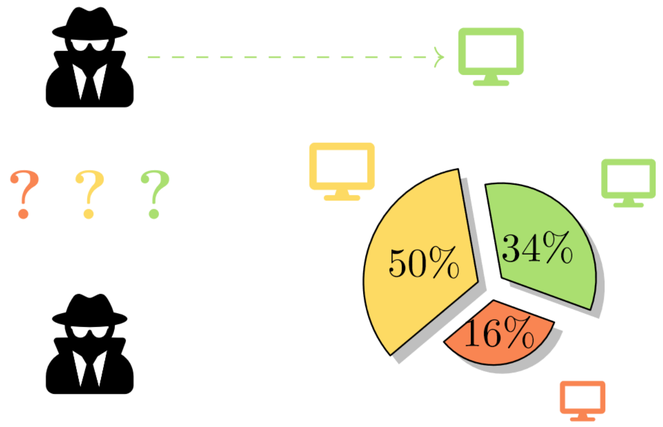
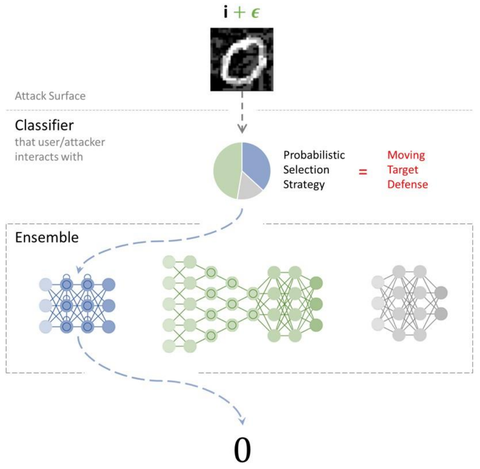
Adaptive Artificial Intelligence: from Adversarial to Assistive Scenarios

Sailik Sengupta *PhD Candidate @ Arizona State University*

What, _When_ and _How_ of MTD Systems Inference and Learning of game-theoretic strategies for of Moving Target Defense (MTD) in cyber-security.[AAMAS'17,'18; GameSec'19'20; IEEE Comm. S&T'20]

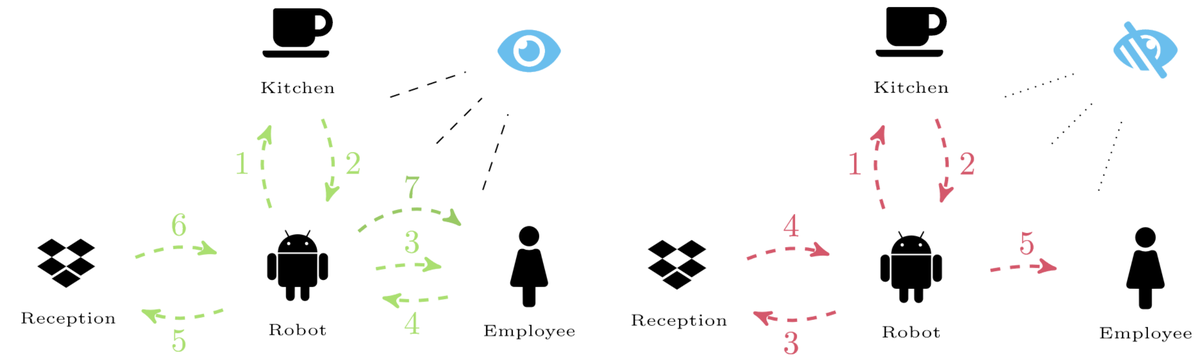




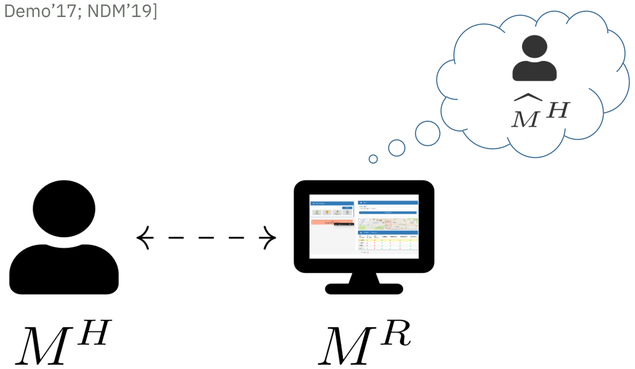
MTDeep Test time randomization as an add-on security against adversarial attacks (beats SOTA). [GameSec'19]

Adversarial/Non-cooperative

To See or Not to See Game-theoretic trust in robot supervision tries to come up with a mixed supervision strategy. Ensures (1) robot does not deviate from expected behavior and (2) saves supervisor's time.[Trust AAMAS Ws'19]

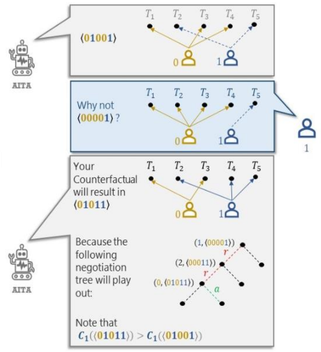


RADAR– Proactive Decision Support Systems Leverages Automated Planning technology and HCI design principles (stages and the ladder of automation) to increase the efficiency and quality of plans in Naturalistic Decision Making scenarios.[AAAI FSS'17; ICAPS Demo'17; NDM'19]



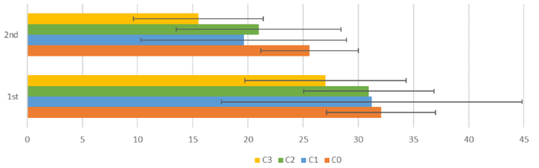
Cooperative

AI knows better Negotiation-aware allocation generation and contrastive explanations when humans have imperfect knowledge and limited compute capability.



Human knows better We train a classifier to align with a human's view of failure modes. These explicable classifiers reduce egregious mistakes (that can have odious societal impacts).
[ICML Ws'20]

Human Subject Evaluations Faster generation of plans, higher satisfaction, better learning.
[HCI Journal'20]



MA-RADAR Decision support for human teams.
[ICAPS Demo'18]

RADAR-X Contrastive Explanations and Preference Elicitation.

https://sailik1991.github.io/ sailik.cse.jdvu@gmail.com https://github.com/sailik1991

Microsoft Research AI Breakthroughs 2020