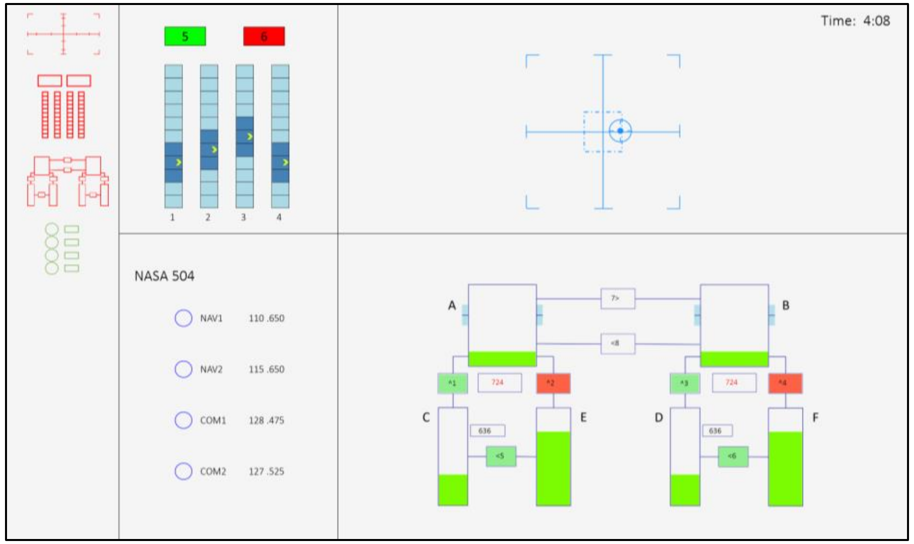
A Human-Aware Decision Making System for Human-Robot Teams

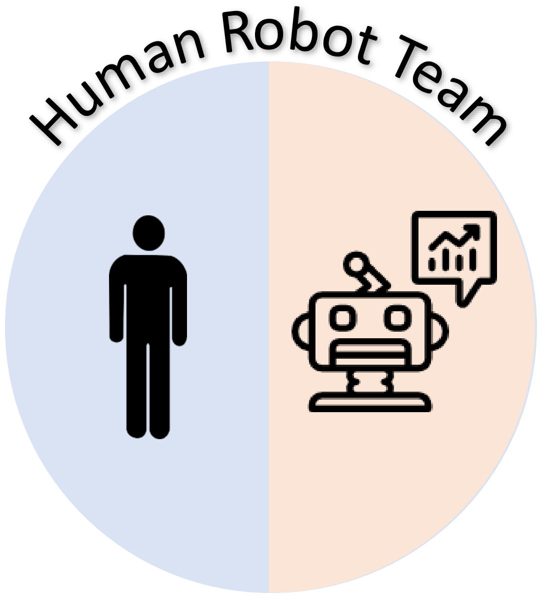
Saurav Singh
Electrical and Computer Engineering Department
Rochester Institute of Technology
Rochester, NY, USA
ss3337@rit.edu

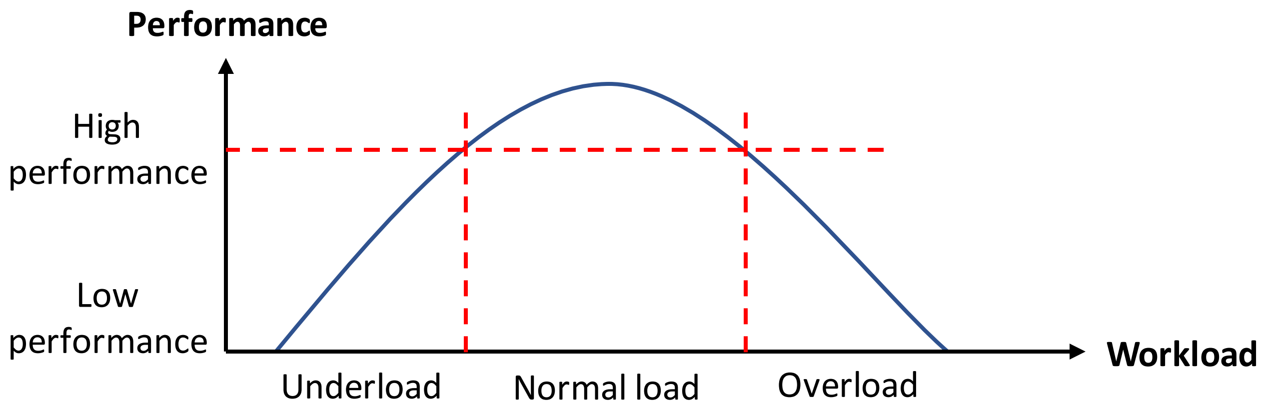
Jamison Heard
Electrical Engineering Department
Rochester Institute of Technology
Rochester, NY, USA
jrheee@rit.edu



Motivation

- Robots can understand what is going on around them using various sensors.
- What about human behavior and human's internal states?
- Human physiological data changes with change in human internal states.
  - ➢ High heart rate under high workload



