Multi-Level Attack and Defense Simulation Environment for Artificial Intelligence Education and Research

Zhou Li, Sergio Gago Masague, Sameer Singh (University of California, Irvine) https://maestro-ai.github.io/

Objectives

- Design a platform that simulates adversarial machine learning tasks of existing research
- Give students hands-on experiences with building robust AI systems
- Enable comprehensive comparison of different attacks and defenses of robust Al





Challenges

- The research efforts in robust AI are \bullet fragmented, lacking a systematic simulation and evaluation framework
- Though AI courses are broadly offered by ulletinstitutions, the topic of robust AI is often overlooked
- No hands-on labs for students to learn \bullet robust Al

Scientific Impacts

- Several papers published in top AI and cyber-security conferences and journals about robust AI (NAACL'21, TIFS, ACSAC'21)
- Platform (Maestro) implementing various attack and defense methods has been developed
- Game-based strategies are developed for the pedagogical activities of robust AI



	A Lea	ttack	Home the CS175 AI P	W(roject (Ork Course	(Dashi
	Hide columns	~		Stude	nt name	Filte	er
Rank	Name	Evaluation Time	Attacker Success Rate	distance	Time	Queries ★	Meets Expectations
Rank 1	Name Madison Caleb Isaac Koelzer	Evaluation Time 2022-04-12-00:02:15	Attacker Success Rate	distance 7.26	Time 132.02	Queries ★ 1624.81	Meets Expectations True
Rank 1 2	Name Madison Caleb Isaac Koelzer Iris Yu	Evaluation Time 2022-04-12-00:02:15 2022-04-12-06:54:11	Attacker Success Rate 100.0 100.0	distance 7.26 6.9	Time 132.02 168.55	Queries ★ 1624.81 2605.76	Meets Expectations True True
Rank 1 2 3	Name Madison Caleb Isaac Koelzer Iris Yu Aayush Atish Bokil	Evaluation Time 2022-04-12-00:02:15 2022-04-12-06:54:11 2022-04-11-17:57:10	Attacker Success Rate 100.0 100.0 100.0	distance 7.26 6.9 7.34	Time 132.02 168.55 69.92	Queries ★ 1624.81 2605.76 2624.81	Meets Expectations True True True
Rank 1 2 3 4	Name Madison Caleb Isaac Koelzer Iris Yu Aayush Atish Bokil Austin J. Nelson	Evaluation Time 2022-04-12-00:02:15 2022-04-12-06:54:11 2022-04-11-17:57:10 2022-04-12-00:29:46	Attacker Success Rate 100.0 100.0 100.0 100.0	distance 7.26 6.9 7.34 6.94	Time 132.02 168.55 69.92 317.24	Queries ★ 1624.81 2605.76 2624.81 4364.49	Meets Expectations True True True True True

Education Setup

Societal Impact

- Raise public awareness of Al robustness
- Foster a workforce with skills of building robust AI systems

University of California, Irvine

Education and Outreach

- A project-based course has been offered at UCI (2022 Winter and Spring) to teach robust AI with the platform
- Tutorials about robust AI have been delivered at AAAI 2021 and CPAIOR 2021

Course Leaderboard

Quantify Impacts

- More than 150 UCI students enrolled in the robust Al course
- Course surveys show ulletstudents benefit from the Maestro platform
- Maestro platform is openulletsourced

The 5th NSF Secure and Trustworthy Cyberspace Principal Investigator Meeting (2022 SaTC PI Meeting) June 1-2, 2022 | Arlington, Virginia