AFRL-AFOSR-JP-TR-2024-0061

A competency-aware multi-agent framework for human-machine teams in adversarial environments

CINDY BETHEL
MISSISSIPPI STATE UNIVERSITY
245 BARR AVE
MISSISSIPPI STATE, MS, 39762
USA

03/24/2024
Final Technical Report

Air Force Research Laboratory
Air Force Office of Scientific Research
Asian Office of Aerospace Research and Development
Unit 45002, APO AP 96338-5002

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
20240324	Final	START DATE 20210618	END DATE 20231217

4. TITLE AND SUBTITLE
A competency-aware multi-agent framework for human-machine teams in adversarial environments

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
	FA2386-21-1-4015	61102F
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Cindy Bethel

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
MISSISSIPPI STATE UNIVERSITY
245 BARR AVE
MISSISSIPPI STATE, MS 39762
USA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AOARD UNIT 45002 APO AP 96338-5002	AFRL/AFOSR IOA	AFRL-AFOSR-JP-TR-2024-0061

12. DISTRIBUTION/AVAILABILITY STATEMENT
A Distribution Unlimited: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This work involved developing a generalized multi-agent human-machine teaming framework that can allow for changes in taskings, number of operators (human and machine), information reliability and operational factors. This model has been adjusted throughout the course of this project to take on new functionality as well as lessons learned from performance analysis. The results are very encouraging and the effort has been the source of multiple papers improving our understanding of the field. The collaboration proved quite successful and their final presentation to supporting members of the effort was well received.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT	b. ABSTRACT	c. THIS PAGE		
U	U	U	SAR	18

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
GEOFFREY ANDERSEN	

Standard Form 298 (Rev.5/2020)
Prescribed by ANSI Std. Z39.18

Annual Report for AOARD Grant FA2386-21-1-4015

"A Competency-Aware Multi-Agent Framework for Human-Machine Teams in Adversarial Environments"

March 15, 2024

Name of Principal Investigators (PI):

- E-mail address: cbethel@cse.msstate.edu
- Institution: Mississippi State University (MSU)
- Mailing Address:
  Computer Science and Engineering, P.O. Box 9637, Mississippi State, MS 39762-9637
- Phone: +1 662-325-2756 or cell: +1 813-316-8136

Period of Performance: 06/18/2023 – 12/17/2023

ABSTRACT

The goal of our research is to create a generalized multi-agent human-machine teaming framework that can be applied to various multi-objective tasks performed in unknown or potentially adversarial-filled environments. In the last six months of the project, we worked on formalizing the framework and began writing a journal article on the formalized framework and to provide examples of how to apply the framework in real-world scenarios. In the last month of the project and the month following, the framework went through another cycle of design changes as a result of attending several tactical training field exercises at the United States Military Academy. Consequently, this update to the framework necessitated major revisions to the journal article, which we aim to finalize and submit by the end of April to give time for all authors to thoroughly review, edit, and provide feedback. Also, in the last six months of the project, our collaborators in Australia underwent a personnel change, which has led to a re-design of the hide-and-seek grid-world for which we had been helping to design for use in testing the framework.

The two MSU PhD students whose dissertation work stems from this framework are currently completing an IRB to perform user studies, that evaluate how team coordination and information sharing affect a human's mental model development during multi-objective human-robot teaming. They will begin collecting data this month and one plans to graduate in December 2024 and the second one in May 2025. Their work focuses on distinct aspects of the human/AI mental models and shared understandings and how those can be shared to improve 1) situation awareness, mental workload, reliance, and team performance and 2) explainability and interpretability between a robot and human. We are interested in continuing our work related to this project and are seeking funding.

I. INTRODUCTION

Future combat teams operating in adversarial environments will utilize trusted autonomous systems to achieve mission objectives, such as identifying, classifying, locating, and suppressing threats while ensuring safety and survival of team members. To achieve this, humans and machines will work in teams, exploiting the unique potential of each team member in completing tasks with competing objectives (such as threat localization and identification, while minimizing detection or ensuring safe egress). Such complex planning tasks cannot feasibly be solved by a single decision maker (e.g. a human team lead), and agents must coordinate behaviors to achieve required mission outcomes, relying on an understanding of

the capabilities of other team members.

Trust between team members, and up the chain of command, is supported by explainable decisions and actions relative to mission objectives, particularly when teams include autonomous machines. In a multi-objective, multi-agent, human-machine planning problem, understanding, and explaining how agent actions impact goal attainment and how actions are inter-dependent between agents will assist command personnel to effectively plan, execute, and evaluate missions in complex environments, and support uptake of trusted autonomous systems in defense teams. The goal of this project was to create a human-machine teaming (HMT) framework that can generalize to incorporate multiple autonomous agent types as well as teaming configurations, e.g., one-to-one, one-to-many, many-to-many, many-to-one, human-UGV-UAV, etc.

II. MODELING HUMAN-MACHINE TEAMING – UPDATES

Since our last report, the shared mental model (SMM)-Centered Framework (Fig. 1) went through another iteration of design changes. As such, the Arbitration AI has a new compilation of capabilities and slight change in objectives. While the Arbitration AI's previous purpose was to oversee communication and delegate tasks, its new main purpose is to oversee facilitate collaboration among agents. The Arbitration AI's new functionality includes providing suggestions for optimizing performance (Multi-Task Optimization), planning alternate routes when requested or as part of performance optimization (Path Planning), calibrating reliance between agents (Reliance Calibration), and computing progress updates at the team and individual agent levels (Progress Evaluation Toward Subgoals/Goal). These changes were made in response to attending and participating in human-robot teaming tactical training

