


BALAJI DARUR

 [balajidarur.github.io](https://github.com/balajidarur)

 [balajidarur](https://www.linkedin.com/in/balajidarur)

 darurlakshmipathibalaji@gmail.com

 [Scholar](#)

EDUCATION

International Institute of Information and Technology, Hyderabad

Dec 2021 – Present

B.Tech in Computer Science and MS in Computational Linguistics by Research

CGPA: 8.5

- **Teaching Assistant** - Responsible & Safe AI Systems under *Dr. PK*; Introduction to NLP under *Dr. Manish*.
- **Courses:** Computer Vision, Advanced NLP, Digital Signal Processing, Probability and Statistics.

PUBLICATIONS

One Identity, Many Roles: Multimodal Entity Coreference for Enhanced Video Situation Recognition [Paper] [Website]
Balaji Darur, Amanmeet Garg, Makarand Tapaswi **CVPR 2026 Findings**

Visual-Aware Speech Recognition for Noisy Scenarios [Paper] [Website]
Balaji Darur, Karan Singla **EMNLP 2025** (Oral, short)

Improving Bias Metrics in Vision-Language Models by Addressing Inherent Model Disabilities [Paper]
Lakshmipathi **Balaji Darur**, S.K Gouravarapu, S. Goel, Ponnurangam Kumaraguru **NeurIPS 2024 Workshop**

RESEARCH EXPERIENCE

Katha-AI Group, CVIT, IIIT Hyderabad

May 2023 – Present

Undergraduate Researcher | Advisor: Prof. Makarand Tapaswi

Hyderabad, India

- Introduced **multimodal entity coreference** resolution for video understanding, tying spatial tracking with semantic roles to consistently identify entities across events. Published at CVPR 2026 Findings.
- Extended this work to VideoLLMs by building new datasets for **coreference-aware grounded dense video captioning**, in collaboration with *Dr. Ravi Kiran*. Currently under review.
- Working on generating and evaluating **audio descriptions** for visually impaired media consumers.
- Previously, improved **emotion recognition in movies** by incorporating audio modality, gaining experience in noise separation and audio-visual processing.

Red Hen Lab, Google Summer of Code

May 2024 – Aug 2024

Open Source Contributor | Mentor: Dr. Karan Singla

Remote

- Developed an **audio-visual transcription** system that integrates visual scene cues with speech to improve transcription in noisy settings. Accepted as an Oral at EMNLP 2025.

Adobe Research, India

May 2025 – Aug 2025

Research Intern, | Mentors: Dr. Sayan Nag, Dr. Tejas I Dhamecha

Bengaluru, India

- Introduced the problem of **style transfer for design documents** and proposed a method using self-supervised learning. Gained experience working with large-scale design datasets and training models at scale.

Precog, IIIT Hyderabad

Jan 2026 – May 2026

Research Student | Mentors: Dr. Sayan Nag, Dr. Sanjoy, and Dr. PK.

Hyderabad, India

- Integrating **Concept Bottleneck models** with grounded vision-language models to improve model explainability by grounding entities in both spatial and semantic dimensions.

PROJECTS

Conditional Code Mixed Generation | *Pytorch, NLP* [Code] [Report]

April 2023

- Built a conditional generation pipeline for producing Hindi-English (Hinglish) code-mixed text from monolingual English input, fine-tuning multilingual seq2seq models (mT5, IndicBART) on parallel corpora.

Greddit | *React, MongoDB, Express, Node* [Code] [Report]

March 2023

- Implemented a clone of the Reddit application both frontend and backend.

HONORS & AWARDS

IndiaAI Fellowship, 2025. Awarded by the Government of India to support promising AI researchers.

Dean's List, M24. Recognized for outstanding academic performance at IIIT Hyderabad.

TECHNICAL SKILLS

Tools: SLURM, GCP, AWS, GitHub

Domains: Computer Vision, NLP, Audio Processing

Languages: Python, C/C++

Frameworks: PyTorch, Huggingface, LLaMA Factory