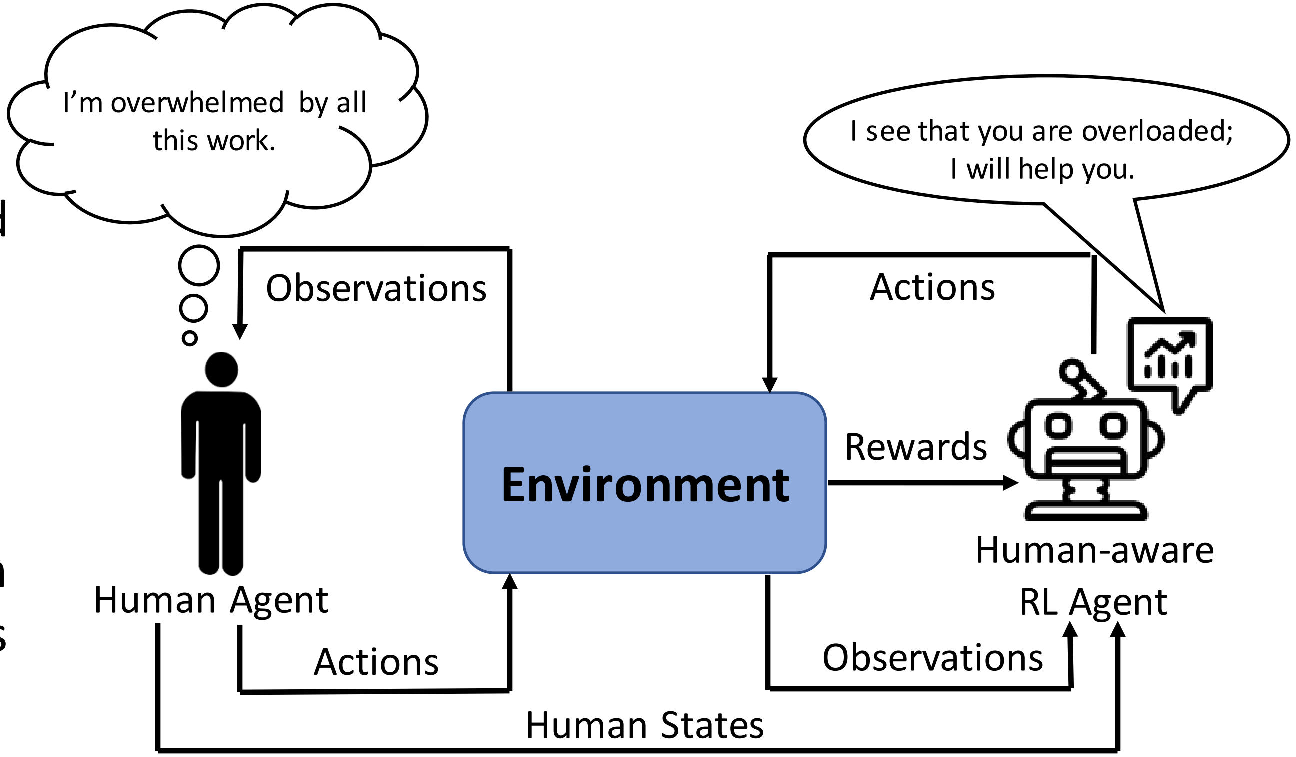


[1] C. D. Wickens, J. D. Lee, Y. Liu, and S. E. G. Becker, An Introduction to Human Factors Engineering, 2nd ed. Pearson Education, Inc., 2004.

Research Question

- How can we leverage the most from estimated human states to adapt robot's behavior?
  - ➢ Existing rule-based implementations

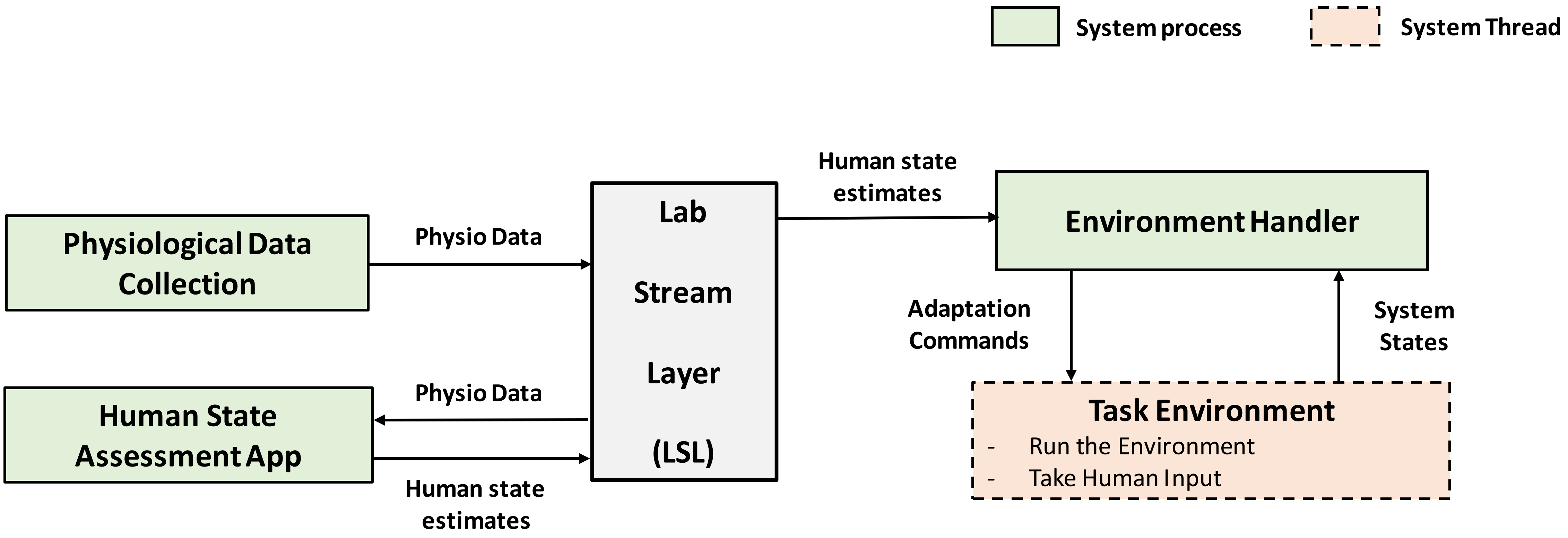
- We need a human-aware decision-making system that can learn the trends between human states and overall team performance.