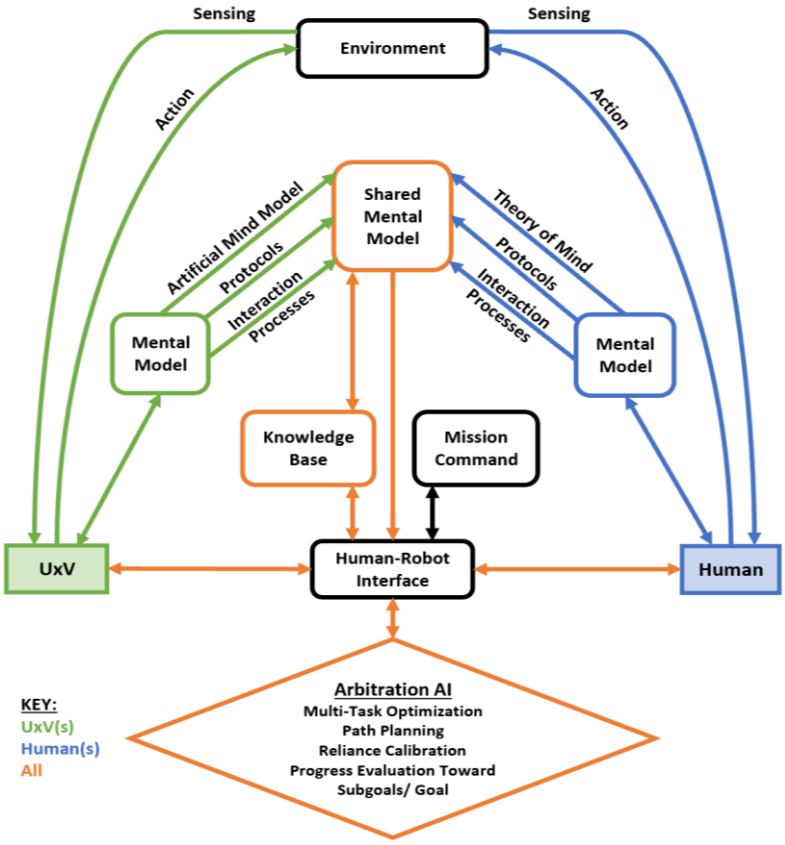

Fig. 1. SMM-Centered Human-Machine Teaming Framework

exercises in the field at the United States Military Academy. Many insights and observations gathered from those experiences have shaped our approach to promoting collaboration and reliance between human and robot teammates.

Because most of the other components in the framework remained the same through this iteration cycle, we do not go into extreme detail on the theoretical definitions of the components (provided in the previous report). We do, however, offer the computational definitions that have since been formalized. We also provide a real-world example that ties this framework with current technologies being used in the field by warfighters (in a later section). This SMM-centered meta-model (model of models) provides a general overview of teaming processes and can be broken down into interdependent sub-processes or components. When particular components are grouped together, the resulting structures of sub-processes conceptualize taskwork and teamwork within a team. The framework is adaptable to incorporate multiple agents, regardless of type (unmanned ground vehicle, unmanned aerial vehicle, unmanned vehicle (UxV), on-board autonomous agent, etc.). `Agent' refers to any member in the team (human, UxV, etc.).

For context when describing the framework's formal and computational functions, imagine an application such as search and rescue or hide that potentially includes adversaries in the field. These types of scenarios can quickly become complex and time restrictive enough to require the deployment of large teams for successful completion. With larger teams, efficiency may increase but so does the risk of losing human lives. That is why reducing the number of humans in the field and improving the outlook on robots as equal partners is so important. As technology advances and becomes more mainstream, warfighters need to be ready to incorporate robots into their teams. Through formally defining this framework, the authors portray how robots can integrate with human teams (utilizing advanced technology) without requiring extra personnel or cognitive capacity to control and oversee robot performance.

For easy understanding of framework definement and explanation clarity of the complex search task, one human and one UxV are a team performing coordinated actions in an environment while monitoring their surroundings for environmental cues (or adversaries) and updating the team's SMM. In such a team, teammates (human and robot) must be able to collaboratively work toward a shared goal, dynamically adjust their plans in response to environmental cues, and adapt their plans to assist their teammate.

A. Knowledge Base

Before starting a mission, Mission Command will provide information compiled during the pre-mission planning phase. Such information, stored in the Knowledge Base (Fig. 2), includes Mission Goals, Protocols, Agent Roles, and Agent Capabilities, e.g. expertise, agent equipment and function. An initial mission plan which details the main goal, subgoals, and paths each agent should take will be input to the Knowledge Base either by Mission Command or agent(s) in a team. Ideally, when mission planning software, such as the Team Awareness Kit (Wintak, ATAK) and the Interactive Tactical Decision Game (ITDG) system, becomes more widely used in the field, the mission plan will be sent from this software. The mission plan will be translated by the Arbitration AI and sent to each agent. Agent Locations will be provided by the technology accompanying each teammate, assuming the human has some technology with GPS capability.

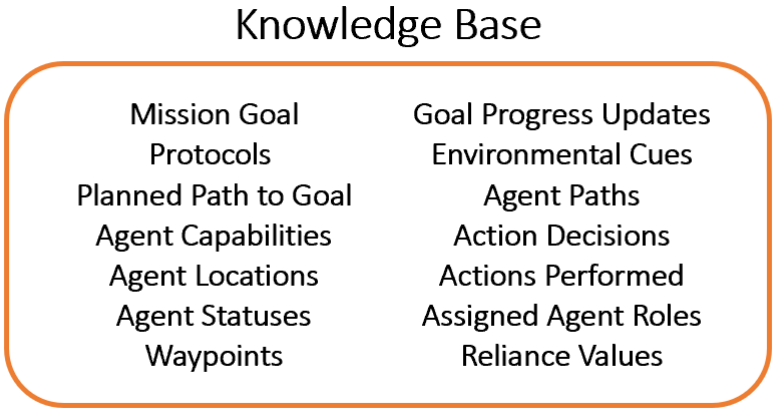

Fig. 2. Knowledge Base: a partial view of the information it stores.

While access to this information might be helpful to teammates under certain circumstances, it could become overwhelming for some, especially in unfavorable conditions. For example, if an agent is stuck making a difficult decision, such as needing an easy, short, and safe alternate path to evacuate an injured person, accessing all the information would help the agent make a well-informed decision. Alternatively, if an agent is cognitively overwhelmed or taking on too many tasks because teammates are not trusted or relied upon, having more information to filter through could incapacitate the agent. These issues were considered when conceptualizing how a team could optimize information sharing and use SMMs. Upon consideration, it was determined that proper communication, reliance, and task expectations need to be established for the team and each agent.

