

Siddharth Yayavaram

syayavar@andrew.cmu.edu | siddharth.yayavaram@gmail.com | (412) 689-0376 | Work Authorization: US Citizen

EDUCATION

Carnegie Mellon University

MS in Computer Science - Machine Learning & Natural Language Processing (**GPA: 4.25/4.0**)
Coursework: Advanced Natural Language Processing, Generative AI, Machine Learning

Dec 2026

Pittsburgh, PA

Birla Institute of Technology and Science, Pilani

BE in Computer Science (**CGPA: 9.97/10**, Institute **Gold** Medalist - Rank 1)

June 2025

Pilani, India

PUBLICATIONS [ALL FIRST/CO-FIRST AUTHOR]

CAIRE: Cultural Attribution of Images by Retrieval-Augmented Evaluation.

CEGIS @ ICCV'25, EACL'26 (Main Conference) | [Paper](#) | [Code](#)

ICCV'25, EACL'26

BERT-based Idiom Identification using Language Translation and Word Cohesion.

Multiword Expressions and Universal Dependencies @ LREC-COLING | [Paper](#) | [Code](#)

LREC-COLING'24

Interpretable Feature Optimization for Sadness Recognition in Speech Emotion Analysis.

IEEE 12th International Conference on Intelligent Systems (IS) | [Paper](#) | [Code](#)

IEEE IS'24

EXPERIENCE

Carnegie Mellon University, Machine Learning Department

Graduate Student Researcher

Pittsburgh, PA

Aug 2025 – Present

- Developing **GameDevBench**, a scalable benchmark for evaluating multimodal LLM and computer-use agents (CUAs) in agentic Godot game development, comprising ~200 tutorial-derived tasks with automated task and test generation.
- Built automated task-quality scoring using pixel-level metadata and VLM-judge-assessment, eliminating manual validation.

Carnegie Mellon University, Language Technologies Institute

Research Intern (Undergraduate Thesis), NeuLab | Advisor: [Prof. Graham Neubig](#) | [Code](#)

Pittsburgh, PA

May 2024 – Mar 2025

- Built **CAIRE**, a retrieval-augmented evaluation system for cultural attribution in images, grounding visual content via large-scale entity linking. Implemented efficient retrieval over a 6M-entity FAISS index with multimodal SigLIP embeddings, outperforming LVM baselines on fine-grained object grounding (**FOCI benchmark**).
- Improved visual entity linking precision by reranking retrieved candidates using text-based semantic disambiguation.
- Showed that **CAIRE** enables open-source VLMs to outperform frontier models on cultural relevance evaluation by conditioning predictions on retrieved cultural context, achieving **+28% F1** and Pearson $r > 0.65$ alignment with human judgments; accepted at **ICCV-W** and **EACL** (Main Conference).

Amazon, Applied Science

Summer Intern | Advisor: [Abhishek Persad](#)

Bangalore, India

May 2023 – Aug 2023

- Shipping cost anomaly detection: trained regression models to estimate expected shipping costs beyond a rule-based heuristic, flagging anomalies via prediction residuals and reducing false negatives by ~25%; deployed via a Django REST API.
- Product entity extraction (NER): fine-tuned a BERT-based token classification model to extract brand and model fields from noisy product webpages, producing structured entities for downstream product knowledge bases.

Nanyang Technological University

Research Intern, SpeechLab | Advisor: [Prof. Chng Eng Siong](#) | [Code](#)

Singapore

Mar 2024 – Sep 2024

- Built a text-based depression detection system by LoRA-fine-tuning LLaMA-3.1-8B on DAIC-WOZ, reformulating prediction as **PHQ-8**-aligned symptom scoring for interpretability and structured reasoning; leveraged transcript preprocessing and LLM-based synthetic dialogue augmentation, achieving +7.1% F1 over prior text-only baselines.

PROJECTS

Hybrid Retrieval RAG System with Qwen2.5

- Built a Qwen2.5-7B RAG system using hybrid retrieval (MXBAI dense + BM25 sparse) with RRF.
- Implemented grid-search evaluation over retrieval hyperparameters using accuracy, BLEU, BERTScore, and LLM-as-Judge.

Representation-Level Unlearning in Multimodal Foundation Models

- Proposed RCRU, a representation-level unlearning method for LLMs with stable, on-manifold forgetting.
- Developed concept-wise visual unlearning for VLMs by steering vision-encoder embeddings toward neutral centroids.

Structured Agentic Reasoning with Diffusion Language Models

- Fine-tuned diffusion language models (Fast-dLLM v2, 1.5B) to act as ReAct agents, generating structured Thought–Action–Observation trajectories and improving tool-call reliability (5% → 60%) while reducing trajectory length (9.2 → 6.4 steps).

SKILLS

Programming & OS: Python, C/C++, Java, SQL, Linux, Git, REST APIs, High Performance Computing Clusters (HPC)

Libraries & Frameworks: PyTorch, Scikit-Learn, HuggingFace, PEFT (LoRA), FAISS, Django, NumPy, Pandas