Can we use human states such as workload, fatigue and comfort in a Reinforcement Learning paradigm to improve human-robot team performance?

[2] A. H. Memar and E. T. Esfahani, "Objective assessment of human workload in physical human-robot cooperation using brain monitoring," J. Hum.-Robot Interact., vol. 9, no. 2, dec 2019.
[3] J. Heard, C. E. Harriott, and J. A. Adams, "A survey of workload assessment algorithms," IEEE Transactions on Human-Machine Systems, vol. 48, no. 5, pp. 434–451, 2018.

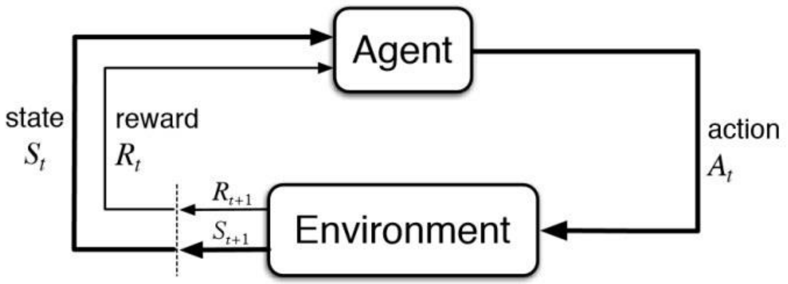
Human-Aware Decision-Making System Architecture

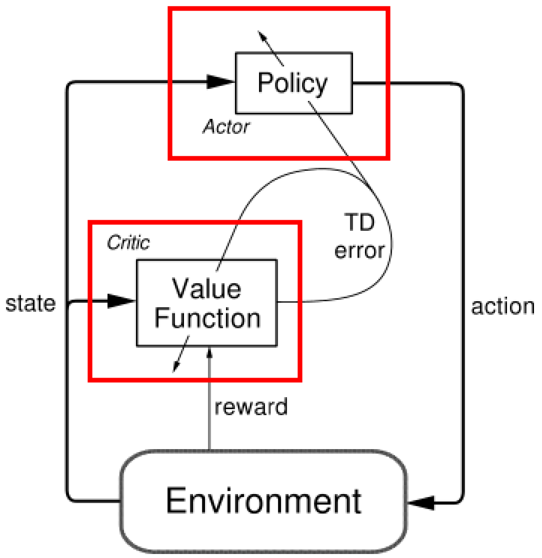


[4] J. Heard, and J. A. Adams, "Multi-Dimensional Human Workload Assessment for Supervisory Human–Machine Teams" in *Journal of Cognitive Engineering and Decision Making*, 2019; vol. 13(3), pp. 146-170. doi:10.1177/1555343419847906

Reinforcement learning

- Formulated as a Markov Decision Process (MDP) with the tuple (S, A, T, R, γ).
- The agent interacts with the environment with a known reward function and computes a policy that maximizes the rewards.





Soft Actor Critic Algorithm:

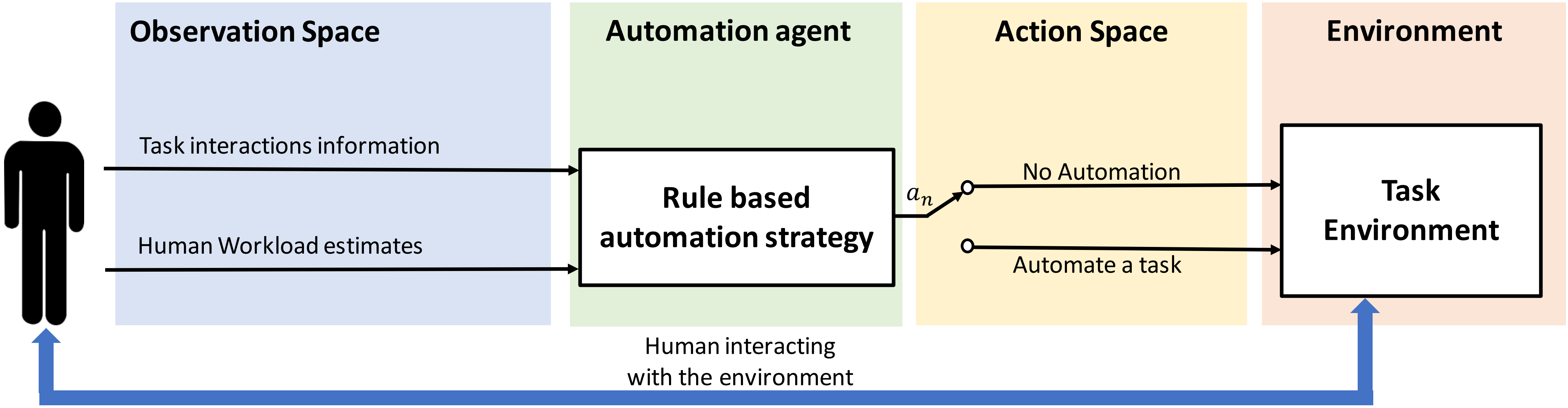
The algorithm is based on a maximum entropy RL where the objective is to find the optimal policy that maximizes the expected long-term reward and long-term entropy.

