Adaptive Cyber Deception, Human-AI Teaming, and Group Problem Solving

Yinuo Du, Ph.D

- School of Computer Science
- Carnegie Mellon University
- Supported by ARL-CRA and MURI-CATCH
- Human and AI Decision-Making
- in Cybersecurity:

Yinuo Du, Ph.D (cont.)

AI INSTITUTE
FOR SOCIETAL
DECISION MAKING

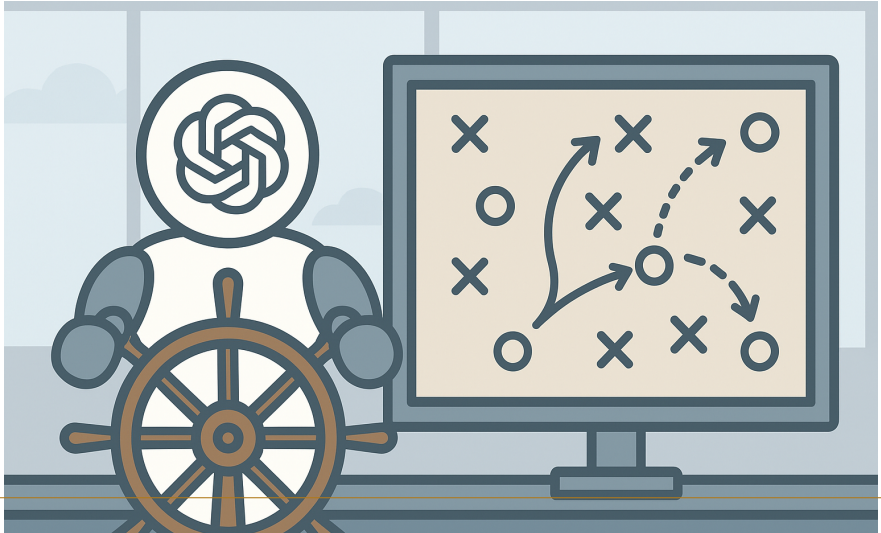
The 2024 threat landscape is fast + adaptive

- Data Breach
- Phishing
- Ransomware
- 4.45 million
- 4.76 million

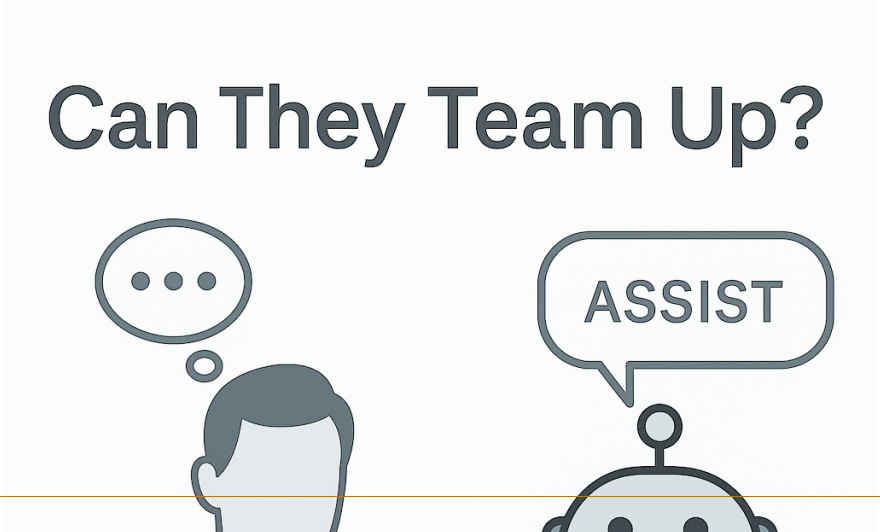
The 2024 threat landscape is fast + adaptive (cont.)

Too Fast for Humans

AI at the Helm?



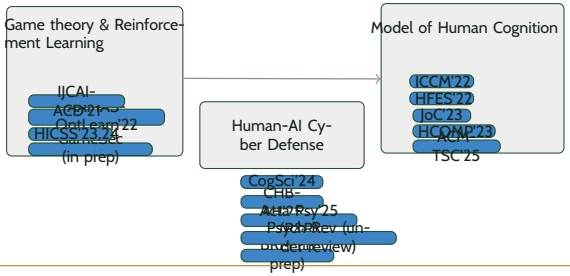
Human-AI Cyber Defense: A Missing Partnership



Research

- Overview
- Game theory
- Reinforcement Learning

Research Overview



RL for Adaptive Cyber-Deception

- Human-AI Team Defense
- LLM for Group Problem-Solving
- Other projects
- Future vision
- Q&A

RL for Adaptive Cyber-Deception (cont.)

Outline

RL for Adaptive Cyber-Deception Human-AI Team Defense LLM for Group Problem-Solving Other projects Future vision Q&A

GameSec (in preparation)

AAMAS-OptLearn'22

CHB-AH'25

CogSci'24

Deception has been used to engage and mislead attackers

- True Configuration
- os]: Red Hat Linux 7

Perturbation & Obfuscation

Honey-X

Fake Configuration -[os]: Red Hat Linux 7 -[web]: Tomcat 8.0.30 -[files]: public

Pawlick, J., Colbert, E., & Zhu, Q. (2019). A game-theoretic taxonomy and survey of defensive deception for cybersecurity and privacy. ACM Computing Surveys (CSUR), 52(4), 1-28.

Stackelberg security games have been used to model the strategic interaction between defender and attacker

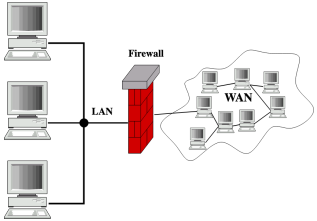
- Cyber Deception Game (Schlenker et al. 2018) models the defender's choice of deceptive network configurations and the attacker's reconnaissance as a zero-sum game.

