SHUBH GARG

shubhgarg265@gmail.com | +91 8171808091 | linkedin.com/in/sgarg3011 | github.com/Shubhf | Google Scholar | shubhf.github.io

Evaluation Reliability · Clinical AI · Neurosymbolic & Embodied AI · AI Safety

RESEARCH SUMMARY

My core empirical finding is that **evaluation protocol is a first-class variable** in AI research: the same model, same layer, and same attribute produce structurally different causal conclusions depending solely on dataset construction. I pursue this question across four domains — clinical AI (federated retinal grading, strabismus quantification), remote sensing and disaster intelligence (multitemporal segmentation, hazard benchmarking), neurosymbolic and embodied AI (KG-grounded EQA agents, drug repurposing), and representation reliability (probe task sensitivity, CLIP concept stability). B.E. ECE, Thapar Institute (CGPA 8.67/10), June 2026. Currently Research Scholar at IAIRO, Gandhinagar, under Prof. **Amit Sheth**.

EDUCATION

B.E. in Electronics and Computer Engineering

Thapar Institute of Engineering and Technology, Patiala | Oct 2022 – Jun 2026

- CGPA: 8.67/10 | Merit Scholarship Awardee (Rs. 1,41,000)
- Relevant coursework: Deep Learning and Applications, Artificial Intelligence, Optimization Techniques, Numerical Analysis, Data Structures and Algorithms

RESEARCH EXPERIENCE

Research Scholar — Neurosymbolic AI and Drug Repurposing

IAIRO (Indian AI Research Organisation) | Prof. Amit Sheth | Gandhinagar | Aug 2026 – Present

- Extending the Hetionet KG-LLM drug repurposing pipeline to identify novel compound-disease associations with structured mechanistic provenance; prior pipeline surfaced Metformin (prostate cancer, 86.1), Lithium (Alzheimer's, 85.0), Celecoxib (colorectal cancer, 85.0) with 96% recall on known associations.
- Contributing to embodied AI and knowledge graph research at India's first dedicated AI research organisation; combining formal KG representations with large-scale LLMs for verifiable, grounded AI.

Research Intern — LLM Alignment, Embodied AI and Knowledge Graphs

AIISC, University of South Carolina | Prof. Amit Sheth | Mar 2026 – Jul 2026

- Built neurosymbolic KG-LLM pipeline on Hetionet for drug repurposing; 96% recall on known compound-disease associations vs. 30% ceiling in prior literature.
- Developed KG-grounded EQA agent (AI2-Thor/ProcThor) with global semantic graph; cross-attention between Qwen3VL and graph embeddings for context-aware long-range navigation.
- Contributing to TIDIER: training-free text-to-image safety framework using Multimodal Knowledge Graph; ~65% reduction in harmful outputs across four benchmarks, +9-10% CLIP score.

Research Intern — Clinical NLP and Algorithmic Fairness

Project ANA (Anorexia in Men Research Initiative) | Jan 2026 – Present

- Building BERTopic NLP pipeline to detect systematic gender bias in eating disorder screening instruments; translating findings into clinician-facing deployment reliability recommendations.

Research Intern — Federated Learning Under Distribution Shift

Ubisys Lab, IIT Jodhpur | Dr. Suchetana Chakraborty | May 2025 – Jun 2025

- Improved minority-class F1 by 18% using personalised CNN-LSTM with temporal attention under non-IID federated settings; reduced communication overhead 60% via selective aggregation.

AI Research Intern — Probabilistic Forecasting and Optimization

Indian Institute of Management Udaipur | Dr. Sumit Kumar | Jan 2025 – May 2025

- Bayesian state-space and Kalman filter models for stochastic sequence prediction with calibrated uncertainty estimates under sparse, irregular observations.

Research Intern — Multimodal Robustness Evaluation

Samsung R&D; Institute (Samsung PRISM) | Oct 2024 – Mar 2025

- Multimodal watermark detection at 91% accuracy under perturbation and compression distribution shifts; transformer inference optimisation via quantisation for low-resource deployment.

Undergraduate Research Assistant — Clinical AI and Medical Imaging

Thapar Institute | Dr. Debabrata Ghosh and Dr. Neeru Jindal | Feb 2024 – Jun 2026

- StrabNet-CQ: automated strabismus classification and continuous prism diopter quantification (Pearson r=0.730 vs. clinical gold standard); Grad-CAM validated against clinician annotations. Published BMC Ophthalmology. 2 patents granted.
- AttentiveHybridNet: CNN-Transformer with cross-attention for brain tumour MRI classification. Published IEEE AI SUMMIT 2025.
- Thalassemia iron overload stratification (AUC 0.83) with sub-second ARM inference for rural screening.

