MEDIUM: Certified Robust Learning for Multi-Agent Planning and Control

PI: Yiannis Kantaros, Dept. of Electrical and Systems Engineering, Washington University in St. Louis

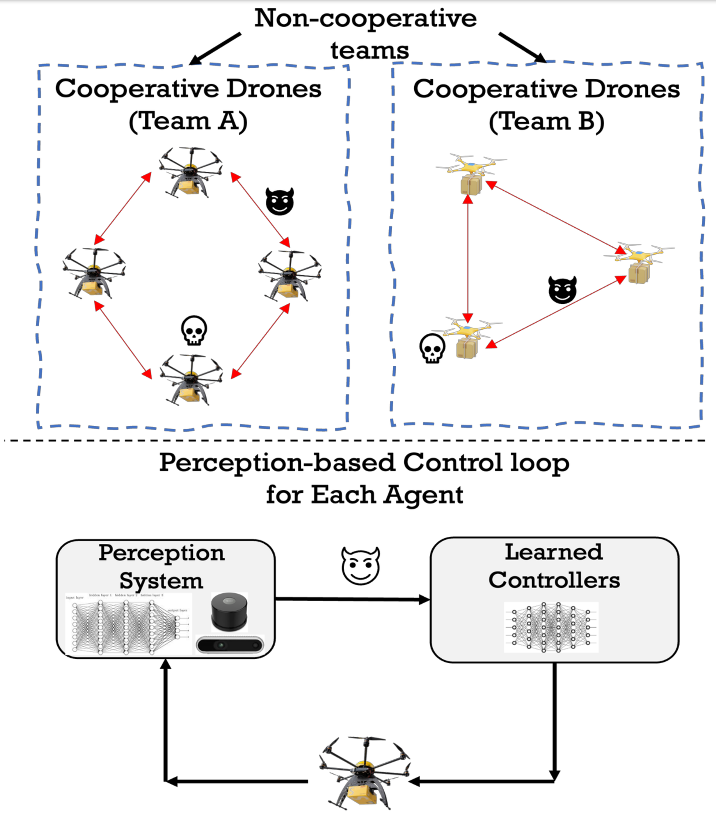
co-PI: Yevgeniy Vorobeychik, Dept. of Computer Science and Engineering , Washington University in St. Louis

co-PI: Hussein Sibai, Dept. of Computer Science and Engineering , Washington University in St. Louis

co-PI: Ning Zhang, Dept. of Computer Science and Engineering , Washington University in St. Louis

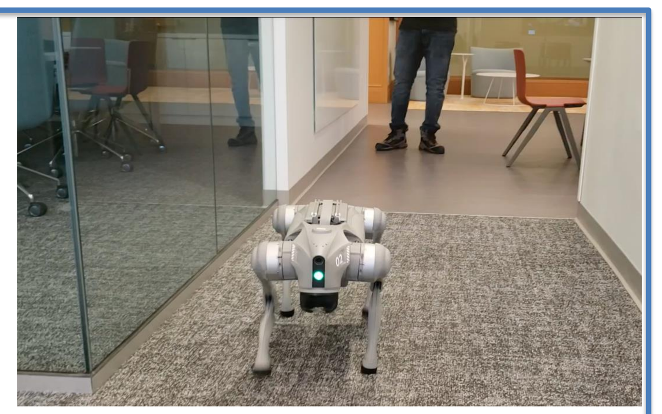
Motivation: Learning-based decision-making algorithms, such as Deep Reinforcement Learning or Neural Model Predictive Control, have been used to control multi-agent systems due to their generalization and real-time performance benefits. However, these algorithms lack robustness to imperceptible input perturbations. Despite impressive progress towards addressing this, existing methods lack rigorous safety and robustness guarantees. Also, lack of robustness becomes more pronounced in multi-agent settings, due to a larger surface of vulnerabilities, which has received significantly less research attention.

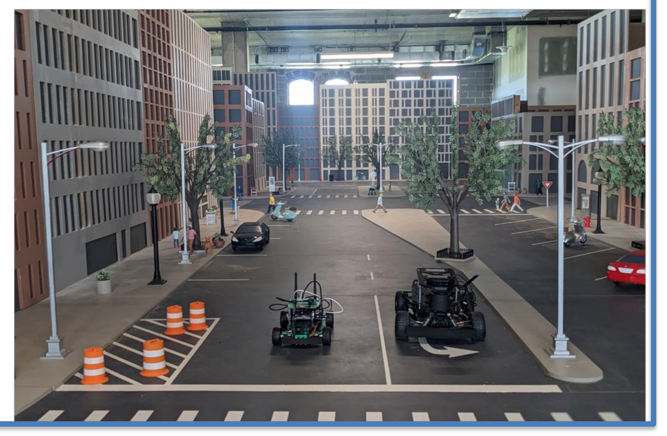
Key Challenges: Consider a multi-agent system and user-specified task and safety requirements ϕ (using e.g., formal languages or reward-based functions). How to design (and/or verify) learning-based controllers that are robust/safe (w.r.t. to satisfaction of ϕ) in the presence of **(i)** perceptual noise; **(ii)** mis-calibrated confidence in predictions; **(iii)** adversarial communications; **(iv)** agent failures; **(v)** non-cooperative agents sharing the same workspace?