Stackelberg security games have been used to model the strategic interaction between defender and attacker

- T he defender and attacker's ability to adapt to real-time observations is under explored

Stackelberg security games have been used to model the strategic interaction between defender and attacker (cont.)

Attackers can gather more information about the network overtime



Defenders are also getting alerts in real time

Security Onion

Alerts

Options

REFRESH

Apt*

01/01/2024 – 01/03/2024

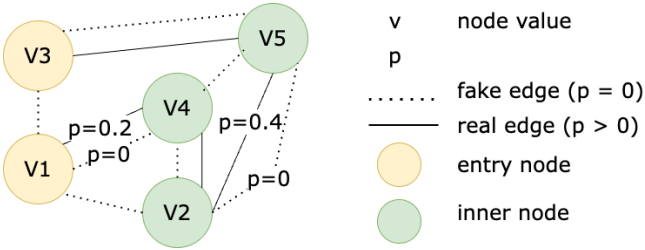
No.		Rule	Module	Date
1	1	Malware Alert	suricata	Jan 3, 2024
2	2	Suspicious Connection	suricata	Jan 2, 2024

We propose an Adaptive Cyber Deception Game

- Observe
- Update defense strategy
- Surveillance
- Update attack strategy

We build our game model on an attack graph that can be interpreted at various granularity

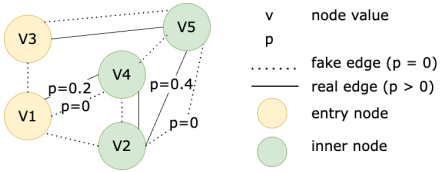
- probability of success



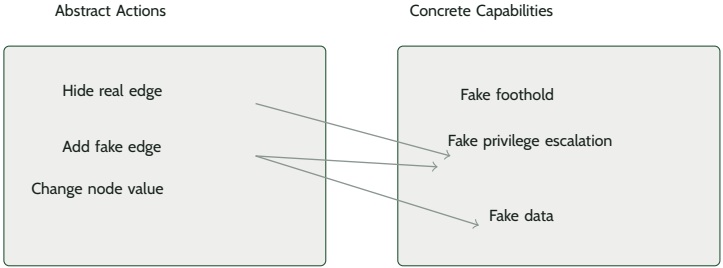
Defenders can take two types of action: deceptive and protective actions

- Deceptive
- Hide a real edge
- Add a fake edge
- Modify the perceived values of a set of nodes
- Protective

Defenders can take two types of action: deceptive and protective actions (cont.)



Each abstract action on the attack graph can be translated to multiple deceptive capabilities



We assume two types of adversary: Powerful and Naive

- Powerful
- Naive

We study the use of a reinforcement learning algorithm Proximal Policy Optimization (PPO) with self-play

- Input:
- Self location
- Perceived node values
- Real edge visibility
- Fake edge visibility

$$\pi_a(\boldsymbol{s}) = \boldsymbol{a}_a$$

We study the use of a reinforcement learning algorithm Proximal Policy Optimization (PPO) with self-play

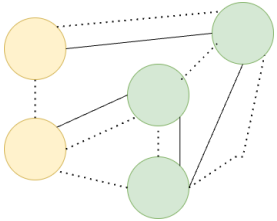
- Input:
- Attacker's location,
- Perceived node values,
- Real edge visibility,
- Fake edge visibility

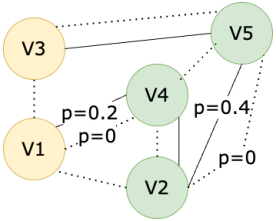
$$\pi_d(\boldsymbol{s}) = \boldsymbol{a}_d$$

Reinforcement learning with self-play

- Attacker's
- Neural Network Policy
- Defender's
- 1. Collect experiences based on the player's current policies
- 2. Run one-step update for the defender/attacker's network parameters separately

Reinforcement learning with self-play (cont.)





$$\pi_a(s) = a_a$$

We compared the PPO defender with a heuristic defender

Training Phase		Evaluation Phase	
Defender	Attacker	Attacker	Defender
PPO	PPO	PPO (PPO-trained)	PPO (trained) Heuristic
Heuristic	PPO	PPO (Heuristic-trained)	
		Heuristic	

We compared the PPO defender with a heuristic defender (cont.)

Powerful/Naive)

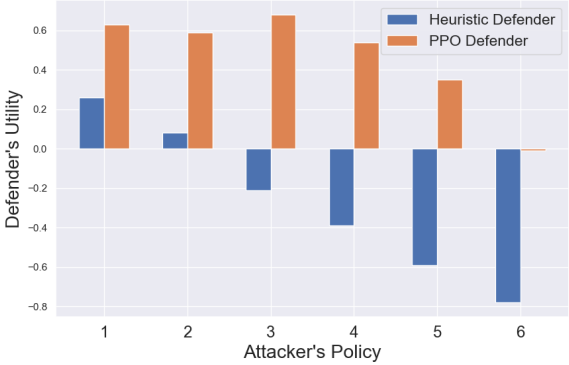
Powerful/Naive)

The PPO Defender always outperforms the heuristic defender.

- Naive
- Powerful

The PPO Defender always outperforms the heuristic defender.

(cont.)



