Integrating Human-AI Teaming into Modeling and Simulation for Live, Virtual, Constructive (LVC) Environments

O. Thomas Holland
Information and Communications Laboratory
Georgia Tech Research Institute
Stafford, VA
thomas.holland@gtri.gatech.edu

Elizabeth Whitaker
Information and Communications Laboratory
Georgia Tech Research Institute
Atlanta, GA
elizabeth.whitaker@gtri.gatech.edu

Margaret Loper
Information and Communications Laboratory
Georgia Tech Research Institute
Atlanta, GA
margaret.loper@gtri.gatech.edu

Larry Kimm
Information and Communications Laboratory
Georgia Tech Research Institute
Stafford, VA
larry.kimm@gtri.gatech.edu

Brian Martens
Information and Communications Laboratory
Georgia Tech Research Institute
Stafford, VA
brian.martens@gtri.gatech.edu

Ethan Trewhitt
Information and Communications Laboratory
Georgia Tech Research Institute
Atlanta, GA
ethan.trewhitt@gtri.gatech.edu

Moon Kim
Information and Communications Laboratory
Georgia Tech Research Institute
Atlanta, GA
moon.kim@gtri.gatech.edu

Clayton Hutto
Applied Systems Laboratory
Georgia Tech Research Institute
Atlanta, GA
clayton.hutto@gtri.gatech.edu

ABSTRACT

The rapid progression of Artificial Intelligence (AI) is redefining human-machine interactions, moving beyond systems as human-controlled tools toward human-AI teaming, where humans and AI collaborate as near-peer partners to achieve superior outcomes. Traditional human-machine collaboration has relied on machines as subordinate aids, but AI's increasing autonomy demands a new paradigm for cooperative performance, especially in complex operational settings. This paper argues for the integration of human-AI teaming concepts into modeling and simulation (M&S) frameworks supporting the Live, Virtual, Constructive (LVC) paradigm, a cornerstone of Department of Defense (DoD) training, evaluation, and mission rehearsal. As the DoD and industries prioritize AI adoption, current M&S approaches and LVC constructs lack the fidelity to represent human-AI team dynamics, risking unpreparedness for real-world applications. Incorporating these concepts is essential to enhance training, optimize system design, and ensure that simulation interoperability standards evolve to meet this emerging need. We propose leveraging LVC environments to explore human-AI teaming through hybrid simulations—combining live human-AI interactions, virtual scenarios, and constructive models—while assessing the adaptability of standards like HLA and DIS. Integrating human-AI teaming into M&S for LVC is expected to bridge critical gaps, enabling more effective preparation for AI-enhanced operations and offering a roadmap for advancing simulation capabilities.

ABOUT THE AUTHORS

O. THOMAS HOLLAND is a Principal Research Engineer at the Georgia Tech Research Institute. He is a technical area expert for modeling and simulation of emergent behavior systems and the development of solutions for modeling systems and managing the design, development, and demonstration of simulation experiments. His current research areas include human-AI collaboration, socio-physical measures of wargaming efficiency, and characterizing emergent phenomena in complex systems. He holds a Ph.D. in Modeling & Simulation from Old Dominion University.

LARRY T. KIMM is a Principal Research Engineer at Georgia Tech Research Institute (GTRI) and Manager of the GTRI Quantico, Virginia Field Office. With over 30 years of military experience (including as Director of the USAF HSI Office), Dr. Kimm specializes in Human Systems Integration, AI Teaming, and Model-Based Systems Engineering. He directs and supports research programs for the USMC, US Army, and DARPA, focusing on areas like robotics, logistics, and human performance in extreme environments. Dr. Kimm is the Course Director for ASE6131, HSI Analysis and Synthesis, at Georgia Tech and a retired Air Force Colonel. He holds a Ph.D. in Environmental/Environmental Health Engineering from the University of Florida.

MARGARET L. LOPER is Associate Director for Operations for the Information & Communications Laboratory at the Georgia Tech Research Institute in Atlanta, Georgia. She has been involved in parallel and distributed simulation for more than 30 years, and has taught simulation courses for both academic and professional education. She holds a Ph.D. in Computer Science from the Georgia Institute of Technology.