[5] 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.* PMLR, 2018, pp. 1861–1870.

Automation Strategies - RB

Rule Based automation strategy **(RB)**:

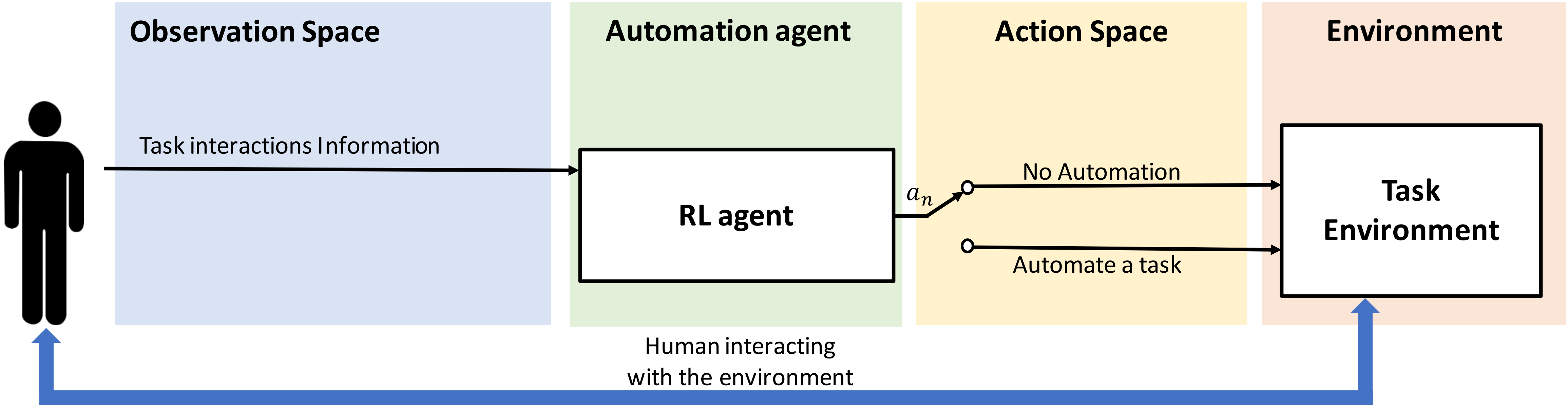
- If under loaded, take away all automation.
- If overloaded, automate the most neglected task.



Automation Strategies - RL

Reinforcement **L**earning agent **(RL)**:

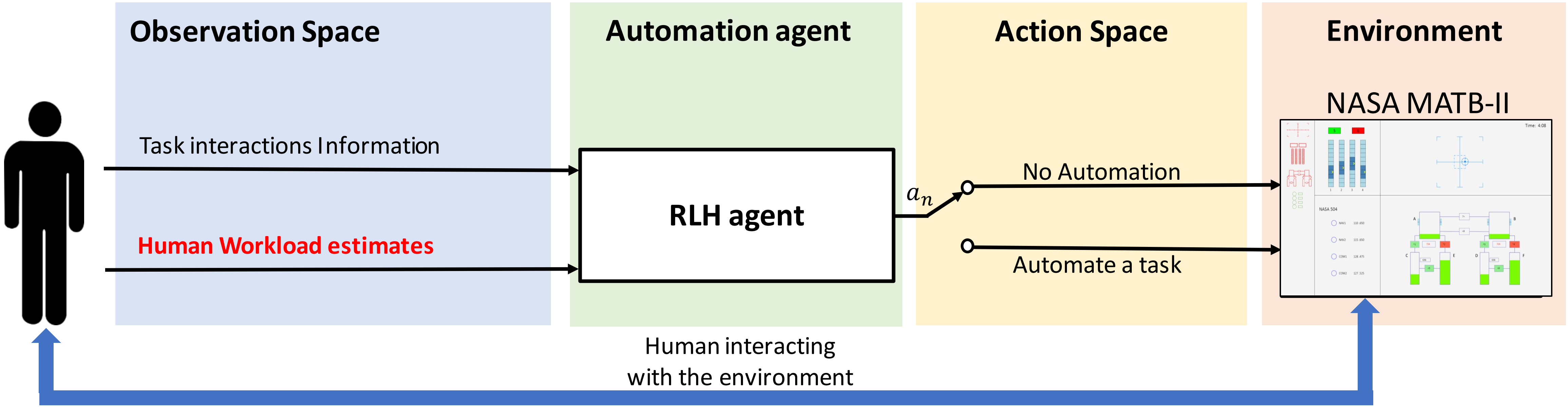
Observation Space: • Task interaction information

