Kaousheik Jayakumar

+1-240-726-8616 kaousheik@gmail.com linkedin.com/in/kaousheik/ kaousheik-26.github.io

College Park, Maryland - 20740

Education

University of Maryland, College Park **August 2025 – May 2027**

MS in Computer Science

Advised by Prof. Dinesh Manocha and Prof. Ramani Duraiswami

Indian Institute of Technology Madras **2018 – 2023**

Dual Degree (B.Tech + M.Tech) in Electrical Engineering, Minor in Artificial Intelligence *CGPA: 8.59/10*

Advised by Prof. S. Umesh

Publications

1. **Audio-Visual Flamingo: Open Audio-Visual Intelligence for Long and Complex Videos**
   Sreyan Ghosh*, Arushi Goel*, **Kaousheik Jayakumar**, Lasha Koroshinadze, Nishit Anand, Zhifeng Kong, Siddharth Gururani, Sang-gil Lee, Jaehyeon Kim, Aya Aljafari, Chao-Han Huck Yang, Sungwon Kim, Ramani Duraiswami, Dinesh Manocha, Mohammad Shoeybi, Bryan Catanzaro, Ming-Yu Liu, Wei Ping
   Work In-progress

2. **Audio Flamingo Next: Next-Generation Open Audio-Language Models for Speech, Sound, and Music**
   Sreyan Ghosh*, Arushi Goel*, **Kaousheik Jayakumar**, Lasha Koroshinadze, Nishit Anand, Zhifeng Kong, Siddharth Gururani, Sang-gil Lee, Jaehyeon Kim, Aya Aljafari, Chao-Han Huck Yang, Sungwon Kim, Ramani Duraiswami, Dinesh Manocha, Mohammad Shoeybi, Bryan Catanzaro, Ming-Yu Liu, Wei Ping
   Arxiv preprint - 2026

3. **Massive Multi-Task Omni Understanding and Reasoning Benchmark for Long Real-World Videos**
   Arushi Goel*, Sreyan Ghosh*, Vatsal Agarwal*, Nishit Anand*, **Kaousheik Jayakumar**, Lasha Koroshinadze, Yao Xu, Katie Lyons, James Case, Karan Sapra, Kevin J. Shih, Siddharth Gururani, Abhinav Shrivastava, Ramani Duraiswami, Dinesh Manocha, Andrew Tao, Bryan Catanzaro, Mohammad Shoeybi, Wei Ping
   Arxiv preprint - 2026

4. **A Closer Look at Failure Modes in Temporal Understanding of Large Audio-Language Models**
   Apoorva Kulkarni, **Kaousheik Jayakumar**, Sreyan Ghosh, Sarah Wiegreffe, Dinesh Manocha, Ramani Duraiswami
   Under review

5. **Do Audio-Visual Large Language Models Really See and Hear?**
   Ramaneswaran Selvakumar*, **Kaousheik Jayakumar***, S Sakshi, Sreyan Ghosh, Ruohan Gao, Dinesh Manocha
   CVPR Findings – Denver, 2026

6. **Sparks of Cooperative Reasoning: LLMs as Strategic Hanabi Agents**
   Mahesh Ramesh, **Kaousheik Jayakumar**, Aswinkumar Ramkumar, Pavan Thodima, Aniket Rege, Emmanouil-Vasileios Vlatakis-Gkaragkounis
   ICML - Seoul, 2026

7. **The Tag-Team Approach: Leveraging CLS and Language Tagging for Enhancing Multilingual ASR**
   Kaousheik Jayakumar, Vrunda N. Sukhadia, A Arunkumar, S Umesh
   Interspeech – Dublin, 2023

8. **Building Robust and Scalable Multilingual ASR for Indian Languages**
   Arjun Gangwar, **Kaousheik Jayakumar**, S Umesh
   Arxiv preprint - 2024

Work Experience

Oracle Bangalore, Karnataka

Member of Technical Staff ***July 2023 – Aug 2025***

- Integrated edge computing solutions into Oracle Cloud Infrastructure (OCI), unifying architectural components to optimize efficiency, reduce latency, and enhance overall system performance across cloud environments.
- Engineered and executed a large-scale OS upgrade to Oracle Linux 8, ensuring cross-compatibility of dependent components, libraries, and middleware and improved stability of rack-mounted infrastructure.
- Engineered a solution to partition large ISO images into smaller chunks for parallelized security scanning, integrating it back into the infrastructure while ensuring compatibility with existing components, storage, and network pipelines.

Oracle Bangalore, Karnataka

Product Development Intern ***May 2022 – July 2022***

- Optimized Docker images by implementing multi-stage builds and minimizing layers, reducing image size and improving deployment speed and runtime efficiency.
- Developed an automated CICD pipeline system that integrated multiple components to generate RPMs and ISO builds
- Automated the configuration management of build environments using Bash/Python scripts, eliminating manual setup overhead and ensuring consistency across development and staging.

