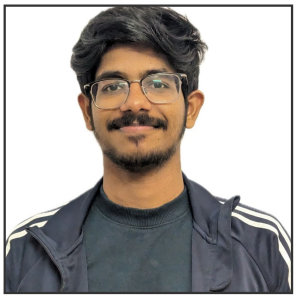
Zuhair Hasan Shaik

+91 9494292277 | zuhashaik12@gmail.com | linkedin.com/in/zuhashaik/ | Zuhair Hasan



Personal Profile

An undergraduate student at IIIT Dharwad with a strong focus on Natural Language Processing (NLP), multimodal systems, and Responsible AI. My research journey spans over two years, with publications in two workshops at top-tier NLP conferences and two core A* conference papers at EMNLP 2024 and WWW 2025. Currently, I am a research intern at Microsoft Research (MSR) India, working with Minghua Ma a Senior Researcher at MSR. Previously I worked in LCS2 lab and FLaMe.nlp as a research intern. My interests lie in developing socially responsible AI solutions, leveraging multimodal learning, and contributing to ethical AI practices, with the long-term goal of pursuing a PhD in Responsible AI and social computing.

Education

Indian Institute of Information Technology Dharwad *Hubli-Dharwad, India*

BTech in Data science and Artificial Intelligence *2021 - Present*

- Specialization in Data science, Machine Learning and Data engineering

Akshara Vidyalaya *Nellore, India*

Schooling *2019 passed out*

- Passed with Distinction
- Specialised in science and Mathematics

Skills

Programming languages	Python, C/C++, Java, SQL.
Frameworks and technologies	Tensorflow, Pytorch, Huggingface, Pandas, Langchain, Azure, KQL, Big-Query, Apache Spark, Hadoop, Tableau, Matplotlib, Seaborn
Soft Skills	Leadership, Time Management, Teamwork, Problem-solving, Engaging Presentation.

Work Experience

Microsoft Research *Banglore, India*

Research Intern *Jan 2025 - Present*

- Working on agents that automate the internal working of Microsoft Incident triage systems. I work with Minghua Ma, who is a senior researcher at MSR Redmond, Washington.

FLaMe.nlp Research Lab *Delhi, India*

Research Intern *May 2024 - Jan 2025*

- Working on a novel framework, M3H, that outperforms SOTA models with a notable margin, designed to tackle dynamic, emerging memes expressing mental health symptoms; which was accepted at WWW 25 as an oral presentation.
- Conducting advanced research on mental health trust analysis and dark humor meme interpretation, targeting publication in TACL 2025 and CVPR 2025, to deepen insights into nuanced social media expressions related to mental well-being.
- Conducting research under the guidance of Md Shad Akhtar.

Vocab.Ai *Hubli, India*

NLP Intern *May 2023 - Oct 2024*

- I am the team lead for the project where Vocab.Ai startup achieved its inaugural revenue milestone.
- I've worked on issue detection and customer sentiment analysis, including the cause for sentiment shifts, within agent-customer conversations, utilizing BERT-based models.
- Implemented RAG on the company's SOPs stored in a vector database, leveraging top NLP models (*llama-70b, zephyr, openchat*), to ensure adherence and streamline information retrieval.

LCS2 Research Lab *Delhi, India*

Research Intern *May 2024 - Sep 2024*

- Worked on emotion recognition in conversational systems and mental health applications. As a result, we outperformed all three datasets in emotion recognition in multiparty conversation, MELD, IEMOCAP, and EmoryNLP. We submitted our work to COLING 2025, which is currently under review.
- Research under the guidance of Md Shad Akhtar.

Research Experience

EMNLP 2024, LaRA: Large Rank Adaptation for Speech and Text Cross-Modal Learning in Large Language Models

ADV: Dr. Deepak KT

- Proposed a novel Large Rank Adaptation (LaRA) method for cross-modal integration of speech and text (speech-to-text or text-to-speech) in large language models, using significantly larger ranks than conventional LoRA.
- This proposed work converts speech to tokens (using HuBERT), fine-tunes an LLM on cross-modal data (using Librispeech and DailyTalk), and can synthesize speech output (using Hi-Fi GAN vocoder).
- Accepted at **EMNLP-2024**.

WWW 2025, Figurative-cum-Commonsense Knowledge Infusion for Multimodal Mental Health Meme Classification

ADV: Dr. Md Shad Akhtar

- Introduced AxiOM, a unique dataset derived from the GAD anxiety questionnaire, specifically tailored for categorizing mental health symptoms in memes into six detailed anxiety symptoms.
- Designed Proposed M3H framework, enhancing Multimodal Language Models' capacity for interpreting figurative language by integrating commonsense knowledge, which improved understanding of mental health symptom expression. Our proposed model also tackles the dynamic social media posts when new memes emerge so our model can be up to date!
- Achieved significant performance improvements with M3H, demonstrating a 4.20% and 4.66% increase on weighted-F1 scores across datasets. Conducted comprehensive experiments and human evaluations, as well as an extensive ablation study on the RESTORE dataset for depressive symptom identification.
- Accepted at **WWW-2025** as an oral presentation (acceptance rate below 19%).

