
CREW: Facilitating Human-AI Teaming Research

Lingyu Zhang, Zhengran Ji, Boyuan Chen

Duke University

{lingyu.zhang, zhengran.ji, boyuan.chen}@duke.edu

<http://generalroboticslab.com/CREW>

Abstract

With the increasing deployment of artificial intelligence (AI) technologies, the potential of humans working with AI agents has been growing at a great speed. Human-AI teaming is an important paradigm for studying various aspects when humans and AI agents work together. The unique aspect of Human-AI teaming research is the need to jointly study humans and AI agents, demanding multidisciplinary research efforts from machine learning to human-computer interaction, robotics, cognitive science, neuroscience, psychology, social science, and complex systems. However, existing platforms for Human-AI teaming research are limited, often supporting oversimplified scenarios and a single task, or specifically focusing on either human-teaming research or multi-agent AI algorithms. We introduce **CREW**, a platform to facilitate Human-AI teaming research and engage collaborations from multiple scientific disciplines, with a strong emphasis on human involvement. It includes pre-built tasks for cognitive studies and Human-AI teaming with expandable potentials from our modular design. Following conventional cognitive neuroscience research, CREW also supports multimodal human physiological signal recording for behavior analysis. Moreover, CREW benchmarks real-time human-guided reinforcement learning agents using state-of-the-art algorithms and well-tuned baselines. With CREW, we were able to conduct 50 human subject studies within a week to verify the effectiveness of our benchmark.

1 Introduction

Over the past decade, significant progress in machine learning has increased the potential and necessity for humans to collaborate and interact with Artificial Intelligence (AI) agents. Human-AI teaming has emerged as a research paradigm to explore the dynamic interactions and collaborative processes between humans and AI. By leveraging the complementary strength of both humans and AI, advancements can significantly enhance the overall team performance.

Unlike traditional AI research, which typically focuses on machine learning algorithms in isolation, Human-AI teaming requires a multidisciplinary approach to incorporate insights from various scientific domains. Numerous studies have examined human-human teaming [29] with cognitive science, neuroscience, and psychology. Machine learning and robotics communities have extensively researched multi-agent machine learning [61], while team dynamics [44] has been explored in complex systems, social science, and network science. Despite the importance and potential of this research paradigm, there is still a lack of a comprehensive and unified platform to benefit research on joint efforts across disciplines and scalable hypothesis verification.

Developing a comprehensive platform for Human-AI teaming research presents several unique challenges. Firstly, the platform needs to support diverse tasks with varying complexities with easy

extensions for new tasks or modifications. While reinforcement learning research platforms [51] have widely adopted this practice, current Human-AI teaming platforms remain limited to single tasks [13, 54]. Secondly, enabling real-time communication through various modalities between multiple humans and heterogeneous AI policies is essential for effective collaboration. However, existing solutions typically support human feedback only through replaying offline trajectories and do not implement real-time feedback mechanisms. Understanding how to build AI that can team with, learn from, be guided by, align with, and augment humans is as crucial as modeling human behavior during interactions with AI. Therefore, the platform must provide interfaces for recording human physiological data alongside agent data, tailored for cognitive science and neuroscience studies. Furthermore, current studies often involve small participant groups, making it difficult to derive generalizable conclusions. Lastly, the absence of a unified platform has limited fully open-sourced studies and the establishment of appropriate benchmarks.

We present CREW, a platform designed to facilitate Human-AI teaming research aiming to address the above challenges. We develop CREW around key principles such as extensible and open environment design, real-time communication support, hybrid Human-AI teaming modes, parallel sessions for scalable experiments, and comprehensive human and agent data collection. Additionally, CREW incorporates highly modular algorithm components compatible with practices in the machine learning community. We demonstrate CREW’s potential by benchmarking real-time human-guided reinforcement learning (RL) algorithms alongside various RL baselines. With CREW, we were able to conduct 50 human subject studies within a week. Moreover, CREW includes a set of cognitive tests to explore how individual differences among humans impact their effectiveness in training AI agents. To our knowledge, CREW is the first platform to unify the desired features for Human-AI teaming research across multiple disciplines. We aim for CREW to serve as an infrastructural foundation for multidisciplinary, reproducible, and scalable Human-AI teaming research.

2 Related Work

Human-AI Teaming Research Extensive research has explored Human-AI teaming across various domains. Machine learning studies have developed algorithms to leverage human expertise to improve the accuracy [33], robustness [8, 63], and interpretability [33, 8, 62] of models. Integrating human feedback can not only improve performance [27, 18] but also align the models with human preference [33, 40, 63]. Human-computer interaction research has created interfaces [56, 37] and workflows [9] that enhance collaboration between humans and AI, combining their strengths to achieve superior performance. Ethical research focuses on understanding and mitigating the societal [24, 17], ethical [19, 59, 43], and technical [22, 49] challenges of the rapid advancement and wide adoption of AI. Many fields, including neuroscience, healthcare, robotics, transportation, education, security, and accessibility, have shown growing interest [38, 25, 6, 39, 42, 32] in Human-AI teaming. Overall, the broad spectrum of interests highlights the need for multidisciplinary collaboration to drive further advancements.

Human-AI Teaming Platform While significant progress has been made in Human-AI teaming research, there remains an absence of a comprehensive research platform. Overcooked Environment [13] is a simplified version of the original popular game to challenge human agents and AI agents in tasks that require close coordination and strategic teamwork. StarCraft II Learning Environment (SC2LE) [54] supports adversarial settings to allow Human-AI interaction and learning from human demonstrations. Rapid Integration and Development Environment(RIDE) [52] focuses on defense-related scenarios, emphasizing rapid development and integration of AI systems for operational purposes. In addition to real-time decision-making tasks, previous research has developed platforms [21, 36, 60] that focus on offline preference or rating settings where humans can provide offline evaluations or corrections with imitation learning or reinforcement learning.

However, existing platforms have more than one of the following limitations that constrain Human-AI teaming research as summarized in Tab. 1. Most environments only support one type of task that can be difficult to extend, and the interactions between humans and machine learning agents are limited to mouse and keyboard operations. Moreover, most of the environments only support the data collection

Tab. 1: Human-AI teaming platforms

Platform	Extensible envs	Real-time interaction	Multiplayer	Multimodal feedback	Physiological data	Fully open-sourced
CREW	✓	✓	No Limit	✓	✓	✓
Overcooked [13]		✓	1+1			
SC2LE [54]		✓	1 v 1			
RIDE [52]	✓	✓	No Limit			
RLHF-Blender [36]		✗*		✓		✓
Uni-RLHF [60]		✗*				

Tab. 2: *Instead of real-time feedback training, RLHF-Blender and Uni-RLHF support online training where humans are presented with data from the replay buffer instead of the current experience.

on the game data, such as state, action, or reward focusing on the machine learning algorithm development, but none of them support the collection and analysis of human physiological data (eye movement, pupillometry, brain activity, heart impulse, or natural language) to understand human cognitive states and different effects on human subjects. Furthermore, the scale of collaboration has been limited to two agents in a collaborative or competitive setting. Notably, most existing Human-AI teaming studies have only been evaluated on a very small number of subjects or among the authors, which greatly limits our understanding of the state-of-the-art performance. There remains a large gap between the existing platforms and the desired environment to facilitate future research.

Real-Time Human-Guided Learning Most real-world decision-making processes require rapid decisions and adaptation in real time. Therefore, real-time human-guided learning is an essential topic in Human-AI teaming research. Previous research has focused on algorithm design to integrate real-time human feedback into policy training. TAMER and Deep TAMER [30, 55] learn to predict human discrete feedback and utilize the feedback model as the value function for policy learning. COACH and DeepCOACH [5, 35] addressed the inconsistency of human feedback by modeling it as the advantage function under an actor-critic framework. Recent progress has further refined these algorithms, addressing various challenges in human feedback integration, such as different feedback modalities feedback stochasticity [4], and continuous action spaces [46]. Other works have explored indirectly inferring human objectives from feedback or preference [26, 57, 18]. CREW provides an extensive environment for real-time human-guided learning with online training and feedback integration, human physiological data collection, and parallel distributed experiment support.