B. Mental Modeling

Instead of approaching mental model design with a specific task or team in mind, a set of inputs and outputs that might influence the development of human mental models, AI mental models, and SMMs, specifically for collaboration, were compiled along with a set of inputs and outputs that might influence their maintenance. By defining mental model requirements this way, reverse engineering can be used to decipher which mental model components are necessary for the development and maintenance of human mental models, AI mental models, and the translation between them. We hypothesize that because humans and AI agents are inherently different beings, they do not need mirror-image mental models to collaborate effectively, and they likely do not require the same components to be present in their respective mental models. Their mental models simply need to serve the same purpose, which is staying up to date on the ongoing taskwork and teamwork.

For our intended purpose of improving collaboration between humans and AI agents (in teams), we propose giving teammates access to a representational form of their team's SMMs and other agent factors in real-time via a human-robot interface. For this application, we do not intend to develop mental models that fully encompass every function and cognitive process required for HMT. We are not presenting novel mental model implementations but highlight the mental model aspects being considered for inclusion in representing a team's SMM. We do this by drawing upon components from other researchers' cognitive modeling/mental modeling approaches to provide examples of the mental model aspects in which we are most interested and how they might be formalized. We view these examples as a starting point from which we can modify and combine components to fulfil our intended purpose.

i. Human Mental Models

A human's mental model is comprised of knowledge and an abstract world model of the surrounding environment and what is happening in that environment. In the human-robot interface, we will visually represent a shared map of the external world that humans will update in a similar manner as performed on an ATAK device. AI agents would have an alternative form of updating the shared map (see next section). It is not intended for a human to update a shared map with every incorrect detail perceived. Humans are expected to share what is important for others to know to make informed decisions.

Knowledge, on the other hand, is a much more difficult concept to materialize and convey to others without explicit communication, which is not optimal for tasks such as search and rescue or hide and seek. (Jodlowski2004) created a Knowledge Scoring Engine that generates a knowledge base for an individual in real-time and interprets actions based on context to update and maintain an accurate knowledge base (in real-time). Jodlowski and Doane (2004) do this by time synchronizing incoming data, parsing the continuous flow of data into meaningful discrete event segments, and making inferences based on the parsed discrete event data (Jodlowski2004). The Knowledge Scoring Engine was integrated with the Adaptive Dynamics and Active Perception for Thought (ADAPT) cognitive architecture, a construction-integration model of cognition, but can easily be integrated with Adaptive Character of Thought – Rational (ACT-R), a rational model of cognition. While the Knowledge Scoring Engine is promising as a solution for measuring an agent's knowledge, it is very dependent on eye-tracking data, which is not currently feasible for real-world applications. We would like to use a similar approach as the Knowledge Scoring Engine, as it provides a promising starting point (Jodlowski2004, Brou2009), but more research needs to be conducted to determine an alternate and effective method for creating propositional thought logic (instead of using eye movement).

ii. AI Mental Models

Many approaches for creating artificial mind models have been reviewed and are discussed in our upcoming journal article. The approach in (Devin2016) for creating an artificial Theory of Mind (ToM) not only models joint activities and the execution of shared plans between human and robot teammates, as many others do, but also considers the knowledge and perspective of other agents. However, this approach is based on building the spatial perspective of other agents by using a robot's incoming sensor data, meaning the robot must always keep `eyes on' its teammates in order to maintain a current world state. This works well in certain situations but is not feasible for all scenarios, such as a search and rescue environment where trees may block a drone's view of a human teammate, or teammates need to collaborate without directly seeing each other. This is where our approach of having a shared world state and sharing a representation of mental model components comes into play.

By creating a world state compiled of information gathered and shared as each agent views and experiences the environment, AI agents will not need to keep `eyes on' their teammates. They will be able to use the information from this world state, which will be shared with humans (via the human-robot interface) as a visual map for easy assimilation and as a set of affordances and relations for AI agents. This form of creating a world state would take the place of (Devin2016)'s approach of using a robot's sensor data to gather observable facts, but this is only part of the adaptation that would need to be made. There are also non-observable facts for each agent that need to be considered. As is stated in (Devin2016), these facts can be shared directly by an agent, or they can be observed and assumed as a side effect from an action being performed. Following our first adaptation,

we plan to share these types of facts through the human-robot interface as well as infer them from agents' actions. More explanation for sharing observable and non-observable facts is provided in the Shared Mental Models section below.

iii. Shared Mental Models

It is hypothesized that incorporating visual representations of SMMs into a human-machine teaming interface (SMM-HMI) will establish improved team interdependency, implicit coordination, trust, and multi-directional reliance. Humans have exceptional pattern recognition capabilities, which makes them proficient at processing spatially-oriented information (Rouse1984). As such, mental models have a form of representation that is spatial or verbal, while their structure is hierarchical or planar. Since there is work suggesting that mental models can be updated via perception (Andrews2022) and that mental imagery is an internalized perception (Hestenes2006), it is reasonable to propose that mental models, at least in some part, are pictorial in form. A research group compared scenes generated from 1) narration and 2) direct observation and found no functional difference in the resulting mental models of the scenes (Tversky2000). They did, however, point out that while perceptions have a fixed linear perspective, inside of mental models, viewpoints can change, making them spatial and more schematic than images.

It is believed that the dimensionality of effective human-machine teaming can be reduced by visually representing abstract team processes and functionality. Using visual representations of how a human-machine team is functioning capitalizes on humans' pattern recognition capabilities and aids them in building more accurate associations among visualized objects and shared/perceived knowledge. In turn, human agents are expected to form context-accurate mental models and learn their teammates' needs, preferences, behaviors, and limitations more easily.

We have compiled a list of factors under consideration for inclusion in the visually-represented SMM. These factors include a shared world view of the environment, shared understandings for task- and teamwork, discrepancies between mental models (not their convergence but the differences), self-awareness statuses for all agents (agent status), mental demand/processing load, and an action state for each teammate. Mental model input/output data types and representative forms will be studied to investigate how using visual SMM data impacts situation awareness, mental workload, and reliance among teammates.