BRIAN D. MARTENS is a Research Scientist at the Georgia Tech Research Institute (GTRI), where he specializes in Model-Based Systems Engineering, Modeling and Simulation, and emerging applications of human-AI teaming. His work centers on developing innovative, scalable solutions that enhance decision-making, operational efficiency, and trust between humans and intelligent systems. Brian holds an M.S. in Applied Systems Engineering from the Georgia Institute of Technology.

ELIZABETH WHITAKER is a Georgia Tech Research Institute (GTRI) Principal Research Engineer (RBW) and GTRI Fellow with over 30 years of AI research experience. She has intelligent tutoring expertise (dissertation and multiple research projects) as well as having researched machine learning, cognitive reasoning and statistical data analysis techniques applied to agent-based modeling and simulation of human behavior. She holds a Ph.D. in Electrical and Computer Engineering from the University of South Carolina.

ETHAN B. TREWHITT is a Senior Research Engineer at Georgia Tech Research Institute (GTRI) and is the Division Chief Scientist of the Software Engineering and Analytics Division (SEAD). He is an expert in knowledge-based artificial intelligence (AI) and cognitive systems. His additional research interests include user-centered design, intelligent tutoring and student modeling, serious games, natural language processing, network systems security, and software-defined radio. He holds an M.S. in Electrical and Computer Engineering from the Georgia Tech.

MOON KIM is a Senior Research Associate at the Georgia Tech Research Institute (GTRI) specializing in human-centered research and design. Since joining GTRI in 2009, she has led projects informing the intersection of people, process, and technology for government clients, including the U.S. Army and the CDC. Her recent work includes modernizing software for Army social work services, implementing data standardizations for the CDC's Data Modernization Initiative, and exploring human-AI teaming. She holds an M.S. in Public Policy and an M.S. in International Affairs from the Georgia Institute of Technology.

INTRODUCTION

The increasing adoption of Artificial Intelligence (AI) has led to a shift from traditional human-machine collaboration to human-AI teaming, where humans and AI systems collaborate as near peers to achieve shared goals. However, it is essential to recognize the distinction between human-machine teams, human-AI teams, and the AI-enhanced environments in which they may operate. The increasing adoption of Artificial Intelligence (AI) is significantly evolving the nature of human-machine interactions. While *human-machine collaboration* has historically involved humans interacting with machines as tools, recent advances in AI has facilitated the evolution towards *human-AI teaming*. In this paradigm, humans and AI systems collaborate as near-peers, leveraging their respective strengths and partnering to achieve shared objectives. It is crucial to acknowledge that human-AI teaming represents a specific form of human-AI collaboration, distinguished by the higher degree of autonomy and shared responsibility exhibited by AI systems within the team.