We measure exploitability to see if the training reached nash equilibrium

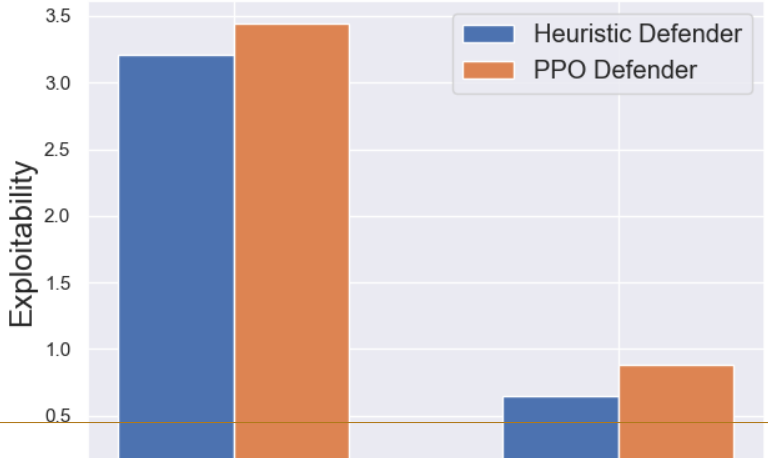
- Exploitability of
- Exploitability of attacker's policy
- Exploitability of defender's policy

$$Exp(\pi_a, \pi_d) = Exp_a(\pi_a, \pi_d) + Exp_d(\pi_a, \pi_d)$$

$$Exp_a(\pi_a, \pi_d) = U_a(\pi_a, \pi_d) - U_a(\pi_a, BR_d(\pi_a))$$

$$Exp_b(\pi_a, \pi_d) = U_d(\pi_a, \pi_d) - U_d(BR_a(\pi_d), \pi_d)$$

The (PPO defender, PPO attacker) pair of policies is closer to a Nash equilibrium



What now? With the promising performance in simulation, we went on deploying RL-based defense strategy in real-world network

rapid7/metasploit-framework

#18621 **Add module f**
Backup Migration W
plugin (CVE-2023-...

2 comments

jheysel-r7 opened on December 15, 2023

Instances (11) Info

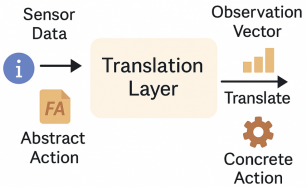
Find instance by attribute or tag (case-sensitive)

Instance state = running | Clear Filters

Name	Instance ID	Instance state
Public Server 1	i-0019dc9fbc9bb81ea	Running
HoneyPot-Pub...	i-0f4012cd80d740b2c	Running
HoneyPot-NTP...	i-04a94d6d7fb9d3066	Running
PC1-WEB	i-078044fd9a42c65b49	Running
PC2-WEB	i-0013de65425006ac3	Running
PC3-WEB	i-0728c5e8f7391614c	Running
Controller	i-04790eba4bba6c27a	Running
Public Server 2	i-031de534e971284af	Running
WEB	i-065a8db824659ae77	Running
NTP	i-0cf451c097f950db6	Running
DB	i-07ab0dcafa14c9bcc	Running

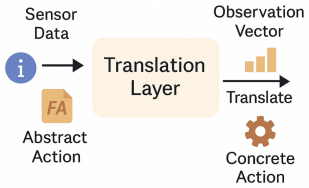
Bridging the Gap: Deploying RL Defenders on AWS

- Observation s

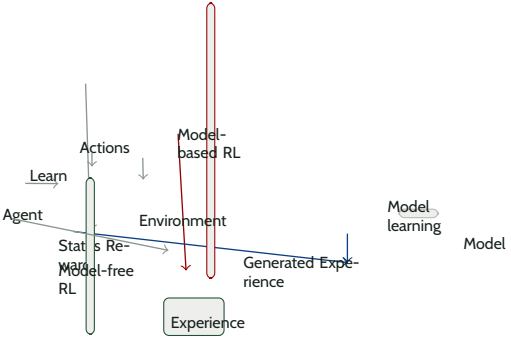


Bridging the Gap: Deploying RL Defenders on AWS

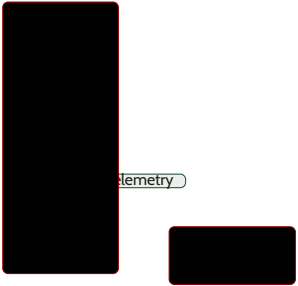
- Policy Improvement
- Reward
- Observation



Model-based reinforcement learning



Latent Graph Dynamics and Opponent Belief Modeling for Strategic Planning



Hafner, D., Lillicrap, T., Norouzi, M., & Ba, J. (2020). Mastering atari with discrete world models. arXiv preprint arXiv:2010.02193.

RL for Adaptive Cyber-Deception

- Human-AI Team Defense
- LLM for Group Problem-Solving
- Other projects
- Future vision
- Q&A

RL for Adaptive Cyber-Deception (cont.)

Outline

RL for Adaptive Cyber-Deception Human-AI Team Defense Problem-Solving Other projects Future vision Q&A

GameSec (in preparation)

AAMAS-OptLearn'22

CHB-AH'25

CogSci'24

Human-AI

- Team Defense
- Security Information and
- Event Management
- SIEM)
- Security Orchestration, Automation, and Response

Human-AI (cont.)

Human-AI Team Defense

Security Information and Event Management (SIEM)

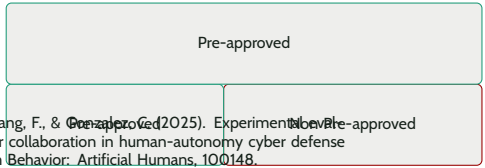
VS

Security Orchestration, Automation, and Response (SOAR)

?

Existing AI techniques for cybersecurity are used as tools rather than teammates of humans

We propose a semi-supervisory Human-AI Teaming (HAT) paradigm



Du, Y., Prbot, B., Malloy, T., Fang, F., & Gonzalez, C. (2025). Experimental Eva-
uation of cognitive agents for collaboration in human-autonomy cyber defense
teams. Computers in Human Behavior: Artificial Humans, 100148.

