Zuhair Hasan

✉ Email in LinkedIn ○ GitHub 🌐 Website Google Scholar 📍 MBZUAI, UAE

EDUCATION

Indian Institute of Information Technology (IIIT) Dharwad — Dharwad, India
Bachelor of Technology (B.Tech) in Data Science and Artificial Intelligence — *2021 – 2025*

- Completed and awarded degree on 04/07/2025

AWARDS AND ACHIVEMENTS

Invited Talk, Alan Turing Institute on AI safety & QMUL	London, UK
NeurIPS Scholar Award 2025	San Diego, USA
Invitation to Google Deepmind Symposium 2025	Banglore, India
MBZUAI NLP Scholar Travel Grant & QMUL	Abu Dhabi, UAE

PUBLICATIONS

1. **Redefining Experts: Interpretable Decomposition of Language Models for Toxicity Mitigation**
Zuhair Hasan Shaik, Abdullah Mazhar, Aseem Srivastava, Md Shad Akhtar
NeurIPS 2025

2. **M-HELP: Using Social Media Data to Detect Mental Health Help-Seeking Signals**
MSVPJ Sathvik*, Zuhair Hasan Shaik*, Vivek Gupta
EMNLP 2025 (Findings)

3. **Figurative-cum-Commonsense Knowledge Infusion for Multimodal Mental Health Meme Classification**
Abdullah Mazhar, Zuhair Hasan Shaik, Aseem Srivastava, Polly Ruhnke, Lavanya Vaddavalli, Sri Keshav Katragadda, Shweta Yadav, Md Shad Akhtar
WWW 2025 (Oral)

4. **LaRA: Large Rank Adaptation for Speech and Text Cross-Modal Learning in Large Language Models**
Zuhair Hasan Shaik, Pradyoth Hegde, Prashanth Bannulmath, Deepak KT
EMNLP 2024 (Findings)

5. **FeedForward at SemEval-2024 Task 10: Trigger and sentext-height enriched emotion analysis in multi-party conversations**
Zuhair Hasan Shaik, D Prasanna, E Jahnavi, Rishi Thippireddy, Vamsi Madhav, Sunil Saumya, Shankar Biradar
SemEval 2024 (Workshop)

6. **IIITDWD-zk@ DravidianLangTech-2024: Leveraging the Power of Language Models for Hate Speech Detection in Telugu-English Code-Mixed Text**
Zuhair Hasan Shaik, Sai Kartheek Reddy Kasu, Sunil Saumya, Shankar Biradar
DravidianLangTech 2024 (Workshop)

** denotes authors who contributed equally

Experience

MBZUAI NLP   Abu Dhabi, UAE
Research Engineer   *Aug 2025 – Present*

- Working on Multimodality and VLM personalization for social impact under the guidance of Dr. Yova Kementchedjhieva

Microsoft Research India   Bengaluru, India
Research Intern   *Jan 2025 – Jul 2025*

- Developed a faster incident triaging system that generates troubleshooting guides automatically (Auto-TSG) using a multi-agent framework trained on an extensive Microsoft's incident data corpus

- Worked under the guidance of Minghua Ma, Senior Researcher at Microsoft Redmond, to create solutions saving significant manual triaging time across 20,000 teams

Laboratory for Computational Social Systems (LCS2)   Delhi, India
Research Intern   *Sep 2024 – Jan 2025*

- Reduced LLM toxicity using interpretable SVD decomposition to identify and dampen toxicity-generating experts with minimal perplexity change and no retraining (**NeurIPS 2025**)

- Conducted benchmark study on trust behavior dynamics between therapists and patients in mental health therapy with extensive qualitative analysis (**under review**)

- Developed M3H framework for dynamic social media analysis and mental health behavior detection in collaboration with the University of Illinois (**WWW 2025**)

Vocab.AI (Research Startup @ IIIT Dharwad)   India
NLP Intern   *May 2023 – Oct 2024*

- Led the project team that achieved Vocab.AI startup's inaugural revenue milestone

- Developed issue detection and customer sentiment analysis systems, including sentiment shift causality analysis, within agent-customer conversations using BERT-based models

- Implemented RAG system on company SOPs stored in vector database, leveraging top NLP models (llama-70b, zephyr, openchat) for adherence and streamlined information retrieval

References

Dr. Yova Kementched-jhieva	**Dr. Md. Shad Akhtar**	**Dr. Sunil Saumya**
MBZUAI ✉	IIIT Delhi ✉	IIIT Dharwad ✉