SELECTED PROJECTS

NaviGraph — Embodied QA via Global Semantic Graph Grounding

AIISC, USC | NeurIPS 2026 NeMo (Accepted Poster) | CoRL 2026 SISREN (Accepted) | ICLR 2027 (Under Review) | 2025 – Present

- "Answering Without Arriving: An Evidence Audit for Embodied Question Answering" (Garg, Pandey, Rawal, Khandelwal, Sheth). NeMo @ NeurIPS 2026. Accept (Poster). Avg. rating 7.00/10 (min 6, max 8).
- 0% closed-loop navigation success in standard EQA; 29.6-point gap between oracle navigation (48.2% QA accuracy) and current agents (18.6%). Extended in "Room to Remember: Knowledge-Graph Memory for Long-Horizon EQA" — ICLR 2027 submission #52972.

Probe Task Sensitivity as a Causal Reliability Diagnostic

ICLR 2027 Target | 2025 – 2026

- Establishes evaluation protocol as a first-class variable in interpretability. Label imbalance produces saturation artifacts across Qwen and Mistral for age, education, gender, and emotion attributes; balanced evaluation recovers model-specific causal geometry.

Concept Stability in Vision-Language Models

Collaboration with Chen Shani, Tel Aviv University (previously Stanford) | 2026

- 500 perturbation triplets on CLIP ViT-B/32 (Visual Genome). Attribute flip similarity: 0.938; spatial relation flip: 0.989. CLIP image encoder actively prefers incorrect caption after spatial flip (delta = -0.0014, $p < 0.0001$). Colour enters representations at L11-L12.

Federated Retinal Foundation Models

WACV 2027 Under Review | First and Sole Author | 2025 – 2026

- First benchmark of federated DR grading with retinal foundation model across four populations (Brazil, India, UAE). FedAvg degrades QWK 0.835 $\rightarrow$ 0.586 with asymmetric grade collapse; post-aggregation personalised fine-tuning closes 90% of the gap without data sharing.

Intermediate DR Grading (ONCL+CMAW)

Collaboration with Dr. Sahil Thakur, Mediwhale/NUS | AIM Under Review | 2025 – 2026

- Ordinal Neighbour Contrastive Learning + Confusion-Matrix Adaptive Weighting across 21,803 fundus images (tabletop, mobile, ultra-widefield; Brazil, France, China). Intermediate F1 0.500; Severe NPDR F1 improved up to 15.7pp over cross-entropy baseline.

DAFSeg — Dual-Stream Adaptive Fusion for Multitemporal Disaster Segmentation

Scientific Reports (Under Review) | BRIGHT Benchmark | 2025 – 2026

- Dual-stream architecture with sample-level fusion selector (concatenation / subtraction / attention-based) for pre-/post-event image pairs. Achieves 63.17% mIoU, 93.85% OA, Destroyed-class IoU 56.40% — highest on BRIGHT under coupled segmentation protocol.
- Ablation across 14 disaster events (Kruskal-Wallis H=188.0, p=4.0×10■³³ event-wise variation) shows no statistically distinguishable advantage of adaptive selection over fixed baselines — less than 1 mIoU oracle headroom. The evaluation was the finding.

CrisisMMD-Seg — Hazard-Centric Benchmark for Ground-Level Disaster Segmentation

Scientific Reports (Under Review) | First Author | 2025 – 2026

- Pixel-level annotations for 5 disaster categories (collapsed structures, debris, fire, flood, vehicle damage) extending CrisisMMD — first benchmark for ground-level disaster segmentation with structured hazard labels.
- UniCrisisNet framework: DSF-ViT for 3-class severity classification (88.1% accuracy across CrisisMMD and BRIGHT), DARFormer with Hazard Interaction Module achieving 51.9% mIoU; consistency analysis via ordinal agreement metrics.

Geometry Without Structure — Biomedical LM Embedding Geometry

ACL 2027 Under Review | 2025 – 2026

- First systematic characterisation of biomedical LM geometry against molecular similarity. Cross-family cosine similarity 0.02 despite identical inputs; SGA fine-tuning ceiling r=0.265. DDI F1 improved 14.2pp over frozen baselines.

