Using Human-Machine Teaming to Accelerate Operational Trust in Mission Autonomy (OPTIMA) for OUSD(R&E)

Ground System Architectures Workshop (GSAW)

Dr. Christopher Tschan1, Dr. Amy O'Brien2, Dr. Michael Langford

Distance, complexity, and scale
of emerging military and space missions necessitate
trustworthy AI/autonomy that human users can adopt with confidence

The Aerospace Corporation has a technical vision to use
Human-Machine Teaming (HMT)
to help ***accelerate development*** and assessment of
and ***build confidence and trust in***
AI/autonomy solutions
to enhance ***Operational Trust in Mission Autonomy (OPTIMA)***,
including future space missions and operations

This presentation, which illustrates Aerospace's vision and
how that vision can benefit our customers, was funded by the
Office of the Undersecretary of Defense for Research and Engineering (OUSD(R&E))
to support the OPTIMA program

Agenda

- What is Human-Machine Teaming (HMT)?
- What is OPTIMA and how can HMT support it?
- Our vision for using HMT to accelerate AI/autonomy confidence for OPTIMA
- Our progress so far
- Summary & Conclusion

What is Human-Machine Teaming (HMT)?

Smartphones offer low-level HMT

- **Brookings**

  "Human-machine teaming is ***a relationship—***made up of at least three equally important elements: ***the human***, ***the machine***, and ***the interactions*** and interdependencies ***between them***" [1]

- **AFRL**

  "Better human-machine teamwork leads to ***faster performance of tasks with fewer errors***" [2]

- **Pacific Northwest National Laboratory**

  "Today's human-machine interaction places a heavy burden on the human. … We recognize a new opportunity ***to build machines that can work with the user to meet their goals, rather than blindly executing tasks***. These new machines can both ***enhance team performance and minimize the work*** required for the human to manage the machine." [3]

Sources:

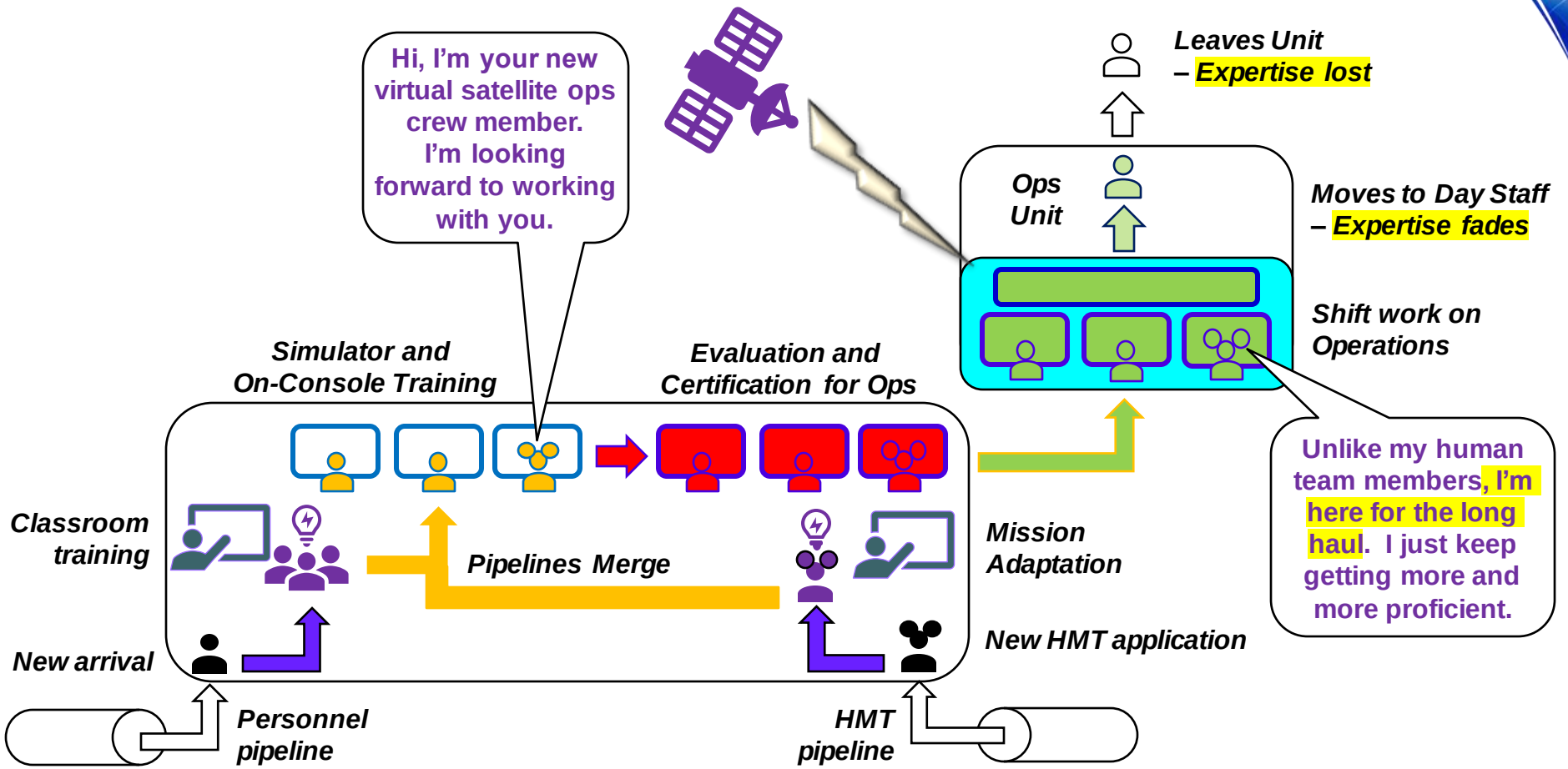
[1] See https://www.brookings.edu/techstream/building-trust-in-human-machine-teams/#:~:text=Human%2Dmachine%20teaming%20is%20a,interactions%20and%20interdependencies%20between%20them.

[2] See https://defenseinnovationmarketplace.dtic.mil/technology-interchange-meetings/autonomy-tim/human-machine-teaming/

[3] See www.pnnl.gov/sites/default/files/media/file/Human-Machine%20Teaming.pdf

Example: How HMT Can Improve Military Space Ops

Contrast with the process if HMT were not available



Operational teams of humans and machines can increase efficiency and expertise

What is OPTIMA, & how can HMT accelerate it?