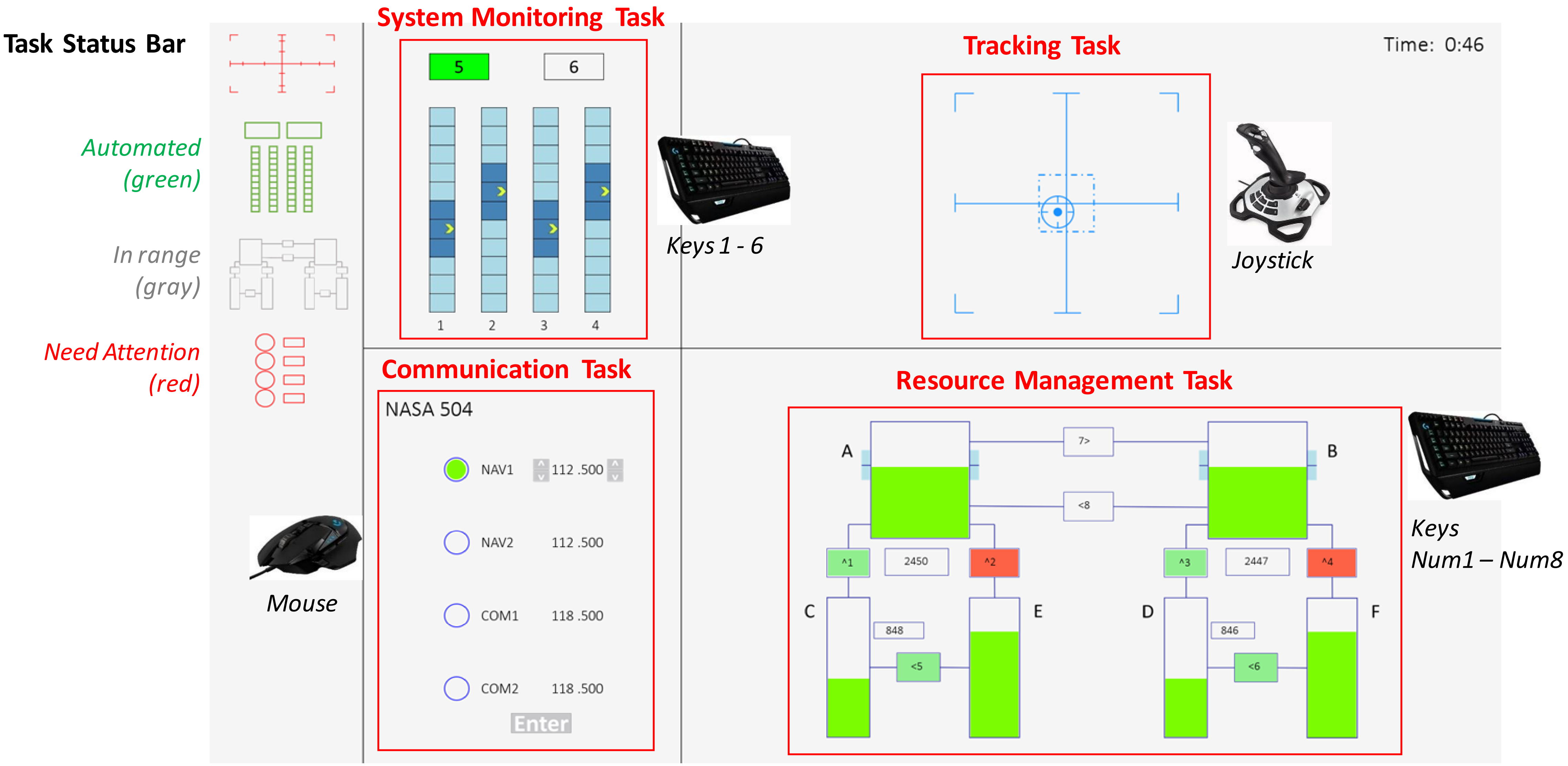


Automation Strategies - RLH

Reinforcement **L**earning agent with **H**uman states **(RLH)**:

Observation Space:
- Task interaction information
- **Human Workload estimates**





NASA Multi-Attribute Task Battery-II (MATB-II)

Reward Structure

$$r = w_0 * r_0 + w_1 * r_1 + w_2 * r_2 + w_3 * r_3 + Interaction_{Human}$$

Bad performance was penalized.

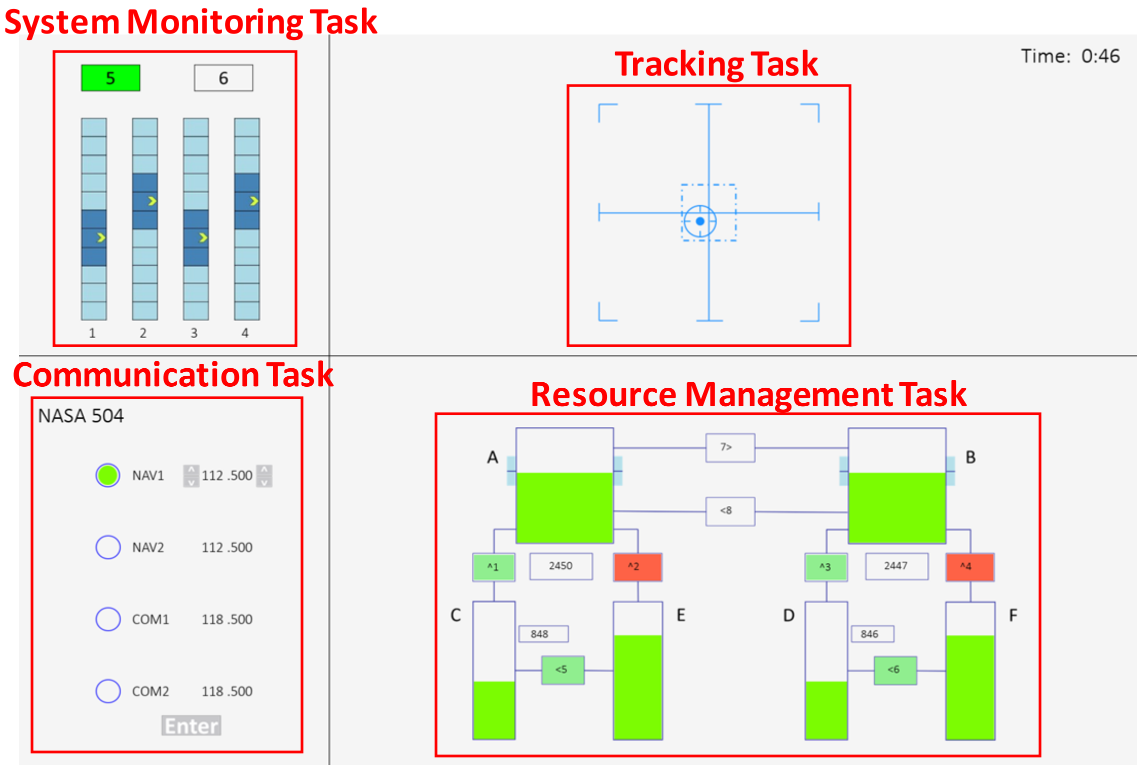
System Monitoring (r_0):
- Response Time for acknowledging an alarm

