Nilesh Sarkar

AI Researcher | Knowledge Distillation, Mechanistic Interpretability, Agentic Systems & Edge Deployment

nileshsarkar.cs@gmail.com | +91-9748480484 | linkedin.com/in/nileshsarkar-ai | github.com/nileshsarkar-ai | nileshsarkar-ai.github.io

SKILLS

ML / DL	PyTorch, TensorFlow, CNNs, Transformers, Diffusion Models, Reinforcement Learning, Representation Learning
LLM & NLP	RAG, LangChain, LangGraph, n8n, Hugging Face, AI Agents, Vector DBs, LoRA, QLoRA, Quantization, Knowledge Distillation, Mechanistic Interpretability
Computer Vision	OpenCV, Object Detection, Depth Estimation, Gesture Recognition
Robotics & Edge	ROS2, Gazebo, Arduino, Raspberry Pi, NVIDIA Jetson Nano, LiDAR, IMU, FPGA, SLAM (coursework)
Tools & Cloud	Git, Docker, GCP, Azure AI, AWS, MCP (Model Context Protocol), MATLAB, Jira
Languages	Python (expert), C++, C; English (fluent), Hindi (native), Kannada (conversational)

RESEARCH EXPERIENCE

Founding Researcher *Mar 2026 – Present*

Erdős AI Lab | Student-founded research lab, incubated at IIT Bombay

- Researching **knowledge distillation (KD)** pipelines for LLMs, compressing teacher models into efficient students; investigating **mechanistic interpretability**: attention patterns and circuit-level behaviour in transformers. Paper submitted to COLM 2026 (under review).
- Exploring **representation learning** for cross-domain transfer and **continual learning** for sequential task acquisition without catastrophic forgetting.

AI & Deep Learning Researcher *Jun 2025 – Apr 2026*

Moog India Technology Centre (Moog Controls), Bengaluru | Internship extended

- Deployed air-gapped **LangGraph/LangChain agentic RAG** pipelines for aerospace engineering; retrieval accuracy **70% → 90%+**; electronics department suite: 55% productivity boost, 94% adoption (100+ users), 99.3%+ reliability across 3–4 agents.
- Built **human-in-the-loop multimodal document intelligence agents** for electronics engineering: autonomously parse engineering diagrams, embedded images, and dense technical docs for downstream analysis, all on-network.
- Built and deployed **AI agents for supply chain operations** across 3 departments, autonomously handling procurement queries, document routing, and status tracking; 2.5× workflow efficiency improvement.
- Built and deployed an internal **MCP (Model Context Protocol) server** exposing company tools as structured interfaces for LLM agents, including Verilog MCP integration for engineering workflow automation; 37% productivity gain.
- Built a **computer vision gesture recognition** system for restricted labs: hands-free authentication and equipment control; 98.6% accuracy, 6× faster login, 1.4% FRR, 99.99% uptime (2-month deployment), 97% adoption.
- Prototyped an **LLM-assisted CAD tool**: natural language to parametric CAD geometry with real-time FEA simulation; continued as independent research project (see below).

Undergrad Student Researcher *Jun 2024 – Present*

Department of AI & Robotics Engineering, Dayananda Sagar University

- Department-led research on **LLMs**, **diffusion modelling**, **knowledge representation for LLMs**, **generative models**, and **energy-based models**; ongoing LLM project work alongside coursework.

RESEARCH PROJECTS

Parametric CAD AI Engine *Feb 2025 – Present*

- Extended from initial Moog prototype into standalone research project: natural language prompt → AI generates production-ready parametric CAD geometry with physics-aware 3D models, real-time FEA & kinematics simulation, and manufacturing-ready exports.

LLM Scaling & Edge Deployment (Personal Research) *2024 – Present*

- Scaling laws study (0.5B–7B params); 72B teacher-student distillation for Hindi/Kannada low-resource fine-tuning. Evaluated QLoRA, LoRA, 4/8-bit quantization; deployed 1B model real-time on **Jetson Nano (4 GB RAM)**.

Medical AI: PCOS Detection *Dec 2025 – Present*

- Deep learning for automated PCOS detection from medical imaging with generative data augmentation; paper submitted for journal review, awaiting decision.

PUBLICATIONS

Sarkar, N., Deka, D.J. *Geometric Limits of Knowledge Distillation: A Minimum Width Theorem via Superposition Theory.* COLM 2026 (under review). [arXiv]

EDUCATION

B.Tech, Artificial Intelligence and Robotics Engineering *2023 – 2027*

Dayananda Sagar University, Bangalore | Co-developed and taught the department's LLM Engineering course and lab curriculum.

AWARDS, CERTIFICATIONS & LEADERSHIP

- **India AI Impact Summit 2026**, Invited Delegate, DSU; demonstrated LLM architectures, medical AI, and autonomous drone systems.

- **IEEE RAS & CIS Exceptional Volunteering Award, 2025.**
- Kaggle ML Certification (2025) | RapidMiner Certified Data Science Professional (2024) | AWS Certified Cloud Practitioner.
- Tech Lead, E-Cell DSU | Co-Founder, RoboVerse Club (100+ members, 30+ workshops) | Executive Committee, IEEE RAS & CIS.