

Aveen Dayal

AI Scientist | Multimodal and Generative AI | ML Systems and Evaluation

aveendayal97@gmail.com | +91 9650150648 | LinkedIn | Research Website | GitHub | Google Scholar

Professional Summary

AI researcher with a Ph.D. in Artificial Intelligence from IIT Hyderabad, where I was a Prime Minister's Research Fellow and received the IIT Hyderabad Excellence in Research Award. I have 6+ years of hands-on AI R&D experience across industry and academia, including experience at Dolby ATG, Microsoft Research, and Adobe Research. Expertise spans multimodal and generative AI, efficient inference, out-of-distribution (OOD) robustness, content protection and attribution, and model evaluation, with experience building end-to-end ML systems from problem formulation and data design to model development, benchmarking, optimization, and edge deployment. Published at top venues like, CVPR, WACV, ECCV, NeurIPS, and IEEE Transactions on Image Processing.

Research and Industry Experience

Senior Multimodal AI Researcher

Sep. 2025 – Present

Dolby Advanced Technology Group India

Bengaluru, India

- Addressing piracy in streaming media content by developing watermarking methods across offline and online setting; contributing to an invention under preparation for patent filing.
- Architecting a human-aligned evaluation framework for generated multimodal content across **5 open-source models** and **20+ quality dimensions**, by leveraging annotated benchmarks to validate model-generated scores.
- Defined and executed a content-attribution study for text-to-image diffusion models, comparing SAM3, DINO, and CLIP-based approaches on Stable Diffusion 1.5 and SDXL with precision-recall evaluation.

Research Intern

Jun. 2025 – Aug. 2025

Adobe Research India

Bengaluru, India

- To make poster generation and editing more controllable for creator workflows with potential application to **Adobe Express**, built an interactive layer-aware pipeline in which a multimodal LLM interpreted user intent and existing design elements before applying localized edits.
- Demonstrated the proof of concept through an end-to-end interactive demo and side-by-side evaluation against GPT-4o. On the evaluated posters, it showed stronger instruction adherence and fewer unintended changes.

Research Intern

Jul. 2024 – Dec. 2024

Microsoft Research India

Bengaluru, India

- To address LLM inference latency, designed and benchmarked a tree-based speculative-decoding method for Vicuna-7B and Llama-3-Instruct-8B on an NVIDIA A100 80GB GPU, achieving **2.91×** and **2.62×** speedups.
- Built a benchmark against popular baselines: EAGLE and EAGLE-2; matched EAGLE's Vicuna-7B speedup and improved accepted length on Vicuna-7B (4.08 vs. 3.94) and Llama-3-Instruct-8B (4.12 vs. 3.65), while characterizing throughput-acceptance trade-offs.
- Designed a distribution shift evaluation study for Phi-3.5-Vision and LLaVA-OneVision across **8 domains** from OfficeHome and TerraIncognita. Quantified deployment-relevant degradation and tested prompt optimization and adversarial LoRA techniques and identified the limits of these methods under domain shift.

Research Intern

Mar. 2023 – Nov. 2023

Autonomous and Cyber-Physical Systems Lab, University of Agder

Grimstad, Norway

- Addressed model failure under unseen environments and heterogeneous data sources by translating theoretical error bounds into trainable adversarial algorithms for domain generalization and multi-source adaptation. Built robust methods and evaluated across five DG and five MDA benchmarks, resulting in first-author publications at **NeurIPS 2023** and **IEEE TIP 2025**.

Visiting Researcher

Jan. 2020 – Jul. 2021

Autonomous and Cyber-Physical Systems Lab, University of Agder

Grimstad, Norway

- Co-developed **six published end-to-end ML systems** spanning sensor data acquisition, dataset curation, lightweight model design, benchmarking, and deployment for biomedical imaging, contactless interfaces, and autonomous perception.
- Converted, profiled, and deployed models on Raspberry Pi 4B, and NVIDIA Jetson AGX Xavier using TensorFlow Lite and ONNX, measuring accuracy, latency, model size, and memory under resource constraints.

