Collaborative Research: Robust and Verifiable Reinforcement Learning in Safety-Critical Cyber-Physical Systems

NSF Directorate for Computer and Information Science and Engineering (CISE)
Division of Computer and Network Systems (CNS)

Project Description

1. Overview

Modern cyber-physical systems (CPS)—ranging from autonomous vehicles to smart power grids—increasingly rely on deep reinforcement learning (RL) to make decisions in dynamic, high-dimensional environments. Despite their empirical success, deep RL policies lack formal safety guarantees. They are prone to catastrophic failures when encountering out-of-distribution scenarios or adversarial disturbances. This proposal outlines a novel framework to bridge the gap between empirical data-driven learning and formal control theory.

The goal of the proposed research is to develop a unified framework for **Robust and Verifiable Reinforcement Learning (RV-RL)** that guarantees safety during both the learning phase and the deployment phase. Our approach leverages control barrier functions (CBFs) as a mathematical foundation to constrain the learning agent's search space, ensuring that safety invariants are never violated.

2. Intellectual Merit

The intellectual merit of the proposed research lies in three key scientific contributions:

1. **Synthesizing Neural Control Barrier Certificates:** We will develop a self-supervised method to learn neural barrier certificates directly from system trajectories, circumventing the need for analytical control models.
2. **Safe Exploration via Differentiable Projections:** We will introduce a differentiable projection layer into the policy network, mapping unconstrained actor outputs to safety-preserving control actions.
3. **Verifiable Sim-to-Real Transfer:** We will establish robust generalization bounds that account for model discrepancies between simulation and hardware deployments, guaranteeing safety despite bounded model mismatches.

By integrating control-theoretic barrier certificates directly into the policy optimization loop, this research addresses a fundamental trade-off between optimization performance and safety constraints.

3. Broader Impacts

The broader impacts of this project cover educational integration, open-source tool development, and industrial technology transfer:

- **Educational Outreach:** The Principal Investigator (PI) will design a new graduate curriculum on "Formal Methods in Autonomous Systems" and recruit underrepresented undergraduate students through the university's Summer Research Experience program.
- **Open-Source Software:** We will release `SafeRL-Suite`, an open-source, PyTorch-based benchmarking environment for evaluating safe RL algorithms on complex physical control tasks.
- **Industrial Collaboration:** The PI will collaborate with local autonomous driving startups to transfer the developed algorithms to hardware testbeds, accelerating the adoption of safe AI in commercial transportation.

4. Research Plan

The proposed research is organized into three interconnected research thrusts over a three-year period, as detailed below.

4.1 Learning Safety-Preserving Control Barrier Certificates

To establish safety guarantees without relying on accurate analytical system models, we formulate the learning of control barrier certificates as an optimization problem. Let the system dynamics be represented by the discrete-time relation $x_{t+1} = f(x_t, u_t)$. We define a neural barrier candidate $h_\theta(x)$ parameterized by weights θ. The certificate condition is defined as:

$$h_\theta(x_{t+1}) - h_\theta(x_t) \geq -\gamma h_\theta(x_t), \quad \forall x_t \in \mathcal{X} \tag{1}$$

where $\gamma \in (0, 1]$ represents the recovery rate. We will design loss functions that penalize violations of this inequality over state trajectories, training $h_\theta(x)$ concurrently with the policy network.

4.2 Robust Policy Optimization with Formal Constraints

Using the learned barrier certificates, we formulate a safety filter that wraps around any standard RL agent. The policy output $\pi(x)$ is projected onto the set of safe controls $\mathcal{C}(x)$ defined by the barrier certificate:

$$u^* = \arg\min_u \|u - \pi(x)\|_2^2 \quad \text{s.t.} \quad \nabla h(x)^T f(x, u) \geq -\gamma h(x) \tag{2}$$

This projection is formulated as a quadratic program (QP) that can be solved differentiably inside the policy network, allowing end-to-end backpropagation of the policy gradient through the safety layer.