OUSD(R&E)'s Operational Trust in Mission Autonomy (OPTIMA)

OPTIMA explores ***Trust as a battlefield asset*** that is ***measurable, auditable*** and provides ***real time, bi-directional and attributable*** characteristics of autonomy to result in ***more effective autonomous system performance*** in HMT, Manned-Unmanned Teaming (MUM-T), and other application(s) of autonomous behavior

> *"'The main future goal (with OPTIMA) is to deliver trusted autonomy on the battlefield for the warfighter in complex and contested environments.' Jaret Riddick, principal director for Autonomy in the Office of the Undersecretary of Defense for Research and Engineering, said. 'To deliver that (trust) as a product so that we can quantify it and use it as a battlefield asset.'"* [1]

Although OPTIMA seeks to improve HMT, Aerospace proposes to use HMT to improve OPTIMA:

- Letting users train/train with the HMT application ***grows trust from inception***
- Such an HMT application progresses to greater levels of autonomy organically

Source:

[1] See https://www.jconline.com/story/news/2022/06/29/dod-summit-looks-quantifying-trust-machines-purpose/7745605001/

Let's use HMT to create trusted AI/Autonomy faster

HMT can move OPTIMA from "concept" to "accelerated"

We propose to use HMT from inception to accelerate trust throughout training, assessment and deployment of AI/autonomy systems

What's needed: a relatively ***simple, generic HMT application***

- *Use the **Aerospace AI/autonomy Solution Architecting** process, which includes the **Aerospace Trusted AI Framework**, to develop a thoughtful, scalable design*
  - To achieve well-defined (e.g., DARPA) levels of autonomy
  - To apply seamlessly to multiple (closed world) technical domains

HMT will help us ***archive, evaluate, and exploit human technical expertise***

Currently, AI/autonomy requires custom, purpose-built code
- E.g., Google's codebase entails ~2 billion lines of code

But creating custom AI/autonomy for every enterprise is not feasible
- High cost, schedule, and technical risk
- Discourages broad interoperability

Harnessing HMT for AI/autonomy builds trust by incorporating human users into the development and training process from inception

HMT for Challenging Domains, like OPTIMA

Our plan

- ***Develop a generic software application*** to
  - — Let humans easily train a virtual team member to perform technical activities
  - — Mimic how humans train a human replacement
  - — Adapt to multiple human users with varied levels of experience
    - Provide information gathered from experienced users to novices
- ***Train the HMT application*** to learn the technical activity until proficient
- ***Test and certify*** the HMT to perform its activity
- Provide variations so the ***HMT application gathers and archives information to improve its performance over time*** as humans do
- The HMT application becomes a ***valued, trusted team member*** working along side humans
- Because the starting HMT application is generic, ***very little code is needed to adapt it*** to a new mission or domain

We will use HMT to mimic human learning & help archive human technical expertise

Our First Focus: Developing Level 0 HMT Autonomy

Levels of Autonomy

(adapted from DARPA's definitions for the Orbital Express program*):

0. No autonomy (fully manual)

- Human uses wizards help adapt the generic HMT application to the specific domain
- Human interacts with machine via computer interface (keyboard and, later, voice) to adapt it to technical tasks
- Human connects the HMT application with data sources needed for the tasks
- The HMT application observes and records human actions to perform the complete technical activity
- The HMT application continually creates evolving data set of technical activities, including metrics
- The HMT application offers the human real-time perspectives re the technical activity based on most-relevant past instances

1. **Sequence autonomy (human authorizes machine to perform sequences within an activity)**
2. **Activity authority (human authorizes machine to perform an activity)**
3. **Supervised autonomy (human in the loop)**
4. **Observable autonomy (human on the loop)**
5. **Full autonomy (machine makes decisions and takes actions independently of human)**

Source:

* For discussion of DARPA's four levels of supervised autonomy, to which Aerospace has added unsupervised levels, see https://digitalcommons.usu.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1768&context=smallsat

Our Progress Developing HMT for OPTIMA

We are developing the generic HMT application

Completed:

☑ ***Select a simplified focusing application***—an intuitive example
- — Easy to prototype
- — Easy to build on, expand, and identify gaps
- — Easy to explain to an audience

☑ ***Develop a rudimentary HMT application*** to demonstrate and build on

In progress:

- Use the ***Aerospace AI/autonomy Solution Architecting (AASA)*** process to develop & assess the prototype
  - — Includes the Slingerland-Perry Aerospace ***Trusted AI Framework***
  - — Leverages our ***crawl-walk-run digital engineering tools & testbeds***
- Identify and document gaps between the first-cut prototype and the needs identified via the AASA process
- ***Demonstrate Level 0 HMT*** to a selected technical domain
- ***Demonstrate easy adaptation*** of the HMT to another technical domain

