-Audio Understanding and Expert Reasoning Abilities
   Sreyan Ghosh, Zhifeng Kong, Sonal Kumar, S Sakshi, Jaehyeon Kim, Wei Ping, Rafael Valle, Dinesh Manocha, Bryan Catanzaro
   arXiv 2025

2. GAMA: A Large Audio-Language Model with Advanced Audio Understanding and Complex Reasoning Abilities
   *Sreyan Ghosh**, Sonal Kumar*, Ashish Seth, Chandra Kiran Reddy Evuru, Utkarsh Tyagi, S Sakshi, Oriol Nieto, Ramani Duraiswami, Dinesh Manocha
   EMNLP 2024 (Oral)

3. Synthio: Augmenting Small-Scale Audio Classification Datasets with Synthetic Data
   Sreyan Ghosh, Sonal Kumar, Zhifeng Kong, Rafael Valle, Bryan Catanzaro, Dinesh Manocha
   ICLR 2025

4. MMAU: A Massive Multi-Task Audio Understanding and Reasoning Benchmark
   S Sakshi*, Utkarsh Tyagi*, Sonal Kumar*, Ashish Seth*, Ramaneswaran Selvakumar*, Oriol Nieto*, Ramani Duraiswami, *Sreyan Ghosh**, Dinesh Manocha
   ICLR 2025 (Spotlight)

5. ReCLAP: Improving Zero Shot Audio Classification by Describing Sounds
   Sreyan Ghosh, Sonal Kumar, Chandra Kiran Reddy Evuru, Oriol Nieto, Ramani Duraiswami, Dinesh Manocha
   ICASSP 2025 (Oral)

6. LipGER: Visually-Conditioned Generative Error Correction for Robust Automatic Speech Recognition
   Sreyan Ghosh, Sonal Kumar, Ashish Seth, Purva Chiniya, Utkarsh Tyagi, Ramani Duraiswami, Dinesh Manocha
   InterSpeech 2024 (Oral)

7. AV-RIR: Audio-Visual Room Impulse Response Estimation
   Anton Ratnarajah, *Sreyan Ghosh*, Sonal Kumar, Purva Chiniya, Dinesh Manocha
   CVPR 2024

8. CompA: Addressing the Gap in Compositional Reasoning in Audio-Language Models
   *Sreyan Ghosh**, Ashish Seth*, Sonal Kumar*, Utkarsh Tyagi, C. K. Evuru, Oriol Nieto, Dinesh Manocha
   ICLR 2024

9. RECAP: Retrieval-Augmented Audio Captioning
   *Sreyan Ghosh**, Sonal Kumar, Chandra Kiran Reddy Evuru, Ramani Duraiswami, Dinesh Manocha
   ICASSP 2024 (Oral)

10. EH-MAM: Easy-to-Hard Masked Acoustic Modeling for Self-Supervised Speech Representation Learning
    Ashish Seth*, Ramaneswaran S, S Sakshi, Sonal Kumar, *Sreyan Ghosh**, Dinesh Manocha
    EMNLP 2024 (Oral)

11. Stable Distillation: Regularizing Continued Pre-training for Low-Resource Automatic Speech Recognition
    Ashish Seth*, *Sreyan Ghosh**, S. Umesh, Dinesh Manocha
    ICASSP 2024

12. FusDom: Combining In-Domain and Out-of-Domain Knowledge for Continuous Self-Supervised Learning
Ashish Seth*, *Sreyan Ghosh**, S. Umesh, Dinesh Manocha
ICASSP 2024

13. AdVerb: Visually Guided Audio Dereverberation
Sanjoy Chowdhury*, *Sreyan Ghosh**, Subhrajyoti Dasgupta, Anton Ratnarajah, Utkarsh Tyagi, Dinesh Manocha
ICCV 2023

14. MMER: Multimodal Multi-task Learning for Speech Emotion Recognition
Sreyan Ghosh, Utkarsh Tyagi, S Ramaneswaran, Harshvardhan Srivastava, Dinesh Manocha
InterSpeech 2023 (Oral)