C. Arbitration AI

In the context of this paper and the HMT framework, the definition of team decision-making branches from the version found in (CannonBowers1993) fits the level of independence the framework provides. *Team decision-making* refers to the process of filtering, synthesizing, and communicating information thought necessary to help the decision maker(s) choose optimal task-relevant decisions that are dependent, independent, or interdependent of other team members. These task-relevant decisions are contingent on implications handled by the Arbitration AI component.

The purpose of the Arbitration AI is to partially remove complex decision-making responsibility from human agents and provide aid by having the Arbitration AI handle decisions based on computation and the information required for making informed decisions. Beneficially, having the Arbitration AI partially handle team decision making should not only reduce a human agent's mental workload but could also reduce the number of humans needed

in the field. By (Cannon-Bowers1993)'s definition, *team decision making* does not require the involvement or consensus of all teammates, but it does necessitate that teammates gather, process, integrate, recommend, and communicate relevant information for making informed decisions (CannonBowers1993). With all that goes into team decision making, it is clear that this teaming process is brittle, and its quality depends solely on a team's ability to maintain effective teamwork. For as soon as a teammate is overwhelmed by a momentary task demand, that teammate is no longer able to contribute new, relevant information or individual expertise to the team decision making process. With this in mind, it is envisioned that the Arbitration AI acts as an on-board AI teammate (OBAI) that gathers, integrates, and processes information from all teammates and uses this information to aid a team in making decisions. With a substantial portion of this task being handled by the Arbitration AI, teammates are free to use their cognitive resources to focus on other tasks.

i. Multi-Task Optimization

The Arbitration AI has many capabilities for aiding teamwork and taskwork, the first of which includes making suggestions of actions to be performed or of actions that should be changed based on inferences made from its access to information flowing among agents in a team. The idea behind multi-task optimization is to optimize tasks simultaneously (Zhao2023). Statistical decision theory can be used to find the maximum possible performance (Mumford2005) and applied in incidental information acquisition (act of monitoring information flow), whose central element is to recognize potentially relevant information (Heinstrom2014). Ideally this information would come from the Knowledge Base, SMM, what is sensed from the environment, and information shared through the human-robot interface. Specifically, the information that would be most helpful includes human and artificial theories of mind (beliefs, needs, preferences), agent capabilities, agent statuses, progress evaluations, reliance scores, agent locations, protocols, overall mission goal, mission sub-goals (if any), and environmental cues.

This Multi-Task Optimization suggestion capability, at the team or individual agent level, is not intended to take the place of Mission Command or the pre-mission planning phase of an operation but is meant to provide decision-making support if a need arises or a team/agent is amenable, e.g. if something unexpected happens in an unknown dynamic environment or an agent(s) is mentally overwhelmed, fatigued, or engaged. At the agent level, the Arbitration AI monitors and provides optimal alternative actions, when appropriate, for tasks being performed by each agent. An example of this would be to provide an alternate route(s) to an agent if new information is received about the environment that makes a route untraversable. There could be a drastic change to the environment that was previously unknown and discovered by another teammate. Rather than the teammate sharing the information via a human-robot interface and the agent having to then determine a new route or course of action, the Arbitration AI's suggestion could save the agent time and mental capacity that could be used to disseminate information or focus on other tasks.

ii. Path Planning

Additionally, and as part of its Multi-Task Optimization capability, the Arbitration AI can provide optimal paths to desired locations, based on pre-mission information and current information being shared between teammates about the environment. The path planning function uses weighted paths, which are calculated using the agent's status, terrain traversal difficulty, agent's current location, primary goal (target location), and an optional secondary goal, such as meeting up with teammate. Dependent upon the various factors that go into weighting a path, the Arbitration AI selects the best path(s) and

presents the path(s) to the agent(s) via the Human-Robot Interface. For this type of path planning, knowing the reason for needing a planned path could influence the type of path chosen by the Arbitration AI. Suppose there are two people in need of rescue - one suffers from multiple broken bones from a fall, the other is lost and dehydrated. The person with multiple broken bones will need to be carried while the dehydrated person could be given fluids on the scene and might be able to walk a short distance. For the person with broken bones, a shorter path with easily traversable terrain would be ideal for carrying the person to safety. An alternate plan of rescue for that person could be using an unmanned ground vehicle, for which flatter smoother terrain would be best. In this case, the path with the smoothest terrain might not be the shortest path to safety. The shortest path might be rocky, narrow, or include a steep hill. Depending on how dehydrated the second rescuee is, the short, narrow, rocky path might be chosen.

iii. Reliance Calibration

It is hypothesized that calculating an agent's reliance on their teammates will provide insight into that agent's cooperative behavior. Following (deFineLicht2021)'s definition of agential reliance, an agent must believe (or accept) that a teammate is competent (able to do the task) and motivated (willing to do the task) if given an opportunity to do the task. From this definition, it seems that reliance is a double-sided measurement between two teammates rather than a single score for individual teammates. One side of the measurement is a teammate's competence and motivation while the other side is a second teammate's acceptance of the first teammate's competence and motivation.

When assistance is needed, an agent notifies the Arbitration AI of the task for which assistance is requested. Rather than having an agent assess its teammates' capabilities and willingness to help, the Arbitration AI evaluates each teammate accordingly. It calibrates reliance by outputting a score reflecting each teammate's ability (competence) and availability (motivation) to successfully help complete a task. These scores are measured in relation to the task for which assistance is being requested. It is assumed that because teammates are all working toward the same goal, they are equally motivated to help each other accomplish tasks to reach the shared overall goal. As such, an agent's *motivation* is regarded as a measure of availability; however, priorities of the tasks each agent is performing at the time of the assistance request are considered. For example, if an agent is medically treating an injured rescuee, the medical treatment task will take priority over the task requiring assistance.

