SHIVANGI TRIPATHI

Ph.D. Candidate in Computer Science | Multimodal AI | Reliable Physical-World AI

San Marcos, TX | tjk86@txstate.edu | LinkedIn | GitHub | Google Scholar | Website

Summary

Ph.D. candidate in Computer Science at Texas State University (graduating May 2027) who builds reliable multimodal AI systems for real-world data. My work covers vision-language grounding for traffic scenes, audio-language models for clinical conversations, evidence-grounded correction of LLM output, adversarial robustness, and benchmark design. Five peer-reviewed papers, including **Findings of EACL 2027**. Experienced in end-to-end experiment design, multimodal datasets, Python/PyTorch pipelines, and large GPU/HPC workloads (**12,400 GPU-hour NAIRR Pilot allocation**).

Education

Doctor of Philosophy (Ph.D.) in Computer Science — **Aug 2023 – May 2027 (Expected)**
Texas State University • Advisor: Dr. Heena Rathore — *San Marcos, TX*

Master of Science (M.S.) in Electrical Engineering — **Jul 2021 – Jul 2023**
Indian Institute of Technology Jodhpur — *Jodhpur, India*

Bachelor of Technology (B.Tech.) in Electronics & Communication Engineering — **Aug 2017 – May 2021**
Guru Gobind Singh Indraprastha University — *Delhi, India*

Technical Skills

Programming & ML: Python, C++, Java, MATLAB; PyTorch, TensorFlow, scikit-learn, Hugging Face Transformers, Keras

Multimodal AI: Vision-language models (VLMs), large language models (LLMs), audio-language models, CLIP, DINO, Qwen2.5-VL, Qwen2-Audio, text-to-image diffusion (FLUX.1-dev), image–text alignment, contrastive and self-supervised representation learning, multimodal fusion, scene graphs, knowledge graphs

Evaluation & Research: Experiment design, dataset curation, benchmark construction, ablation studies, error analysis, NLI and semantic evaluation, retrieval-augmented generation (RAG), hallucination detection, grounding and faithfulness evaluation

Systems: Linux, Git, Docker, CUDA, multi-GPU and HPC workflows, AWS, GCP, FAISS, Pandas, NumPy

Research Experience

Doctoral Researcher and Instructional Assistant — **Aug 2023 – Present**
Human AI Synergy Lab, Texas State University — *San Marcos, TX*

- Lead independent research on **reliable multimodal AI** across vision, language, and audio: grounding, hallucination detection and correction, robustness, and evaluation.
- Design and run reproducible **Python/PyTorch/Hugging Face** experiments, multimodal data pipelines, benchmarks, and ablations on multi-GPU servers and the **NCSA Delta** supercomputer.
- Published five first-author papers (EACL Findings, IEEE CCNC, IEEE AITest, IEEE TPS-ISA, IEEE MobiSecServ) and delivered guest lectures on LLM security and hallucination.

Research Assistant — **Aug 2022 – May 2023**
Wireless Communications Lab, Indian Institute of Technology Jodhpur — *Jodhpur, India*

- Built MATLAB/Simulink simulation and analysis pipelines for mission-critical mobile edge computing and high-mobility 6G wireless systems, studying trade-offs between latency, reliability, and resource allocation.

Selected Research Projects

MS-CAPS: Minimal Sufficient Captions for Safety-Critical Traffic Scenes — *2025 – Present*

- Built a knowledge-graph-guided framework that shortens long traffic-scene captions to a **minimal sufficient** form while keeping safety-critical objects, attributes, relations, traffic controls, and multi-agent interactions.

- Evaluated on **150 real-world scenes** (50 each from Vehicle-centric, CoDA-LM corner cases, and NuInteract) with **Qwen2.5-VL, GPT-4o, and Gemini**, using **FLUX.1-dev** image reconstruction and image-, caption-, and scene-graph-level metrics to measure what information survives compression.
- Reached up to **98% acceptance (49/50 scenes)** on the Vehicle-centric set with **62.5% average caption reduction** for accepted GPT-4o outputs; now testing whether compressed captions preserve **ego-vehicle action and primary-hazard** decisions.

Clinical Audio Grounding and Hallucination Attribution *2025 – Present*

- Built an evaluation pipeline on **57 doctor–patient consultations (PriMock57)** comparing cascade **audio → transcript → LLM** systems with end-to-end audio-language models for clinical note generation.
- Trace errors across **recognition, recall, and reasoning** stages with claim-level labels (supported, contradicted, omitted, altered, unsupported); pilot runs showed about **75% hallucination** and **80% topic-mismatch** rates, motivating stage-wise attribution.
- Measuring which architecture introduces the fewest clinically significant errors, and how mistakes in medications, numbers, negation, and symptoms carry into the final note.

REVISE: Evidence-Grounded Correction of LLM-Generated Text *2024 – Present*

- Developed a paragraph-level framework that splits generated text into verifiable claims, retrieves and ranks evidence, checks each claim with **NLI**, and rewrites only the unsupported claims instead of regenerating the full response.
- Evaluated on **1,400+ misinformation-containing passages** from Natural Questions and SQuAD across several generative models; reached up to **76.8% factual accuracy** on NQ-Mis while preserving the original paragraph's structure and meaning.

Adversarial Robustness of Open LLMs *2024*

- Tested Vicuna-7B and LLaMA-2-7B-chat under weak-semantic and out-of-distribution prompt attacks; attacks induced hallucinations with success rates up to **92.53%** (IEEE MobiSecServ 2024).

Publications

- **S. Tripathi**, H. Rathore. "MS-CAPS: Minimal Sufficient Image Captions for Safety-Critical Traffic Scene Understanding." *Findings of the Association for Computational Linguistics: EACL*, 2027.
- **S. Tripathi**, T. Jennifer, H. Griffith, H. Rathore. "Paragraph-Level Hallucination Detection and Correction for Trustworthy Large Language Models in Networked Systems." *IEEE 23rd Consumer Communications & Networking Conference (CCNC)*, 2026.
- **S. Tripathi**, H. Rathore. "Detecting and Correcting Hallucinations in Paragraph-Level Text with Ensemble-Based Evaluation." *IEEE 7th International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA), DISTILL Workshop*, 2025.
- **S. Tripathi**, T. Jennifer, H. Griffith, H. Rathore. "REVISE: A Framework for Paragraph-Level Misinformation Correction in Large Language Models." *IEEE International Conference on Artificial Intelligence Testing (AITest)*, 2025.
- **S. Tripathi**, H. Griffith, H. Rathore. "Assessing Hallucination in Large Language Models under Adversarial Attacks." *Ninth International Conference on Mobile and Secure Services (MobiSecServ)*, 2024.

Teaching

Guest Lecturer, Department of Computer Science, Texas State University

- **Cryptography**: security attacks on large language models *Fall 2026*
- **Computer System Security (CS 4371)**: three lectures, incl. LLM hallucination detection *Fall 2025 – Fall 2026*

Awards & Training

- **NAIRR Pilot Compute Allocation**: 12,400 GPU hours on NCSA Delta for trustworthy VLM research *2026 – 2027*
- **NVIDIA Deep Learning Institute**: Building AI Agents with Multimodal Models *2025*