Scientific Impact

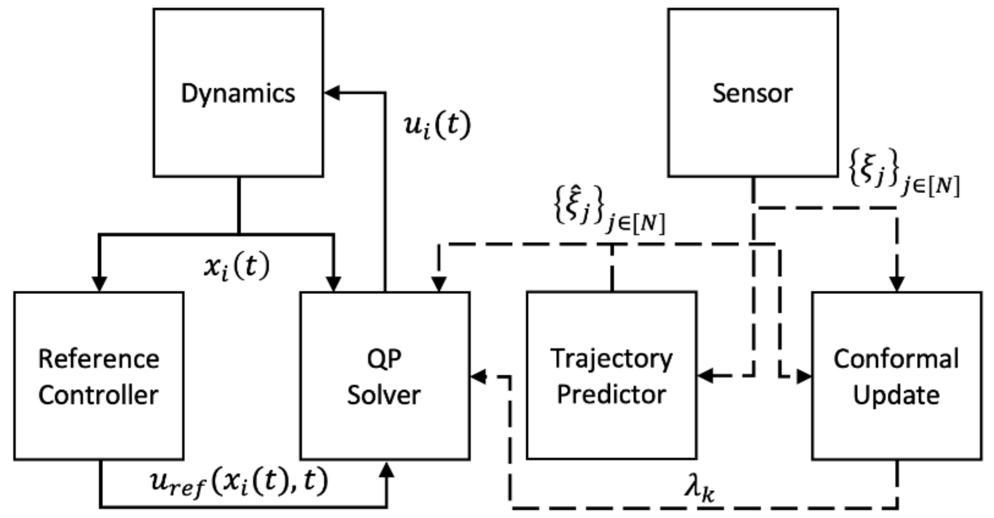
Provide the first theoretically-grounded algorithms to *certify*, *verify*, and *train* safe and robust multi-agent learning-enabled decision-making algorithms against (i)-(iv).





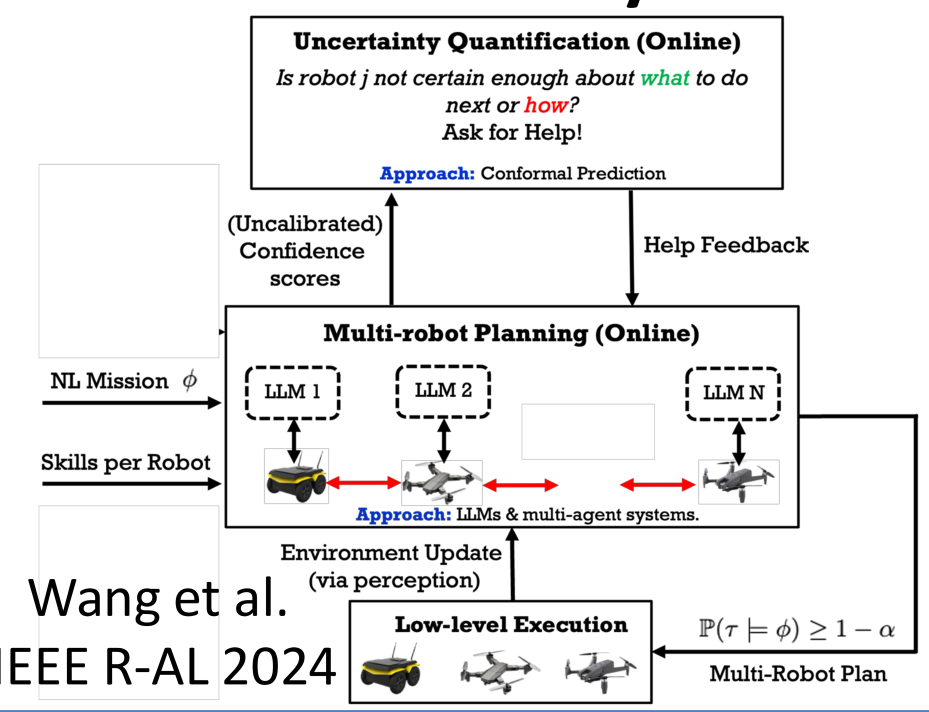
New Contributions:

