Gaurav Hadavale

hadavalegaurav@gmail.com | LinkedIn | GitHub

EDUCATION

Sardar Patel Institute of Technology (University Of Mumbai) Mumbai, India
B.Tech in Electronics & Telecommunication Engineering Aug 2023 – June 2027 (Expected)

- **Minor in AI & Machine Learning:** CGPA **9.0/10** (Major CGPA: 7.58/10)
- **Relevant Coursework:** Deep Learning, Medical Image Analysis, Computer Vision, Linear Algebra, Probability & Statistics.

TECHNICAL SKILLS

- **Explainable AI (XAI):** Mechanistic Interpretability, Activation Steering, Causal Interventions, Grad-CAM/++, RISE, SHAP, Concept Bottlenecks, Shortcut Learning Diagnosis.
- **Data Analysis & Statistics:** Statistical Modeling, Hypothesis Testing, A/B Testing, Distribution Analysis, Youden's Index, Spectral Analysis, Experimental Design, Variance Analysis.
- **Computer Vision:** Vision Transformers (ViT), CNNs (DenseNet, ResNet, ConvNeXt), Frequency Analysis (FFT), Medical Image Analysis, Image Forensics, Saliency Mapping.
- **Deep Learning:** Variational Autoencoders (VAE), Self-Supervised Learning, Random Forests, XGBoost, SVM, PCA, Clustering, Transfer Learning, Optimization Algorithms.
- **NLP:** Transformers (BERT, GPT-2, T5), LLMs, Attention Mechanisms, Residual Stream Analysis.
- **Languages/Tools:** Python, Julia, SQL, C++, PyTorch, TransformerLens, Captum, OpenCV, NumPy, Pandas, SciPy, Scikit-learn.

RESEARCH PROJECTS

Beyond Accuracy: An Interpretability-Driven Audit of Deep Learning Models for Pneumonia Detection from Chest X-Rays *PyTorch, RISE*

- Audited SOTA DenseNet121 models to diagnose "Clever Hans" artifacts; revealed predictions statistically correlated with radiographic text markers rather than pathology.
- Engineered a **statistical audit pipeline** using RISE and Grad-CAM++ to analyze High-Confidence False Positives, identifying a critical failure mode (Specificity: 0.49).
- Derived a targeted data-cleaning protocol based on audit insights and calibrated decision thresholds using **Youden's Index**, restoring Specificity to 0.8675.
- **Publication:** "Beyond Accuracy: An Interpretability-Driven Audit..." **Abstract Submitted to CARS / ECR 2026**.
  [Read on TechRxiv] | *[Code on GitHub]*

Causal Auditing of Latent Affect in Language Models *Python, TransformerLens*

- Conducted a mechanistic interpretability study on GPT-2 to isolate latent "anger" representations in the residual stream (Layer 8) using **Contrastive Activation Analysis**.
- Performed **inference-time activation steering** to causally modulate discourse metrics (sentence length, negation rate) without retraining the model.
- Designed a statistical validation framework using **norm-matched negative control vectors**, demonstrating that arbitrary perturbations induce incoherent noise while concept-aligned steering produces systematic structural shifts.
- Analyzed distribution shifts via box-plots/variance analysis, proving latent interventions affect expressive structure even when standard emotion classifiers saturate.

- *[Preprint Available]*

Deepfake Forensic Detection using Frequency Analysis *PyTorch, Fourier Transform*

- Developed a forensic detection pipeline using ConvNeXt-Base on the CIFAKE dataset, achieving 97% validation accuracy via Test-Time Augmentation (TTA).
- Conducted **Spectral Analysis (FFT)** combined with Score-CAM to visualize frequency domain discrepancies, proving the model detected high-frequency GAN artifacts invisible to the human eye.

Latent-Space Counterfactual Explanations *PyTorch, VAE*

- Developed a Variational Autoencoder (VAE) to analyze the latent manifold of risk data.
- Formulated a latent-space optimization objective to generate minimal counterfactual perturbations, quantifying the exact feature changes required to reverse a classifier's decision.

EXPERIENCE

College Mini Project | Prof. Milind Paraye Mumbai, India
Smart Helmet for Crash Detection (IoT & Edge ML) Aug 2025 – Nov 2025

- Deployed a lightweight Edge ML model on an STM32 microcontroller to classify crash severity from real-time accelerometer data.

LEADERSHIP & ACHIEVEMENTS

- **Amazon ML Summer School:** Selected for a competitive nationwide mentorship program focused on applied Deep Learning and industry best practices.
- **Technical Lead, IEEE AESS:** Orchestrated technical workshops on aerospace systems; led a team of 10+ students to manage event logistics and technical demos.
- **Campus Ambassador, IIT Bombay Techfest:** Spearheaded outreach and logistics for Asia's largest science and technology festival.
- **Competitive Exams: JEE Advanced Qualified** (Top 2.5% nationwide); **JEE Mains:** 96.34%ile.