Achyuthan Sivasankar

achyuthan.sivasankar@gmail.com | linkedin.com/in/achyuthan-s | github.com/achyuthan-s | achyuthan.dev | New York, NY

ML Research • Adaptive Computation, Sparse Routing & Mechanistic Interpretability • MoE / KAN Systems • NYU MS CS (2025–2027)

Education

New York University | M.S. Computer Science, GPA 4.0 — Aug 2025 – May 2027
Vellore Institute of Technology | B.Tech Computer Science, GPA 3.92 / 4.0 — Aug 2021 – May 2025

Technical Skills

ML / Research: PyTorch, JAX, HuggingFace, MoE & KAN architectures, LoRA/PEFT, RL (REINFORCE, Meta-World), mechanistic interpretability, self-supervised learning, RAG
Languages & Infra: Python (expert), C/C++, Go, SQL, Bash; Docker, FastAPI, AWS (S3, EC2), Weights & Biases, Git, Linux

Research

Circuit Synchronization Precedes Generalization: Causal Evidence from Fourier Structure in Grokking Transformers *First author • arXiv:2606.12966*

- Introduced the **Frequency Synchronization Degree (FSD)**, a permutation-tested metric that predicts grokking 500–3,000 steps in advance across nine configurations (mean lead +1,722; sign-test $p \approx 0.004$), beating Nanda et al.'s excluded-loss baseline in all nine.
- Causal weight-decay intervention isolates the regularisation rate as the sole control on the generalisation delay: inverse-λ timing law $\Delta t \propto 1/\lambda$ ($R^2 = 1.00$), replicated across three primes.

KAN-Multi: Adaptive Multi-Basis Routing with Emergent Specialization *First author • in preparation*

- Statistics-driven routing layer selecting among 6 function bases (Fourier, Wavelet, Chebyshev, RBF, Rational, Sigmoid) with zero routing supervision; emergent specialization measured via per-layer Shannon entropy.
- +6.8% over MLP baseline on CIFAR-100; a task-complexity scaling law ($r = 0.972$) predicts *when* adaptive routing helps — characterising regime boundaries rather than chasing SOTA.

MoE-Bench: Open Benchmark for Expert Collapse & Routing in Sparse MoE LLMs *First author • open-source toolkit + paper* • GitHub

- Pip-installable toolkit for routing entropy, expert utilisation, and collapse across OLMoE, JetMoE, and Qwen1.5-MoE; found 50% layer-collapse on OLMoE (math) vs. 0% on the others — collapse is architecture-dependent, not universal.

AutoMoE: Meta-Learning Expert Topologies for Multi-Task Robotic Manipulation *Research project*

- Evolutionary REINFORCE meta-learner over discrete MoE topologies (Meta-World ML10); discovered topology achieves 100% task success at **2× parameter efficiency** vs. a fixed hierarchical MoE.

Experience

Research Assistant — Prof. Anna Choromanska's Lab, NYU Tandon — May 2026 – Present

- Building **AD-LiST-JEPA**, a Joint-Embedding Predictive Architecture for automotive LiDAR perception and occupancy completion/forecasting (OCF) on the Waymo Open Dataset.
- Completed the full self-supervised pretraining pipeline (30 epochs); running downstream OCF finetuning toward JEPA-based world-model evaluation.

Indian Institute of Science (IISc) — Research Intern — Bangalore, India — Jun 2024 – Aug 2024

- GNN-based graph-theoretic optimisation on large-scale RF mesh networks (1k+ nodes); routing efficiency improved 42% over the shortest-path baseline.

National University of Singapore (NUS) — AI Research Intern — Singapore — Dec 2023 – Feb 2024

- Real-time multimodal stress-detection system; 99.67% accuracy on SWELL-KW (prior SOTA ~97%), with a multi-threaded architecture cutting inference latency 35%.

Selected Project

ACGA: Adaptive Corrective Graph-Augmented RAG Python, FastAPI, Neo4j, ChromaDB, Docker • GitHub
Query router selects among vector, graph, and keyword retrieval per query type; 4-layer memory; 85ms latency, 15.5× cached speedup; MIT-licensed.

Earlier Publications & Patent

- **IEEE (×3, 2024)** — video steganography (AES-256/RSA-4096); CNN-based acoustic event detection (MFCC); retail ML demand forecasting.
- **Patent (2025)** — "ML-Enhanced Intrusion Detection for Blackhole Attacks in IoT RPL Networks," VIT IPRTT Cell, India.