3 CREW: Design and Components

3.1 Platform Vision and Design Philosophy

We design CREW to facilitate Human-AI teaming research. Our vision is to construct a unified platform for researchers from diverse backgrounds, allowing them to join forces from human-AI interaction, machine learning algorithms, workflow design, as well as human analysis and training. To achieve this, CREW incorporates the following capabilities, as illustrated in Fig. 1.

Extensible and open environment design. CREW provides built-in tasks for rapid development and allows users to integrate customized tasks to accommodate the limitless applications of Human-AI teaming.

Real-time communication. While some Human-AI interaction tasks, such as human preference-based fine-tuning, can be performed offline, many applications require online real-time interaction. Whether it is training decision-making models with real-time human guidance or general human-AI collaboration tasks, the ability to convey messages with minimum delay is essential. Synchronizing data flow between human interfaces, AI algorithms, and simulation engines necessitates the establishment of a real-time communication channel.

Hybrid Human-AI teaming support. Teaming is an essential aspect of our daily jobs. Our vision extends this concept to Human-AI teaming, where both humans and AI operate in teams. Increasing interest in the organization, dynamics, workflow, and trust in multi-human and multi-AI teams highlights the need for a platform capable of distributing and synchronizing tasks, states, and interactions across multiple environment instances and even across physical locations.

Parallel sessions support. A key bottleneck for human-involved AI research is the requirement to conduct experiments with dozens or hundreds of human subjects to obtain trustworthy and

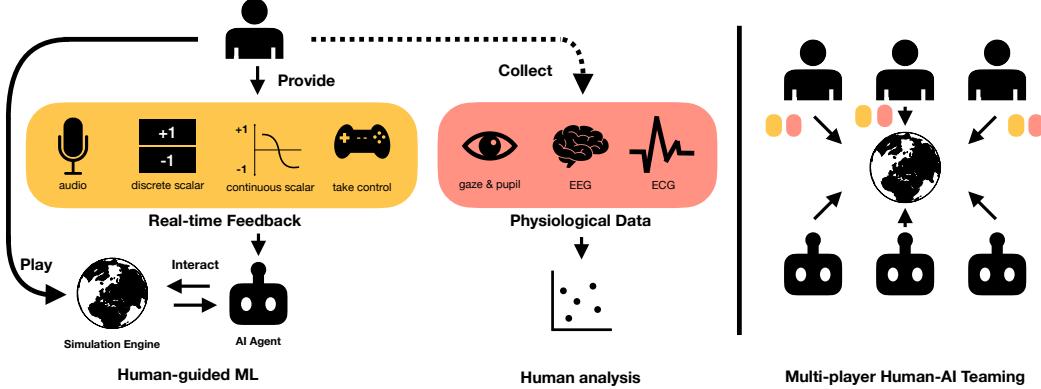


Fig. 1: **CREW** is a platform to facilitate Human-AI teaming research. CREW is designed under the vision of multidisciplinary collaboration research from both human and AI science. CREW allows real-time interaction among multi-human and multi-agents while enabling extensive data collection on both AI agents and human agents.

reliable conclusions. Such a process can be tedious and time-consuming. To enhance efficiency and scalability, CREW supports multiple independent parallel sessions of the same setting, unconstrained by geographical locations, to obtain the “crowd-sourcing” effects of large-scale experiments. This capability enables experimenters to collectively share experimental data and results.

Comprehensive human data collection. Though human plays an important role in Human-AI teaming, our understanding of human behaviors remains limited and under-explored in existing studies. Therefore, CREW offers interfaces to simultaneously collect multi-modal human data, ranging from active instructions and feedback to passive physiological signals.

ML community-friendly algorithm design. The choice of programming language and libraries should align with the customs and preferences of the ML community. The system design should be modular to allow seamless transitions between tasks and algorithms.

3.2 Environments

Tasks We select Unity as the simulation engine for CREW due to its popularity in game design and AI research to allow extensible and open environment design. We have implemented four challenging tasks as examples. Multi-player tasks are designed for multi-agent and multi-human teaming research, and single-player tasks are designed for human-guided AI agent learning studies. For each task, we provide both visual and structured state input options. The detailed settings are summarized in Tab. 3.

Bowling is a modified version of Atari bowling where each round consists of 10 rolls and the score for each roll is the number of pins hit. Bowling is designed to have the simplest dynamics among our tasks to serve as a rapid test for training a single agent. **Find Treasure** (Fig. 2A) is a single-player embodied navigation task where the agent’s goal is to explore a maze and reach the treasure with randomized initial and goal locations. **1v1 Hide-and-Seek** [14, 15] (Fig. 2B) is a one-on-one hide-and-seek task where the seeker learns to explore the maze and catch a moving hider that follows a heuristic policy for obstacle avoidance, and run away from the seeker within line of sight. We introduce this task as an adversarial competition setting. **NvN Hide-and-Seek** (Fig. 2C) is a multiplayer version where multiple seekers and hiders can coordinate, collaborate, and compete. The hiders and seekers can either be controlled by humans or heterogeneous AI policies.

Tab. 3: Task Specifications

Tasks	Visual Observation	State Observation	Action Space	Reward
Bowling	grayscale	ball pos, pin pos, pin status	release pos, length before steer, steer direction	# pins hit
Find Treasure	rgb	agent pos, treasure pos	next loc x, next loc y	-1 / step, +10 treasure found
1v1 Hide-and-Seek	rgb	seeker pos, hider pos	next loc x, next loc y	-1 / step, +10 hider caught
NvN Hide-and-Seek	rgb	seekers pos, hiders pos	next loc x, next loc y	user define

Visual Observations Different visual observations create various challenges in visual embodiment learning for both humans [53, 20] and AI agents [7, 50]. We provide various camera configurations

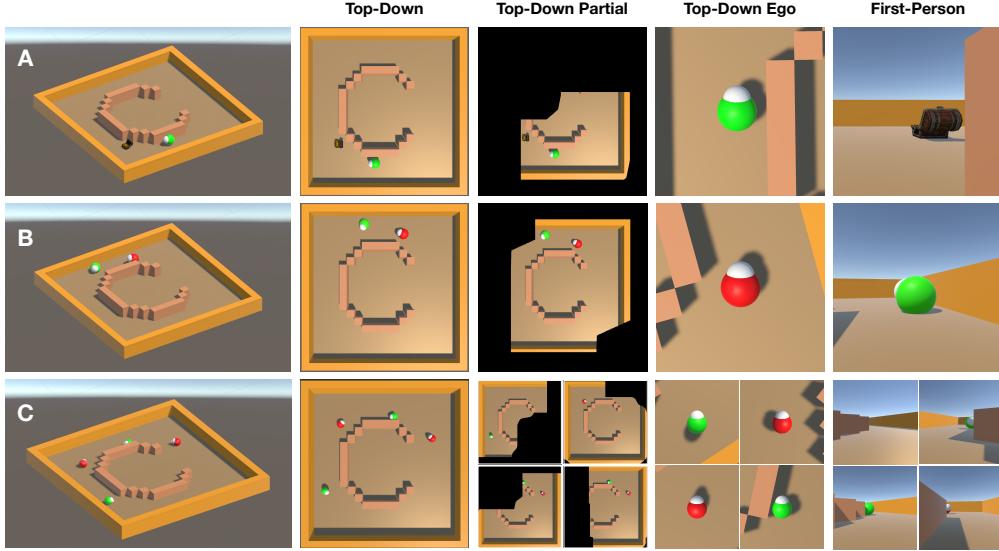


Fig. 2: CREW supports multiple tasks from single agent (A: Find Treasure) to multi-agent competitive setting (B: 1v1 Hide-and-Seek), and multi-agent collaborative and competitive setting (C: NvN Hide-and-Seek). We also show different camera views supported by CREW for perceptual-motor research. As shown in Fig 2, for all our navigation and competitive tasks, we implemented visual observations from multiple views for the users’ selection: a top-down fully observable view, a top-down accumulated partially observable view similar to SLAM in robotics [48], a top-down egocentric view, and a first-person view.