Competence, as it applies to calibrating reliance, stems from (CannonBowers1993)'s definition of task competence and refers to the set of knowledge, skills, and abilities a teammate possesses in relation to an assigned task or role. Competency and motivation ratings for agents in a team will depend on a set of factors determined as necessary for providing help or completing a requested task. The Arbitration AI will make implications based on how these factors align for each teammate. Factors considered for measuring competence include a teammate's abilities, distance to get to desired location, time to get to desired location, difficulty to get to desired location, current agent status, momentary (mental) task demand, and momentary task priority. Motivation or availability factors for consideration include (task/team) protocols, momentary (mental) task demand, distance to get to desired location, time to get to desired location, difficulty to get to desired location, momentary task priority, and current agent status. Due to overlapping factors influencing the measurements of competence and motivation, the sets of factors were merged to create a combined score that reflects the reliability of each teammate to assist in completing a task, i.e. *competence + motivation = reliability*. The following list contains an example of

factors and implications the Arbitration AI might make based on possible factor values (factor → implication):

- protocols → follows protocol / violates protocol
- abilities → match / partial match / no-match
- distance → too far / within reason / NA
- time → too long / within reason / NA
- difficulty → detrimental / within reason / NA
- agent status → optimal / within reason / poor
- momentary task demand → overloaded / balanced / underloaded
- momentary task priority → high priority / low priority

No single implication made can determine whether a teammate will be able and willing to help if requested, but when all implications are considered, the picture of which teammate will be able and willing to help becomes clearer. For instance, if *protocols* and *abilities* dictate a set of teammates who can help with a specified task, which teammate would most likely be willing or available to help? Eliminate teammates with poor agent statuses. Determine how quickly the request for assistance needs to be completed. Consider the remaining teammates' distances (to where they need to perform task), time to travel, and difficulty of their travel. Also consider whether this request for assistance is of a lower or higher priority than the tasks being performed by teammates. Then consider the remaining teammates' momentary task demand. It might not be the best decision to interrupt a teammate that is using a high cognitive load or processing load to perform a task. Alternatively, if the priority of the task requiring such a heavy load is of a low priority, the teammate's time and `mental' resources would be better spent assisting with the requested task.

For the Arbitration AI to calibrate reliance, let T be the set of all possible tasks for which an agent could need assistance, and a task $t \in T$ is defined as $t = \{task_t, Ab_t, Loc_t, Pr_t, TC_t\}$ and consists of the elements required of an agent in order for the agent to be capable of providing assistance. These required elements include the task name $task_t$, abilities Ab_t, location of the task Loc_t, priority Pr_t, and time constraint TC_t. The Arbitration AI will refer to these task requirements when assessing the reliability of each agent. The reliability scores are represented as a set RA_T, and a reliability score $ra \in RA_T$ is defined as $ra = \{P_{ra}, A_{ra}, DL_{ra}, Ti_{ra}, Df_{ra}, AS_{ra}, MTD_{ra}, MTP_{ra}\}$, where P_{ra} indicates whether the agent would be following (true) or breaking (false) task protocols to provide assistance. A_{ra} reflects if the agent has the ability to assist. DL_{ra} is the distance the agent must travel to where assistance is needed (Loc_t). Ti_{ra} is the time it would take the agent to travel to where assistance is needed. Df_{ra} is the difficulty of traveling to where assistance is needed, which could depend on terrain difficulty or amount of battery required to travel the distance. For the instance of battery level, a check will need to be performed to ensure that the robot teammate remains capable (enough remaining battery) of providing assistance after having traveled the distance. AS_{ra} is the agent's current status. MTD_{ra} is the agent's momentary task demand, which could be reflected as processing load or mental workload; and MTP_{ra} is the momentary task priority for the task the agent is currently performing.

Reducing an agent's competence and motivation to a list of relevant elements provides a clearer depiction of how to assess a teammate's reliability. However, there remains an essential piece for measuring reliance between teammates, acceptance. *Acceptance* is the act of embracing an idea, thing, or person without attempting to

change, control, or avoid it (McAndrews2019). Instead of trying to predict or perceive a teammate's acceptance of an agent's reliability, we attempt to fill the gap in our two-sided reliance score by having the Arbitration AI compute a confidence score for each agent. The basis of the confidence score stems from the following question: *Would you rather have help from an agent that is more willing (available) or more able?* The answer that was determined was, *it depends*, and, as such, the confidence score is task dependent.

The confidence score CS_T is dependent upon the priority of the task requiring assistance and whether that task is more constrained by time or ability. While the confidence score is computed for each agent and included in the final reliance score RL_T it is especially useful for breaking a tie between two or more agents and finding alternative agent pairs that may be better suited to assist a teammate rather than a single agent. The final reliance scores are represented as a set RL_T, where a reliance score $rl \in RL_T$ is defined as $rl = \{CS_T, RA_{Tbest}\}$. In this equation, RA_{Tbest} is the agent or agent pair best suited to assist a teammate in a given task. For instances of a tie in reliability between two or more agents, the confidence score will identify which agent is better suited to assist with the task, whether it is the agent with the lowest time of travel or the agent that is better matched in ability.

For example, an agent (unmanned aerial vehicle (UAV)) requests assistance for a low battery level. It does not have enough remaining battery to fly to the nearest teammate, and, consequently, sends the Arbitration AI the task request, an estimation of time (in minutes, e.g. 5 min.) remaining on battery life, and the location it will land. In this example, assume that recovery of the UAV has a priority of 8, meaning the UAV should not be left on the ground for an extended period of time. The Arbitration AI determines that, for the $\mathrm{CHANGEBATTERY}\{\mathrm{replaceBattery}, \mathrm{hasBattery}\}$ task, three agents have the same reliability scores and are as follows (note: example shows partial list of elements with arbitrary values):

Agent 1 {HUMAN}
- partial abilities (can change battery, has 0 UAV batteries)
- distance = is close-by
- travel time = 6 min.
- momentary task priority = 5

Agent 2 {Unmanned Ground Vehicle (UGV)}
- partial abilities (cannot change battery, has 3 UAV batteries)
- distance = is close-by
- travel time = 7 min.
- momentary task priority = 4

Agent 3 {HUMAN}
- complete abilities (can change battery, has 2 UAV batteries)
- distance = is farther away
- travel time = 13 min.
- momentary task priority = 7

The Arbitration AI considers the high priority of the task requiring assistance along with the time constraint accompanying the task and determines that the first two agents

(human and UGV), being closest to UAV landing location and jointly able to assist, are the best agents to help their UAV teammate. In the unlikely instance of a tie in reliability that results from equivalent element values between two or more agents, the Arbitration AI will let the teammate requesting assistance choose the agent.