Tracking Task (r_1):
- Error between the cross-head and the target

Communication Task(r_2):
- Response Time for following the audio command

Resource Management Task(r_3):
- Tanks out of range.

Human Idle time ($Interaction_{Human}$)



Experimental Design

Recruited 9 participants (5 males and 4 females; average age of 26.3)

Mixed design experiment

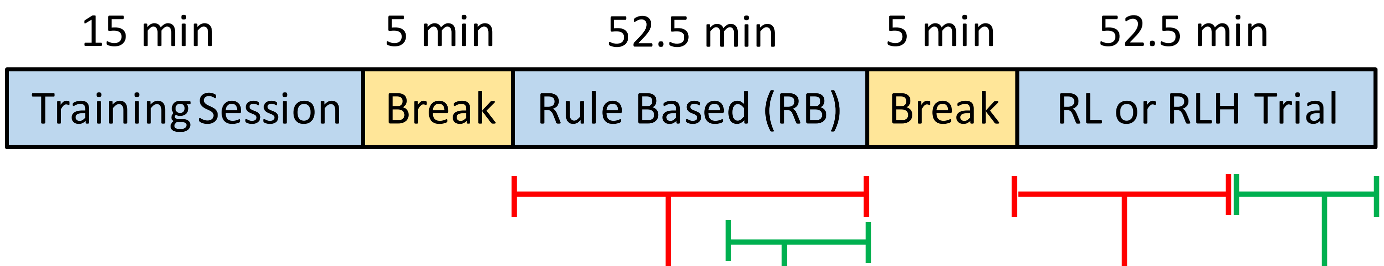
Independent variables:

- Workload Condition (within-subjects variable)
  - ➢ Underload (UL)
  - ➢ Normal Load (NL)
  - ➢ Overload (OL)
- Automation strategy type (between-subjects variable)
  - ➢ Rule Based (RB)
  - ➢ Reinforcement Learning agent (RL)
  - ➢ Reinforcement Learning agent with human states (RLH)

Dependent Variables:

- Estimated Workload
- Rewards
- Individual Task Performance
- Automation Time

Experiment Sessions



Participants completed the NASA Task Load Index (NASA-TLX) after each trial.

Pre-training:

- Offline with expert human data.

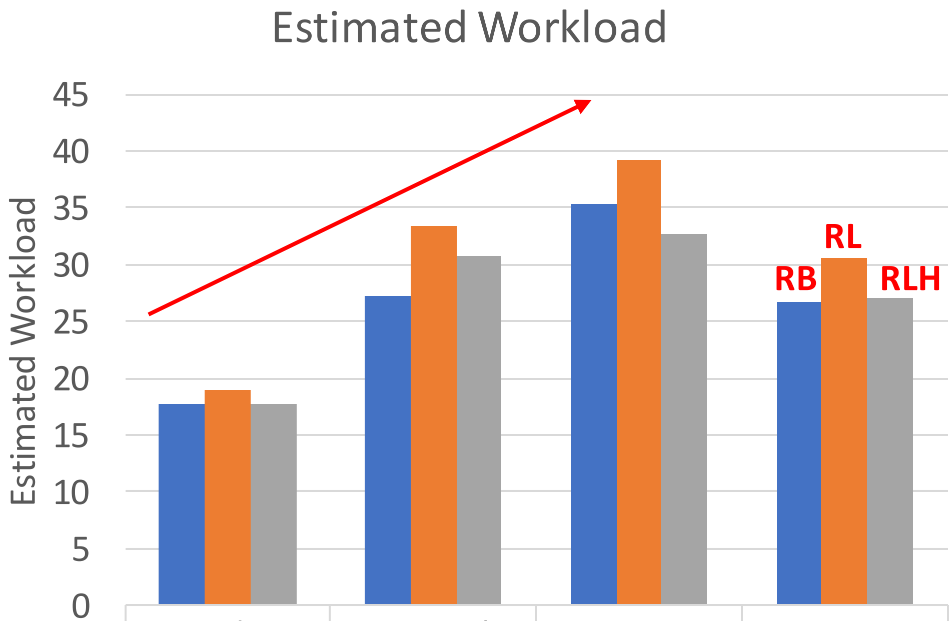
Training:

- During the RB trial without exploration.
- During the first 29.5 minutes of RL/RLH with exploration.