Procedural Generation Learning robust, generalizable, and scalable AI agents requires diverse training environments. Procedural generation allows for the creation of a wide range of environment patterns and terrain types. We provide randomized maze patterns where the grids are guaranteed to be connected (Fig. 3A) and procedural-generated terrains for more complex simulations as in Fig. 3B.

3.3 Human and Agent Role Assignment

Humans and AI agents often have complementary strengths. For example, humans are generally better at exploring and adapting to new situations, while AI agents are good at repetitive exploitation and precise calculation. Naturally, a team consists of humans and AI agents should have various roles to be effective. Different roles can also be assigned within AI agents to study multi-agent machine learning with heterogeneous policies. To facilitate these experiments, we provide the role assignment feature in CREW.

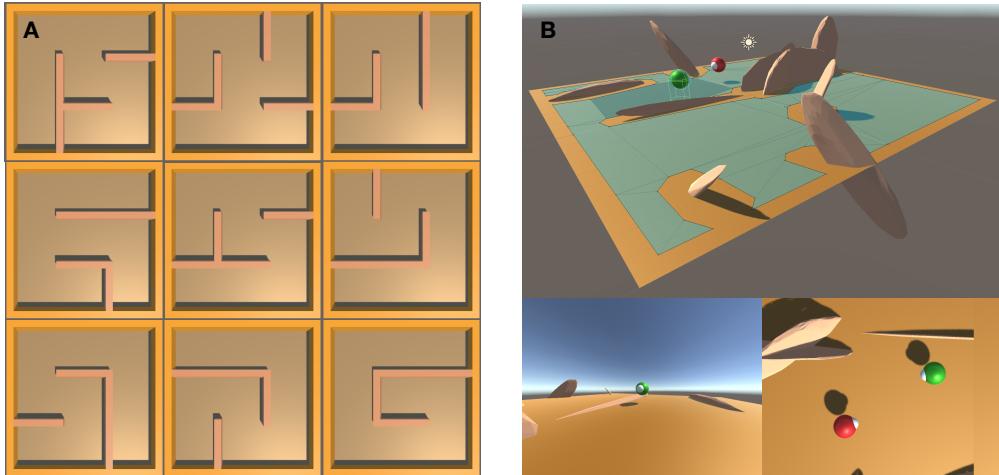


Fig. 3: Environment generation in CREW. (A) Randomized maze. (B) Procedure generated terrains.

“Player” allows humans to directly control an agent. “Viewer” allows humans to observe and provide feedback signals as guidance to an AI agent. “Server” role allows humans to host the entire task by monitoring it without direct participation. “AI Agent” assigns different learning policies to each agent.

Human Feedback Interface We provide a user interface to allow the Viewer role to provide direct feedback to AI agents shown in Fig. 4. Scalar feedback is the most common feedback type used in human-guided RL [30, 55, 57]. Conventionally, a human chooses to provide a positive or negative rating to the state-action pair of an agent policy. Additionally, CREW offers an interface that allows humans to provide feedback that is continuous in value and time, enabling granular feedback information. Moreover, our interface also allows humans to directly take control over agents and perform teleoperation.

3.4 Multiplayer and Parallel Sessions

Enabling multi-human multi-agent sessions requires robust networking solutions (Fig. 5). We use Unity Netcode [2] for game state synchronization, and Nakama [1] as the networking server. In CREW, a server instance hosts the task, runs the simulation, and handles agent policy training, which can be executed on a powerful headless GPU server. Human participants can connect via client instances on less powerful machines, which display synchronized game states and collect human input. CREW is cross-platform, allowing participation from Linux, Windows or MacOS machines.

3.5 Human and Agent Data Collection

Data collection is at the core of Human-AI teaming research. CREW includes a pipeline for thorough data collection on both the human side and AI agent side.

Human data Besides the feedback interfaces that collect feedback signals of multiple modalities and teleoperation actions, we also provide interfaces for collecting audio, eye gaze, pupillometry, electroencephalogram (EEG), and electrocardiogram (ECG) physiological responses as in Fig. 6.

Agent data including the policy weights, observations, actions, rewards, feedback received, and loss values at every time step can all be saved for further analysis. Users also have the option to enable experiment monitoring and logging by Weights & Biases [3]. As all of our tasks include vision-based settings, we also provide implementations of a set of vision encoder architectures.

3.6 Designing Algorithms

Algorithms research is crucial for Human-AI teaming. We designed the algorithm component of CREW to be extensible and accessible to the ML community. Algorithms are implemented in Python

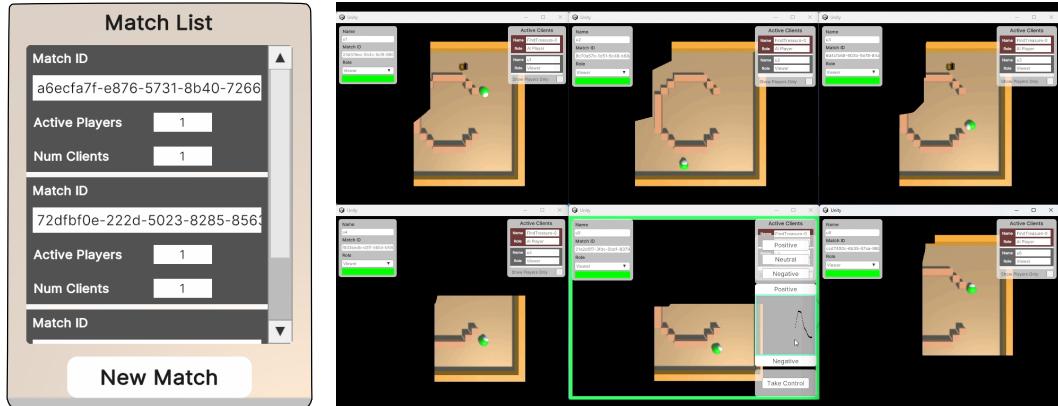


Fig. 5: In CREW, it is possible to connect and join multiple instances of tasks. It is as simple as typing in an IP address and selecting the task one wants to join.

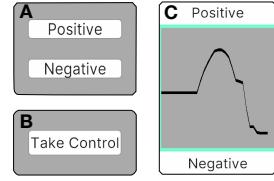


Fig. 4: (A) Discrete scalar feedback. (B) Option to take control of the agent and teleoperate. (C) Continuous scalar feedback: the human can hover the mouse over this window to provide per-step feedback.

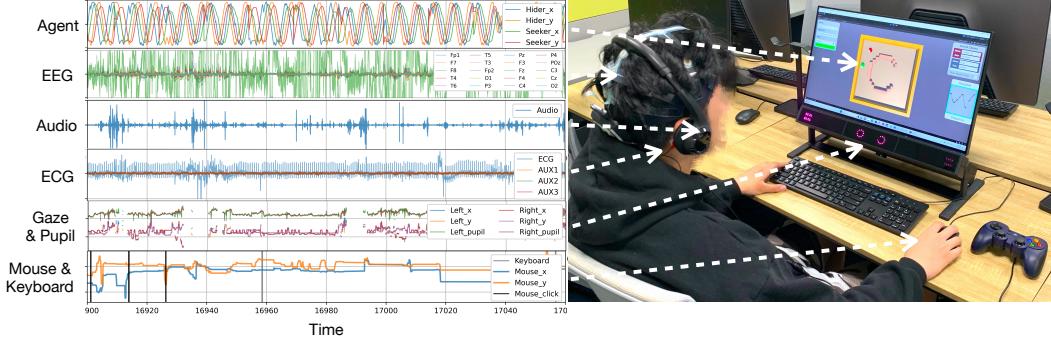


Fig. 6: Data collection in CREW. All human and agent data, including game states, feedback, mouse and keyboard, audio, gaze and pupil data, EEG and ECG data are all streamed and synchronized through Lab Streaming Layer [31].

with PyTorch [41] and TorchRL [12] which is part of the PyTorch ecosystem, making it easy for researchers to design and deploy new algorithms. We provide implementations of state-of-the-art human-guided machine learning algorithms and strong reinforcement learning baselines. The API for communication between a Unity instance and Python algorithm is implemented with MLAgents [28]. All environments follow a uniform communication protocol for the observation and action data across tasks. Our modular design allows smooth switching between tasks and algorithms. Real human experiments are time-consuming and costly. To ease algorithm debugging, we implemented simulated human feedback providers as surrogates for real humans for complex tasks. These simulated feedback uses heuristics based on prior task knowledge and is not available in novel tasks in the real world; hence, they should be used solely as debugging tools, not as replacements for real human evaluations.