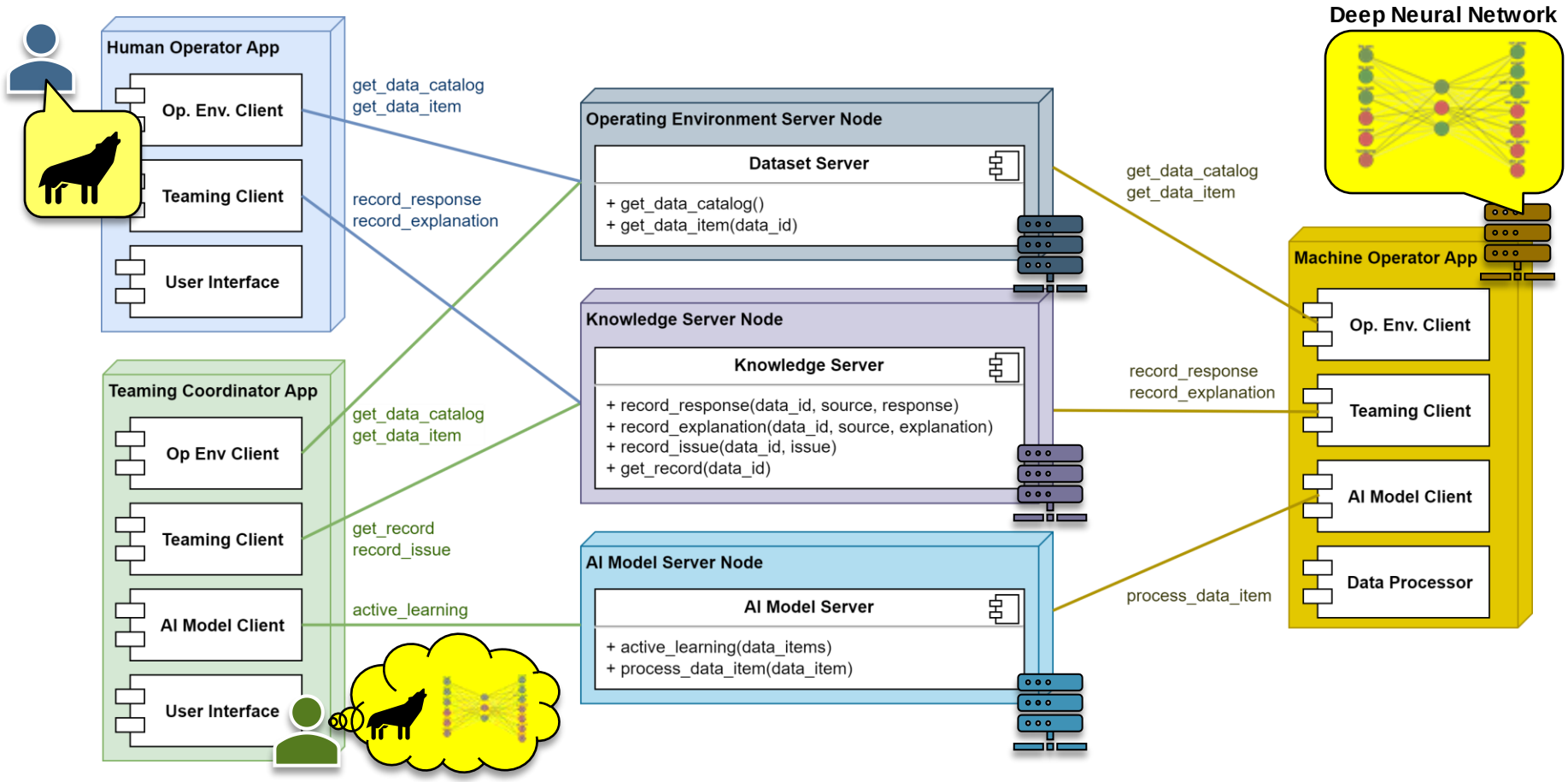
We and our customers can use the generic HMT application like a reference architecture—a starting point, an illustration of scope—for vendors to build on

Suppose Little Red Riding Hood is a machine…

*…needing a human expert to teach it what to **do** based on what it **sees***

Example Problem: both ***human*** and ***machine*** choose actions based on observations

- Observable feature inputs describe ***wolf***, ***grandmother***, or ***woodcutter***
- Friend-or-foe action outputs include activities such as ***approach, offer food, scream, run away***, etc.



Data-centric HMT: operators respond to data inputs/triggers

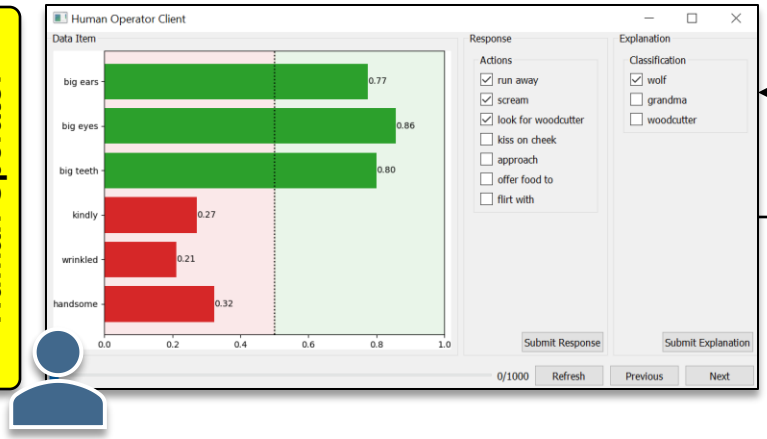
Our HMT Can Teach Little Red Riding Hood

*Which **features** do you see? Which **entity** is it? **Why** do you think so?*

Both human and machine operators interact with data from the environment to provide

- ***Responses:*** resulting analysis/evaluation of observed data (e.g., ***actions***)
- ***Explanations:*** reasoning for the provided response (e.g., ***classifications***)

Human Operator



Data Item
Features: [*big ears*, *big eyes*, *big teeth*]

Operating Env. Node
get observable features

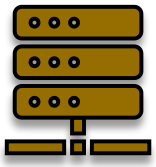
Response
Actions: [*run away*, *scream*, *look for woodcutter*]

Knowledge Node
record response/explanation

Explanation
Classification: [*wolf*]

Human operators can ***observe*** features, ***respond*** with actions, and ***explain*** by classifying features

Machine Operator



Automated Data Process

Data Item
Features: [*big ears*, *big eyes*, *big teeth*]

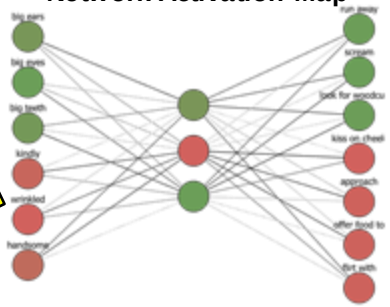
Operating Env. Node
get observable features

Response
Actions: [*action probabilities*]

Knowledge Node
record response/explanation

Explanation
Reason: [*activation map*, *network weights*]