iv. Progress Evaluation

The Progress Evaluation component assesses performance at the team level and agent level. From its output, a person should be able to get an idea of how much of the task a team has completed (%) and how an individual agent's actions and decision-making are affecting taskwork and teamwork. A team's evaluation can be viewed as how much of the primary area A_p, secondary area A_s, and total environment A_t have been search and if the target's location has been realized $TargetLocationRealized$. These values are reflected as a set of updates, U, where an update $u \in U$ is defined as $\mathrm{u} = \{\mathrm{A_{pu}}, \mathrm{A_{su}}, \mathrm{A_{tu}}, \mathrm{TargetLocationRealized}\}$. The evaluation of an agent is more focused on how an agent's performed actions affect task completion and the agent's status. This includes an agent's proximity to the target (main goal) and the resulting change in an agent's status after performing an action. An agent's evaluation is reflected as a value pair that provides a `task' score and `agent' score, such as (task, agent). The following factors are used for computing an agent's progress: agent's distance from previous location to main goal D_{Pi}, agent's distance from current location to main goal D_{Ci}, main goal (target) location L_T, agent's previous status S_{Pi}, and agent's current status S_{Ci}.

In evaluating a team's progression, a clear understanding of the task must be realized, making the progress evaluation context-specific regarding the task at hand. For this scenario, a team has identified primary and secondary search areas where they believe they will find the target, an injured person. The main goal is to rescue the target from the environment. A team accomplishes this goal by safely and efficiently removing the target from the environment, presumably so that the target can seek expert medical attention if necessary. The main goal is broken down into three subgoals: locate target, reach target, rescue target. The following example walks through a simplified set of questions and subgoal checks, presented in a sequence, which reflects a team's natural order of movement progression through a search and rescue task/environment.

TEAM EVALUATION
$TargetLocationRealized$
while FALSE, how much of total environment has been searched?
    has primary area been searched?
        if NO, how much has been searched?
        if YES, $TargetLocationRealized$?
    has secondary area been searched?
        if NO, how much has been searched?
        if YES, $TargetLocationRealized$?
while TRUE, $isRescued$?
    if YES, exit
    while NO, $isReached$? (has agent reached target?)
        if YES, exit
        while NO, moving to L_T,
            $isReached$? (if YES, exit)

In the example above, the progress of a team is evaluated as the team moves through an environment. The Arbitration AI computes and records the percentages of the total environment, primary search area, and secondary search area that have been searched. It also records when an area has been completely searched and if and where the target was located ($TargetLocationRealized$). Upon locating the target, teammates discontinue their current searches and begin to either move toward the target or perform some other task. This continues until 1) the appropriate agent(s) has reached the target and 2) the target has been rescued.

Similar to the team evaluation, progress evaluation at the agent level is context-specific. It will eventually be role-specific as well, but, currently, we are focusing on evaluating an agent's progress as a result of its decision-making. In the studies we have planned for the near future, we reduced search and rescue to a search task in order to simplify the problem space. For this reason, the following example only outlines the search part of search and rescue. Consequently, for the purpose of our studies, the team evaluation will not contain the $isRescued$ component, but it will contain the $isReached$ component. Having the appropriate agent reach the target will signify the end of the search task, which is why an agent's role will need to be considered.

As previously stated, there are two components in evaluating an agent that will be scored, the `task' and the `agent'. For this evaluation, the `task' will be reflected as the agent's proximity to the target. For the `agent' score, the change in an agent's status after performing an action (movement in the environment) will be denoted by a numerical value. Using values to delineate change in agent statuses affords a partial view of decision effect, i.e. did the action performed by an agent have a positive effect on the task, team, and/or agent, or is the agent's decision-making causing harm to the task, team, and/or agent. For instance, if an agent moved closer to the target, but the agent's status dropped from "good" to "fair" due to expending energy, the status change is reflected as a 1. In the search task, the agents want to reach the target, hence, using energy to move closer is rewarded rather than penalized. Alternatively, if an agent's proximity to the target does not change or gets worse, and the agent's status change was due to expending energy, the agent is neither

rewarded nor penalized. The agent could be moving to assist a teammate. The agent could be changing routes due to an impasse. The agent could also have gone rogue or been compromised, but those possibilities are unaccounted for at this time and will be included in future work. The possibilities we are currently considering for inclusion in status-changing instances are internal damage (e.g., software failure in UxV or ATAK device), external damage (e.g., physical crash, dehydration), and acts of reliance (requesting/ providing aid).

AGENT EVALUATION (once $TargetLocationRealized$)
did agent move closer to L_T? ($D_{Pi} - D_{Ci}$)
if YES (positive difference (1)):
change in agent status denotes some number:
if $S_{Pi} \rightarrow S_{Ci}$ = no change (1), improved (1)
if $S_{Pi} \rightarrow S_{Ci}$ = worsened:
- was it due to using energy/ battery for traversal? (1)
- status-changing instance (UxV)? e.g., software failure (0), physical crash (1), low battery level (1), helping teammate or target (1)
- status-changing instance (human)? e.g., software failure (0), medical attention (dehydration -1), low energy (1), helping teammate or target (1)

if NO (zero (0) or negative difference (-1)):
change in agent status denotes some number:
if $S_{Pi} \rightarrow S_{Ci}$ = no change (0), improved (1)
if $S_{Pi} \rightarrow S_{Ci}$ = worsened:
- was it due to using energy/ battery for traversal? (0)
- status-changing instance (UxV)? e.g., software failure (0), physical crash (-1), low battery level (0), helping teammate or target (1)
- status-changing instance (human)? e.g., software failure (0), medical attention (dehydration -1), low energy (0), helping teammate or target (1)

The next step in the progress evaluation update at the agent level is to include a variable for agents involved in an act of reliance (requesting aid, assisting teammate). This will change how an agent's progress is evaluated to ensure agents are not penalized for requesting help or providing assistance to a teammate.