3.7 Modular Pipeline Design for Quantifying Human Characteristics

Individual differences among humans can significantly affect their teaming with AI agents. To support research along this line, CREW supports a set of cognitive tests to quantify these differences shown in Tab. 4. We provide a modular and convenient pipeline (Fig. 7) for executing cognitive tests and Human-AI experiments. The framework integrates various media files (e.g., instruction videos or pictures), inter-trial intervals, executable Python scripts, and Unity builds, ensuring a smooth and effective workflow. The pipeline requires minimal effort from researchers during proctoring, as a single click initiates the sequential execution of experiments. The pipeline allows restart from interruption points.

4 Benchmarking Study

As an example of running experiments with CREW, we benchmark a state-of-the-art real-time human-guided RL framework, Deep TAMER [55], along with strong RL baselines. In the original Deep TAMER, the framework was only tested on Atari Bowling with 9 human subjects. With CREW, this is the first time it is possible to systematically conduct human-guided RL benchmarking across multiple environments on a larger population. We summarize our findings as well as insights on the scalability of the framework in this section. We also discuss the relationship between human characteristics and guided agent performance.

4.1 Experiment Setup

Tasks We selected 3 single-player games: Bowling, Find Treasure, and 1v1 Hide-and-Seek for this benchmark. For Find Treasure and Hide-and-Seek, each episode has a time limit of 15 seconds. All algorithms directly learn from visual inputs with the top-down accumulated partially observable view.

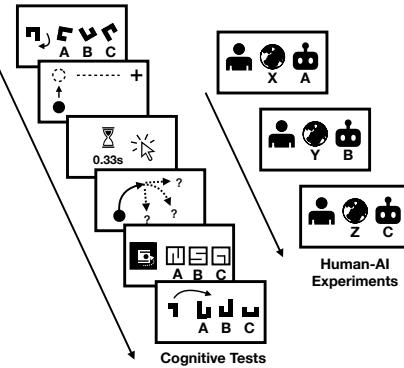


Fig. 7: Modular pipeline for experiment setup. Experimenters can freely select and organize the order of cognitive tests and Human-AI experiments with CREW’s deployment pipeline.

Tab. 4: Cognitive Tests Specifications

Tests	Rule	Score Metric
Eye Alignment (eye) [45]	Align a ball on the left side of the screen with a target on the right side as accurately as possible within five seconds.	Negative average of the distances between the ball and target's horizontal positions across trials.
Reflex (reflex) [11]	Click the screen as quickly as possible when it changes from yellow to green.	Negative average of the response times.
Theory of Behavior (theory) [16]	Observe a red ball moving in an unknown but fixed pattern for 5s. When the ball pauses, predict the ball's position one second after it resumes moving.	Negative average of the distances between the ball's actual and predicted positions across trials.
Mental Rotation (rotation) [47]	Identify the piece among three similar pieces that cannot be rotated to match the target piece.	Accuracy of the subject's identifications across all trials.
Mental Fitting (fitness) [47]	Identify the only piece among three similar pieces that can fit with the target piece.	Accuracy of the subject's identifications across all trials.
Spatial Mapping (spatial) [10]	A video of an agent navigating a maze with a restricted field of view is presented. Identify the maze from a selection of three similar mazes.	Accuracy of the subject's identifications across all trials.

Human Trainers We recruited 50 human subjects for the experiments under the approval of the Institutional Review Board. For each 1.5-hour session, the subject will receive \$20 compensation. Benefiting from CREW’s unique feature to host experiments with parallel sessions, we were able to conduct all 50 experiments within one week. Prior to guiding the agents, the participants were asked to complete all our cognitive tests in Tab. 4.

Deep TAMER leverages human feedback as value functions. During training, a human trainer provides positive or negative discrete feedback based on the agent’s behavior. This feedback is assigned to relevant state-action pairs through a credit assignment window. A neural network is trained to predict human feedback given state-action pairs, and the policy chooses actions that maximize this predicted feedback. Originally, Deep TAMER was limited to discrete actions. We extend it to continuous action spaces using an actor-critic framework, termed as c-Deep TAMER. For bowling, the agent is trained for 5 minutes, and for 10 minutes in Find Treasure and Hide-and-Seek.

Baseline RL We selected two state-of-the-art RL algorithms: Deep Deterministic Policy Gradient (DDPG) [34] and Soft Actor-Critic (SAC) [23].

Heuristic feedback We also evaluate simulated feedback-guided RL. We simply add the feedback signals as additional dense rewards to the environment reward. DDPG is selected as the backbone RL algorithm as the state-of-the-art transitioned from SAC to DDPG [58] in visual control tasks.

Evaluation We checkpoint model weights every 2 minutes and evaluate on unseen test environments. For bowling, every checkpoint is evaluated for 1 game (10 rolls). For Find Treasure and 1v1 Hide-and-Seek, the checkpoints are evaluated for 100 episodes.

4.2 Results

Agent training performance We hypothesize that subjects with higher cognitive tests can lead to higher-performing agents. Therefore, we show the agent performances guided by the 15 subjects who scored the highest in our cognitive tests side by side with the performance of all 50 subjects in Fig. 8. As shown in the results, the agents guided by the top 15 subjects exhibit higher performance than the overall average. In particular for the top 15 subjects, on the simple bowling task, c-Deep TAMER surpasses RL baselines by an average of 10 scores given the same training time. On Find Treasure, heuristic feedback achieved the highest performance as expected, showing the upper bound performance with accurate and non-delayed dense rewards. c-Deep TAMER also demonstrated strong performance with faster learning trends than RL baselines. On 1v1 Hide-and-Seek, c-Deep TAMER performed similarly to RL baselines, suggesting that c-Deep TAMER has difficulty scaling to tasks with higher complexity. Similar conclusions still hold for all 50 subjects.

Analysis of Individual Differences Due to the cognitive test feature and emphasis on Human-AI teaming, we can deepen our understanding of how individual human differences can affect the performance of human-guided agents. We calculated the correlation between human subjects’ cognitive test scores and c-Deep TAMER training results in Fig. 9. The cognitive test scores are normalized by the mean and variance over the subjects through z-score, and outliers with scores $1.5 \times$

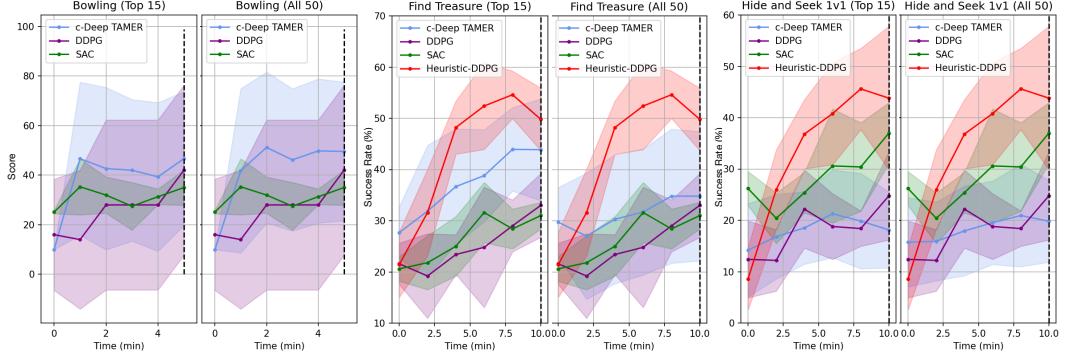


Fig. 8: The average result of the human subjects with the top 15 cognitive test scores and the average result across all 50 subjects. Subjects scoring higher on cognitive tests are generally better at guiding RL. We also find that c-Deep TAMER has difficulties scaling to more complex tasks.