4.3 Hardware-in-the-Loop Validation on Autonomous Ground Vehicles

We will validate our framework on a custom fleet of scale-model 1/10th autonomous ground vehicles (AGVs) equipped with LiDAR, depth cameras, and onboard Jetson Orin processing units. The validation scenarios will include obstacle avoidance in dynamic environments and multi-agent intersection coordination.

4.4 Project Management, Timeline, and Milestones

The research activities will proceed according to the schedule in Table 1.

Table 1: Timeline and Milestones for the proposed project (Y1–Y3).

Research Thrust	Year 1	Year 2	Year 3
Thrust 1: Safe RL	Formulate differentiable safety projection layers.	Implement scalable QP solvers for real-time control.	Extend formulation to multi-agent environments.
Thrust 2: Certificates	Train neural barrier candidate networks.	Verify certificate robustness under disturbances.	Derive formal generalizability bounds.
Thrust 3: Validation	Build 1/10th AGV scale testbed.	Validate single-agent obstacle avoidance.	Validate multi-agent intersection scenarios.
Dissemination	Launch `SafeRL-Suite` open repository.	Publish intermediate results at conferences.	Host workshop on Safe Autonomy at RSS/ICRA.

5. Results from Prior NSF Support

PI Smith: Award IIS-2201948 (Total: $450,000, Period: 2022–2025).
Title: Data-Driven Formal Verification of Cyber-Physical Systems.

Intellectual Merit: The prior project developed data-driven techniques for learning Lyapunov and barrier functions from noisy execution traces. Key results included the first algorithm to synthesize verified control policies for hybrid dynamical systems using deep learning, leading to a 35% reduction in verification time compared to traditional semi-definite programming (SDP) methods.

Broader Impacts: The award supported two PhD dissertations and three undergraduate research projects. The PI developed a middle-school robotics module that was integrated into the local school district's STEM curriculum, reaching over 400 students.

Publications resulting from this award:

[1] Smith, J., and Doe, A. "Synthesis of Neural Barrier Certificates for Nonlinear Control Systems." *IEEE Transactions on Automatic Control*, 68(4), 2105–2118, 2023.

[2] Smith, J., Johnson, K., and Doe, A. "Safe Exploration in Model-Free Reinforcement Learning via Barrier Envelopes." *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 2024.

6. Budget Summary

A summary of the requested budget is provided in Table 2, supporting one PI (one month of summer salary per year) and two Graduate Research Assistants.

References

[1] A. Ames, X. Xu, J. Grizzle and P. Tabuada, "Control barrier function based quadratic programs with application to adaptive cruise control," *IEEE Transactions on Automatic Control*, vol. 62, no. 8, pp. 3861-3876, 2017.

Table 2: Budget Summary (Direct and Indirect Costs).

Category	Year 1 ($)	Year 2 ($)	Year 3 ($)	Total ($)
Senior Personnel (PI Summer Salary)	18,500	19,055	19,627	57,182
Graduate Research Assistants (2 GRA)	64,000	65,920	67,898	197,818
Fringe Benefits (PI & GRAs)	19,800	20,394	21,006	61,200
Equipment (Onboard Jetson	12,000	0	0	12,000
Travel (Conferences, Workshops)	5,000	5,500	6,000	16,500
Other Direct Costs (Publications, Fees)	4,500	4,500	4,500	13,500
Total Direct Costs	123,800	115,369	119,031	358,200
Indirect Costs (F&A, 52% MTDC)	64,376	59,992	61,896	186,264
Total Proposal Budget	**188,176**	**175,361**	**180,927**	**544,464**

[2] L. Gu, J. Smith and J. Kober, “A survey of safe reinforcement learning in robotics,” *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 5, pp. 411-435, 2022.

[3] S. Zhao, X. Zhou, and H. Wang, “Safe control of unknown systems via neural barrier functions,” *International Conference on Machine Learning (ICML)*, pp. 1104-1115, 2023.

[4] K. Johnson and J. Smith, “Sim-to-real transfer of safe neural policies under parameter uncertainty,” *IEEE Robotics and Automation Letters*, vol. 9, no. 2, pp. 890-897, 2024.