Saptang Labs Chennai, Tamil Nadu

Software and Cybersecurity Intern ***Jun 2021 – Jul 2021***

- Developed an internal tool to detect vulnerabilities on open ports of any ASN / Organization
- Enhanced machine learning algorithms that were automated into cyber-security solutions for detecting anomalies

Projects and Research Experience

TiDAN: RL-based framework for safety alignment and adaptable Guard-railing of LLMs | *CMSC723*

- Designed and implemented TiDAN, a closed-loop adversarial training framework that co-evolves an AutoDAN-based genetic algorithm attacker with a GRPO-fine-tuned defender, achieving up to 92% reduction in Attack Success Rate on AdvBench (28% $\rightarrow$ 2.3%) for Qwen3-0.6B across five jailbreak benchmarks (JailBreakBench, InTheWild, AdvBench, StrongREJECT, HarmBench).
- Designed a dual-objective LLM-as-a-Judge reward system that scores safety on malicious prompts and helpfulness on benign prompts (from Tulu-3), preventing over-refusal during alignment. Trained iteratively for 10 GRPO rounds on 4×A6000 GPUs, demonstrating that stronger reward models yield proportionally better safety alignment.
- Extended the framework into an adaptable guardrailing pipeline that translates user-defined domain guidelines (e.g., healthcare protocols) into adversarial training signals, demonstrating an 18.5% ASR reduction on HarmBench and 42.8% on AdvBench for domain-specific safety alignment without manual annotation.

Interpreting the modality bias in Audio-Visual Language models | *CVPR 2026*

- Conducted mechanistic interpretability analysis of Audio-Visual LLMs, revealing 56% performance degradation on counterfactual audio-visual inputs and identifying systematic visual bias in cross-modal information processing through attention pattern analysis.
- Developed novel evaluation framework using counterfactual audio-visual samples and LLM-based judges, demonstrating that models encode rich audio semantics (61% latent understanding) but fail to surface them in generation.
- Performed causal mediation experiments via attention knockouts across transformer layers, proving that blocking visual pathways in deeper layers recovers 50% audio understanding by eliminating cross-modal interference.

Enhancing ASR systems for low-resource languages | *Dual Degree Thesis, IIT Madras*

- Exploited phonemic similarities among various Indian languages and enhanced the performance of multilingual ASR systems for 10 low-resource languages, reducing WER by **2-15%** over the corresponding monolingual baselines.
- Analysed the effect of introducing language ID as an additional learning task which showed further **2-5%** improvement.

National Language Translation Mission (Speech & ASR) | *Bhashini, Govt. of India*

- Developed multilingual Speech systems (ASR, Speech Translation, Synthesis) across multiple Indian languages.
- Curated **4,000 hours** of labelled data from NPTEL lectures for low-resource Indian languages by audio-text alignment.
- Deployed real-time code-switched streaming models for ASR, optimizing inference latency for live-transcription.

Video-to-Video translation pipeline with Lip-Syncing | *IIT Madras*

- Engineered a video-to-video translation pipeline to synchronize lip movements of lecturers with translated audio tracks.
- Implemented and fine-tuned generative models (Wav2Lip/GANs) on technical lecture videos, ensuring temporal consistency and realistic mouth articulation.
- Addressed challenges in pose preservation and face restoration, integrating face detection landmarks to maintain visual fidelity.

Bridge - Multiplayer Co-op Card Game | *Interactive Web-Application*

- Developed a real-time multiplayer application for the classic card game Bridge using React, NodeJs and Socket.IO.

Skills

Languages: Python, C++, JavaScript, TypeScript, SQL, MERN full-stack

Libraries and Frameworks: PyTorch, Tensorflow, Keras, Numpy, Pandas, NodeJS, Django, React, Angular, React Native, NextJS, GraphQL

Databases: Postgres, MySQL, MongoDB

Software and Toolkits: Git, Linux, Shell / Bash scripting, REST API, Docker, Kubernetes, LaTeX, ESPNet, Fairseq

Relevant Coursework: Multimodal Foundational Models, Natural Language Processing (NLP), Probability, Statistics, Computer Vision, Linear Algebra, Data Structures and Algorithms, Estimation Theory, Multi-Arm Bandits

Miscellaneous

Teaching Assistant - Signals and Systems, IIT Madras — Jan 2023 – May 2023

Teaching Assistant - Digital Signal Processing, IIT Madras — July 2022 – Nov 2022

ICPC Regionalist - Secured 120th position out of 721 teams in the Asia Amritapuri regionals — Sept 2020