Interquartile Range above the third quartile or below the first quartile are removed. We found that non-outlier subjects performed equally well on the spatial mapping test so we do not include this for correlation analysis. The heatmap shows the sign of correlation coefficients from linear regression and the statistical significance (*).

The Find Treasure score has a significant positive correlation with theory of behavior, mental rotation, and fitness scores, suggesting that subjects who performed better on these tests also trained better agents. This is likely due to the need for good spatial reasoning and future behavior estimation in Find Treasure. The 1v1 Hide-and-Seek score shows a significant positive correlation with reflex score, likely due to the rapid changes in environment and task dynamics in the complex adversarial setting. Subjects with faster reaction speeds provided timely feedback to align with relevant state-action pairs. Overall, the total score (i.e., the sum of three game scores) has significant positive correlations with rotation, fitness, and overall score, suggesting that these cognitive skills are most relevant to the performance of c-Deep TAMER-guided agents.

5 Conclusion, Limitation, and Future Work

We introduce CREW for facilitating Human-AI teaming research from diverse human and machine learning scientific communities. CREW offers extensible environment design, enables real-time human-AI communication, supports hybrid Human-AI teaming, parallel sessions, multimodal feedback, and physiological data collection, and features ML community-friendly algorithm design. We also provide a set of built-in tasks and baseline algorithm implementations. Using CREW, we benchmarked a state-of-the-art human-guided RL algorithm against baseline methods involving 50 human subjects and provided insights into the relationship between individual human differences and agent-guiding performance.

CREW is still in the early efforts among several critical aspects. Future work will explore building more diverse and challenging tasks, including multiplayer tasks with complex strategies and robotics environments requiring an understanding of physics. While we have only benchmarked several algorithms, we hope CREW can help benchmark many existing algorithms that were not fully open-sourced in a unified environment. Finally, more supports on human physiological data processing and analysis shall be investigated and supported in CREW.

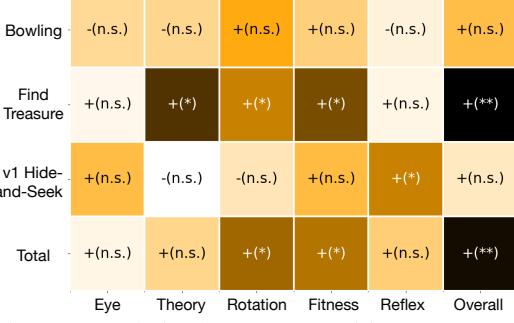


Fig. 9: Correlation between cognitive test scores and c-Deep TAMER training performance. The darker the color, the more statistically significant the correlation. “+” or “-” means positive or negative correlation.

6 Acknowledgements

We thank the members at Duke General Robotics Lab, Sida Zhu and Harish Yerra for early development on the platform, and Jiaxun Liu and Easop Lee for beta testing. This work is supported in part by ARL STRONG program under awards W911NF2320182 and W911NF2220113.

References

- [1] Nakama. URL <https://github.com/heroiclabs/nakama?tab=readme-ov-file>.
- [2] Unity netcode. URL <https://unity.com/products/netcode>.
- [3] Weights & biases. URL <https://wandb.ai/site>.
- [4] R. Arakawa, S. Kobayashi, Y. Unno, Y. Tsuboi, and S.-i. Maeda. Dqn-tamer: Human-in-the-loop reinforcement learning with intractable feedback. *arXiv preprint arXiv:1810.11748*, 2018.
- [5] D. Arumugam, J. K. Lee, S. Saskin, and M. L. Littman. Deep reinforcement learning from policy-dependent human feedback. *arXiv preprint arXiv:1902.04257*, 2019.
- [6] S. Atakishiyev, M. Salameh, H. Yao, and R. Goebel. Explainable artificial intelligence for autonomous driving: A comprehensive overview and field guide for future research directions. *arXiv preprint arXiv:2112.11561*, 2021.
- [7] A. Bandini and J. Zariffa. Analysis of the hands in egocentric vision: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 45(6):6846–6866, 2020.
- [8] Y. Bao, S. Chang, M. Yu, and R. Barzilay. Deriving machine attention from human rationales. *arXiv preprint arXiv:1808.09367*, 2018.
- [9] D. Bau, H. Strobelt, W. Peebles, J. Wulff, B. Zhou, J.-Y. Zhu, and A. Torralba. Semantic photo manipulation with a generative image prior. *arXiv preprint arXiv:2005.07727*, 2020.
- [10] M. Berkowitz, A. Gerber, C. M. Thurn, B. Emo, C. Hoelscher, and E. Stern. Spatial abilities for architecture: Cross sectional and longitudinal assessment with novel and existing spatial ability tests. *Frontiers in psychology*, 11:609363, 2021.
- [11] C. J. Boes. The history of examination of reflexes. *Journal of neurology*, 261(12):2264–2274, 2014.
- [12] A. Bou, M. Bettini, S. Dittert, V. Kumar, S. Sodhani, X. Yang, G. De Fabritiis, and V. Moens. Torchrl: A data-driven decision-making library for pytorch. *arXiv preprint arXiv:2306.00577*, 2023.
- [13] M. Carroll, R. Shah, M. K. Ho, T. Griffiths, S. Seshia, P. Abbeel, and A. Dragan. On the utility of learning about humans for human-ai coordination. *Advances in neural information processing systems*, 32, 2019.
- [14] B. Chen, S. Song, H. Lipson, and C. Vondrick. Visual hide and seek. In *Artificial Life Conference Proceedings 32*, pages 645–655. MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . . , 2020.
- [15] B. Chen, Y. Hu, R. Kwiatkowski, S. Song, and H. Lipson. Visual perspective taking for opponent behavior modeling. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 13678–13685. IEEE, 2021.
- [16] B. Chen, C. Vondrick, and H. Lipson. Visual behavior modelling for robotic theory of mind. *Scientific Reports*, 11(1):424, 2021.

- [17] S. Chowdhury, P. Budhwar, P. K. Dey, S. Joel-Edgar, and A. Abadie. Ai-employee collaboration and business performance: Integrating knowledge-based view, socio-technical systems and organisational socialisation framework. *Journal of Business Research*, 144:31–49, 2022.
- [18] P. F. Christiano, J. Leike, T. Brown, M. Martic, S. Legg, and D. Amodei. Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30, 2017.
- [19] N. Ezer, S. Bruni, Y. Cai, S. J. Hepenstal, C. A. Miller, and D. D. Schmorow. Trust engineering for human-ai teams. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 63, pages 322–326. SAGE Publications Sage CA: Los Angeles, CA, 2019.
- [20] H. D. Fishbein, S. Lewis, and K. Keiffer. Children’s understanding of spatial relations: Coordination of perspectives. *Developmental psychology*, 7(1):21, 1972.
- [21] P. Freire, A. Gleave, S. Toyer, and S. Russell. Derail: Diagnostic environments for reward and imitation learning. *arXiv preprint arXiv:2012.01365*, 2020.
- [22] B. Green and Y. Chen. The principles and limits of algorithm-in-the-loop decision making. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–24, 2019.
- [23] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning*, pages 1861–1870. PMLR, 2018.
- [24] P. Hemmer, M. Westphal, M. Schemmer, S. Vetter, M. Vössing, and G. Satzger. Human-ai collaboration: The effect of ai delegation on human task performance and task satisfaction. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, pages 453–463, 2023.
- [25] K. E. Henry, R. Kornfield, A. Sridharan, R. C. Linton, C. Groh, T. Wang, A. Wu, B. Mutlu, and S. Saria. Human–machine teaming is key to ai adoption: clinicians’ experiences with a deployed machine learning system. *NPJ digital medicine*, 5(1):97, 2022.
- [26] J. Huang, J. Hao, R. Juan, R. Gomez, K. Nakamura, and G. Li. Gan-based interactive reinforcement learning from demonstration and human evaluative feedback. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4991–4998. IEEE, 2023.
- [27] B. Ibarz, J. Leike, T. Pohlen, G. Irving, S. Legg, and D. Amodei. Reward learning from human preferences and demonstrations in atari. *Advances in neural information processing systems*, 31, 2018.
- [28] A. Juliani, V.-P. Berges, E. Teng, A. Cohen, J. Harper, C. Elion, C. Goy, Y. Gao, H. Henry, M. Mattar, et al. Unity: A general platform for intelligent agents. *arXiv preprint arXiv:1809.02627*, 2018.
- [29] R. Klimoski and S. Mohammed. Team mental model: Construct or metaphor? *Journal of management*, 20(2):403–437, 1994.
- [30] W. B. Knox and P. Stone. Interactively shaping agents via human reinforcement: The tamer framework. In *Proceedings of the fifth international conference on Knowledge capture*, pages 9–16, 2009.
- [31] C. Kothe, S. Y. Shirazi, T. Stenner, D. Medine, C. Boulay, M. I. Crivich, T. Mullen, A. Delorme, and S. Makeig. The lab streaming layer for synchronized multimodal recording. *bioRxiv*, pages 2024–02, 2024.
- [32] N. Kumar and A. Jain. A deep learning based model to assist blind people in their navigation. *J. Inf. Technol. Educ. Innov. Pract.*, 21:95–114, 2022.

