

Aman Singh

Curriculum Vitae · Machine Learning Engineer & Researcher

2amansingh2@gmail.com | github.com/amanyagami | linkedin.com/in/amanyagami | aman-ai-portfolio.netlify.app

Summary

Machine learning engineer and researcher working on reliable AI systems. My work examines model behavior under adversarial and out-of-distribution inputs, representation-level signals of failure, scalable inference, and practical methods for dependable evaluation and deployment. [Education](#)

M.S., Computer Engineering [Arizona State University](#) · **Jan 2024** Graduate work in scalable AI deployment, model robustness, generative AI, and large language models.

B.E., Electrical and Electronics Engineering [National Institute of Technology Karnataka](#) · **2018 – 2022**
[Research Interests](#)

Reliable Machine Learning Adversarial robustness, out-of-distribution detection, feature-space modeling, dependable inference, model evaluation, and systems for reproducible ML research.

Research Experience

Graduate Researcher [MPS Lab at Arizona State University](#) · [Tempe, AZ](#) · **Mar 2024 – May 2026**

- Led research on machine-learning robustness, adversarial evaluation, and out-of-distribution detection.
- Developed high-performance GPU training and inference systems with sub-10 ms latency on NVIDIA A6000 and A100 GPUs.
- Fine-tuned modern vision models on CIFAR-100 to 92.55% validation accuracy with H100 GPUs.
- Collaborated with Yonsei University on two ESWEEK 2026 submissions.

Industry Experience

Software Development Engineer [Amazon](#) · [India](#) · **Jun 2022 – Mar 2023**

- Owned peak-planning systems and reliability improvements for high-volume infrastructure.
- Reduced operational costs by 30% while supporting a $1.8\times$ increase in client traffic during peak periods.
- Improved deployment processes through CI/CD, reducing incidents by 60%.

Selected Research Software

DrDNA — Silent Data Corruption Detection Post-hoc detection and mitigation framework for soft errors in deep networks using layer-wise activation statistics.

Feature-Space Geometry Analysis of deep feature spaces for class separability, representation collapse, and adversarial robustness.

ViT Fine-Tuning Benchmarks Reproducible fine-tuning benchmarks for vision transformers and modern CNNs on CIFAR-100.

Nandi — SLM Research Assistant Comparative PDF question answering with direct prompting and retrieval-augmented generation.

Make Presentation Simple Serverless multimodal pipeline for turning PDFs into presentation decks.

Open-Source Contributions

PyTorch torchao **PyTorch** Contributed fixes across quantization primitives, FP8 training kernels, and quantization-aware training behavior.

LLVM / MLIR **LLVM** Contributed compiler-correctness fixes for dialect transforms, lifetime handling, and malformed-input diagnostics.

ML Systems **Open source** Contributions to SGLang, VeRL, Megatron-LM, MLX-LM, and arapuca across inference, distributed training, checkpointing, and sandboxing.

Technical Areas

RAG, LLM inference, PyTorch, TensorFlow, Hugging Face, CUDA, Google Cloud Platform, Kubernetes, AWS, distributed training, mixed precision, GPU inference, evaluation and benchmarking, Docker, Python, C++, Java, Go, C#, and Rust.