D. Human-Robot Interface

For formalizing and validating the framework, the initial focus has been on determining the information needed for teammates to form context-accurate SMMs and improve their situational awareness, how and when to best share such information, and methods for promoting reliance and trust among teammates. If done in such a way that transparency is increased between human teammates and AI teammates, then human teammates may begin to view and trust AI agents more as equal partners than as tools. With enhanced information sharing, mental model development, and team functionality, agents (human, AI) of a team can learn to anticipate their teammates' behaviors, preferences, and needs as well as understand their capabilities and limitations. In designing an interface to support this type of collaborative teaming, there is a need to determine the most effective way of visually representing knowledge, mental models, and shared understandings to human teammates. This research is

expected to enable better team coordination and performance through improved shared understandings and reliance by incorporating visual representations of knowledge, shared mental models, and team functionality into a human-machine interface. This work falls under one of the PhD student's dissertation work and thus will be continued as future work.

Visualizations are effective at conveying knowledge between teammates while summarizing complex problem spaces (Eppler2007, Card1999). They reflect structure and associations among data, promote rapid information assimilation, and provide insight beyond the information being displayed. In examining the state of the art, very few researchers have created visualizations of mental model components or SMMs for use by human teammates during human-machine teaming. One group created a shared belief map that assisted humans in sharing information and revealed abstracted SMMs forming within a team (Fan2011). The SMM representations helped to identify discrepancies in the information each team member possessed and the information needs of each teammate (Fan2011). Other researchers implemented a similarity matrix visualization of previous experiences to improve a decision maker's understanding of a decision space (Hanratty2009). Their results showed that the visualization positively affected the decision maker's understanding. These results indicate the potential of what visual SMMs can do for human-machine teams; however, further investigation is needed to understand how to effectively display representations of SMMs.

Visual information is processed firstly by orientation, color, and texture, secondly by object identification and spatial localization. Sketches are versatile, help with reflection and communication, support reasoning, and can quickly visualize an idea (Burkhard2005). Sketches are heavily used in the military planning process and, in turn, are used to map strategic planning in current mobile interfaces, such as in the Android Team Awareness Kit (ATAK). Diagrams are visually intuitive structures that are good at conveying relationships, complex and abstract concepts, analytical knowledge, and systematic information. Maps are good at showing the whole picture, details, and the relationships in between. They can graphically represent knowledge as well as common stories (Burkhard2005). From reading about the various types and forms of visualization along with the preconceived ideas of what should be represented in the human-robot interface, it has been predetermined that a combination of maps and diagrams would be most effective at visually and spatially conveying important teaming factors, such as shared understanding, human cognitive load, robot processing load, teammate status, and progress updates.

Leveraging a human's ability to easily recognize patterns and process spatial information, a novel visual representation interface for displaying shared understandings, differences in mental model convergence, and team functionality is currently under development for use in human-machine teaming. Using spatial and representational forms, teammates will be able to see real-time calculations of various teaming factors and functions, including mental workloads, shared understandings, mission progress updates, situation awareness, and self-awareness (agent statuses). The next steps in collecting visual interface requirements are to construct a hierarchical task analysis (HTA) diagram and mental model (MM) diagram for the search tasks found in search and rescue and hide-and-seek, merge the mental model diagram with the HTA diagram, and conduct an in-depth analysis of the resulting HTA-MM diagram to learn about gaps in the current technology's interface and to identify needs and expectations not being met.

End Goal: design and implement our human-robot interface as a plugin for ATAK devices.

III. EXPERIMENT

Our current/new hide-and-seek environment (still under development by Australian collaborators)

Study Plan

In an initial study, a human user will navigate three 2D grid-world environments of equal difficulty, performing the seeking portion of a hide-and-seek task. In one environment, the human user must work alone under the constraint of knowing very little information about the environment. In the other two environments, the human user will work with a drone teammate, following two different sets of task protocols (one per environment) and utilizing two different sets of drone capabilities (one per environment). The two task protocols are coordinated together-movement, meaning the human and drone teammate move through the environment together, versus a divide-and-conquer approach. For the scope of this study, the divide-and-conquer movement will be synchronous (via waypoints) to more easily observe the information sharing and communication stages of the task. The drone teammate will either have 1) basic sensing functionality that uses way-point navigation or 2) path planning capabilities and optimal path recommendations in addition to basic sensing and way-point navigation. In a within-between subjects design, participants will be randomly assigned (and counterbalanced) to one of the task protocols but will experience both sets of drone capabilities within their assigned task protocol.

This initial study will collect baseline data for various factors of team functionality, including blind trust, reliance, situation awareness, and mental workload. Mental models for human users will be generated to study how the structures and relationships are affected by different task protocols, agent capabilities, and team configurations (human, human+drone).

The primary path of investigation will be the information and interactions needed to establish shared understandings among agents for effective collaborative human-machine teaming. Mental model development along with other factors of team functionality will also be analyzed to answer the following research question: *How do information sharing, task protocols, and teammate capabilities influence various factors of team functionality, including performance, trust, reliance, mental workload, situation awareness, and mental model development?* To validate human mental models, a cognitive mapping methodology, such as Pathfinder, will be used to create graphical representations. Once created, both elicitation (content) and representation (relationships) of mental model elements will be assessed and compared to a global reference for accuracy and to those from other participants' mental models for evaluating similarities and differences (Mohammed2000). Because mental models are typically thought of as being image-like (Rouse1984), it is believed that team members will have a relatively easy time of developing their mental model depictions of the scenario from a 2D grid-world. This study will help with determining the minimum requirements to be included in visual representations of a human-machine team's knowledge and mental models.

IV. SUMMARY

In conclusion, we revised many components of the human-machine teaming framework, particularly the Arbitration AI Module, and have begun formalizing several components. We plan to continue progressing the framework as a whole but will focus on the reliance calibration, mental model explainability/interpretability, and SMM-HMI through two PhD dissertations. We are continuing work on a journal article that provides extensive detail of our framework design and how we plan to computationalize the many processes of human-machine teaming.