- [33] P. Lertvittayakumjorn, L. Specia, and F. Toni. FIND: human-in-the-loop debugging deep text classifiers. *CoRR*, abs/2010.04987, 2020. URL <https://arxiv.org/abs/2010.04987>.
- [34] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.
- [35] J. MacGlashan, M. L. Littman, D. L. Roberts, R. Loftin, B. Peng, and M. E. Taylor. Convergent actor critic by humans. In *International Conference on Intelligent Robots and Systems*, 2016.
- [36] Y. Metz, D. Lindner, R. Baur, D. Keim, and M. El-Assady. Rlhf-blender: A configurable interactive interface for learning from diverse human feedback. *arXiv preprint arXiv:2308.04332*, 2023.
- [37] M. A. Neerincx, J. van der Waa, F. Kaptein, and J. van Diggelen. Using perceptual and cognitive explanations for enhanced human-agent team performance. In *Engineering Psychology and Cognitive Ergonomics: 15th International Conference, EPCE 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15–20, 2018, Proceedings 15*, pages 204–214. Springer, 2018.
- [38] I. R. Nourbakhsh, K. Sycara, M. Koes, M. Yong, M. Lewis, and S. Burion. Human-robot teaming for search and rescue. *IEEE Pervasive Computing*, 4(1):72–79, 2005.
- [39] H. S. Nwana. Intelligent tutoring systems: an overview. *Artificial Intelligence Review*, 4(4):251–277, 1990.
- [40] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744, 2022.
- [41] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.
- [42] A. D. Pazho, C. Neff, G. A. Noghre, B. R. Ardabili, S. Yao, M. Baharani, and H. Tabkhi. Ancilia: Scalable intelligent video surveillance for the artificial intelligence of things. *IEEE Internet of Things Journal*, 2023.
- [43] M. Pflanzer, Z. Traylor, J. B. Lyons, V. Dubljević, and C. S. Nam. Ethics in human–ai teaming: principles and perspectives. *AI and Ethics*, 3(3):917–935, 2023.
- [44] E. Salas, D. L. Reyes, and S. H. McDaniel. The science of teamwork: Progress, reflections, and the road ahead. *American Psychologist*, 73(4):593, 2018.
- [45] D. Scharre. Sage: A test to detect signs of alzheimer’s and dementia. *The Ohio State University Wexner Medical Center*, 2014.
- [46] I. Sheidlower, E. S. Short, and A. Moore. Environment guided interactive reinforcement learning: Learning from binary feedback in high-dimensional robot task environments. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, pages 1726–1728, 2022.
- [47] R. N. Shepard and J. Metzler. Mental rotation of three-dimensional objects. *Science*, 171(3972):701–703, 1971.
- [48] R. C. Smith and P. Cheeseman. On the representation and estimation of spatial uncertainty. *The international journal of Robotics Research*, 5(4):56–68, 1986.
- [49] B. C. Stahl, A. Andreou, P. Brey, T. Hatzakis, A. Kirichenko, K. Macnish, S. L. Shaelou, A. Patel, M. Ryan, and D. Wright. Artificial intelligence for human flourishing—beyond principles for machine learning. *Journal of Business Research*, 124:374–388, 2021.

- [50] A. P. Tarun, N. M. Baig, J. Chang, R. Tanvir, S. Shihipar, and A. Mazalek. Third eye: Exploring the affordances of third-person view in telepresence robots. In *Social Robotics: 11th International Conference, ICSR 2019, Madrid, Spain, November 26–29, 2019, Proceedings 11*, pages 707–716. Springer, 2019.
- [51] M. Towers, J. K. Terry, A. Kwiatkowski, J. U. Balis, G. d. Cola, T. Deleu, M. Goulão, A. Kallinteris, A. KG, M. Krimmel, R. Perez-Vicente, A. Pierré, S. Schulhoff, J. J. Tai, A. T. J. Shen, and O. G. Younis. Gymnasium, Mar. 2023. URL <https://zenodo.org/record/8127025>.
- [52] USC Institute for Creative Technologies. Rapid integration & development environment (ride), 2024. URL <https://ride.ict.usc.edu/>. Accessed: 2024-05-27.
- [53] K. M. Vander Heyden, M. Huizinga, M. E. Raijmakers, and J. Jolles. Children’s representations of another person’s spatial perspective: Different strategies for different viewpoints? *Journal of experimental child psychology*, 153:57–73, 2017.
- [54] O. Vinyals, T. Ewalds, S. Bartunov, P. Georgiev, A. S. Vezhnevets, M. Yeo, A. Makhzani, H. Küttler, J. Agapiou, J. Schrittwieser, et al. Starcraft ii: A new challenge for reinforcement learning. *arXiv preprint arXiv:1708.04782*, 2017.
- [55] G. Warnell, N. Waytowich, V. Lawhern, and P. Stone. Deep tamer: Interactive agent shaping in high-dimensional state spaces. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.
- [56] J. D. Weisz, M. Muller, S. Houde, J. Richards, S. I. Ross, F. Martinez, M. Agarwal, and K. Talamadupula. Perfection not required? human-ai partnerships in code translation. In *26th International Conference on Intelligent User Interfaces*, pages 402–412, 2021.
- [57] B. Xiao, Q. Lu, B. Ramasubramanian, A. Clark, L. Bushnell, and R. Poovendran. Fresh: Interactive reward shaping in high-dimensional state spaces using human feedback. *arXiv preprint arXiv:2001.06781*, 2020.
- [58] D. Yarats, R. Fergus, A. Lazaric, and L. Pinto. Mastering visual continuous control: Improved data-augmented reinforcement learning. *arXiv preprint arXiv:2107.09645*, 2021.
- [59] C. Yin, B. Zhou, Z. Yin, and J. Wang. Local privacy protection classification based on human-centric computing. *Human-centric computing and information sciences*, 9(1):33, 2019.
- [60] Y. Yuan, J. Hao, Y. Ma, Z. Dong, H. Liang, J. Liu, Z. Feng, K. Zhao, and Y. Zheng. Unirlfhf: Universal platform and benchmark suite for reinforcement learning with diverse human feedback. *arXiv preprint arXiv:2402.02423*, 2024.
- [61] K. Zhang, Z. Yang, and T. Başar. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook of reinforcement learning and control*, pages 321–384, 2021.
- [62] J. Zhou and F. Chen. Towards trustworthy human-ai teaming under uncertainty. In *IJCAI 2019 workshop on explainable AI (XAI)*, 2019.
- [63] D. M. Ziegler, N. Stiennon, J. Wu, T. B. Brown, A. Radford, D. Amodei, P. Christiano, and G. Irving. Fine-tuning language models from human preferences. *arXiv preprint arXiv:1909.08593*, 2019.