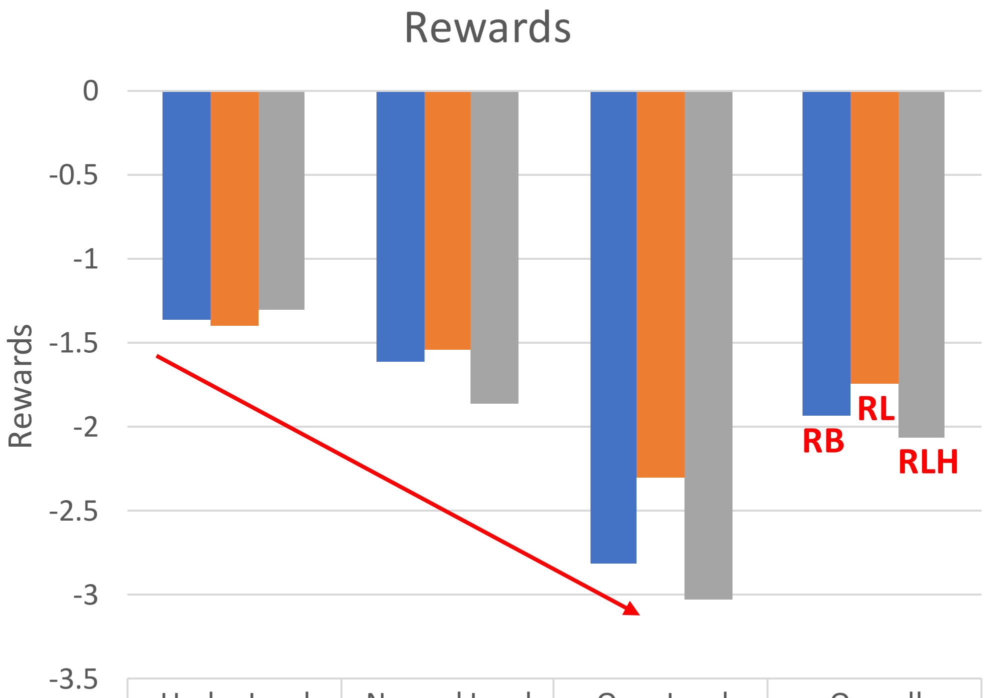
Evaluation:

- During the last 23 minutes.
- No training during evaluation.

Results



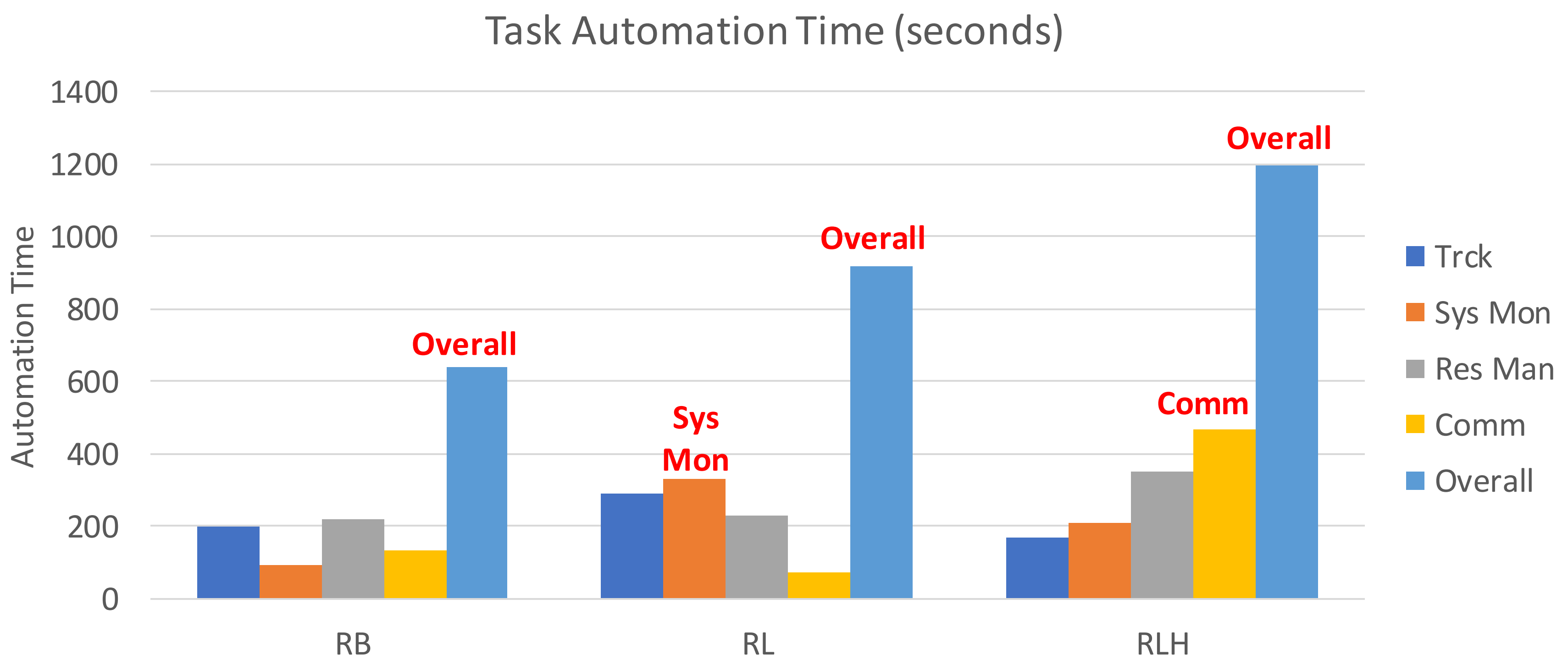
	Under Load	Normal Load	Over Load	Overall
RB	17.79	27.26	35.35	26.8
RL	18.93	33.49	39.2	30.54
RLH	17.78	30.8	32.67	27.08



