

SUNDAR SRIPADA V. S.

✉ sundarsripada26@gmail.com 🌐 github.com/ss26 🌐 ss26.github.io in linkedin.com/in/sundarsripada26

EDUCATION

The University of Texas at Austin
Master of Science in Engineering in Electrical and Computer Engineering
Graduate Teaching Assistant at **Department of Physics**

Austin, TX, USA
Aug. 2022 – Dec. 2024
Jan. 2024 – Dec. 2024

EXPERIENCE

ML Research Engineer
AI Safety Camp

Jan. 2025 – Present
Remote

- Conducting empirical research on adversarial behaviors in LLMs trained with PPO and DRRN policies, aiming to classify and counteract adversarial reward optimization strategies
- Developing an actor-critic Python PPO agent with Causal LM tokenizers for the Machiavelli dataset (**merged PR**)
- Evaluating agents' propensity to optimize for adversarial rewards on the Machiavelli and AgentHarm benchmarks by running PPO and PPO-LLM experiments with **Center for AI Safety's compute clusters**

Software Engineer Intern
Hewlett Packard Enterprise

Jun. 2023 – Aug. 2023
San Jose, CA, USA

- Enhanced scale characterization of HPE Aruba's network management tool by developing Virtual Machine (VM) simulation support in Python, doubling the tool's device handling capacity by up to 2,000,000 simulated devices
- Developed CLI-based VM simulation functionality that enables users to simulate VMs within Docker containers deployed with Kubernetes, achieving fast on-boarding (<1 minute per VM) to enhance scalability
- Collaborated in a cross-functional team writing SQL code for extracting VM system information, integrating it into the simulation framework to enable scale simulation and support data-driven decisions for cloud infrastructure

Graduate Research Assistant
Swarm Lab, University of Texas at Austin

Aug. 2022 – Jun. 2023
Austin, TX, USA

- Built TensorFlow-based activity recognition model with 87% classification accuracy across 4 tool-use behaviors, enhancing ergonomic study of user-tool interaction. **Paper**
- Led a team of 3 students in a data collection project using a 9-DOF IMU on a robot arm that emulates human performance.
- Engineered a data processing pipeline with Pandas to analyze a dataset of over 10,000 entries for activity classification.
- Implemented an ML pipeline in TensorFlow Lite (TFLite) and MLFlow for logging, utilizing convolutional neural networks (CNN) for real-time edge inference on a Raspberry Pi, achieving 162 ms per prediction for activity classification.

Research Engineer
Robotics Research Center, International Institute of Information Technology

Oct. 2020 – Jul. 2022
Hyderabad, India

- Managed a large-scale data mining and collection pipeline using Python scripts with the CARLA simulator's API (autonomous systems) to gather contrived scenes of environments containing over 100,000 LIDAR data points for SLAM
- Ported LOAM's C++11 source code to C++14, enabling usage in the new ROS version (**GitHub**).
- Designed classification and regression models in scikit-learn to predict the presence and magnitude of drift in self-driving vehicles, achieving prediction accuracy of up to 92% with PyTorch
- Devised a robust data processing pipeline by developing API functions and automation scripts in Python to transform raw LIDAR data into custom datasets for PyTorch's DataLoader, significantly enhancing the efficiency of model training
- Formulated a reinforcement learning model using Proximal Policy Optimization (PPO) to reduce drift in highly dynamic autonomous driving scenes. **Paper, GitHub**
- Built a custom learning-to-rank triplet loss function to train a multi-modal CNN in PyTorch, effectively minimizing drift in autonomous driving and improving performance by up to 76.76% over previous SOTA. **Paper, GitHub**

SKILLS

Programming	Python, C++, C, Bash, R, MySQL, MATLAB, Java, Julia, JAX, HTML, CSS, JavaScript, TeX
Frameworks	PyTorch, TensorFlow, TFLite, LangChain, Gymnasium, ROS, MLFlow, InspectAI (AISI), Apache Spark, AWS (S3, EC2, SageMaker), CARLA, Gazebo
Libraries	numpy, pandas, polars, matplotlib, seaborn, OpenCV, HuggingFace Transformers, HuggingFace TRL, Stable Baselines3, CleanRL, RLLib, Pillow, scikit-learn, Keras
Tools	git, Docker, VS Code, Google Colab, Linux, Windows 10/11, MacOS, Microcontrollers
Certificates	AI Safety Fundamentals
Academic Service	Reviewer for: CASE 2022, CASE 2023, IROS 2023, ICRA 2024