Network Activation Map



AI Model Node
inference with AI model

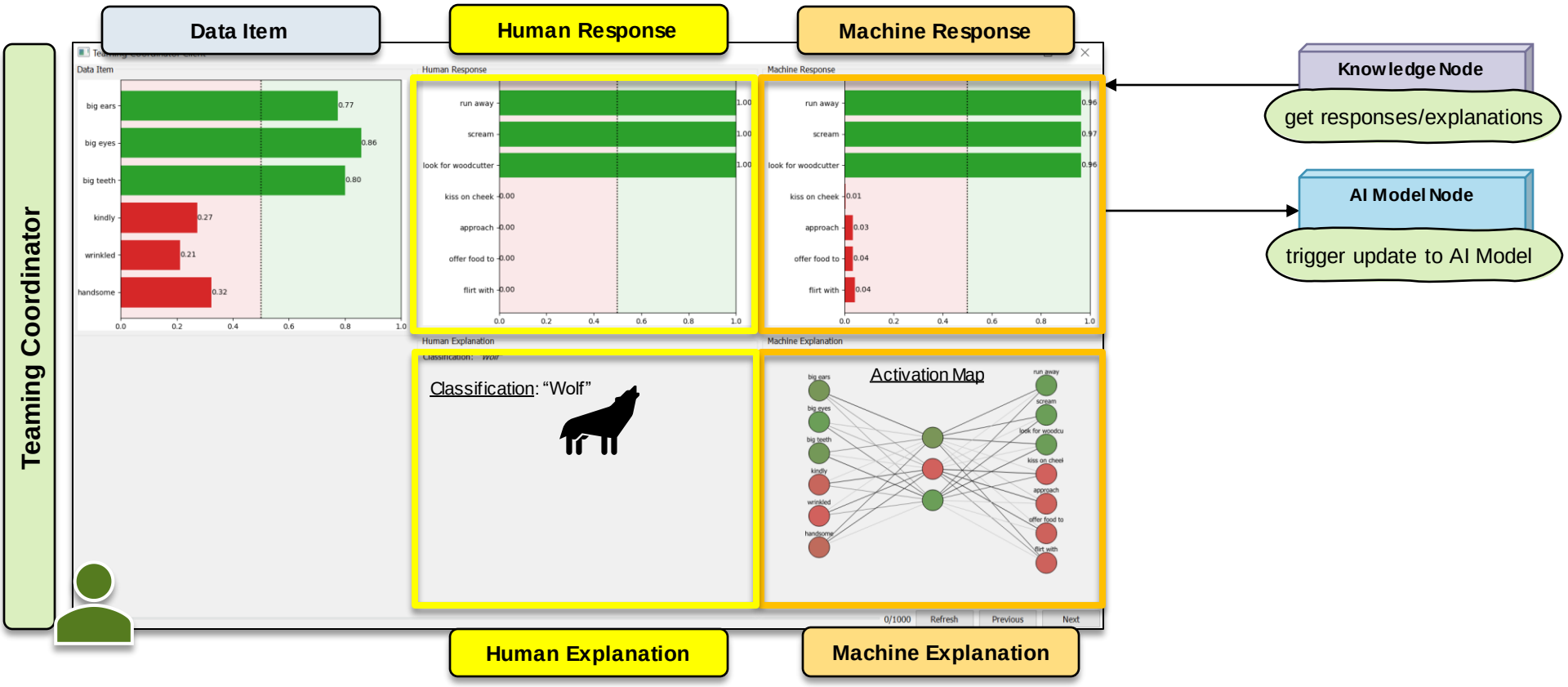
Machine operators use AI models (e.g., *DNNs*) to ***observe*** features, ***respond*** with model predictions, and ***explain*** model behaviors (e.g., *activations maps*)

How do you deconflict human-machine answers?

*One solution: a human **Teaming coordinator***

Teaming coordinators can **review responses/explanations** and **take action**

- Identify ***deviations*** and ***trust issues***
- Update AI models, via ***active learning*** of prioritized data [1]



Teaming coordinators can observe responses/explanations for both humans and machines to ***record issues***, ***collect data***, and ***perform analytics*** to improve both human and machine performance

Source:

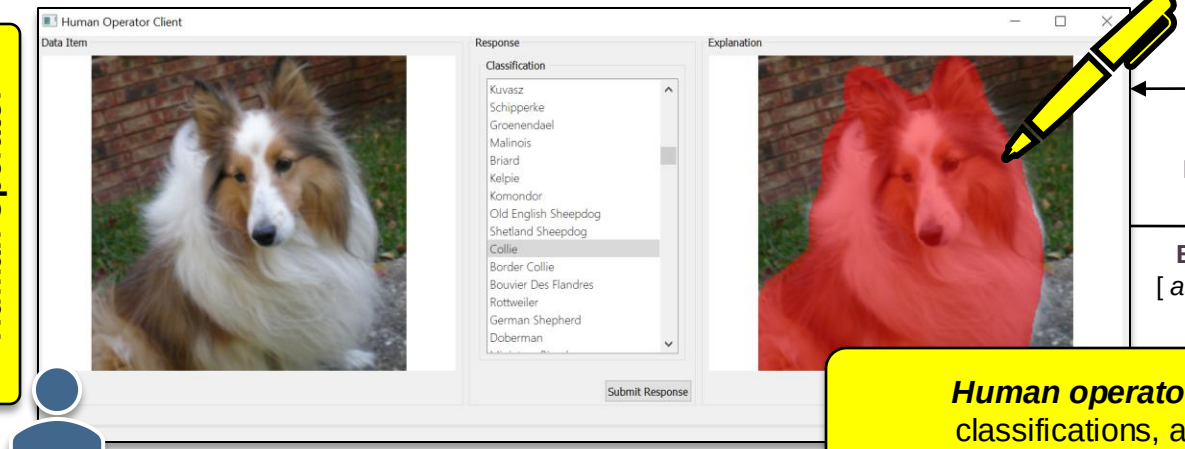
[1] See References

Extending to other Problem Domains

Our prototype HMT can do image recognition (e.g., ImageNet)

Consider ***image recognition*** (*identifying imaged objects*) for computer vision: operators can ***label images*** and explain by ***highlighting regions of interest***

Human Operator



Data Item [*image*]

Operating Env Node

get observable features

Response [*collie*]

Knowledge Node

Explanation [*attention map*]

record response/explanation

Human operators can ***observe*** images, ***respond*** with classifications, and ***explain*** by drawing attention maps

Machine Operator

Automated Data Process

Data Item [*image*]

Operating Env Node

get observable features

Response [*class probabilities*]

Knowledge Node

Explanation [*class activation map*]

record response/explanation

AI Model Node

inference with AI model

Class Activation Map

Grad-CAM [2]