	Under Load	Normal Load	Over Load	Overall
RB	-1.36	-1.61	-2.82	-1.93
RL	-1.4	-1.54	-2.3	-1.74
RLH	-1.3	-1.86	-3.03	-2.06

- The addition of **human states** resulted in a **lower overall workload** but with **worse rewards**.
- RLH agent may have **picked up this trend** to reduce human workload, however it **could not achieve rewards** due to a much more complex state space compared to the RL agent.

Results



- The RL agent automated the System Monitoring task the most.
- The RLH agent automated the communications task the most.
- The RLH agent tends to automate most frequently.
- Very high or very low automation time **does not imply** better overall team performance.

Limitation

- The addition of **human states** resulted in a **lower overall workload** but with **worse rewards**.
- Addition of human states made the **state space more complex** to explore.
- **Longer training times may be needed** for the RLH agent learn an optimal action policy.
- **Discretizing human states** will also simplify the observation space for RLH agent.

Conclusion

- Presented a human-aware decision-making system that determines appropriate robot adaptations using task and human workload information.
- Presented a reinforcement learning-based approach that can learn trends between human states and team performance.
- Developing a human-aware system architecture can help in achieving a more fluent team collaboration.
- Improving the team performance in dynamic multitask environments.

Thank You!!

Any questions

Please feel free to reach out to us at ss3337@rit.edu

Appenndix

Workload Condition	Trial	Workload	Rewards	Tracking Error (pixels)	Gauges RT (seconds)	Lights RT (seconds)	Tanks in Range (%)	Comms RT (seconds)
Underload (UL)	RB	17.79	-1.36	**18.70**	-	**1.60**	86.41	-
	RL	**18.93**	-1.40	22.54	-	2.13	84.68	-
	RLH	17.78	**-1.30**	19.95	-	2.07	**87.52**	-
Normal load (NL)	RB	**27.26**	-1.61	25.28	3.07	3.13	84.30	12.13
	RL	33.49	**-1.54**	**22.62**	**1.80**	**1.60**	**98.14**	**11.99**
	RLH	**30.80**	-1.86	23.18	4.72	2.82	84.33	13.57
Overload (OL)	RB	35.35	-2.82	**27.44**	4.17	4.10	73.98	12.84
	RL	39.20	**-2.30**	28.51	**2.82**	**2.96**	**87.00**	12.43
	RLH	**32.67**	-3.03	**27.64**	3.29	3.55	75.00	**11.48**
Overall	RB	**26.80**	-1.93	**23.81**	4.01	3.79	81.57	12.72
	RL	30.54	**-1.74**	24.58	**2.69**	**2.67**	**89.94**	12.38
	RLH	**27.08**	-2.06	**23.59**	3.51	3.30	82.28	**11.98**

Workload Condition	Trial	Automation Time (seconds)				Interaction Time (seconds)			
		Tracking	Sys. Mon.	Res. Man.	Comms.	Tracking	Sys. Mon.	Res. Man.	Comms.
Underload (UL)	RB	**58.28**	15.00	10.57	19.71	137.28	0.85	6.28	0.14
	RL	89.33	**110.33**	80.00	23.00	114.33	6.66	12.00	0.00
	RLH	60.75	87.50	100.50	**156.00**	147.75	0.75	4.75	0.00
Normal load (NL)	RB	**60.57**	45.57	**59.71**	50.85	136.42	7.14	6.71	9.14
	RL	**108.00**	**111.00**	77.00	14.66	115.33	12.33	6.00	9.33
	RLH	61.75	64.00	127.75	**151.75**	122.5	10.25	5.00	6.0
Overload (OL)	RB	77.57	29.85	**146.28**	63.14	96.57	37.57	1.57	39.57
	RL	94.66	**106.66**	70.66	32.00	106.33	38.00	0.33	38.33
	RLH	47.50	57.50	124.25	**160.00**	93.00	39.75	0.25	29.25
Overall	RB	196.42	90.42	**216.57**	133.71	370.28	45.57	14.57	48.85
	RL	292.00	**328.00**	227.66	69.66	336.33	57.00	18.33	47.66
	RLH	170.00	209.00	352.50	**467.75**	363.25	50.75	10.00	35.25

Results

NASA-TLX Ratings

	RB	RL	RLH
Overall WL	67.8 (8.2)	49.9 (0.5)	71.2 (7.3)
Mental Demand	67.5 (20.7)	53.3 (13.1)	78.7 (21.5)
Physical Demand	40.6 (21.4)	40.0 (20.4)	45.0 (4.1)
Temporal Demand	77.5 (10.0)	43.3 (8.4)	81.2 (21.5)
Performance	43.7 (20.4)	20.0 (10.8)	60.0 (8.9)
Effort	74.3 (13.0)	68.3 (16.4)	71.2 (5.5)
Frustration	68.7 (18.9)	48.3 (11.7)	67.5 (5.4)