List of Publications and any Significant Collaborations that resulted from your AOARD-supported project <u>for this year of the project</u>: In standard format showing authors, title, journal, issue, pages, and date, for each category list the following:

a) papers published in peer-reviewed journals: **one article to be submitted April 2024**
b) papers published in peer-reviewed conference proceedings:

A. L. Aldridge and C. L. Bethel, "M-OAT Shared Meta-Model Framework for Effective Collaborative Human-Autonomy Teaming," ***In Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23),*** **Association for Computing Machinery, New York, NY, USA. 663–666, 2023. https://doi.org/10.1145/3568294.3580169**

B. Mooers, A. L. Aldridge, A. Buck, C. L. Bethel, D. T. Anderson, "Human-robot teaming for a cooperative game in a shared partially observable space," ***Proc. SPIE 12525, Geospatial Informatics XIII, 125250B,*** **June 2023. https://doi.org/10.1117/12.2663430**

c) papers published in non-peer-reviewed journals and conference proceedings - **None**
d) conference presentations without papers - **None**
e) manuscripts submitted but not yet published – **None-submitting in 202404.**
f) list any interactions with Air Force Research Laboratory, other US scientists/institutions, significant collaborations and/or AOARD-funded exchanges/visits that resulted from this work:

This grant had an unfunded collaborator, Derek Anderson at University of Missouri, with whom the Mississippi State University research team has extended our collaborations on this research. PI Bethel and collaborator Derek Anderson are further developing these models for use in search and rescue applications in military environments using multiple humans, unmanned ground vehicles, and unmanned aerial vehicles performing multi-objective tasks. We are interested in securing funding to expand this research direction.

g) any awards, promotions, or notable achievements – **Two PhD student dissertations in progress:**

Audrey L. Aldridge – titled: "Visual Representation of Shared Mental Models to Promote Collaboration and Omni-Directional Reliance in Human-Robot Teams" – planned graduation in December 2024.

Logan Cummins – title: still to be determined – planned graduation in May 2025.

REFERENCES:

1. R. W. Andrews, J. M. Lilly, D. Srivastava, and K. M. Feigh, "The role of shared mental models in human-AI teams: a theoretical review," *Theoretical Issues in Ergonomics Science*, 2022, doi: 10.1080/1463922X.2022.2061080.
2. R. J. Brou, S. M. Doane, and G. L. Bradshaw, "Real-time generation of representations for cognitive models," *Behavior Research Methods*, 41(3), pp. 633-638, 2009, doi: :10.3758/BRM.41.3.633.
3. D. J. Bryant, B. Tversky, and M. Lanca, "Retrieving Spatial Relations From Observation and Memory," *Cognitive Interfaces: Constraints on Linking Cognitive Information*, E. van der Zee and U. Nikanne (eds.), Oxford University Press, ch. 6, 2000.
4. R. A. Burkhard, "Towards a Framework and a Model for Knowledge Visualization: Synergies Between Information and Knowledge Visualization," in *Lecture Notes in Computer Science*, pp. 238-255, 2005, doi: 10.1007/11510154_13.
5. S. K. Card, J. D. Mackinlay, and B. Shneiderman, *Readings in Information Visualization: Using Vision to Think*, Morgan Kaufmann Publishers Inc., San Francisco, CA, 1999.
6. J. A. Cannon-Bowers, E. Salad, and S. Converse, "Shared mental models in expert team decision making," in *Individual and group decision making: Current issues*, J. N. J. Castellan (ed.), Lawrence Erlbaum Associates, Inc., ch. 12, 1993.
7. K. de Fine Licht and B. Bulde, "On Defining "Reliance" and "Trust": Purposes, Conditions of Adequacy, and New Definitions," *Philosophia*, 49(515), pp. 1981-2001, 2021, doi: 10.1007/s11406-021-00339-1.
8. S. Devin and R. Alami, "An Implemented Theory of Mind to Improve Human-Robo Shared Plans Execution," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 319-326, 2016, doi: 10.1109/HRI.2016.7451768.
9. M. Eppler and R. A. Burkhard, "Visual representations in knowledge management: Framework and cases," *Journal of Knowledge Management*, 11(4), pp. 112-122, 2007, doi: 10.1108/13673270710762756.
10. X. Fan and J. Yen, "Modeling Cognitive Loads for Evolving Shared Mental Models in Human-Agent Collaboration," in *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(2), pp. 354-367, 2011, doi: 10.1109/TSMCB.2010.2053705.
11. T. Hanratty et al., "Knowledge visualization to enhance human-agent situation awareness within a computational Recognition-Primed Decision System," pp. 1-7, 2009, doi: 10.1109/MILCOM.2009.5379847.
12. J. Heinström, "Openness to experience – the exploration dimension", in J. Heinström (ed.), *From Fear to Flow: Personality and information interaction, Chandos Information Professional Series*, Chandos Publishing, pp. 15-37, 2010, doi: 10.1016/B978-1-84334-513-8.50003-0.
13. D. Hestenes, "Notes for a Modeling Theory of Science, Cognition and Instruction," in *Proc. 2006 GIREP Conf. Modelling in Phys. and Phys. Edu.*, 2006.
14. M. T. Jodlowski and S. M. Doane, "A Knowledge Scoring Engine (KSE) for Real-Time Knowledge Base Generation Use in Intelligent Tutoring Systems," in *Proceedings of the 37th Hawaii International Conference on System Sciences*, 2004, doi: 10.1109/HICSS.2004.1265330.
15. Z. McAndrews, J. Richardson, and L. Stopa, "Psychometric properties of acceptance measures: A systematic review," *Journal of Contextual Behavioral Science*, 12, pp. 261-277, 2019, doi: 10.1016/j.jcbs.2018.08.006.
16. S. Mohammed, R. Klimoski, and J. R. Rentsch, "The Measurement of Team mental Models: We Have No Shared Schema," *Organizational Research Methods*, 3(2), pp. 123-165, 2000, doi: 10.1177/109442810032001.
17. M. D. Mumford and L. E. Leritz, "Heuristics," in K. Kempf-Leonard (ed.) *Encyclopedia of Social Measurement*, Elsevier, pp. 203-208, 2005, doi: 10.1016/B0-12-369398-5/00168-7.
18. W. Rouse and N. Morris, "On Looking Into the Black Box. Prospects and Limits in the Search for Mental Models," *Psychological Bulletin*, 100(3), 1984, doi: 10.1037/0033-2909.100.3.349.
19. H. Zhao, X. Ning, X. Liu, C. Wang, and J. Liu, "What makes evolutionary multi-task optimization better: A comprehensive survey," *Applied Soft Computing*, 145, 110545, 2023, doi: 10.1016/j.asoc.2023.110545.