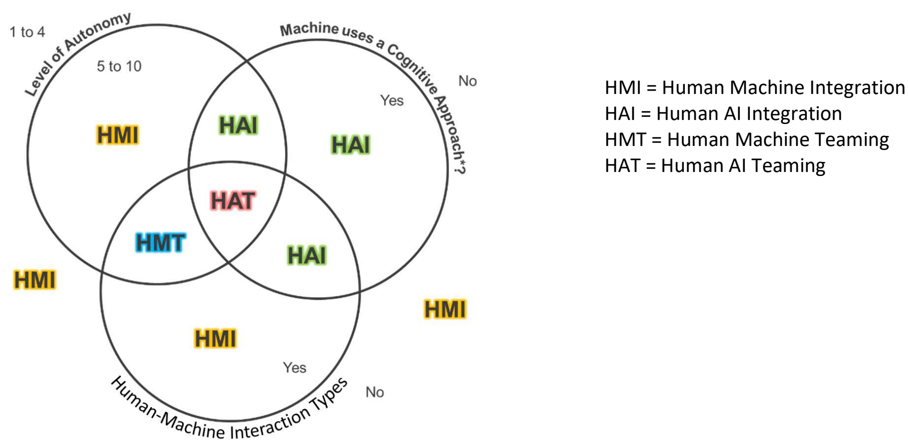

Figure 1. The intersectionality of human-AI teaming

Humans and machines may function within contexts significantly shaped by AI, such as simulations with AI-controlled entities, AI-driven decision support systems, or real-world settings with integrated AI systems (e.g., autonomous vehicles, smart cities). In this paper we focus on that intersection where humans and AI rely on each other to fulfill their roles and complete tasks toward collective goals. Although humans may team with various machines of varying sophistication or capability, we are interested in AI systems that understand context, make decisions, and learn from experiences, and so we distinguish them from the more basic automated systems or machines that operate without these cognitive functions. Figure 1 illustrates the relationship between different types of human-machine interactions, particularly focusing on the role of AI. The diagram depicts three key dimensions that define the spectrum of human-AI collaboration:

1) **Level of Autonomy**: This dimension represents the degree of independence the machine possesses in performing tasks.
2) **Machine Uses a Cognitive Approach**: This dimension signifies whether the machine employs cognitive processes, such as learning, decision-making, and problem-solving, in its operations.
3) **Human-Machine Interaction Types**: This dimension encompasses:
   - HMI (Human-Machine Integration): Basic interactions where humans utilize machines as tools.
   - HAI (Human-AI Integration): Interactions involving AI technologies, where the AI primarily assists human tasks.
   - HMT (Human-Machine Teaming): Collaborative interactions where humans and machines work together as a coordinated team, but the machine's role remains primarily supportive.
   - HAT (Human-AI Teaming): The highest level of collaboration, characterized by near-peer relationships between humans and AI systems, involving shared decision-making and responsibility.

Complicating this is the fact that these distinctions are not always clear-cut, and there can be overlap between these different types of human-AI interaction. For example, the specific characteristics of each type of interaction will vary depending on the use-case, the specific AI technologies involved, and the desired level of human involvement. Our primary focus lies on the HAT-type collaboration, particularly its implications within what we might call AI-enabled Live, Virtual, Constructive (LVC) environments. It is important to understand that while interaction is a necessary component of teaming, not all interactions constitute teaming. When teaming, humans and AI systems function as true partners, sharing goals, responsibilities, and decision-making authority (O'Neill, Flathmann, McNeese, & Salas, 2023). AI systems that serve as true collaborators have agency and the ability to learn and adapt. As such, activities, such as LVC training events, that expect human-AI teaming must emphasize interdependency, mutual trust, and shared understanding. In this context, human-AI teaming is a richly collaborative relationship between humans and AI systems, where both entities work together to achieve a common goal. To distinguish HAT from other forms of human-AI collaboration, we must consider:

- Level of AI Autonomy: Human-AI teaming involves higher levels of AI autonomy and shared responsibility compared to human-autonomy interaction.
- Focus of Collaboration: Human-AI teaming emphasizes true collaboration and partnership, while human-autonomy interaction and machine-machine teaming with human oversight prioritize human control and system reliability.
- Decision Making Authority: In human-AI teaming, decision-making authority is often shared between humans and AI, whereas in other models, humans retain primary control.
- Complexity: Human-AI teaming in both training and operational environments presents unique challenges due to the increased sophistication of the AI entities, the interactions between human and AI decision making elements, and the need for robust trust and collaboration.

Effective human-AI teaming in AI-enhanced environments requires a deep understanding of human cognition, behavior, and social dynamics, as well as the capabilities and limitations of AI systems and their impact on the surrounding environment (National Academies of Sciences, Engineering, and Medicine, 2021). Prior research in teaming performance suggests, "For an AI system to be a part of a team, it must be capable of interdependence in its operations, as well as a degree of autonomy in its execution" (National Academies of Sciences, Engineering, and Medicine, 2021) (Reyes, Dinh, & Salas, 2019) (Salas, Cooke, & Rosen, 2008). Achieving this level of collaboration and autonomy demands a nuanced understanding of how humans and AI systems can work together effectively within these dynamic and increasingly complex environments.