15. data2vec-aqc: Search for the right Teaching Assistant in the Teacher-Student training setup
Lodagala V S V Durga Prasad*, *Sreyan Ghosh**, S. Umesh
IEEE ICASSP 2023

16. MAST: Multiscale Audio Spectrogram Transformers
*Sreyan Ghosh**, Ashish Seth*, S. Umesh, Dinesh Manocha
IEEE ICASSP 2023

17. SLICER: Learning universal audio representations using low-resource self-supervised pre-training
Ashish Seth*, *Sreyan Ghosh**, S. Umesh, Dinesh Manocha
IEEE ICASSP 2023

18. Decorrelating Feature Spaces for Learning General Purpose Audio Representations
*Sreyan Ghosh**, Ashish Seth*, S. Umesh
IEEE JSTSP Special Issue on Self-Supervised Learning for Speech and Audio Processing

19. PADA: Pruning Assisted Domain Adaptation for Self-Supervised Speech Representations
Lodagala V S V Durga Prasad, Sreyan Ghosh, S. Umesh
IEEE SLT 2022

20. CCC-wav2vec 2.0: Clustering aided Cross Contrastive Self-supervised learning of speech representations
Lodagala V S V Durga Prasad, Sreyan Ghosh, S. Umesh
IEEE SLT 2022

21. Span Classification with Structured Information for Disfluency Detection in Spoken Utterances
Sreyan Ghosh, Sonal Kumar, Yaman Kumar Singla, Rajiv Ratn Shah, S. Umesh
Interspeech 2022

22. DeToxy: A Large-Scale Multimodal Dataset for Toxicity Classification in Spoken Utterances
Sreyan Ghosh, Samden Lepcha, Sakshi, Rajiv Ratn Shah, S. Umesh
Interspeech 2022

23. End-to-end Named Entity Recognition from English Speech
Hemant Yadav, Sreyan Ghosh, Yi Yu, Rajiv Ratn Shah
Interspeech 2020

24. ProSE: Diffusion Priors for Speech Enhancement
*Sreyan Ghosh**, Sonal Kumar*, Utkarsh Tyagi, Purva Chiniya, Anton Jeran Ratnarajah, Chandra Kiran Reddy Evuru, Ramani Duraiswami, Dinesh Manocha
NAACL 2025 (Oral)

Publications (NLP & Vision)

1. Visual Description Grounding Reduces Hallucinations and Boosts Reasoning in LVLMs
*Sreyan Ghosh**, C. K. Evuru*, Sonal Kumar*, Utkarsh Tyagi, O. Nieto, Z. Jin, Dinesh Manocha
ICLR 2025

2. A Closer Look at the Limitations of Instruction Tuning
   *Sreyan Ghosh**, C. K. Evuru*, Sonal Kumar*, Ramaneswaran S, D. Aneja, Z. Jin, R. Duraiswami, Dinesh Manocha
   ICML 2024
   Media Coverage 1

3. ABEX: Data Augmentation for Low-Resource NLU via Expanding Abstract Descriptions
   *Sreyan Ghosh**, Utkarsh Tyagi*, Sonal Kumar, Chandra Kiran Reddy Evuru, Ramaneswaran S, S Sakshi, Dinesh Manocha
   ACL 2024

4. ASPIRE: Language-Guided Augmentation for Robust Image Classification
   *Sreyan Ghosh**, C. K. Evuru*, Sonal Kumar, S. Sakshi, Utkarsh Tyagi, Dinesh Manocha
   ACL 2024 (Findings)

5. DALE: Generative Data Augmentation for Legal NLP
   *Sreyan Ghosh**, C. K. Evuru*, Sonal Kumar, S. Sakshi, Utkarsh Tyagi, Dinesh Manocha
   EMNLP 2023

6. CoSyn: Detecting Implicit Hate Speech in Online Conversations Using a Context Synergized Hyperbolic Network
   Sreyan Ghosh, Manan Suri, Purva Chiniya, Utkarsh Tyagi, Sonal Kumar, Dinesh Manocha
   EMNLP 2023