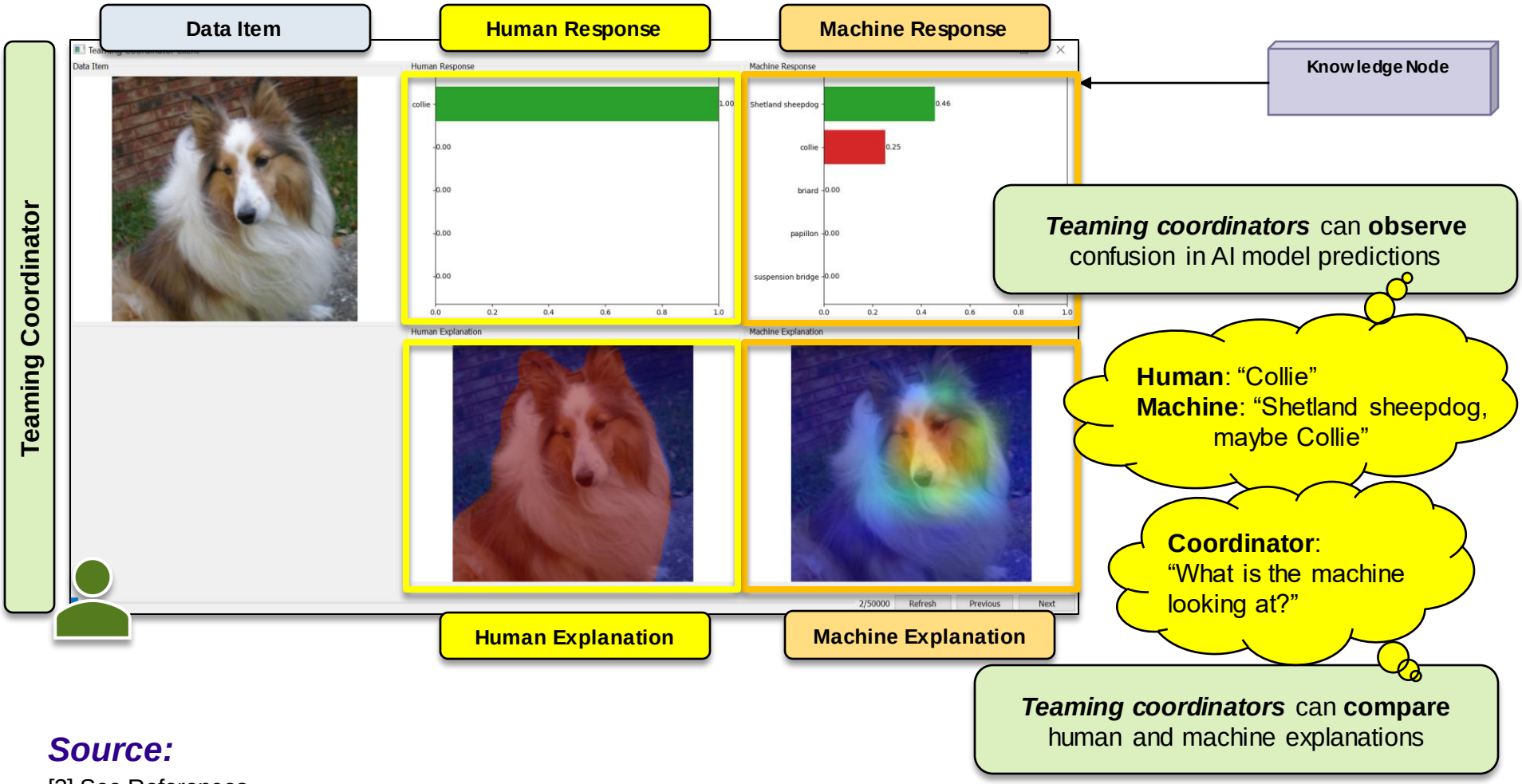
Machine operators use CNNs to ***observe*** images, ***respond*** with classification predictions, and ***explain*** with visualizations of class activation maps

Our HMT Can Do Image Recognition

A Teaming coordinator can validate & evaluate Trusted AI/autonomy

Our HMT enables a multi-faceted analysis of AI behavior with respect to human operations

- ***Explainable AI (XAI)*** [3] is a subset of TAI Framework ***Interpretability***



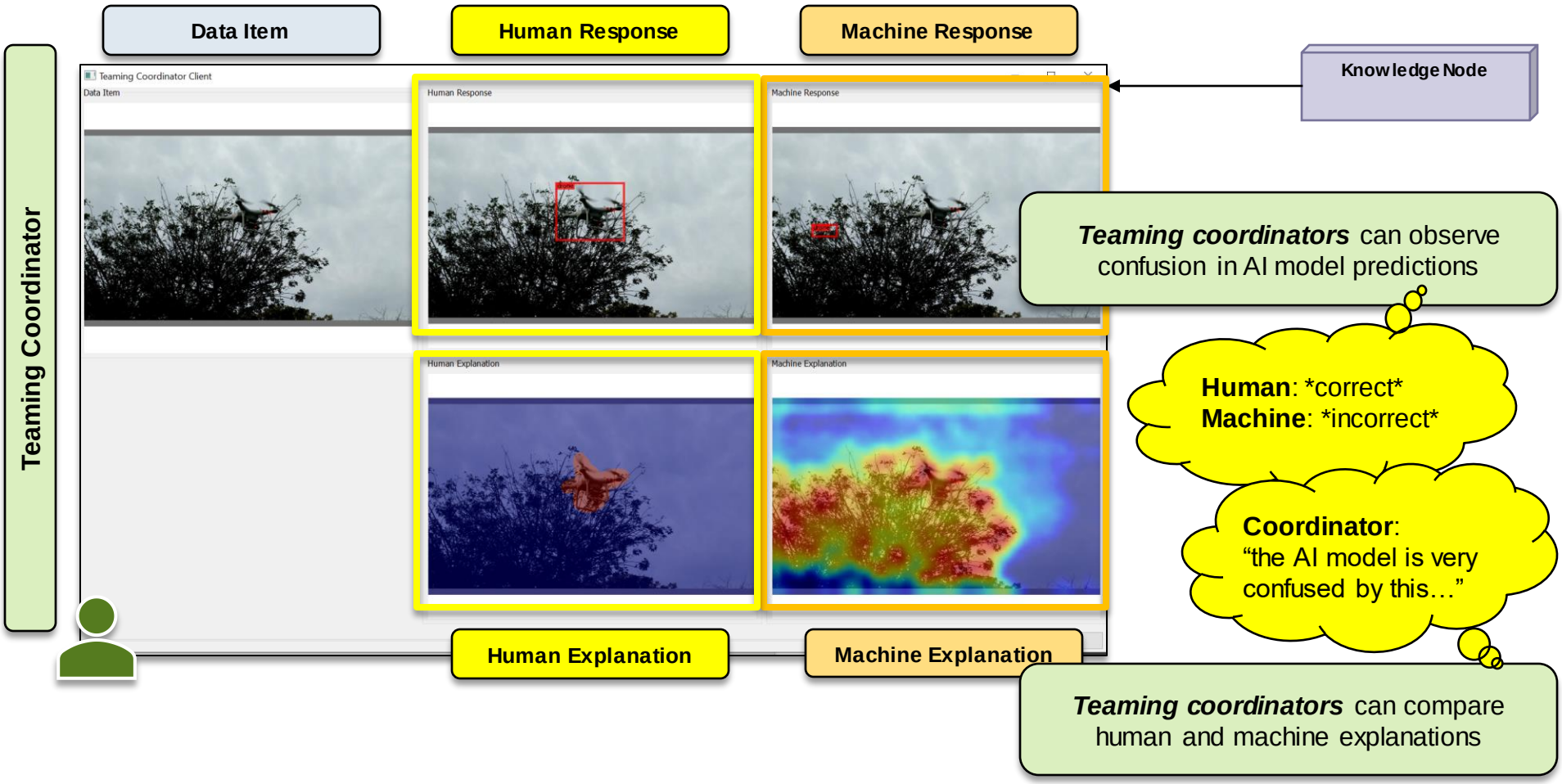
Source:

[3] See References

Our HMT Can Do Drone Detection

Teaming coordinators can record & prioritize examples of interest

Consider ***object detection*** (*both identifying and locating imaged objects*):
operators can ***annotate images*** and explain by ***highlighting regions of interest***

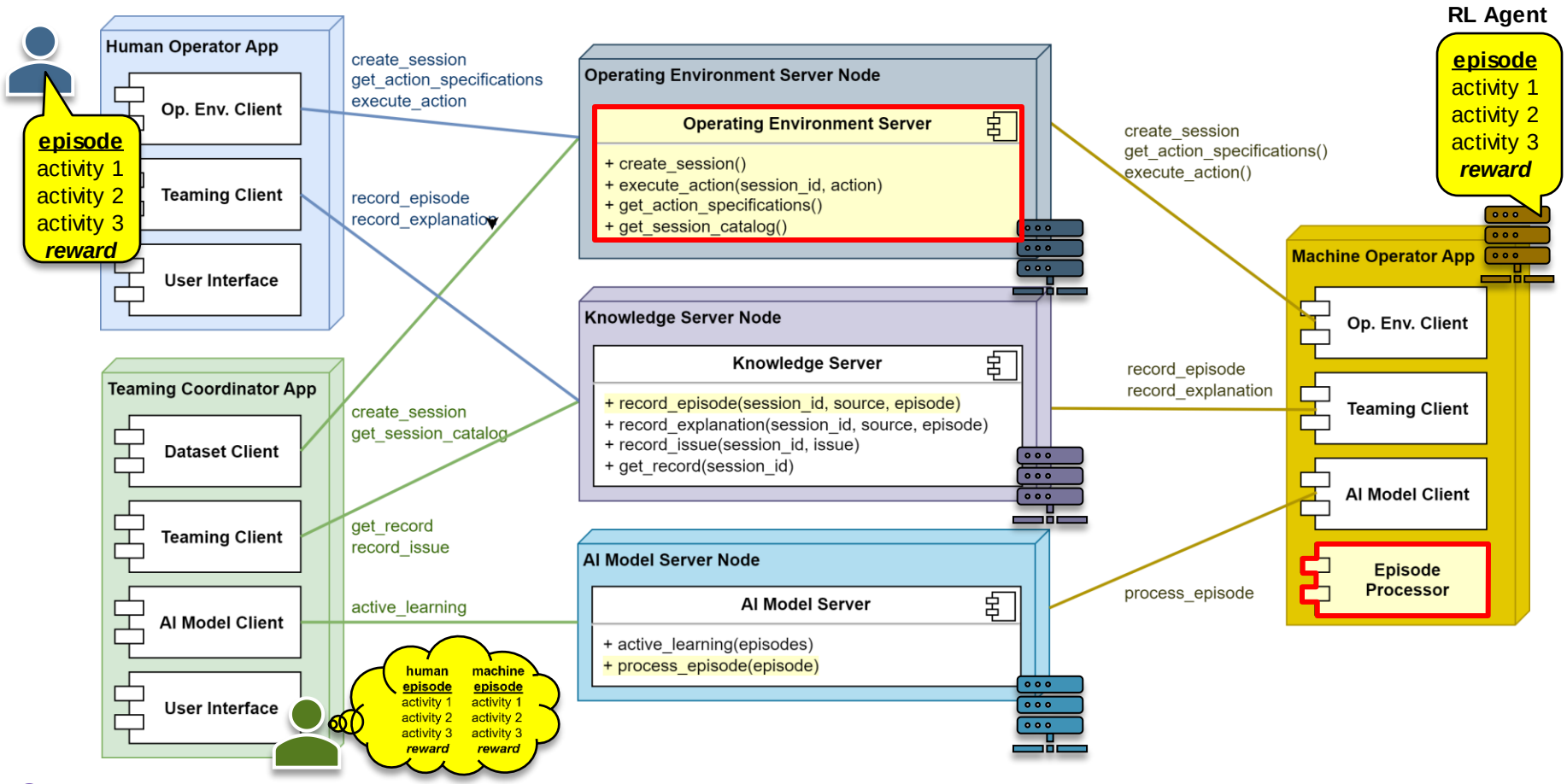


Our Vision: HMT Will Process Data & Do Operations

Reinforcement Learning [4]: episodic HMT interacting through episodes

Human and Machine interact with Operating Environment through **episodes of activity**

- Operators make ***observations***, respond with ***actions***, and evaluate ***rewards***
- Knowledge base stores ***episodes of activity*** with ***explanations*** and ***rewards***



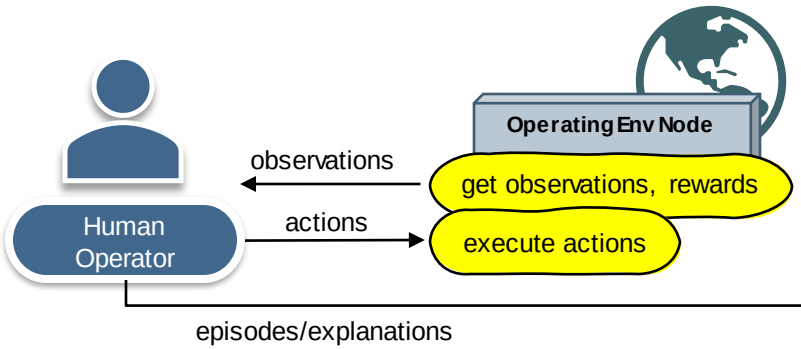
Source:

[4] See References

Our HMT Will Do Operations

Reinforcement learning facilitates team interaction & learning

Human Operator

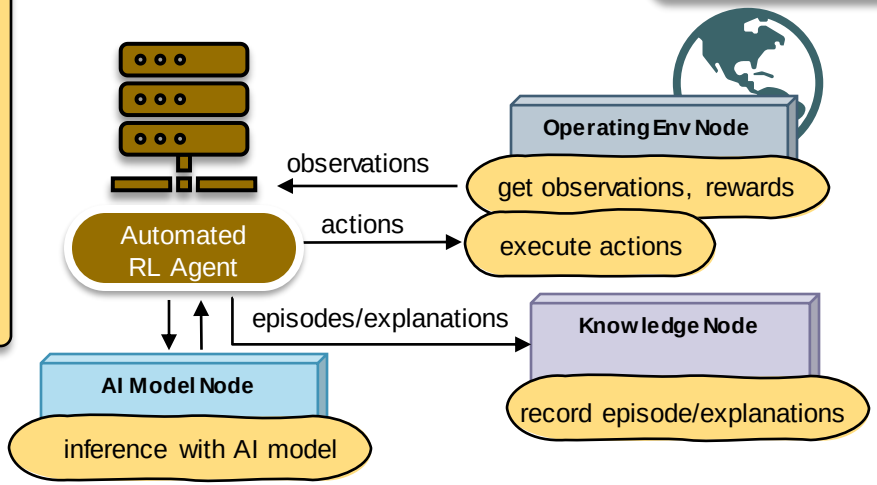


Episode

Time	Obs.	Reward	Action
t=1	O_0	R_0	A_1
t=2	O_1	R_1	A_2
…	…	…	…
	O_{t-1}	R_{t-1}	A_t

Episodes comprise a ***sequence of observation states,*** corresponding ***rewards*** (i.e., goal satisfaction), and ***chosen actions*** (i.e., operator response)

Machine Operator



RL Policy (e.g., Q-Table)

	Action 1	Action 2	Action 3
State 1	0.99	0.15	0.45
State 2	0.65	0.75	0.33
State 3	0.15	0.33	0.82
State 4	0.50	0.50	0.99

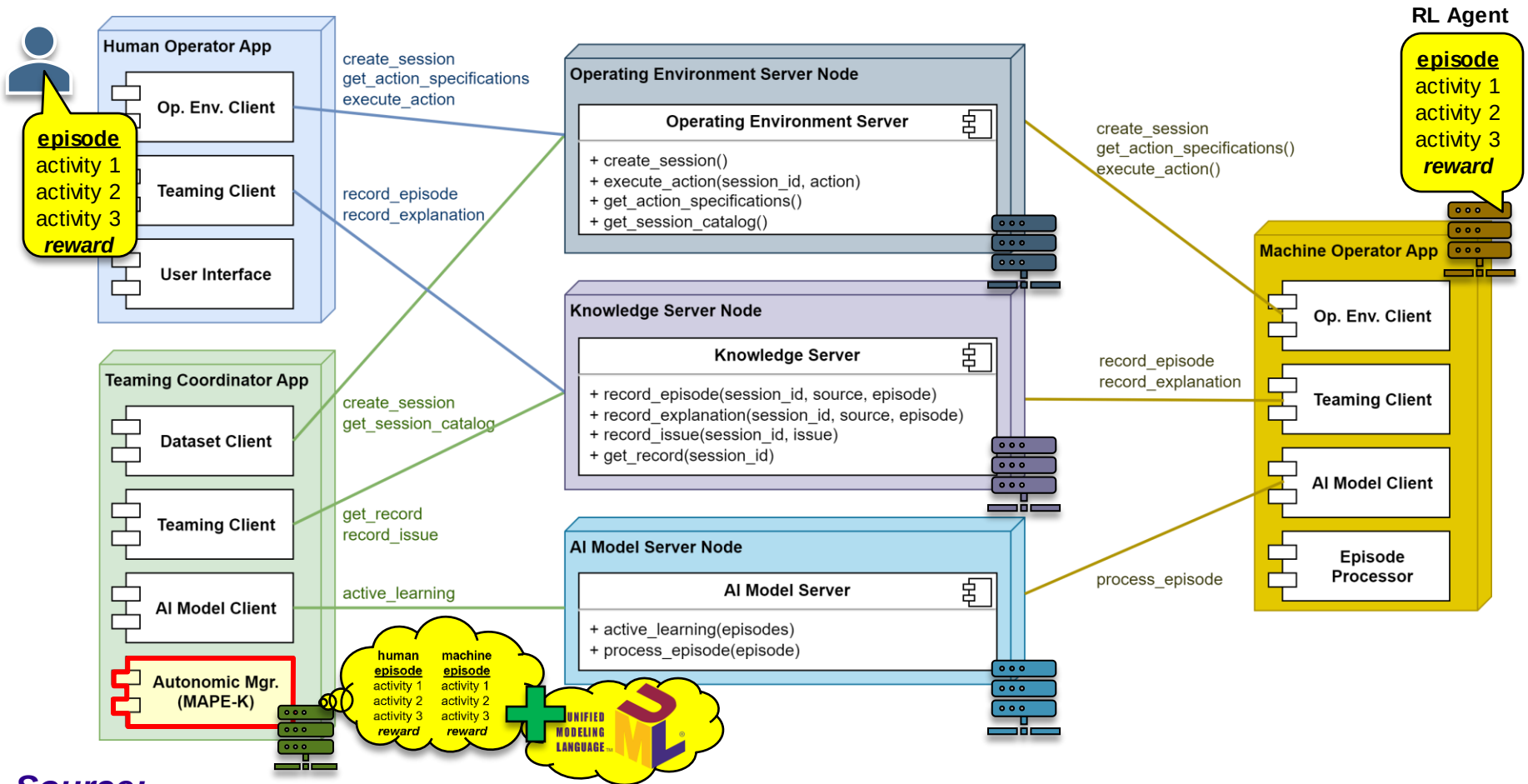
Machine operators use learned ***policies*** to interact with environment, act out episodes, and update knowledge base.