Education

Indian Institute of Technology Hyderabad

Aug. 2021 – Dec. 2025

Ph.D. in Artificial Intelligence, GPA: 9.19/10

Hyderabad, India

Prime Minister's Research Fellow (PMRF); Excellence in Research Award, 2024

BML Munjal University

May 2020

B.Tech. in Computer Science and Engineering, GPA: 9.44/10

Gurgaon, India

Selected Publications

- **A. Dayal** et al., “Bridging the domain gap in small multimodal models: A dual-level alignment perspective,” *IEEE/CVF WACV*, 2026.
- S. VCR, R. Lalla, **A. Dayal** et al., “Foundation model priors enhance object focus in feature space for source-free object detection,” *IEEE/CVF CVPR*, 2026.
- **A. Dayal** et al., “Leveraging mixture alignment for multi-source domain adaptation,” *IEEE TIP*, 2025.
- **A. Dayal** et al., “Improving unsupervised domain adaptation: A pseudo-candidate set approach,” *ECCV*, 2024.
- **A. Dayal** et al., “MADG: Margin-based adversarial learning for domain generalization,” *NeurIPS*, 2023.
- **A. Dayal** et al., “Adversarial unsupervised domain adaptation for hand gesture recognition using thermal images,” *IEEE Sensors Journal*, 2023.
- **A. Dayal**, L. R. Cenkeramaddi, and A. Jha, “Reward criteria impact on the performance of reinforcement learning agent for autonomous navigation,” *Applied Soft Computing*, 2022.
- N. Paluru, **A. Dayal** et al., “Anam-Net: Anamorphic depth embedding-based lightweight CNN for segmentation of anomalies in COVID-19 chest CT images,” *IEEE TNNLS*, 2021.

Selected End-to-End ML Systems

Label-Efficient Adaptation of Cloud-Deployed Multimodal Language Moels *Multimodal AI / WACV 2026*

- Addressed a practical deployment challenge in which multimodal models lose accuracy when applied to new image and question distributions. Across 12 cross-domain VQA settings, LLaVA-OneVision model achieved **54.9%** average accuracy, compared with **76.8%** when retrained using fully labeled data (cost-intensive and infeasible) from each target.
- Developed DuAA, a two-stage adaptation approach that combines existing labeled data with unlabeled target examples, allowing the model to adjust both visual understanding and image–language reasoning without requiring a newly labeled dataset for every deployment environment.
- Improved average accuracy to **57.9%**, closing **15.5% of the gap** to the fully supervised oracle. The adaptation required a one-time **8 GPU-hour** run on an NVIDIA A100 40 GB while preserving inference at **0.50 s per sample**, adding no recurring latency or per-request compute overhead.

Contactless Gesture Recognition and Authentication

Edge AI / Four Published Studies

- Developed the following edge-AI pipelines for reliable contactless interaction under poor or absent lighting: **(1)** created a dataset of 3,200 low-resolution thermal images of sign-digit data and deployed a lightweight CNN on Raspberry Pi, achieving **99.52% test accuracy** with **30 ms inference per image**; **(2)** extended the system to high-resolution thermal sensing using 14,400 images from 30 participants and deployed a 2 MB TensorFlow Lite model on NVIDIA Jetson AGX, achieving **98.81% accuracy** with **75.1 ms inference**.
- Translated the recognition capability into practical applications by **(3)** building a Raspberry Pi-based prototype that captured sign-language digits from a live camera feed and assembled them into an n -digit contactless authentication code, using a 0.97M-parameter, 4 MB TensorFlow Lite model with **99.1% classification accuracy** and **280 ms inference per digit**; and **(4)** addressing the scarcity of labeled thermal data through adversarial RGB-to-thermal domain adaptation, achieving **91.32%** and **80.91% target accuracy** on sign-digit and alphabet-gesture recognition, respectively.

Deployment Optimization for Point-of-Care Medical Imaging

Healthcare AI / IEEE TNNLS and TUFFC

- Optimized two medical-imaging models for resource-constrained deployment: Anam-Net for CT lung-abnormality segmentation and Mini-COVIDNet for ultrasound-based classification of COVID-19, pneumonia, and healthy cases. Benchmarked mobile and embedded execution across Raspberry Pi, NVIDIA Jetson, and Android platforms.
- Anam-Net used **7.8× fewer parameters than U-Net** and reduced Raspberry Pi inference time from **43.3 s to 23.3 s**; Mini-COVIDNet achieved **83.2% accuracy** with a **51.29 MB footprint**, **4.39× fewer parameters** than the closest-performing baseline, and under 30 minutes of Jetson training time.

mmWave Radar Target Classification

ADAS and traffic management / Electronics 2021

- Co-developed an outdoor measurement-to-model pipeline that removed radar clutter, extracted statistical features from range FFT peaks, and compared logistic regression, Naive Bayes, SVM, and LightGBM for car, pedestrian, and UAV classification. LightGBM achieved the best performance of **95.6% accuracy**.

Technical Skills

Programming and ML: Python, PyTorch, TensorFlow, TensorFlow Lite, scikit-learn, LightGBM, NumPy, OpenCV

Applied ML: LLMs and multimodal models, diffusion models, speculative decoding, LoRA/parameter-efficient fine-tuning, domain adaptation/generalization, computer vision, sensor fusion, reinforcement learning

ML Systems and Evaluation: Distributed training/inference, model profiling and optimization, ONNX, APIs and model serving, Linux, Git, experiment tracking, OOD evaluation, precision–recall analysis, human-aligned benchmarking