TIDIER — Semantic Safety Margin for Text-to-Image Diffusion

AIISC, USC | NeurIPS Target | 2025 – Present

- Training-free safety via Semantic Safety Margin over a Multimodal Knowledge Graph; dual-space alignment (prompt-space repair + trajectory-space intervention). ~65% reduction in harmful outputs across four benchmarks; +9-10% relative CLIP score gain.

PUBLICATIONS AND PATENTS

Preprint

- Garg S, Shani C. When VLMs Destroy What They Know: Spatial Representations Fall Below Lexical Priors. *arXiv preprint arXiv:2610.13910*. 2026 Oct 9. **[Under review, NAACL 2027]**

Journal Articles

- Garg S, Sachdeva A, Veer Singh K, Jindal N, Kumar M, Singh M. Machine learning techniques for electric vehicles battery charging: a decade of revolutionary progress (2015–2025). *Ionics*. 2026 May 14:1–22. Springer Nature.
- Garg S, Sunkarapalli A, Ghosh D, Kaur S, Sukhija J. StrabNet-CQ: an integrated deep learning framework for automated strabismus classification and quantification using ocular landmark detection. *BMC Ophthalmology*. 2025 Oct 23;25(1):595.
- Garg S, Parakh R, Khandelwal J, Dhasaratahan C, Jindal N. Deep Learning–Based Stratification of Iron Overload in Thalassemia Patients. In: *Artificial Intelligence for Predictive Healthcare*. Auerbach Publications (Taylor & Francis); 2026. pp. 119–137.

Conference Papers

- Garg S, Pandey D, Rawal N, Khandelwal V, Sheth A. Answering Without Arriving: An Evidence Audit for Embodied Question Answering. *NeurIPS 2026 Workshop on Neuro-Symbolic Embodied Intelligence (NeMo)*. **Accept (Poster).** Avg. rating: 7.00 (min 6, max 8); avg. confidence: 3.33.
- Garg S et al. Do Embodied VLMs Actually Use Their Scene-Graph Memory? *CoRL 2026 Workshop on Scene and Situation Representations for Embodied Navigation (SISREN)*. **Accepted.**
- Osho O, Garg S, Chakraborty S, Das SK. PCIAFL: Personalized and Class Imbalance-Aware Federated Learning for Driver Behavior Classification. In: *Proceedings of the 27th International Conference on Distributed Computing and Networking (ICDCN 2026)*. 2026 Jan 6; pp. 128–132.
- Kaler P, Garg S, Jindal N. Beyond Peaks: Deep Temporal Models for Ocular Electrophysiology. In: *International Conference on Innovative Computing and Communication (ICICC 2026)*. Cham: Springer Nature Switzerland; 2026 Feb 6; pp. 132–140.
- Garg S, Kaler P, Samiksha S, Jindal N. AttentiveHybridNet: A CNN-Transformer Architecture with Cross-Attention for Robust Brain Tumor Classification from MRI Scans. In: *2025 2nd Global AI Summit–International Conference on Artificial Intelligence and Emerging Technology (AI Summit)*. IEEE; 2025 Nov 19; pp. 874–879.
- Garg S, Singh N, Kaler P, Jindal N. AutoML-Enhanced Smart Grid Energy Forecasting for IoT-Enabled Homes. In: *2025 IEEE International Conference on Intelligent Signal Processing and Effective Communication Technologies (INSPECT)*. IEEE; 2025 Nov 7; pp. 1–6.
- Garg P, Gupta B, Garg S, Jindal N, Mandia S, Singh K. A Scalable Ensemble Framework for Robust Image Steganography and Data Restoration Under Distortion Attacks. In: *2025 IEEE INSPECT*. IEEE; 2025 Nov 7; pp. 1–6.
- Garg S, Singh N, Jindal N, Singh M. SEFO-GB: Smart Energy Forecasting and Optimization for Green Buildings. In: *2025 IEEE 5th International Conference on Sustainable Energy and Future Electric Transportation (SEFET)*. IEEE; 2025 Jul 9; pp. 1–7.
- Gupta B, Arora M, Garg S, Ghosh D. Enhancing Strabismus Diagnosis from Detection to Classification with Deep Learning. In: *2025 3rd International Conference on Artificial Intelligence and Machine Learning Applications (AIMLA)*. IEEE; 2025 Apr 29; pp. 1–6.

