Shubh Garg

+91 8171808091 | shubhgarg265@gmail.com | sgarg4_be22@thapar.edu | LinkedIn | GitHub

Vision–Language Models & NLP • Multimodal Reasoning & Generative AI • Interpretable & Alignment-Aware Learning • Contrastive & Representation Learning • Federated & Distributed AI Systems

Summary: AI researcher focusing on vision–language models, NLP, and interpretable multimodal systems. Experienced in generative reasoning pipelines, retrieval-augmented inference, and alignment-aware modeling for language and vision. Developed cross-modal architectures with contrastive and triplet learning, and deployed scalable models across cloud, edge, and federated environments. Author of 10+ peer-reviewed papers and 2 patents spanning multimodal learning, safety-aligned AI, and the study of discourse and knowledge flows across digital platforms.

Education

Thapar Institute of Engineering and Technology — Patiala,Punjab
Bachelor of Engineering in Electronics and Computer Engineering — *Oct. 2022 – June 2026*
- **CGPA**: 8.6

Neerja Modi School — Jaipur,Rajasthan
Grade XII — *April. 2020 – May 2021*
- **Percentage** : 94.4%

Experience

Research Intern — Quantitative ML — Apr 2025 – Jun 2025
AstratInvest — *Remote (Mumbai)*
- Engineered **LLM-augmented pipelines** for multimodal time-series signals (RSI/OFI), validating robustness with a **Sharpe ratio of 1.4** under noisy, real-world behavioral data.
- Built a **vectorized simulation framework** with latency, slippage, and turnover modeling, accelerating Monte Carlo experimentation by 40% for agent-based learning systems.
- Designed **volatility-gated sequential models** within distributed simulators, illustrating parallels with multi-agent decision-making and reinforcement learning.

AI Research Intern — Jan 2025 – May 2025
Indian Institute of Management (IIM) — *Udaipur*
- Implemented **Bayesian state-space models with Kalman filters** for stochastic sequence prediction, linking control theory with reinforcement learning.
- Developed **probabilistic forecasting pipelines** with Bernstein polynomial interpolation and uncertainty quantification for adaptive decision support.
- Prototyped **stochastic decision agents** validated via posterior scoring and causal inference, aligning with reinforcement learning and behavioral modeling.

Research Intern — Federated Multimodal Learning — May 2025 – Jun 2025
Ubisys Lab, IIT Jodhpur — *Jodhpur, India*
- Designed a **personalized CNN–LSTM with temporal attention** for behavioral sequence modeling, achieving **92.6% accuracy** under non-IID federated settings.
- Enhanced minority-class **F1-score by 18%** using cost-sensitive loss and adapter tuning, improving fairness in multimodal federated pipelines.
- Reduced **federated sync cost by 60%** via selective aggregation and GPU-accelerated inference, highlighting scalable distributed learning.

Data Science Intern — Jun 2025 – Jul 2025
Celebal Technologies — *Remote (Jaipur)*
- Built **cloud-native ML pipelines** (XGBoost + PyTorch) integrated with AWS S3/Lambda for hybrid batch + streaming workloads.
- Automated **CI/CD retraining and A/B testing**, reducing drift latency by 15% and strengthening adaptability in live systems.

Samsung PRISM Research Intern — Oct 2024 – Mar 2025
Samsung R&D — *Remote (Bengaluru)*
- Developed **RAG-based watermark detection engines** with CLIP and LLaMA, achieving 91%+ accuracy in multimodal vision–language inference—advancing **AI Security**.
- Optimized **transformer pipelines** through quantization and Triton-backed serving, cutting memory footprint and boosting throughput for edge-scale deployments.

Undergraduate Research Assistant — Feb 2024 – Present
Thapar Institute of Engineering & Technology (TIET) — *Patiala, India*
- Built **real-time multimodal forecasting models** on 500K+ samples with 97% test accuracy, leveraging PyTorch Lightning for distributed training.
- Engineered **lightweight edge-AI systems** for medical imaging, enabling ARM-based inference under 1 second—bridging accessibility in low-resource settings.
- Led **explainable multimodal ML research** (Grad-CAM, PCA/UMAP), advancing interpretability for vision-language and biomedical pipelines.

