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 2026*
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*

Selected Publications on Audio (Speech, Sounds & Music) - All Publications

1. Audio Flamingo 3: Advancing Audio Intelligence with Fully Open Large Audio Language Models
   *Sreyan Ghosh**, Arushi Goel*, Jaehyeon Kim, Sonal Kumar, Zhifeng Kong, Sang-gil Lee, Chao-Han Huck Yang, Ramani Duraiswami, Dinesh Manocha, Rafael Valle, Bryan Catanzaro
   NeurIPS 2025 (Spotlight)

2. 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
   ICML 2025

3. Music Flamingo: Scaling Music Understanding in Audio Language Models
   Sreyan Ghosh, Arushi Goel, Lasha Koroshinadze, Sang-gil Lee, Zhifeng Kong, Joao Felipe Santos, Ramani Duraiswami, Dinesh Manocha, Wei Ping, Mohammad Shoeybi, Bryan Catanzaro
   arXiv 2025

4. 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)

5. MMAU-Pro: A Challenging and Comprehensive Benchmark for Holistic Evaluation of Audio General Intelligence
   Sona Kumar et al., Sreyan Ghosh, Ramani Duraiswami
   AAAI 2026

6. 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)

7. OmniVinci: Enhancing Architecture and Data for Omni-Modal Understanding LLM
   Hanrong Ye, ⋯, *Sreyan Ghosh*, et al.
   arXiv 2025

8. UALM: Unified Audio Language Model for Understanding, Generation and Reasoning
   Jinchuan Tian, Sang-gil Lee, Zhifeng Kong, *Sreyan Ghosh*, Arushi Goel, Chao-Han Huck Yang, Wenliang Dai, Zihan Liu, Hanrong Ye, Shinji Watanabe, Mohammad Shoeybi, Bryan Catanzaro, Rafael Valle, Wei Ping
   arXiv 2025

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

10. Failing Forward: Improving Generative Error Correction for ASR with Synthetic Data and Retrieval Augmentation
    Sreyan Ghosh, Mohammad Sadegh Rasooli, Michael Levit, Peidong Wang, Jian Xue, Dinesh Manocha, Jinyu Li
    ACL 2025

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

12. 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

13. Several other works on self-supervised speech and sound representation learning, speech error correction, RIR generation, captioning, etc, published at ACL/ICLR/ICASSP/InterSpeech: here
*Sreyan Ghosh** et al.

Selected Publications on NLP & Vision - All Publications

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

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

5. Several other works on synthetic data generation published at ACL/NAACL/EMNLP: here
*Sreyan Ghosh** et al.

Internships

NVIDIA Santa Clara, CA, USA
Research Scientist Intern *August 2024 – Present*
- I lead the Flamingo series at Nvidia, a family of fully-open frontier multi-modal models with audio, visual, and text understanding and generation capabilities.
- Published several papers at ICLR, ICMl, and NeurIPS.

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. Paper accepted to ACL 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

Audio Flamingo GitHub | ⋆ 915

- Audio Flamingo is a series of fully open advanced audio understanding (large) language models, developed in collaboration with Nvidia.

GAMA GitHub | ⋆ 141

- GAMA is a Large Audio-Language Model 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.

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.