A Accessing CREW

CREW is a platform for cross-disciplinary human-AI teaming research. The codebase for our platform, including environments, interfaces, and algorithms can be accessed through <https://github.com/generalroboticslab/CREW>. Videos, documentation, and tutorial are hosted on <http://generalroboticslab.com/CREW>. Our platform is fully open-sourced for academic use.

B Platform details

The environments of CREW is implemented using Unity 2021.3.24f1, with packages ML Agents 2.3.0-exp.3 [28], Netcode for GameObjects 1.3. [2] and Nakama Unity 3.6.0. [1]. Algorithms are developed with torchrl 0.3.0 [12].

C Additional Results

C.1 Feedback Visualization

In Fig 10 we show examples of the state-action pairs and human assigned feedback value during c-Deep TAMER training. Since humans trainers do not see the action (next navigation destination) explicitly, we show consecutive frames instead.

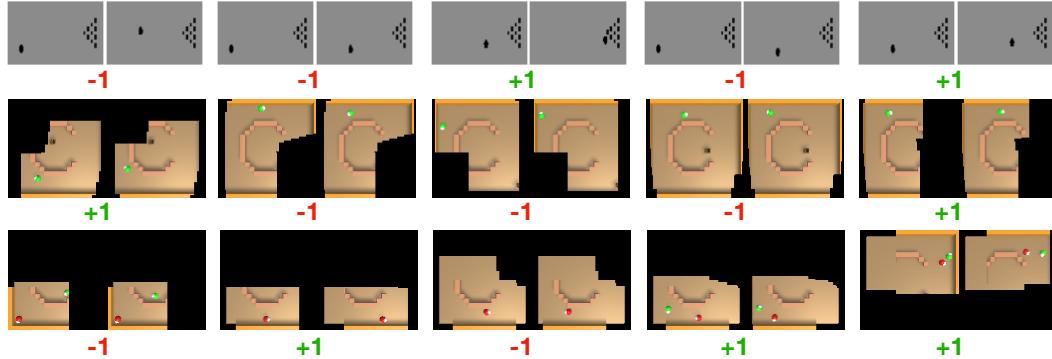


Fig. 10: c-Deep TAMER training examples.

C.2 Full Cognitive Test Analysis

We include the full results for the cognitive test and c-Deep TAMER score analysis. All linear regression plots including coefficients and p-values and summarized in Fig 11.

D Computational Resources

All human subject experiments were conducted on desktops with one NVIDIA RTX 4080 GPU. All evaluations were run on a headless server with $8 \times$ NVIDIA RTX A6000 and NVIDIA RTX 3090 Ti.

E Benchmarking Experiment Details

E.1 Implementation details

c-Deep TAMER Algorithm The detailed c-Deep TAMER algorithm is summarized in Alg 1. It is modified from the original Deep TAMER integrated with an actor network for action selection. The actor is updated by gradient ascent that outputs the actions that maximizes the predicted human feedback.

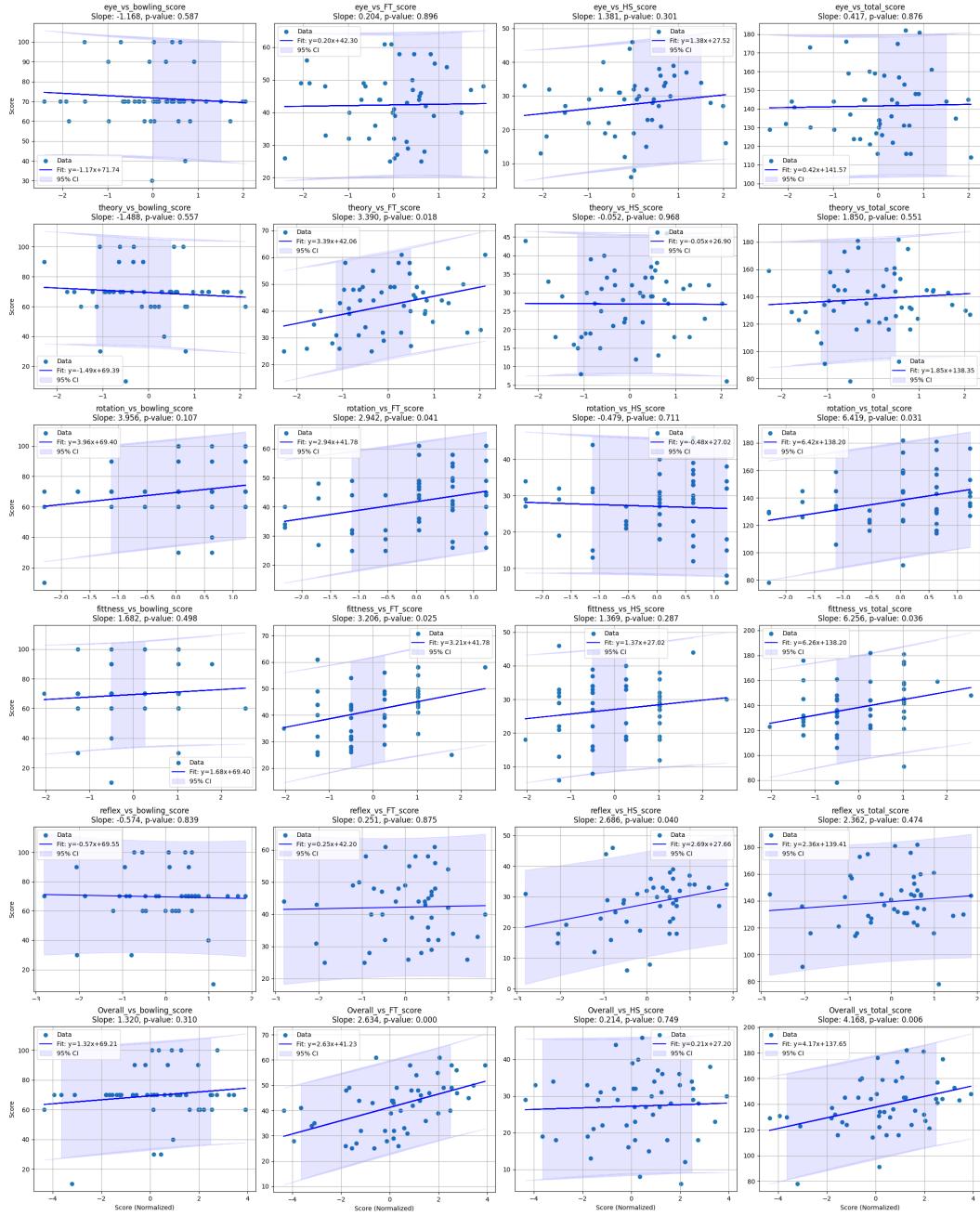


Fig. 11: Linear Regression Plots of Correlation Analysis

Processing Inputs All tasks in our experiments directly learn from pixels. The frames rendered from the environment are first resized to 100×100 pixels. For Find Treasure and 1v1 Hide-and-Seek, we stack the past 3 frames as input. We then apply simple random shift to the frame stack, as it has been shown to be an effective augmentation strategy for visual reinforcement learning. The shifting factors along the height and width are uniformly sampled from $[0, 0.08]$.

Model architecture We use a 3-layer CNN as the vision encoder, each having 64 channels and followed by batch normalization and ReLU activation function. All actor-critic frameworks uses a 3-layer MLP with 256 neurons in each layer for both the actor and critic network.

E.2 Hyperparameters

The hyperparameter settings for our experiments is summarized in Tab 5.

Algorithm 1 The c-Deep TAMER algorithm.