Patents

A Method for Detection and Quantification of Strabismus Using Deep Learning Tools	Patent Published
A Novel AI-Assisted Framework for Smart Optical Glass Development	Patent Published

Research & Publications

Accepted Peer-Reviewed Publications

• *Enhancing Strabismus Diagnosis from Detection to Classification with Deep Learning.* IEEE AIMLA 2025.
• *SEFO-GB: Smart Energy Forecasting and Optimization for Green Buildings.* IEEE SEFET 2025.
• *AutoML-Driven Smart Grid Energy Forecasting for IoT-Enabled Homes Using AutoGluon.* IEEE INSPECT 2025.
• *A Scalable Ensemble Framework for Robust Image Steganography: Neural and Traditional Methods Under Attack.* IEEE INSPECT 2025.
• *AttentiveHybridNet: A CNN–Transformer Architecture with Cross-Attention for Robust Brain Tumor Classification from MRI Scans.* IEEE AI SUMMIT 2025.
• *Stratification of Iron Overload in Thalassemia Patients.* CRC Press, Forthcoming
• *PCIAFL: Personalized and Class Imbalance-Aware Federated Learning for Driver Behavior Classification.* ICDCN 2026.
• *DisasterNet: Joint Learning of Tweet and Image Features for Damage Severity Classification.* CVIP 2025.
• *StrabNet-CQ: An Integrated Deep Learning Framework for Automated Strabismus Classification and Quantification Using Ocular Landmark Detection.* BMC Ophthalmology.

Preprints / Under Review

• *Neuromorphic Computing using AI: A Strategic Survey of the Last Decade (2015–2025).* Submitted to IEEE Access.
• *Machine Learning for Ultrasound Report Generation: A Decade Review of Techniques, Challenges, and Translational Potential in Low-Resource Settings.* Submitted to ACM CSUR.
• *Cognitive Computing in Healthcare Crisis Simulations: A Decade-Long Systematic Review (2015–2025).* Submitted to Scientific Reports
• *Can We Bridge Severity Classification and Hazard Segmentation for Real-World Disaster Response?* . In Preparation for CVPR 2026
• *Can We Disentangle Biomedical Embeddings? A Comparative Clustering Study of BioBERT and SciBERT* . In Preparation for EMNLP 2026

Projects

MediGlot 2.0: Biomedical RAG & Embedding Platform | *PyTorch, Transformers, FAISS, UMAP, HDBSCAN*

- Built a **retrieval-augmented generation (RAG) pipeline** with BioBERT/SciBERT + FAISS, applied to multilingual biomedical corpora—supporting analysis of how knowledge is accessed and communicated online.
- Implemented an **embedding audit framework** with coherence scores, HDBSCAN clustering, and UMAP visualizations—revealing semantic shifts and anomalies across language communities.
- Prototyped a **human-in-the-loop QA interface** for clinical discourse, generalizable to studying trust, alignment, and misinformation in health-related online behavior.

Achievements and Contributions

- **Merit Scholarship Awardee** — Awarded Rs 1,41,000 for academic excellence in 2022–23 at Thapar Institute of Engineering and Technology.
- **Student Placement Representative**, TIET — Spearheaded coordination between 100+ students and 20+ companies during campus placement season.
- **Hacktoberfest 2024** — Successfully completed 4 PRs in key OSS projects, showcasing team-based development and open collaboration skills.
- **Top 100 (67th) Rank**, NKSr Hackathon — Developed a prototype under constrained time and data, outperforming 500+ teams in a national innovation sprint.

Skills

Core ML & Programming: Python, C++, PyTorch, TensorFlow, NumPy, Pandas, Hugging Face, Scikit-learn
Language & Multimodal AI: LLMs (GPT, LLaMA), VLMs (CLIP, ViT), Retrieval-Augmented Generation (RAG), Diffusion Models, Cross-Modal Fusion, Contrastive & Triplet Learning
NLP & Representation Learning: Text Mining, Biomedical Embeddings (BioBERT, SciBERT), Word2Vec, FastText, Topic Modeling (LDA), Graph-based Analysis (GraphSAGE, NetworkX)
Responsible & Distributed AI: Federated Learning, Explainability (Grad-CAM, SHAP), Bayesian Inference, AutoML, Model Compression (Quantization, Pruning)
Deployment & Systems: FastAPI, Docker, FAISS, REST APIs, AWS, GCP, Real-Time Edge/Cloud Pipelines
Research Tooling: LaTeX, Weights & Biases, UMAP, HDBSCAN, Kalman Filters, Probabilistic Programming (Pyro, NumPyro)