Krishiv Agarwal

Fairfax, VA | (703) 589-0362 | Krishivagarwalk@gmail.com | linkedin.com/in/krishiv-agarwal |https://github.com/Kraga922

EDUCATION

Bachelor of Science in Computer Science **May 2027**
Courses: Data Structures, Autonomous Robots, Machine Learning Theory, Algorithm Design, Operating Systems, Computer Architecture, Linear Algebra
University of Florida, **Gainesville, FL** **GPA: 3.9/4.0**

WORK EXPERIENCE

LLM Researcher **May 2025 – Present**
Stanford Research Institute (SRI International) *Arlington, VA*

- First name author on **NeurIPS publication** on interpretability and safety for foundation models up to 120B parameters, optimizing activation steering.
- Innovated dual-vector jailbreaking methodology combining activation steering and token injection to penetrate test LLMs, achieving **10,000+ white-hat jailbreaks** across Llama, GPT-oss, Qwen, and Phi families with **86-94% success rate**, exposing critical safety vulnerabilities in production LLMs.
- Pioneered token-space projection for steering vectors, successfully decoding neuron activations into correlated tokens for effective prompt injection.
- Built a systematic framework to expose safety gaps in foundation models and establish deployable benchmarks for LLM evaluation.

Software Lead **Jan 2025 – Present**
Dream Team Engineering Gainesville*, FL*

- Developed **CI/ CD ML Ops pipeline** using **AWS SageMaker, S3 and EC2** to preprocess **2.4 TB of clinical samples** to create a hybrid 1D CNN and Transformer model for real-time blood pressure prediction from wearable sensors data.
- Led a team of 4 to design and deliver end-to-end smartwatch hardware and software prototype for **continuous blood pressure monitoring**.
- Architected **BLE data streaming** and mobile app enabling low-latency data transmission, real-time biometric data capture and BP estimation.

AI and Robotics Developer **May 2024 – August 2024**
Stanford Research Institute (SRI International) *Arlington, VA*

- Engineered a procedural generation system in NVIDIA Omniverse to create **10,400 m² hi-fi Isaac Sim worlds**, slashing design from 30 days to 1 minute.
- Designed **ROS2 Visual SLAM & OpenCV pipelines** for navigation, object detection, and mapping with **95%+ accuracy** across dynamic simulations.
- Enabled cross-platform **Omniverse** development by integrating **AWS EC2, Docker & NoMachine**, unlocking distributed team workflows on any OS.
- Accelerated robotics R&D by generating scalable, interactive Isaac Sim environments for high-fidelity training of autonomous agents and control systems

Artificial Intelligence Research Intern **May 2023 – November 2023**
Stanford Research Institute (SRI International) Menlo Park, CA

- Designed a **Reinforcement Learning system** using **Proximal Policy Optimization (PPO)** to generate **10,000+** autonomous ground vehicle CAD designs, optimizing chassis, wheel geometry, and power under real-world constraints, engineering creative exploration beyond traditional designs.
- Simulated and validated designs across procedurally generated terrains (rain, sand, hills, road) in Webots, **achieving a 20x improvement** over baseline.
- Created custom ETL pipelines, control policies, and hybrid testing frameworks to support a **DARPA** autonomous mobility project.
- Recognized as the **first-named author** on a published research paper and spotlighted by SRI for contributions to AI-generated vehicle design.

PROJECTS

Autonomous Robotic Object Retriever | ROS 2, RVIZ, SLAM, OpenCV, Hugging Face, Python

- Architected and deployed an autonomous robotic system using **ROS 2** and **Python** to enable TurtleBot 4 to perform **frontier navigation** and **SLAM**, dynamically mapping unknown environments and creating a semantic object map via QR code detection.
- Developed and integrated a **computer vision pipeline** with a **Large Language Model (LLM)** to enable a natural language interface, allowing users to request tasks and be guided to contextually relevant objects, while receiving intelligent, task-specific recommendations.

AI Physical Therapist (Start Up CEO) | OpenCV, MediaPipe, Streamlit, Python, Agile, Project Management

- Founded and led the development of GatorAID, an AI-powered physical therapy platform, winning UF's hackathon and business pitch competition.
- Engineered a real-time motion tracking and feedback system using computer vision and pose estimation to guide and assess rehabilitation exercises.
- Secured pilot program interest from Shands Hospital, validating the platform's potential for real-world adoption in clinical rehabilitation settings.

Monkeypox Classifier | Python, Keras, Tensorflow, OpenCV

- Engineered a **deep convolutional neural network (CNN)** that achieved **93% diagnostic accuracy** in identifying Monkeypox rashes from limited data.
- Designed a comprehensive **computer vision pipeline** using OpenCV for medical image cleaning, segmentation, and extensive augmentation to enhance dataset diversity, address class imbalance, and ensure robustness across varying lighting and skin tones.
- Applied dropout, cross-validation, and reproducibility protocols to prevent overfitting and compliance with validation standards.

SKILLS

Languages: Python, Java, C++, C, Typescript, SQL | **Tools:** AWS, Docker, Git, Linux, Hugging Face, Jupyter, NVIDIA Omniverse, CUDA, ROS 2, Webots
Libraries: Transformers, Scikit-learn, OpenCV, Pandas, PyTorch, TensorFlow, Keras, NumPy, MediaPipe, Streamlit, React Native, PostgreSQL, Agile
Awards & Publications: NeurIPS (Lock-LLM) Publication, 2x Hackathon Winner, 2x Business Pitch Finalist, Eagle Scout