Require: initial policy A , feedback model \hat{H} , empty replay buffer \mathcal{D} , step size η , buffer update interval b , credit assignment window w

Init: $j = 0, k = 0$

- 1: **for** $i = 1, 2, \dots$ **do**
- 2: **observe** state s_i , time t_i
- 3: **execute** action $a_i = \text{clip}(A(s_i) + \epsilon, a_{Low}, a_{High})$, where $\epsilon \sim \mathcal{N}$
- 4: $x_i = (s_i, a_i, t_i, t_{i+1})$
- 5: **if** new feedback y **then**
- 6: $j = j + 1$
- 7: $y_j = y$
- 8: $\mathcal{D}_j = \{(x, y_j) \mid w(x, y_j) \neq 0\}$
- 9: $\mathcal{D} = \mathcal{D} \cup \mathcal{D}_j$
- 10: **update** \hat{H} by one step of gradient descent using
$$\nabla \hat{H} \frac{1}{|\mathcal{D}_j|} \sum_{(x,y) \in \mathcal{D}_j} (\hat{H}(s, a) - y)^2$$
- 11: **update** A by one step of gradient ascent using
$$\nabla A \frac{1}{|\mathcal{D}_j|} \sum_{(x,y) \in \mathcal{D}_j} (\hat{H})(s, A(s))$$
- 12: **update** target networks
- 13: $k = k + 1$
- 14: **end if**
- 15: **if** $\text{mod}(i, b) == 0$ and $\mathcal{D} \neq \emptyset$ **then**
- 16: Randomly sample a batch of transitions, $B = \{(s, a, y)\}$ from \mathcal{D}
- 17: **update** \hat{H} by one step of gradient descent using
$$\nabla \hat{H} \frac{1}{|\mathcal{D}_j|} \sum_{(x,y) \in \mathcal{D}_j} (\hat{H}(s, a) - y)^2$$
- 18: **update** A by one step of gradient ascent using
$$\nabla A \frac{1}{|\mathcal{D}_j|} \sum_{(x,y) \in \mathcal{D}_j} (\hat{H})(s, A(s))$$
- 19: **update** target networks
- 20: $k = k + 1$
- 21: **end if**
- 22: **end for**

E.3 Human subject experiment details

Overview We recruit the human subject on campus by using flyers and emails. For every human subject, the experiment time is approximately 40 minutes without interruptions. The experiment starts with cognitive tests (10 minutes) and is followed by the human guiding the agent using the c-Deep TAMER framework (30 minutes). The order of the cognitive tests is Eye Alignment, Reflex, Theory of Behavior, Mental Rotation, Mental Fitting, and Spatial Mapping. There are detailed instruction videos for each test before the test starts. As for the human guiding agent part, each human subject guides the agent to play 3 games for a total of 30 minutes (5 minutes for Bowling, 10 minutes for Find Treasure, and 10 minutes for 1v1 Hide-and-Seek). Before each game, there is a detailed instruction video that describes the rule of the game and how human subjects can give feedback to the agent, which we included below.

Tab. 5: Hyperparameters

	c-Deep TAMER	DDPG	SAC
γ	0.99	0.99	0.99
learning rate	1e-4	1e-4	1e-4
max_grad_norm	0.1	0.1	0.1
batch size	16	240	240
frames per batch	8	240	240
alpha_init	-	-	0.1
target entropy	-	-	-6.0
actor scale_lb	-	-	1e-4
# Q value nets	-	2	2
target update polyak	0.995	0.995	0.995
actor exploration noise	$\mathcal{N}(0, 0.1)$	$\mathcal{N}(0, 0.1)$	-
credit assignment window	bowling[0.2, 4], others[0.2, 1]	-	-

E.3.1 Instruction Video Script

General Introduction:

Welcome to the General Robotics Lab's Human AI Collaboration Experiment. Today, our session will start with preliminary cognitive and gaming proficiency tests to gauge your initial skills. Following this, we will delve into the main experiments where you will interact with AI algorithms through a series of engaging games. Our session will conclude with a short survey to capture your feedback on the experience. Your participation is invaluable in advancing our understanding of human AI interactions. Thank you for joining us today, and let's begin.

Cognitive Test Introduction:

Welcome to the Cognitive Test segment of our experiment. In this session, you will participate in five interactive games designed to assess various cognitive skills for about seven minutes. These tests will challenge your precision, reflexes, predictive abilities, problem-solving skills, and spatial awareness through engaging activities. Each test is brief, and you'll receive clear instructions before each one begins. Between each trial of each game, there will be a three-second intertrial interval where the computer screen will turn white with a Gray cross in the middle. Please focus on the center of the cross as much as possible during this time. Let's get started and see how you do.

Eye Alignment Instruction:

In this experiment, your goal is to align the ball positioned on the left side of the screen with the square on the right as accurately as possible. Each trial lasts for five seconds, and you'll have a total of six trials. Use your mouse to drag the ball during each trial. The time bar at the top center of the screen shows how much time you have for each trial. As time goes down, the bar gets smaller and changes color from green to red.

Reflex Instruction:

In this experiment, after a three-second countdown, the screen turns yellow. You must click as quickly as possible when the screen changes from yellow to green. Clicking during the yellow phase will result in a failure, and failing to click within two seconds after the screen turns green will also fail. You'll have a total of 6 trials.

Theory of Behavior Instruction:

In this experiment, watch the red ball moving for five seconds, then guess where it will be one second after it pauses. Click on the map to mark your prediction. You have only one chance to click. The closer you are, the higher your score. Remember, you only have two seconds to make your guess after the ball pauses. There will be 6 trials in total. The time bar at the top left shows the time for each trial's observation and prediction as time decreases. The bar shrinks and changes from green to red.

Mention Rotation and Mental Fitting Instruction:

In this experiment, you will need to answer 12 questions in total, each with only one correct answer. Click the button to choose the option that you think best answers the question. For each question, you have 8 seconds to view and respond. The time bar at the top right of the screen indicates how much time you have for each question. As time decreases, the bar shrinks and changes color from green to red.

Spatial Mapping Instruction:

In this experiment, you will need to answer six questions in total, each with only one correct answer. Watch the video and click the button to choose the option that you think best answers the question. For each question, you are free to answer while the video is playing, and you will also have three extra seconds after the video is paused to answer the question. The time bar at the top right of the screen indicates how much time you have for each question. As time decreases, the bar shrinks and changes color from green to red.

Human Guiding AI Introduction:

Welcome to the Human-Guiding AI section of the experiment. You will guide the AI to play three games for about 30 minutes in total in this section. You'll give feedback to the AI agents based on their performance in each game. Before starting each game, please enter your name in the name box and click on the Connect button. A match list will appear. Click on the Match button in the list. It will turn brown when clicked. Then click the Join Match button to enter the match. Once you've joined the match, select the AI agent in the Active Clients panel at the top right corner of the screen by clicking on it. The agent button will turn brown to indicate that you've selected it successfully. Now you can observe the agents and provide feedback. You'll constantly click on the positive and negative buttons to inform the agent of its performance. Positive means it's doing well. Negative means it's doing poorly.

Bowling Instruction:

Each episode of this bowling game consists of 10 rolls. At the start of each roll, there are 10 pins positioned on the right side of the screen. The agent then launches a ball in an attempt to knock down the pins, with the number of pins knocked down determining the score for that roll. The total score for the episode is calculated as the sum of the scores from all 10 rolls. Please guide the AI agent to maximize the total score across all rolls.

Find Treasure Instruction:

In this game, the AI agent, represented by the green character, begins with a limited field of view, only able to perceive surroundings, while the rest

of the map is obscured by shadow. As the agent navigates the map, areas it has visited become revealed within its field of view. Please guide the AI agent to locate the treasure and navigate to it, represented by the brown chest, as quickly as possible.

1v1 Hide-and-Seek Instruction:

In this game, the AI agent, represented by the red character, begins with a limited field of view, only able to perceive surroundings, while the rest of the map is obscured by shadow. As the agent navigates the map, areas it has visited become revealed within its field of view. Guide the AI agent to chase the hider, represented by the green character, and catch it as quickly as possible.

E.3.2 Compensation

We pay each human subject \$20 for participation.

F Author Statement

The authors bear all responsibility in case of violation of rights.