Muhammad Haseeb

haseeb099m@gmail.com | github.com/ha405 | linkedin.com/in/muhammad-haseeb

RESEARCH INTERESTS

Mechanistic Interpretability: Analyzing internal circuit evolution for domain generalization and to understand representational collapse in Federated Learning.

Domain Generalization: Adapting Models to distribution shifts through pruning to identify sparse subnetworks that preserve domain-invariant representations.

Efficient ML: Compression methods (pruning, quantization) to save compute and memory for LLMs without loss in performance.

EDUCATION

Lahore University of Management Sciences (LUMS) Sep 2022 – May 2026
B.S. Computer Science
Relevant Coursework: Machine Learning, Deep Learning, Computer Vision, AI on Edge Devices, Advanced Topics in ML, LLM Systems, Linear Algebra, Probability.

PUBLICATIONS

– **BaCP: Backbone Contrastive Pruning for Preserving Representations in Extremely Sparse Neural Networks (Submitted to *AAAI 2025*)**.
Mohammad Haroon Khawaja, **Muhammad Haseeb**, Mohammad Fatim Shoaib, Muhammad Tahir.

RESEARCH EXPERIENCE

ML Research Assistant Jan 2025 – Present
Centre for Urban Informatics, Technology and Policy (CITY), LUMS
Advisors: Dr. Muhammad Tahir, Dr. Zubair Khalid

– **Backbone Contrastive Pruning (BaCP):**
  ◦ Contributed to a generalized pruning framework to merge standard pruning criteria with contrastive learning to prevent representational collapse.
  ◦ Designed a **multi-objective loss function** that aligns sparse embeddings with pretrained, fine-tuned, and historical snapshot references to maintain feature consistency.
  ◦ Maintained accuracy comparable to dense baselines at up to **99% sparsity**, outperforming standard unstructured pruning approaches across Vits/CNNS/Language Models.
– **Mechanistic Analysis of Federated Learning:** (*Actionable Interpretability ICML 2026, Target Submission*)
  ◦ Studied FedAvg through lens of Interpretability to analyze the performance loss under Non-IID conditions. We used circuit discovery, sparse autoencoders (SAEs), and linear probes to examine client and global representations.
  ◦ Compared class-specific circuits across clients and the global model and observed circuit drift and interference patterns during aggregation in non-iid conditions while iid circuits behaved in a perfect constructive manner.
  ◦ Used Universal SAE-based concept analysis and linear probing to show that useful latent features similar to iid-model remain present even when global accuracy degrades under non-iid conditions.
– **Domain Generalization & Interpretability:**
  ◦ Working on reliable AI by constructing models from understood, generalizable building blocks (circuits) rather than treating them as black boxes.
  ◦ Decomposing deep neural networks into constituent circuits to investigate whether domain shift failures arise from core pathways relying on shortcuts or from noise in non-essential pathways disrupting performance.
  ◦ Exploring the use of cross-domain data as a diagnostic probe to reveal generalizable, core circuits within a standard pre-trained model.
  ◦ Previously, developed a framework using visual queries in ViTs trained via RL (group relative query optimization) to improve domain generalization, achieving a **3% accuracy gain** on the PACS dataset compared

to empirical risk minimization.

– **Quantization of Diffusion Models:**
  - Proposed **Log-SNR Conditioned Temporal Dynamic Quantization** to handle non-stationary activations, achieving a **55% reduction in LPIPS** on PixArt-α by conditioning predictor networks on physical signal dynamics..
  - Implemented Hessian-based optimization (GPTQ and QroNos) for **4-bit weight compression**, utilizing second-order curvature information to mitigate quantization errors.

AI/SWE Research Intern May 2025 – Aug 2025
University of Illinois Urbana-Champaign (UIUC)

– Investigated LLM-based heuristics for automated `#ifdef` guard insertion in C codebases.
– Explored the use of large language models for software debloating and code dependency analysis.
– Contributed to the development of a VS Code extension supporting a C-to-Rust translation pipeline.

ML Research Assistant Jan 2025 – Aug 2025
Computer Vision & Graphics Lab, LUMS

– **KL Aware Quantization (KLAWQ):**
  - Proposed an augmented GPTQ framework that integrates KL divergence and second-order curvature information to optimize a combined MSE+KL+CE objective, achieving a **30% reduction in perplexity** over standard baselines in LLMs.
– **Single-Image Camera Calibration (SOFI-UGCL):**
  - Developed a hybrid method combining a Multi-Scale Deformable Transformer with geometric post-processing to recover full camera projection matrices (intrinsics and extrinsics) from single image.

INDUSTRY EXPERIENCE

OSS Engineer Dec 2025 – Feb 2026
DatacurveAI (YC S24)

– Solved bugs and added new features across various open-source machine learning repositories to test and train LLMs on working with large codebases.

Machine Learning Engineer Jul 2025 – Oct 2025
Innova Tech

– Automated data annotation pipeline, refined training configurations, and fixed model architectures getting a **4% accuracy gain**. Also, implemented inference optimization with TensorRT for speedup and various quantization techniques to reduce memory footprint by **40%** on edge devices.

TEACHING EXPERIENCE

Teaching Assistant Sep 2025 – Dec 2025
CS436: Computer Vision, LUMS

– Designed and supervised assignment on transfer learning and multithreaded C++ object detection pipeline. Also, mentored 40 student groups in building a virtual tour application using Structure from Motion.

SKILLS & OPEN-SOURCE

Technical Skills: Python, C, C++, Rust,SQL; PyTorch, Transformers, ONNX, TensorRT, diffusers, adapters.

Open-Source Contributions:

- **pytorch-image-models:** Added F1, precision, and recall metrics to evaluation and training pipelines for both single-GPU and distributed setups.
- **adapters:** Implemented PEFT support for Group Query Attention models, fixing tensor mismatch issues.