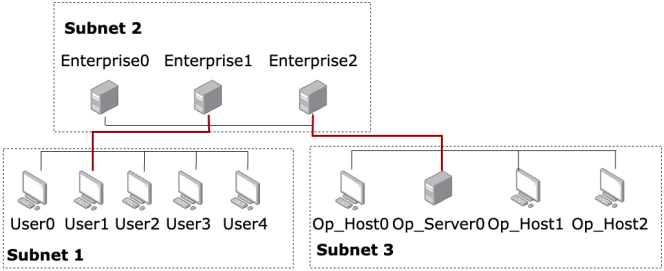
Monitor Remove Restore Misinform

Monitor Remove Restore Misinform

The HAT defend an enterprise network together

- Du, Y., Prbot, B., Malloy, T., Fang, F., & Gonzalez, C. (2025). Experimental evaluation of cognitive agents for collaboration in human-autonomy cyber defense teams. Computers in Human Behavior: Artificial Humans, 100148.

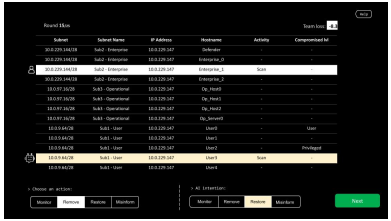
The HAT defend an enterprise network together (cont.)



HAT's goal: minimize cumulative loss

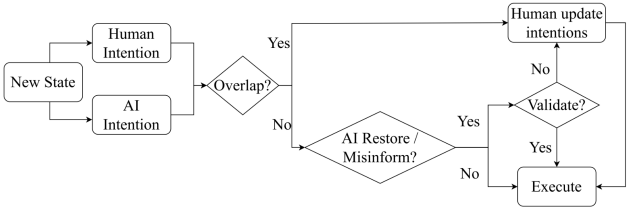
Event or Action	Defender's Loss
Attacker has administrator access on a Host	0.1
Attacker has administrator access on a Server	1
Attacker successfully Impact Op_Server0	10
Defender restore a Host or Server	1
Defender deploy a honeyfile on a host	0.5

Slide 36


Help
Round 15min
Team loss
Subnet
Subnet Name
IP Address
Hostname
Activity
Compromised Id
10.0.229.144/28
Sub2 - Enterprise
10.0.229.147
Defender
-
-
10.0.229.144/28
Sub2 - Enterprise
10.0.229.147
Enterprise_0
-
-
10.0.229.144/28
Sub2 - Enterprise
10.0.229.147
Enterprise_1
Scan
-
10.0.229.144/28
Sub2 - Enterprise
10.0.229.147
Enterprise_2
-
-
10.0.97.16/28
Sub1 - Operational
10.0.229.147
Op_Host0
-
-
10.0.97.16/28
Sub1 - Operational
10.0.229.147
Op_Host1
-
-
10.0.97.16/28
Sub1 - Operational
10.0.229.147
Op_Host2
-
-
10.0.97.16/28
Sub1 - Operational
10.0.229.147
Op_Server0
-
-
10.0.9.64/28
Sub1 - User
10.0.229.147
User0
-
User
10.0.9.64/28
Sub1 - User
10.0.229.147
User1
-
-
10.0.9.64/28
Sub1 - User
10.0.229.147
User2
-
Privileged
10.0.9.64/28
Sub1 - User
10.0.229.147
User3
Scan
-
10.0.9.64/28
Sub1 - User
10.0.229.147
User4
-
-
> Choose an action:
Monitor
Remove
Restore
Misinform
> All Intention:
Monitor
Remove
Restore
Misinform
Next

Human is responsible for resolving conflicts and validating risky AI actions

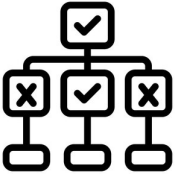
- Du, Y., Prbot, B., Malloy, T., Fang, F., & Gonzalez, C. (2025). Experimental evaluation of cognitive agents for collaboration in human-autonomy cyber defense teams. Computers in Human Behavior: Artificial Humans, 100148.



We compared HATs with cognitive agent, heuristic agent, and random agent

- Du, Y., Prbot, B., Malloy, T., Fang, F., & Gonzalez, C. (2025). Experimental evaluation of cognitive agents for collaboration in human-autonomy cyber defense teams. Computers in Human Behavior: Artificial Humans, 100148.
- Cognitive
- Heuristic
- Random

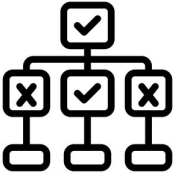
We compared HATs with cognitive agent, heuristic agent, and random agent (cont.)



We compared HATs with cognitive agent, heuristic agent, and random agent

- Cognitive
- Heuristic
- Random
- Du, Y., Prbot, B., Malloy, T., Fang, F., & Gonzalez, C. (2025). Experimental evaluation of cognitive agents for collaboration in human-autonomy cyber defense teams. Computers in Human Behavior: Artificial Humans, 100148.

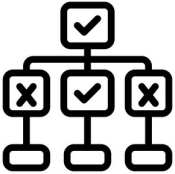
We compared HATs with cognitive agent, heuristic agent, and random agent (cont.)



We compared HATs with cognitive agent, heuristic agent, and random agent

- Cognitive
- Heuristic
- Random
- Du, Y., Prbot, B., Malloy, T., Fang, F., & Gonzalez, C. (2025). Experimental evaluation of cognitive agents for collaboration in human-autonomy cyber defense teams. Computers in Human Behavior: Artificial Humans, 100148.
- Adaptive, Competent

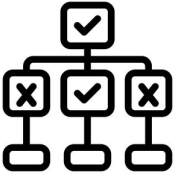
We compared HATs with cognitive agent, heuristic agent, and random agent (cont.)



We recruited human participants from Amazon Mechanical Turk (MTurk) to team up with the agents

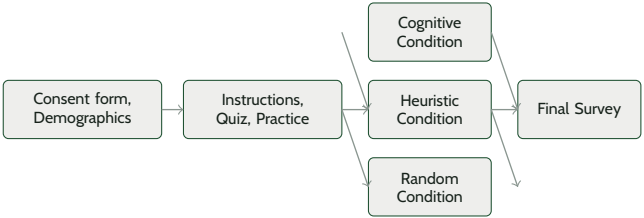
- Cognitive 48 Participants
- Heuristic 42 Participants
- Random 66 Participants
- Du, Y., Prbot, B., Malloy, T., Fang, F., & Gonzalez, C. (2025). Experimental evaluation of cognitive agents for collaboration in human-autonomy cyber defense teams. Computers in Human Behavior: Artificial Humans, 100148.