Safe Decentralized ML-based Multi-agent Control

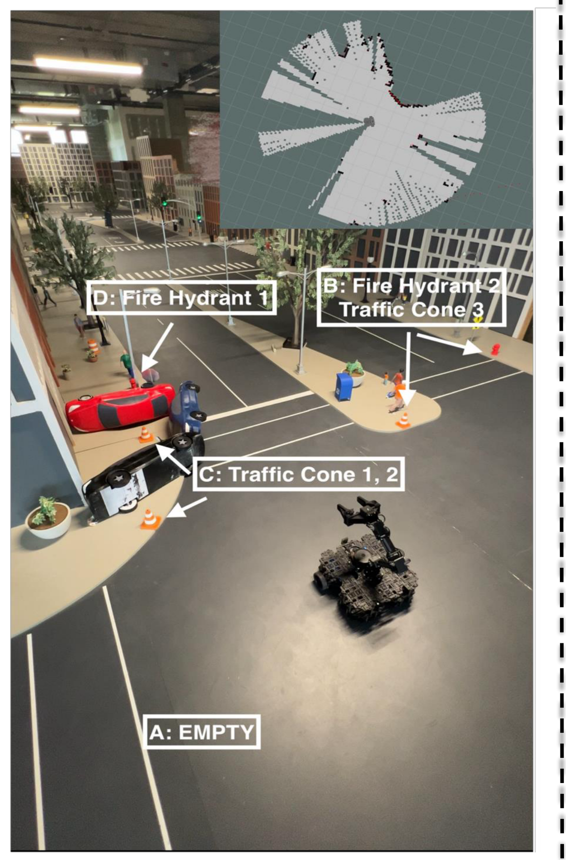


Huriot et al ICRA 2025

Safe LLM-based Planning for Multi-Robot Systems



Wang et al. IEEE R-AL 2024



Verified Safe RL-based Control

Train controllers that can be efficiently verified for safety.

$$\begin{aligned}
&\max_{\theta} \mathcal{J}(\pi_\theta) \\
\text{s.t. } &\mathcal{J}_{C_i}(\pi_\theta) \leq d_i, \quad \forall i \in [m] \quad \text{(empirically satisfied)} \\
&s_t \in \mathcal{S}_{\text{safe}}, \quad \forall t \in [K] \quad \text{(mathematically verified)} \\
&s_{t+1} = F(s_t, a_t), a_t = \pi_\theta(s_t), s_0 \in \mathcal{S}_0 \subseteq \mathcal{S}_{\text{safe}}
\end{aligned}$$

Obstacle Avoidance (quadrotor)

	Verified-50(↑)
PPO-Lag	0.0
PPO-PID	0.0
CAP	57.1
MBPPO	0.0
CBF-RL	0.0
RESPO	0.0
VSRL	**100.0**

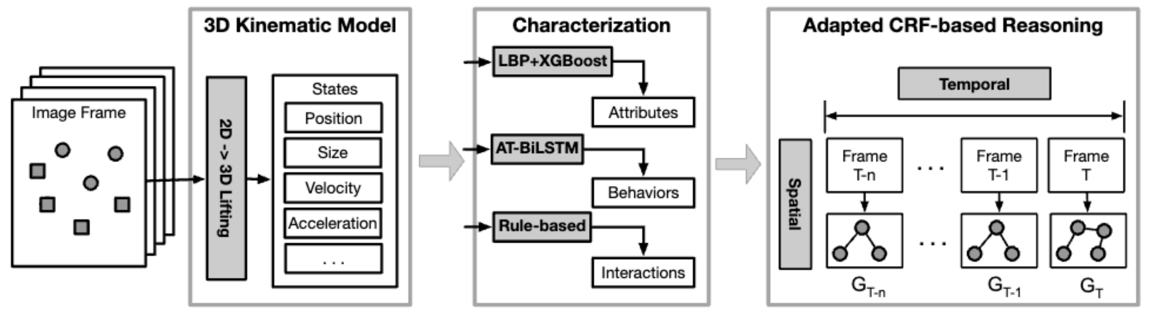
Vehicle Avoidance

	Verified-50(↑)
PPO-Lag	72.8
PPO-PID	72.0
CAP	73.3
MBPPO	82.6
CBF-RL	73.0
RESPO	74.5
VSRL	**100.0**

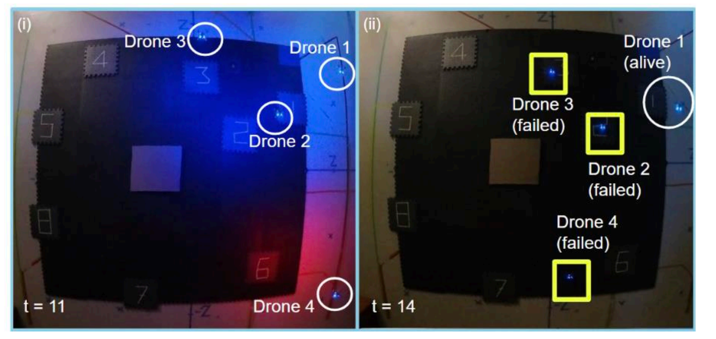
Wu et al, NeurIPS 2024

Defenses against Physical Adversarial Examples in Autonomous Systems

Yu et al, CCS 2024



Reactive Multi-Robot Planning to Robot Skill Failures



Kalluraya et al (under review)

Broader Impacts:

- Enable more reliable and robust learning-enabled multi-agent systems that can safely perform complex tasks in uncertain, adversarial, and dynamic environments.
- Applications: delivery, transportation, manufacturing, search-and-rescue.
- Research opportunities to K12, UG, MS, PhD students and the WashU Robotics Club.
- Design new graduate courses (e.g., *Learning and Planning in Robotics, Trustworthy Autonomy)*
- Release open-source software and demonstrations

Email: ioannisk@wustl.edu

Start date: 06/15/2024

End date: 05/31/2027

Award ID#: CNS 2403758