

Akshith Acharya

@ acharyaakshith@gmail.com | [LinkedIn](#) | [GitHub](#) | [Website](#) | [London, UK](#)

AI researcher (PhD) working on **robustness under distribution shift, shortcut learning, interpretability-driven auditing, and fairness**. 3.5+ years industrial AI research experience.

EDUCATION

King's College London

PhD (UKRI/DRIVE Health CDT), Machine Learning

London, UK

Oct 2024 – Present

- Research on shortcut learning, auditing under shift, and bias mitigation; building evaluation/monitoring for reliable deployment.

Birla Institute of Technology and Science (BITS Pilani)

M.Sc. Mathematics

Pilani, India

Aug 2016 – Jul 2020

- Training in probability, optimisation, and statistical inference; thesis awarded highest grade.

SELECTED PUBLICATIONS

- **Acharya, A et al.** *Multi-Way Representation Alignment*. ReAlign @ ICLR (2026).
- **Acharya, A et al.** *Right Regions, Wrong Labels: Semantic Label Flips in Segmentation under Correlation Shift*. CAO @ ICLR (2026).
- **Acharya, A et al.** *Localising Shortcut Learning in Pixel Space via Ordinal Scoring Correlations for Attribution Representations (OSCAR)*. Preprint.
- **Acharya, A. et al.** *Invisible Attributes, Visible Biases: Exploring Demographic Shortcuts in MRI-based Alzheimer's Disease Classification*. FAIMI @ MICCAI (2025).
- **Acharya, A et al.** *Watching the AI Watchdogs: A Fairness and Robustness Analysis of AI Safety Moderation Classifiers*. NAACL (2025).
- **Acharya, A et al.** *Efficient Biomedical Entity Linking: Clinical Text Standardization with Low-Resource Techniques*. BioNLP @ ACL (2024).

PATENTS (US APPLICATIONS)

- **Acharya, A., et al.** *Automated clinical trial matching system*. US Patent App. 18/952,925 (2025).
- Sasidharan, S., **Acharya, A., et al.** *Systems and methods for intelligent medical editors*. US Patent App. 18/906,528 (2025).
- Ghose, S., Sasidharan, S., Cho, S., **Acharya, A., et al.** *Natural language based clinical recommendation system*. US Patent App. 18/468,593 (2025).

EXPERIENCE

PhD Researcher (Robustness & Responsible AI)

King's College London

Oct 2024 – Present

- Mechanism-aware auditing under spurious correlations: explanation-derived statistics and spatial analyses to identify failure patterns beyond output uncertainty.
- Shift diagnostics for segmentation and structured prediction; emphasis on failure mode characterization and monitoring signals.
- Bias mitigation and evaluation: group-robust objectives (e.g., GroupDRO-style) and worst-group risk concentration analysis.

Research Engineer (Applied AI Research)

GE Research / GE Healthcare

Bangalore, India

Apr 2021 – Jul 2024

- **Vision & structure monitoring:** developed ML methods for reliability/quality monitoring in imaging workflows (CT/X-ray), including failure-mode flagging and risk scoring from structured signals and model outputs.

- **Runtime AI / machine-event monitoring:** analysed real-world machine event logs and telemetry to detect and predict potential faults; developed learning-based triage and early-warning signals, including graph-based methods (e.g., **GNNs**) over event relations.
- **Clinical NLP (oncology portfolio):** built and evaluated pipelines for information extraction, normalisation, and summarisation of patient reports; performed clinician-in-the-loop error analysis to address high-impact mistakes (e.g., negation, temporality, fine-grained qualifiers).
- **Clinical trial matching (research lead):** led development of an end-to-end matching system (trial ingestion, eligibility criteria structuring, patient entity linking, matching/ranking), with emphasis on error analysis and robustness of matching decisions.
- **Leadership & external engagement:** represented GE Healthcare as part of the core group at **G20** events (Goa and Gandhinagar, India), presenting AI initiatives and engaging with cross-stakeholder audiences.

ADDITIONAL RESEARCH & OPEN SOURCE

- **Google Summer of Code (2020):** Developed [rminizinc](#) (R interface to MiniZinc); solver backends via Rcpp.
- **TUM (2019):** Built [mirtargetbenchmark](#) to benchmark miRNA target prediction tools using expression-derived ground truth.
- **Machine Vision Lab, CSIR-CEERI (2019):** Developed an RGB-D respiratory pattern classifier (VGG-16; ImageNet init) to support breath-aware X-ray timing and reduce unnecessary exposure to healthy tissue; collected sensor data, built preprocessing, and validated the approach experimentally.

SKILLS

Programming: Python, C/C++, Java, R, SQL, Git, Docker

ML/DS: PyTorch, PyTorch Lightning, TensorFlow, Scikit-learn; evaluation under shift; interpretability tooling

Topics: Robustness & distribution shift, shortcut learning, responsible AI, NLP for clinical text

Coursework: Measure-Theoretic Probability, Statistical Inference, Geometric Deep Learning

Languages: Hindi (Native), English (Professional)

SERVICE

- **Reviewer:** NeurIPS 2024, ICML 2025; **Top Reviewer** (NeurIPS 2025), AISTATS 2026.