We recruited human participants from Amazon Mechanical Turk (MTurk) to team up with the agents (cont.)



Experiment Procedure

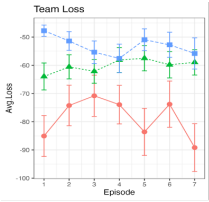
- Instructions,
- Quiz, Practice



Du, Y., Prbot, B., Malloy, T., Fang, F., & Gonzalez, C. (2025). Experimental evaluation of cognitive agents for collaboration in human-autonomy cyber defense teams. Computers in Human Behavior: Artificial Humans, 100148.

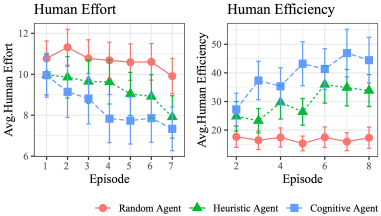
Cognitive partner had the highest performance than alternatives

- Cognitive Agent
- Team loss
- Avg. loss
- Episode



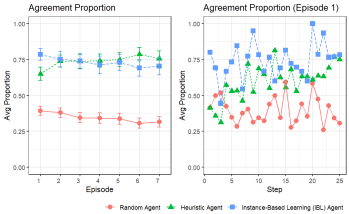
Team Loss Comparison

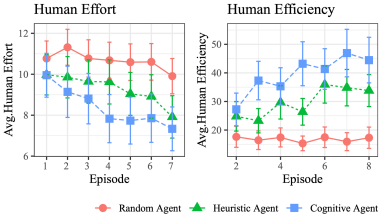
- IBL teammate agent had the highest performance across alternatives.
- IBL team loss somewhat decreased over time, heuristic somewhat increased.



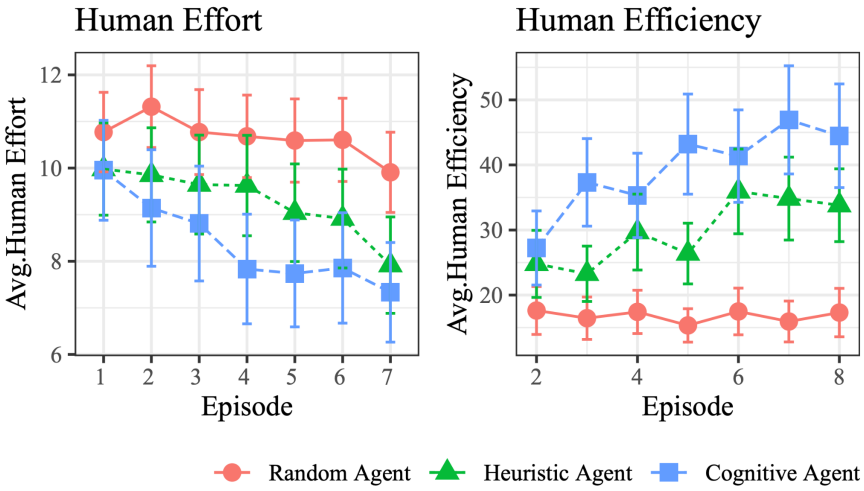
Humans quickly learn to trust competent agents

- Cognitive Agent
- Agreement Proportion
- Avg. Proportion
- Episode





Humans working with cognitive agents achieve more while doing less



Perception of agents is affected by high expectations, which later led to blind trust or great disappointment

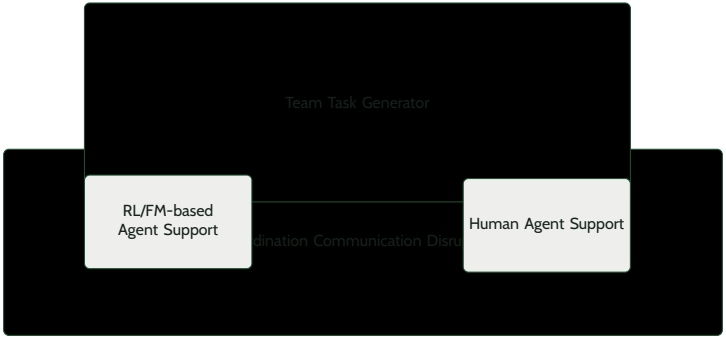
What now? Cyber defense requires coordinated, multi-step strategies

- Strategic collaboration is crucial
- A deeper understanding of task dynamics is needed

Teamwork

- Coordination
- Communication
- Disruption & Recovery
- In our ongoing work, we expand TDG platform to enable automated generation of team tasks and supports experiments with real agents.

Teamwork (cont.)



RL for Adaptive Cyber-Deception

- Human-AI Team Defense
- LLM for Cyber Group Consensus
- Other projects
- Future vision
- Q&A

RL for Adaptive Cyber-Deception (cont.)

