

Utkarsh Tyagi

☎ +1 240 501 3163 ✉ utkarsht@umd.edu 🌐 [utkarsh4430.github.io](https://github.com/utkarsh4430) 🔗 [Linkedin](#) 🎓 [Google Scholar](#)

EDUCATION

University of Maryland (Advised by [Prof. Dinesh Manocha](#))

College Park, Maryland

M.S. in Computer Science, Advised by - GPA 3.9/4.0

08/2023 – 05/2025

Delhi Technological University (Formerly Delhi College of Engineering)

Delhi, India

B.Tech in Computer Science - GPA 8.7/10, Major GPA 9.05/10

2017 – 2021

PUBLICATIONS

MULTIMODAL LEARNING

- ProSE: Diffusion Priors for Speech Enhancement
NAACL 2025 (Under review)
- MMAU: A Massive Multi-Task Audio Understanding and Reasoning Benchmark
ICLR 2025 (Under review)
- VDGD: Mitigating LVLH Hallucinations in Cognitive Prompts by Bridging the Visual Perception Gap
ICLR 2025 (Under review)
- CompA: Addressing the Gap in Compositional Reasoning in Audio-Language Models
ICLR 2024
- GAMA: A Large Audio-Language Model with Advanced Audio Understanding and Complex Reasoning Abilities
EMNLP 2024
- Do Vision-Language Models Understand Compound Nouns?
NAACL 2024
- LipGER: Visually-Conditioned Generative Error Correction for Robust Automatic Speech Recognition
InterSpeech 2024
- AdVerb: Visually Guided Audio Dereverberation
ICCV 2023
- MMER: Multimodal Multi-task Learning for Speech Emotion Recognition
InterSpeech 2023

NATURAL LANGUAGE PROCESSING (NLP)

- ABEX: Data Augmentation for Low-Resource NLU via Expanding Abstract Descriptions
ACL 2024
- ASPIRE: Language-Guided Data Augmentation for Improving Robustness Against Spurious Correlations
ACL 2024
- CoDa: Constrained Generation based Data Augmentation for Low-Resource NLP
NAACL 2024
- ACLM: Selective-Denoising based Data Augmentation for Low-Resource Complex NER
ACL 2023
- DALE: Generative Data Augmentation for Low-Resource Legal NLP
EMNLP 2023
- BioAug: Conditional Generation based Data Augmentation for Low-Resource Biomedical NER
SIGIR 2023

INDUSTRY EXPERIENCE

Samsung SARC/ACL

San Jose, California

Research Intern - Efficient GenAI Inference

May 2024 – August 2024

- Optimizing state-of-the-art Large Language Models (LLMs) for deployment on resource-constrained devices.
- Finetuning LLMs and optimizing their performance through sophisticated Retrieval-Augmented Generation (RAG) techniques leveraging internal documentation databases.

Atlassian India LLP

Bangalore, Karnataka

Software Development Engineer 2

July 2021 – August 2023

- Worked on the JIRA Issue Create Experience to deliver a reliable and performant interface, which helped improve key business metrics and customer satisfaction ratings.
- Achieved **99.996% frontend reliability** and scaled the experience for over 250k customers with more than 2M issues created every day. Improved performance TTI from 3.9s to **1.2s** using modern technology stack.

Software Engineer Intern

Jan 2021 – June 2021

- Identified limitations and bottlenecks in the existing legacy API implementation and proposed comprehensive improvements for system revamping. Utilized this performant API to develop the user interface.
- Reduced the **latency** of modal load by **95.57%** by avoiding redundant server-side database queries and scaled to 1M customers.

Samsung R&D Institute Bangalore

Bangalore, Karnataka

Research Intern - Speech AI

May 2020 – July 2020

- Worked with the **Voice Intelligence** team of Samsung to improve Bixby's wakeup word detection.
- Researched keyword spotting techniques and built an efficient on-device LSTM-CTC based, vocabulary-independent keyword spotter

RESEARCH EXPERIENCE

GAMMA Lab, University of Maryland

College Park, Maryland

Research Assistant

August 2022 – Present

- Currently researching multimodal learning; working on video representation learning for long-form video understanding
- Also focus on low-resource (labeled data and compute) learning with speech, NLP, or vision applications. In this area, I solve problems using self-supervised learning, synthetic data augmentation, etc.
- Published at **ICLR, ICCV, ACL, EMNLP, InterSpeech, SIGIR, NAACL**
- Advised by [Prof. Dinesh Manocha](#)

Multimodal Digital Media Analysis Lab

IITD, Delhi

Machine Learning Researcher

April 2022 – March 2023

- Worked under the supervision of [Dr. Rajiv Ratn Shah](#) in the areas of Speech and Language Processing.
- Worked on multi-lingual automatic speech scoring systems for low-resource Indian languages.
- Also explored novel architectures to detect implicit hate speech in online conversations. Paper accepted at **AAAI 2023 Defactify**

PATENTS

- Apparatuses, methods, and computer program products for generating an abstract context summary scheduling interface configured for scheduling and outputting abstract context summaries for multi-party communication channels
U.S. Patent Application No. 17/936,695 (Patent Pending)
- Apparatuses, methods, and computer program products for generating and selectively outputting abstract context summaries for multi-party communication channels
U.S. Patent Application No. 17/936,705 (Patent Pending)

TECHNICAL SKILLS

Languages: (Highly Proficient) Python, JavaScript, C++ (Moderate) Java, SQL

Frameworks: Pytorch, FastAI, Tensorflow, Scikit-Learn, ReactJS, Jest, Enzyme, NLTK

ACHIEVEMENTS

- Awarded **\$7,000 Fellowship** at UMD: Project on EEG-based Brain-Computer Interaction using LLMs
- **Reviewer** for ICLR 2025, ACL ARR 2024 (Feb, April, June, October), ICASSP 2025, EMNLP 2024
- **People's choice award 2021 & 2022:** Atlassian's internal hackathon
- **Winner** of Digital Management Inc. India Hackathon, 2019
- **2nd position** in Student Hackday 2019 organised by Skillenza
- **Barclays India Hackathon 2019 Top 4**, All India
- Achieved **5 star rating** on **CodeChef** platform for competitive programming
- **JEE Mains 2017: 99.49%** percentile, Pan India