7. ACLM: A Selective-Denoising based Generative Data Augmentation Approach for Low-Resource Complex NER
   *Sreyan Ghosh**, Utkarsh Tyagi*, Manan Suri, Sonal Kumar, S Ramaneswaran, Dinesh Manocha
   ACL 2023

8. Do Vision-Language Models Understand Compound Nouns?
   Sonal Kumar*, *Sreyan Ghosh**, S Sakshi, Utkarsh Tyagi, Dinesh Manocha
   NAACL 2024

9. CoDa: Constrained Generation based Data Augmentation for Low-Resource NLP
   Chandra Kiran Reddy Evuru*, *Sreyan Ghosh**, Sonal Kumar, Ramaneswaran S, Utkarsh Tyagi, Dinesh Manocha
   NAACL 2024 (Findings)

10. BioAug: Conditional Generation based Data Augmentation for Low-Resource Biomedical NER
    *Sreyan Ghosh**, Utkarsh Tyagi*, Sonal Kumar*, Dinesh Manocha
    SIGIR 2023

Internships

NVIDIA — Santa Clara, CA, USA
Research Scientist Intern — *August 2024 – Present*

- Working as a research scientist intern at the Audio Understanding and Generation team. Working on scaling audio generation and understanding with LLMs.
- Submitted a paper to ICLR 2025 on synthetic data for audio classification.

Microsoft — Redmond, WA, USA
Research Scientist Intern — *May 2024 – August 2024*

- Worked as a research scientist intern at the Speech and Audio team at Microsoft Research.
- Developed a synthetic data generation pipeline to train robust generative error correction models for Automatic Speech Recognition models. Will submit our findings to NAACL 2025.

Adobe — Seattle, WA, USA
Research Scientist Intern — *May 2023 – December 2023*

- Worked as a research scientist intern at the Video Understanding group.
- My primary project involved investigating and improving instruction tuning for Large Language Models. We published our findings at ICML 2024.
- Another side project involved evaluating and improving compositional reasoning in audio-language models. We published our work at ICLR 2024.

Google Summer of Code Remote

Open Source Developer *April 2022 – August 2022*

- Working on building deep learning based NLP (speech and text) notebooks using Tensorflow and Keras.
- Link to PRs and code contributed on personal website.

Cisco Systems Bangalore, India

Software Developer Intern *January 2020 – June 2020*

- Worked on a project, End-to-End Named Entity Recognition from English Speech, under the guidance of **Dr. Rajiv Ratn Shah** as part of my bachelor's thesis. Paper accepted at **Interspeech 2020**.

MIDAS Labs, IIIT-Delhi Delhi, India

Research Intern *January 2020 – June 2020*

- Worked on building a VOIP (Voice Over IP) Traffic Analyzer to detect anomalous SIP messages using machine learning.

Noodle.ai Bangalore, India

Data Science Intern *December 2019 – December 2019*

- Worked on multivariate time-series anomaly detection in high-frequency IoT sensor data obtained from steel manufacturing machines.

TEG Analytics Bangalore, India

Data Science Intern *April 2019 – May 2019*

- Worked under the healthcare intelligence division to provide insights from insurance plan enrollment data, for private insurance companies in the US.
- Used Machine Learning and Deep Learning techniques to predict plan enrollment for insurance companies.

Professional Work Experience

NVIDIA Bangalore, India

Research Scientist *April 2022 – August 2022*

- Worked as a senior solutions architect in the professional services team at NVIDIA. Responsible for delivering deep-learning-based NLP solutions to NVIDIA's premier customers around the globe.
- Contributed to AI R&D at NVIDIA. Published 2 papers at **IEEE SLT 2022**.