Outline

RL for Adaptive Cyber-Deception Human-AI Team Defense Group Consensus Other projects Future vision Q&A

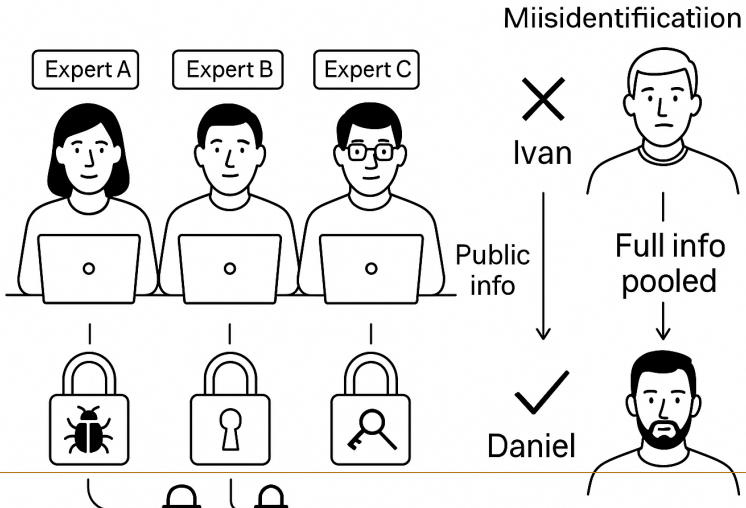
GameSec (in preparation)

AAMAS-OptLearn'22

CHB-AH'25

CogSci'24

Attacker profiling is a group task



Individuals often fail to recognize the relevance of their own knowledge and are reluctant to express disagreement

We start by exploring similar group problem solving tasks with public datasets – Winter Survival Task

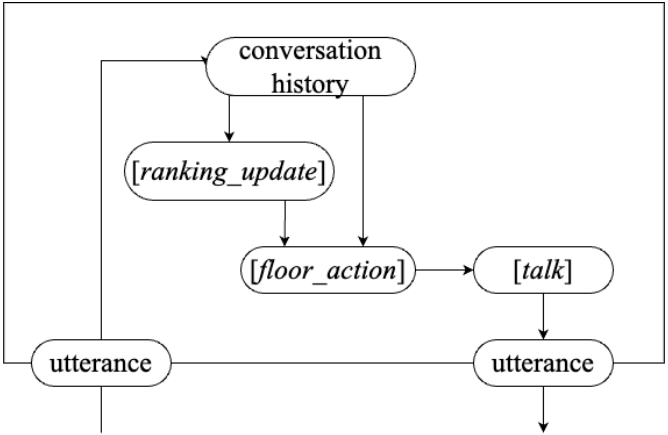
A peek into the human group conversation

So what did everyone do as one"	Pink
I did uh cigarette lighter"	Blue
For one"	Blue
Mm okay I did knife"	Pink
Knife"	Green
Okay well do knife then"	Blue
Yeah yeah makes sense"	Blue
Well I think it that and and um the cigarette lighter are all pretty important"	Pink
So I would say cigarette lighter is two"	Pink
I didn't write that down but I kinda made a connection afterwards"	Pink

we utilized the 0613 version of GPT-4-32k

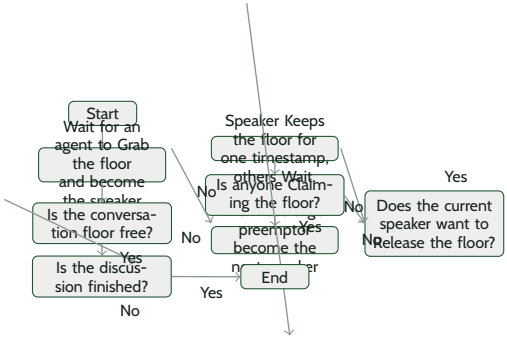
- speaker id, text)
- Update ranking based on new information
- Decide whether to grab the conversation floor
- Generate natural language to express its opinion
- Can LLM agents do better? We designed a large language model based agent for this task

we utilized the 0613 version of GPT-4-32k (cont.)



We proposed an algorithm to enable free-form conversation in groups of LLM agents

- Is anyone Claiming the floor ?



Capturing How Humans Communicate and Feel

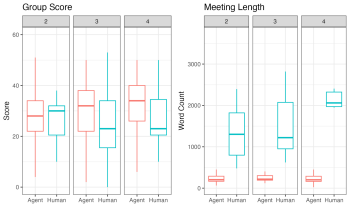
- Dialogue Acts
- Proposal
- Agreement
- Disagreement
- Confirmation

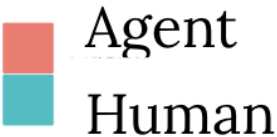
Metrics

- Performance Metrics
- Group Score — distance from expert ranking
- Meeting Length — total word count or time
- Affective Metrics
- Positivity / Negativity — based on sentiment annotations

Agent groups achieved higher scores in shorter time frames

- Group Score
- Meeting Length
- Score
- Word COunt

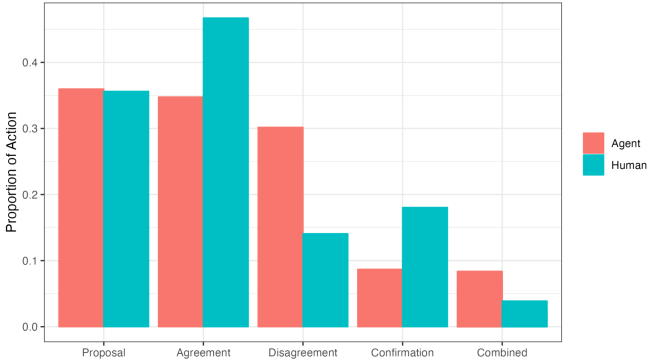




Agent discussions reveal greater disagreement among agents within a group compared to human groups.

- Agent
- Human
- Proposal
- Agreement
- Disa greement

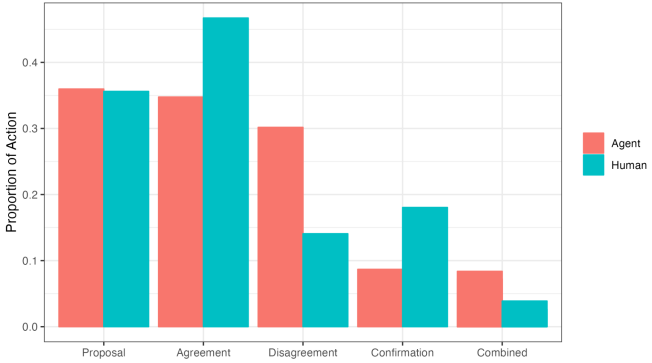
Agent discussions reveal greater disagreement among agents within a group compared to human groups. (cont.)