Under Review

- Garg S. Federated Retinal Foundation Models: Cross-Population DR Grading Benchmark. *WACV 2027*. Under Review. First and sole author.
- Garg S, Pandey D, Khandelwal V, Rawal N, Sheth A. Room to Remember: Knowledge-Graph Memory for Long-Horizon Embodied Question Answering. *ICLR 2027 Conference*. Under Review. Submission #52972.
- Kaippilly Mana RP, Pawar AP, Shyalika C, Garimella R, Garg S, Sheth A. TIDIER: Text-to-Image Alignment via Knowledge-Guided Hate Detection, Interpretation, and Repair. *ICLR 2027 Conference*. Under Review. Submission #53252.
- Garg S et al. Geometry Without Structure: Biomedical LM Embedding Geometry. *ACL 2027*. Under Review.
- Garg S et al. Intermediate DR Grading via ONCL+CMAW. *AIM Journal*. Under Review.
- Garg S et al. The Cognitive Capability Spectrum of AI in Medicine. *IEEE JBHI*. Under Review.
- Garg S et al. DAFSeg: Dual-Stream Adaptive Fusion for Multitemporal Disaster Segmentation. *Scientific Reports*. Under Review.
- Garg S et al. UniCrisisNet: Multimodal Severity Estimation and Hazard Segmentation. *Pattern Recognition*. Under Review.

Patents

- A Method for Detection and Quantification of Strabismus Using Deep Learning Tools. Patent No. 202511042181. Issued Jul 4, 2025. [Published]
- A Novel AI-Assisted Framework for Smart Optical Glass Development. Patent No. 202511051934. Issued Jun 13, 2025. [Published]
- A Portable Histopathology Device, System and Method. Patent Application No. 202511127973. [Filed]

TALKS AND PRESENTATIONS

Invited Student Speaker

Global Resilience Summit 2026 | University of Colorado Colorado Springs / HealersHarbor International Conference 2026 | 2026

- Delivered invited talk on integrated, decision-aware AI systems in healthcare: uncertainty handling, population-scale deployment, and interpretability under distribution shift.

HONORS AND AWARDS

- Merit Scholarship, Thapar Institute — Rs. 1,41,000
- Top 67th of 500+ teams, NKSr Hackathon
- NeMo Workshop at NeurIPS 2026 — Average review score 7.0/10 (min 6, max 8)

TECHNICAL SKILLS

Clinical and Healthcare AI: Medical imaging (Grad-CAM, UMAP, segmentation), clinical NLP (BERTopic, BioBERT/SciBERT), federated learning, ARM edge inference, fairness evaluation, clinical deployment reliability

Neurosymbolic and Embodied AI: Knowledge graph construction (Hetionet, MMKG, Semantic Scene Graphs), KG-LLM cross-attention pipelines, EQA agents (AI2-Thor/ProcThor), long-range navigation, multimodal knowledge graphs, drug repurposing, FAISS, RAG

AI Safety and Evaluation Reliability: Activation steering, linear probing, causal intervention, probe task design, reliability diagnostics, counterfactual evaluation, distribution shift analysis, hallucination detection

LLMs and Alignment: HuggingFace Transformers, LoRA/PEFT, RAG pipelines, constrained decoding, responsible AI evaluation, Qwen3VL

Remote Sensing and Disaster Intelligence: Multitemporal semantic segmentation, change detection, BRIGHT benchmark, CrisisMMD, disaster damage assessment, DeepLabV3, SegFormer/DARFormer, class-weighted focal loss, Kruskal-Wallis ablation design

Multimodal and Vision: CLIP, ViT, SegFormer, multimodal fusion, contrastive learning, cross-domain robustness evaluation, text-to-image safety

Systems and Stack: PyTorch, TensorFlow, Scikit-learn, Bayesian inference, Kalman filters, FastAPI, Docker, AWS/GCP, Weights and Biases, LaTeX, exo distributed inference (contributor)

Languages: Python (Expert), C++, SQL, LaTeX

REFERENCES / RECOMMENDERS

Prof. Amit Sheth: Founding Director, IAIRO / AIISC, University of South Carolina — Neurosymbolic AI and embodied AI research

Dr. Debabrata Ghosh: Assistant Professor, DECE, Thapar Institute of Engineering and Technology — Clinical AI and medical imaging supervision

Dr. Chen Shani: Tel Aviv University (previously Stanford) — Co-author, VLM spatial probing and evaluation (arXiv:2610.13910)

Dr. Sahil Thakur: Mediwhale / National University of Singapore — Collaboration on intermediate DR grading (ONCL+CMAW)