Cisco Systems Bangalore, India

Software Engineer II *Aug 2020 – March 2022*

- Worked as a senior software engineer in the automation and orchestration team under the Customer Experience BU. Built network assurance solutions for Cisco's telecom customers, leveraging state-of-the-art algorithms for anomaly detection at scale. Built a critical component in Cisco's first telemetry-based network assurance solution.
- Lead the development of an AI-based network security system for one of Cisco's telecom customers.
- Was part of the AI team that developed Cisco's first contact center solution, leveraging state-of-the-art NLP algorithms.
- Contributed to AI R&D at Cisco by representing Cisco at various conferences.

Awards & Achievements

- **Winner of NVIDIA Graduate Fellowship 2025 (10/600).**
- **Winner of Apple Graduate Fellowship 2025 (20/1000).**
- **Outstanding Graduate Assistant Award by UMD for the academic year 2023.**
- **Recognised by Cisco CX CTO** and higher management on multiple occasions for my research and innovation initiatives.
- Awarded the **Graham Bell Award** for being one of the most competitive undergraduates to have graduated in the year 2020.
- **Winner of Cisco Collab Hacks 2020.**
- **Winner of P&G Global Innovation Challenge 2020.**
- Appeared on the cover page of Analytics India Magazine twice for winning national level hackathons in 2020 (TEG Analytics and Uber Hackathon)
- **Winner of Hindustan Unilever BFS Technology Hackathon.**
- **Winner of various inter-college and intra-college hackathons sponsored by MNCs and the Government (Including a bronze medal at Kaggle).**

Selected Software Releases

GAMA GitHub | ⋆ 60

- GAMA is a Large Audio-Language Model (LALM) capable of responding to user queries about a user input audio. GAMA has been trained to complete foundational audio processing tasks like audio classification, captioning, etc., and can also respond accurately to complex, open-ended queries about audio with advanced reasoning. The repository contains all training and inference codes, including checkpoints for GAMA. GAMA has outperformed all prior LALMs on various open benchmarks.

MMER GitHub | ⋆ 60

- MMER is a state-of-the-art model for emotion recognition from spoken utterances. MMER is built on a novel cross-modal architecture and employs data augmentation together with combining 3 different types of contrastive losses.

LAPE GitHub | ⋆ 27

- LAPE is an easy-to-use toolkit for audio processing. In its initial release, LAPE supports Self-Supervised Learning (SSL)-based Upstream Pre-training and Downstream Fine-tuning. LAPE, originally introduced in this paper, integrates all our research on low-resource audio processing in one unified framework. We open-source LAPE to promote more research in this space.

ACLM GitHub | ⋆ 17

- ACLM is a synthetic data generation methodology for the task of complex named entity recognition. The repository contains all codes for reproducing ACLM.

CompA GitHub | ⋆ 10

- CompA proposes a novel benchmark for evaluating compositional reasoning in audio-language models. In addition, CompA also proposes novel training techniques for improving compositional reasoning in audio-language models. The repository contains the benchmarks, data, and all code to reproduce CompA.

Skills

Programming: Python, Java, C, MySQL, HTML, CSS

Frameworks: PyTorch, Keras, Tensorflow, Django, Flask, Spark

Tools: GIT, Android, Tableau, Power BI, AWS, GCP, Rest API, Docker, K8s

Concepts: Speech and Natural Language Processing, Software Development, Functional programming, Object-oriented programming, Machine Learning, Deep Learning, Image Processing, Cloud Computing

Community Service

Organized: IEEE ICASSP 2025 SALMA Workshop, DCASE 2025 Task 5, JSALT 2025 Topic on Advancing Audio-Language Models

Reviewer for: CVPR 2025, ICLR 2025, NeurIPS 2024, ECCV 2024, ACL 2024, NAACL 2024, InterSpeech 2024, ICASSP 2024, AAAI 2024, EMNLP 2023. ACL 2023, ICASSP 2023, InterSpeech 2023, AAAI 2023, ACL 2021

Team Lead and Co-founder for: Neuron, Christ University's first AI club focused on research, served as the first Vice President of the club.

Lecturer of: SLP at University of Buffalo, New York.