Agent groups also tend to produce more complex statements that contain both agreements and disagreements.

- Agent
- Human
- Proposal
- Agreement
- Disagreement

Agent groups also tend to produce more complex statements that contain both agreements and disagreements. (cont.)

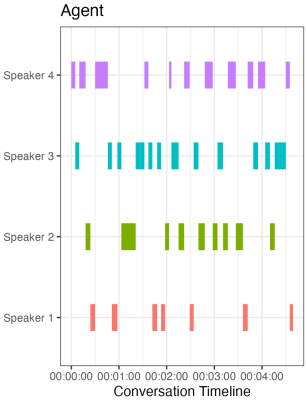


Agents participated in the discussion more equally than humans

- Speaker 1
- Speaker 2
- Speaker 3
- Speaker 4
- Agent

Agents participated in the discussion more equally than humans

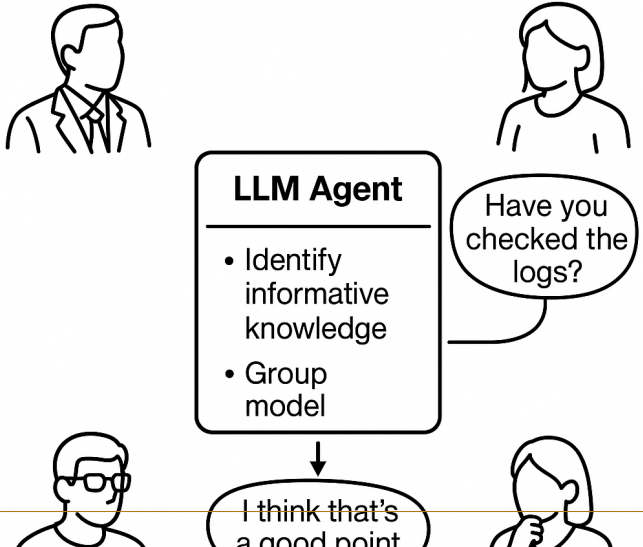
(cont.)



Moving Forward: LLM Agent -assisted group problem solving – Detecting Knowledge Gap & Group Dynamics

- Identify knowledge gap
- Model the group dynamic

Moving Forward: LLM Agent -assisted group problem solving – Detecting Knowledge Gap & Group Dynamics (cont.)



RL for Adaptive Cyber-Deception

- LLM for Group Decision-Making
- Human-AI Team Defense
- Other projects
- Future vision
- Q&A

Outline

RL for Adaptive Cyber-Deception LLM for Group Decision
Team Defense Other projects Future vision Q&A

AAMAS-OptLearn'22

GameSec (in preparation)

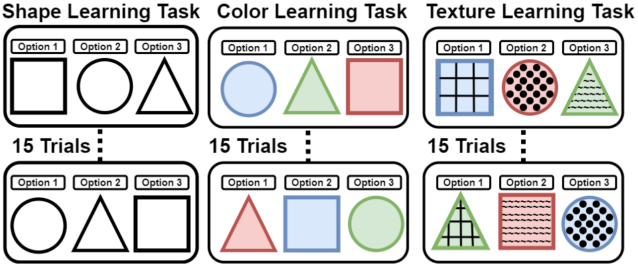
CogSci'24

CHB-AH'25

Contextual Bandits & Human Transfer-Learning Modeling

- Malloy, T., Du, Y., Fang, F., & Gonzalez, C. (2023, November). Accounting for Transfer of Learning Using Human Behavior Models. In Proceedings of the AAAI Conference on Human Computation and Crowdsourcing (Vol. 11, pp. 115-126).

Contextual Bandits & Human Transfer-Learning Modeling (cont.)

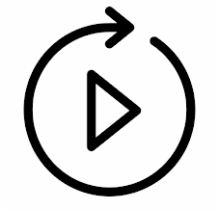


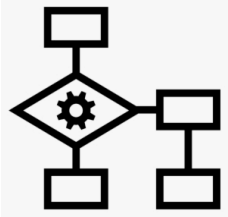
Stackelberg Security Games & Human Adversary Modeling

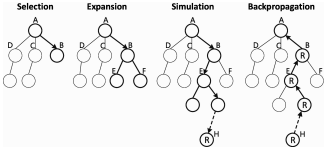
- Du, Y., Prebot, B., Malloy, T., & Gonzalez, C. (2024). A Cyber-War Between Bots: Cognitive Attackers are More Challenging for Defenders than Strategic Attackers. ACM Transactions of Societal Computing.

	Attack Replay	Rule-based Playbook	Monte-Carlo Planning	Ethical
Customizable				
Adaptive				
Scalable				

Stackelberg Security Games & Human Adversary Modeling (cont.)



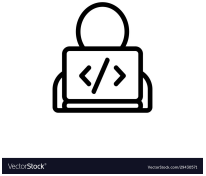


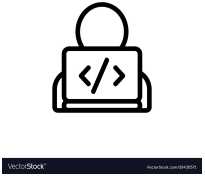


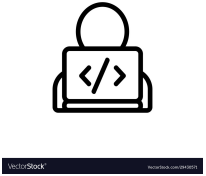
Iterated Prisoner's Dilemma & Human Groupmate Modeling

- Share?
- Not Share?
- Du, Y. , Singh, K., Aggarwal, P., Fang, F., & Gonzalez, C. Emergent Cooperative Decision-Making in Triadic Prisoner's Dilemma: Effects of Incentives and Information. Available at SSRN 5136406.

Iterated Prisoner's Dilemma & Human Groupmate Modeling (cont.)