Autonomic Management of the AI Model [5,6]

*Make the teaming coordinator **self-managing, system models-driven***

Introduce **autonomic computing** to **automate the Teaming coordinator**

- ***Replace Teaming coordinator*** with ***MAPE-K*** process (*Monitor-Analyze-Plan-Execute w/ Knowledge*)
- Use ***system requirements and goal models*** (*e.g., UML)* to trigger updates to AI models

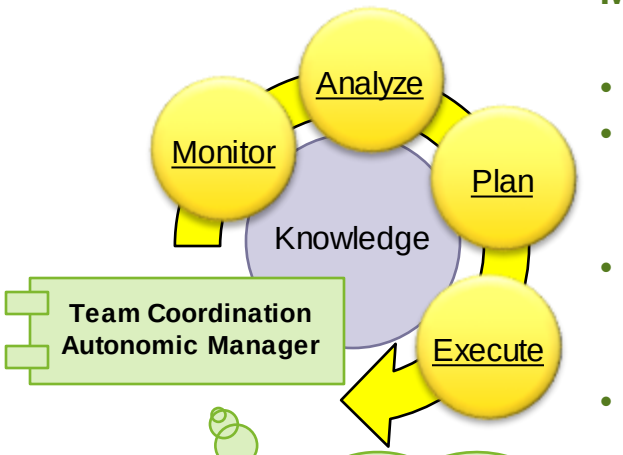


Source:

[5,6] See References

Our Vision for Automating the Teaming Coordinator

Autonomic teaming coordination can make AI models self-managing

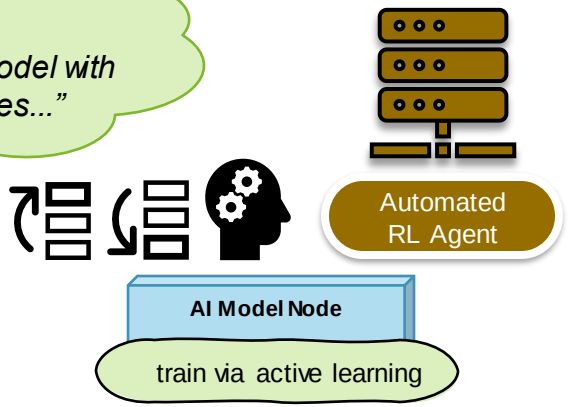


MAPE-K *(MAPE with shared Knowledge)*

- **Monitor:** observe human/machine responses and explanations
- **Analyze:** goal models, evaluation metrics, trust metrics, etc.
  - Focus on ***digital engineering*** with a common modeling language to ***describe problem domain*** (e.g., *UML*, *SysML*, etc.)
- **Plan:** determine appropriate coordination efforts
  - ***Devise suggestions*** to human operator
  - ***Devise update strategy*** for AI model (e.g., *active learning* [1])
- **Execute:** execute coordination plan

Suggestion:
"Performance may improve if you perform these actions…"

Update AI Model:
"Resume training of AI model with focus on selected episodes..."



An automated Teaming coordinator can make recommendations to improve human performance and update the AI model to improve machine performance

HMT Prototype Summary

We prototyped a service-oriented generic HMT application

Key Concepts

- Service-oriented architecture enables ***modularity***, ***scalability***, ***reusability***
- **Front-end:** ***human/machine operators*** and ***teaming coordinator***
- **Back-end:** ***operating environment***, ***knowledge base***, and ***AI models***

Key Approaches

- **Data-centric HMT**
  - *Coordinators easily **review human vs. machine operations** with both responses and explanations for proper **AI auditing and trust analytics***
- **Episodic HMT**
  - ***Machine operators learn to emulate episodes of activity** from human operations*
- **Autonomic HMT**
  - ***Self-management of AI models** via an autonomous teaming coordinator*
  - *Driven by **digital engineering** and **system requirements/domain/goal models** (e.g., UML, etc.)*

Summary

We believe HMT can revolutionize automation for military & space

- We introduced HMT and its benefits
  - *Accelerate trust in autonomy by integrating with human user from inception*
  - *Archive human expertise over time*
- We introduced OUSD(R&E)'s OPTIMA program and shown HMT's benefits
- We outlined our vision, plan, and progress in HMT for OPTIMA
  - *Prototype generic Level 0 HMT application and start building on it using the AASA process, which includes the TAI Framework*
- We presented our generic Level 0 HMT prototype
  - *Intuitive Little Red Riding Hood "toy" problem demonstrates training*
  - *ImageNet object classification & drone detection demonstrate interpretability*
- Although our prototype HMT application is simplified for demonstration, our infrastructure supports broader use in military and space operations

References

[1] R. Pengzhen, Y. Xiao, X. Chang, P.-Y. Huang, Z. Li, B. B. Gupta, X. Chen, and X. Wang, "A Survey of Deep Active Learning," in *ACM Comput. Surv*, vol. 54, no. 9, 40 pages, 2021, doi: 10.1145/3472291.

[2] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in *Int Conf on Computer Vision (ICCV '17),* 2017, pp. 618-626, doi: 10.1109/ICCV.2017.74.

[3] A. B. Arrieta et al. "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI." *Information fusion,* vol. 58, pp. 82-115, 2020.

[4] R. S. Sutton and A. G. Barton, *Reinforcement Learning*, 2nd ed., MIT Press, 2018.

[5] J. O. Kephart and D. M. Chess, "The Vision of Autonomic Computing," in *Computer*, vol. 36, no. 1, pp. 41-50, Jan. 2003, doi: 10.1109/MC.2003.1160055.

[6] J. Cleland-Huang, A. Agrawal, M. Vierhauser, M. Murphy, and M. Prieto, "Extending MAPE-K to Support Human-Machine Teaming," in *Proc 17th Symp on Software Engineering for Adaptive and Self-Managing Systems (SEAMS '22)*, pp. 120–131, 2022, doi: 10.1145/3524844.3528054.

[7] Slingerland, P. and Perry, L., "Trusted AI Framework—Simplified Presentation." Aerospace Technical Report ATR-2022-01436, based on [8]

[8] Slingerland, P.C. and Perry, L.H., "A Framework for Trusted Artificial Intelligence in High-Consequence Environments." Aerospace Technical Report ATR-2021-01456 Rev A.

[9] Lutton, S., O'Brien, A., Schoenberger, R. (2021) "AI/autonomy Solution Architecting for OSAM as a Use Case," The Aerospace Corporation, OTR 2021-00800