EDiReF, SemEval-2024 at NAACL

ADV: Dr. Sunil Saumya

- SemEval 2024 EDiReF focused on recognizing emotions and triggers for emotion shifts in code-mixed Hindi-English conversations, including multi-party dialogues.
- Our team achieved remarkable success, securing **5th**, **3rd**, and **3rd ranks** in tasks 1, 2, and 3 respectively, with the paper accepted by SemEval-2024 at **NAACL 2024**.

HOLD-Telugu, DravidianLangtech-2024 at EACL

ADV: Dr. Sunil Saumya

- The HOLD-Telugu shared task at DravidianLangTech@EACL 2024, focusing on innovative solutions for identifying hate speech and offensive language in Telugu codemixed text, a challenge due to the complexities of multilingual expressions and language blending.
- Paper has accepted and pusblished in **EACL 2024**, DravidianLangTech workshop.

ICML 2025, Redefining Experts: Layer-Wise Semantic Analysis for Enhanced Interpretability in Deep Networks

ADV: Dr. Md Shad Akhtar

- Moved from neuron-centric to layer-wise semantic analysis for improved interpretability in deep networks, correcting a misguided direction in the existing research literature.
- Proven Improvements – Achieved 15.84% and 7.43% AUROC gains on Jigsaw and ToxiCN datasets for toxicity detection.
- Scalable Framework – Developed a clustering-based method revealing hierarchical toxicity encoding in deep layers.
- Submitted to **ICML 2025** currently under-review.

TACL 2025, Trust Modeling in Counseling Conversations: A Benchmark Study

ADV: Dr. Md Shad Akhtar, Dr. Tanmoy Chakraborty

- Proposed a novel approach to assess therapeutic trust in counseling conversations by introducing a dynamic, observable trust metric based on patients' willingness and openness to express themselves.
- Introduced the MENTAL-TRUST dataset comprising 212 annotated counseling sessions with expert-verified ordinal trust levels, used for evaluating trust progression in therapeutic interactions.
- Defined the problem as an ordinal classification task and developed TrustBench, a benchmark for evaluating trust quantification using classical and state-of-the-art language models.
- Submitted to **TACL** currently under-review, pre-print available: here

***IEEE T AFFECT COMPUT*, GenZ-ERC: Advancing ERC with LLMs, Commonsense Reasoning and Semantic-Enriched Memory Networks**

ADV: Dr. Md Shad Akhtar, Dr. Sunil Saumya

- Crossed popular datasets baselines and stood top in emotion recognition in the conversation for three datasets: MELD, IEMOCAP, and EmorynlP.
- Explored large language models, LSTM memory networks, and common sense reasoning for emotion recognition.
- Submitted to **IEEE T AFFECT COMPUT**, currently under review.

***KDD 2025*, Understanding Dark Humor on Social Media**

ADV: Dr. Md Shad Akhtar, Dr. Sunil Saumya

- Proposed **D-MEMEs**, the first annotated dataset focused on dark humor in memes, including dark attributes, meme targets, and darkness intensity ratings.
- Developed **DarkNet**, a novel multimodal framework combining SAM for segmentation and ODI-RAG for dynamic meme interpretation, outperforming competitive baselines.
- Conducted interpretability analysis, qualitative assessments, and human evaluations to enhance understanding of dark humor in social media memes.
- Submitted to **KDD 2025 Research Track**, currently under review.

***ACL 2025 (ARR)*, IVD Indic Vulgar Detection: A vulgar detection dataset for Hindi and Telugu languages**

ADV: Dr. Md Shad Akhtar, Dr. Sunil Saumya

- Crossed popular datasets baselines and stood top in emotion recognition in conversation for three datasets: MELD, IEMOCAP, and EmorynlP.
- Explored large language models, LSTM memory networks, and common sense reasoning for emotion recognition.
- Submitted to **ACL 2025 (ARR)**, currently under review.

***ACL 2025 (ARR)*, TD-CoT: Enhancing Small Language Models through Thought Distillation and Chain of Thoughts**

- Introduced **TD-CoT**, a novel method enabling smaller language models to leverage the reasoning strategies of larger models for improved logical reasoning.
- Conducted extensive experiments across three diverse datasets, showcasing significant performance gains in resource-constrained environments.
- Performed human evaluation on generated samples and provided a balanced analysis, identifying strengths and limitations of the TD-CoT approach.

References

Dr. Sunil Saumya
HoD of DSAI and Assistant Professor, IIIT Dharwad (*Advisor*)
sunil.saumya@iiitdwd.ac.in

Dr. Md. Shad Akhtar
Assistant Professor, IIIT Delhi (*Internship advisor*)
shad.akhtar@iiitd.ac.in

Dr. Deepak KT
CEO of IDRP and Assistant Professor, IIIT Dharwad (*Advisor*)
deepak@iiitdwd.ac.in