RL for Adaptive Cyber-Deception

- LLM for Group Decision-Making
- Human-AI Team Defense
- Other projects
- Future vision
- Q&A

Outline

RL for Adaptive Cyber-Deception LLM for Group Decision Team Defense Other projects Future vision Q&A

AAMAS-OptLearn'22

GameSec (in preparation)

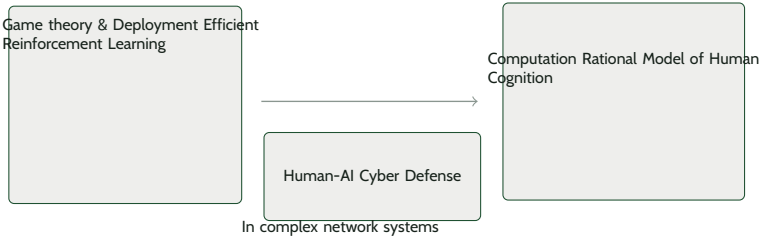
CogSci'24

CHB-AH'25

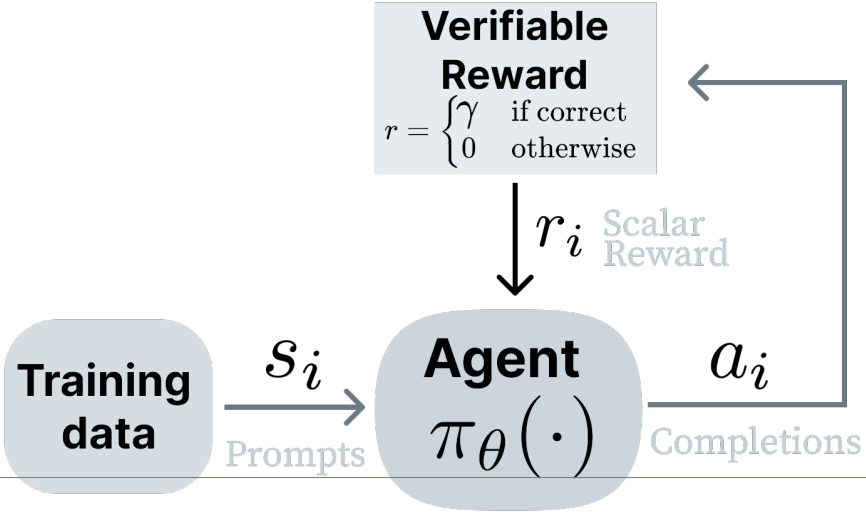
Research Vision

- Research
- Vision
- Game theory
- Deployment Efficient Reinforcement Learning

Research Vision (cont.)

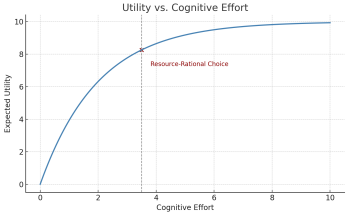


Deployment-Efficient Reinforcement Learning for Autonomous Defense



Computation Rational Behavior Models for Behavioral Cybersecurity

- Cognitive Effort
- Expected Utility



$$\max_{\pi} \mathbb{E}\left[\text{Utility}(\pi) - \lambda \cdot \text{Cost}(\pi)\right]$$

Thank you! Questions?

Funding opportunities

- Grant experience
- PhD research supported by MURI-CATCH, and ARL-CRA
- Participated in proposal writing for
- IARPA-ReSCIND
- Cylab Future Enterprise Seed Grant

Backup Slides

What now ? Cognitive attacker can capture human learning behavior, but not the biases

- Default Setting Bias
- Sunk Cost Fallacy

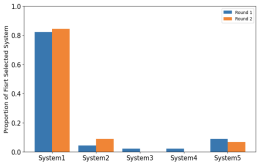


Figure 1. Proportion of participant's first system selection during the network reconnaissance phase.

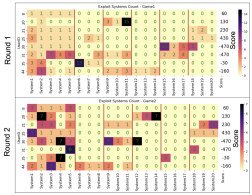
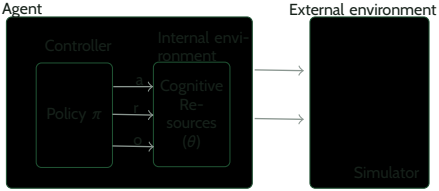


Figure 2. The number of attempts to exploit systems in Round 1 and Round 2.

In our ongoing work, we build computationally rational models of human behavior constrained by their internal environment, that is, cognition.

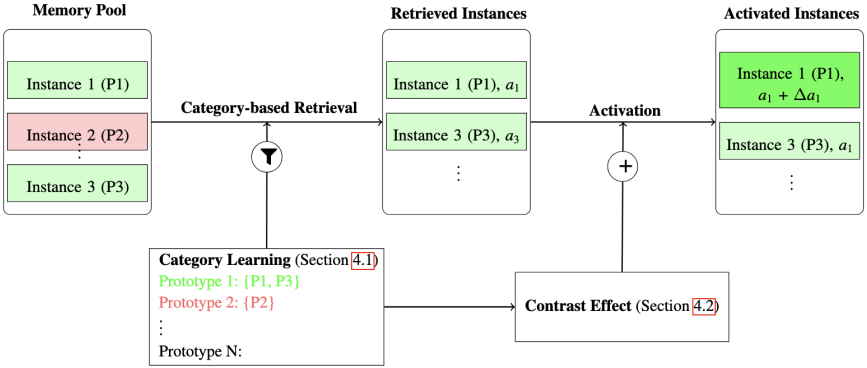
- Cognitive Resources
- (θ)
- Internal



Composition Reasoning, Causal Inference, & Optimal Forgetting

$$R(x) = \frac{1}{n}\sum_{i=1}^{n} R_i(x[S_i])$$

$$Q(x[S_i]) \leftarrow Q(x[S_i]) + \alpha(Q(x[S_i]) - \Phi_i o_{m,n})$$

$$\Phi_i \leftarrow \bar{\Phi}_i + \omega(\bar{\Phi}_i - o_{m,n})$$

Dynamic Weighting, Category Learning, & Contrast Effect