Human-AI Teaming and LVC

Human-AI interaction in AI-enhanced environments requires a structure framework to guide the design, development, testing, and evaluation of effective collaboration. Unlike human-controlled tools, AI systems designed to cooperate with humans achieve greater capabilities through partnership. These systems vary in adaptability and learning potential, requiring intentional design to ensure effective teaming.

Evaluating human-AI teaming is complex and often infeasible, presenting challenges in designing systems that operate safely as intended. Deliberate exploration of various teaming forms is crucial as the U.S. DoD seeks to develop new human-AI capabilities. Effective training and evaluation are essential for these systems to function safely in operational environments. Currently, there is a capability gap in designing human-AI teams and using training to drive innovative system design and refinement.

The DoD’s LVC paradigm, known for training, evaluation, and mission rehearsal, can address these challenges but hasn’t been fully adapted for human-AI teaming. Standards for LVC environments are crucial to ensure seamless interaction among live, virtual, constructive, and AI-enhanced systems. These standards enable consistent evaluation metrics, validating the performance and safety of these systems.

LVC AND THE CHALLENGES OF HUMAN-AI TEAMING

The Department of Defense (DoD) has established standards for Live, Virtual, and Constructive (LVC) training environments to ensure interoperability, reusability, and effectiveness (Allen, Lutz, & Richbourg, 2010). The DoD's LVC standards are primarily guided by the High-Level Architecture (HLA) and the Distributed Interactive Simulation

(DIS) protocols. HLA is a standard for distributed simulation, which enables different simulations to interoperate and exchange data in a standardized way. The DoD has also established the Test and Training Enabling Architecture (TENA) as a common architecture for LVC environments. TENA provides a standardized framework for integrating live, virtual, and constructive elements, enabling more effective and efficient training. Additionally, the DoD's Modeling and Simulation (M&S) community has developed the M&S Master Plan, which outlines the vision, goals, and objectives for M&S in the DoD, including LVC standards. The LVC standards also emphasize the importance of data interoperability, security, and accreditation to ensure that LVC environments can be used effectively and securely across different domains and classifications and has most recently been updated to include consideration of AI as stated in DoD Manual 5000.101,

> "Live, virtual, constructive (LVC) technology to support and augment human-machine teaming, interoperability, and live testing of AI-enabled and autonomous DoD systems; enable replication of specific test scenarios; and reduce the risk of bridging the AI model testing with system integration" (Office of the Director of Operational Test and Evaluation, 2024).

The LVC construct, which categorizes the way humans interact with simulations, first originated in 1991. Inspired by the Army's challenge to increase the effectiveness of training methods to meet the challenges posed by full spectrum warfare, GEN Paul F. Gorman presented a paper to the Society for Computer Simulation that argued "…most military training could be advantaged by Tactical Engagement Simulation (TES) in any or all of its three forms, Constructive, Virtual and Subsistent, and that, ideally, all three forms would be used interactively" (Gorman P. , 1991). As documented in his Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) Fellow's paper, the terms in the 1991 paper morphed over time: "subsistent" became "live" and Seamless TES became "blended training" (Gorman P. F., 2011). Later, the Defense Science Board (DSB) Task Force on Simulation Readiness & Prototyping, on which GEN Gorman was a member, solidified the concept of LVC, by asserting "Everything is simulation except combat" and classified the whole of Modeling, Simulation and Gaming as Live, Constructive, and Virtual.

A commonly used figure capturing the LVC taxonomy emerged in 2013, as shown in Figure 2 (Interservice/Industry Training, 2013). "Live" refers to Modeling & Simulation (M&S) involving real people operating real systems (e.g., a pilot flying a jet) for a simulated mission. "Virtual" refers to real people operating simulated systems (e.g., a pilot in a flight simulator). "Constructive" refers to simulations that involve simulated people operating simulated systems (e.g., a simulated pilot flying a simulated jet). According to the matrix, there is no name for simulated people operating real equipment.

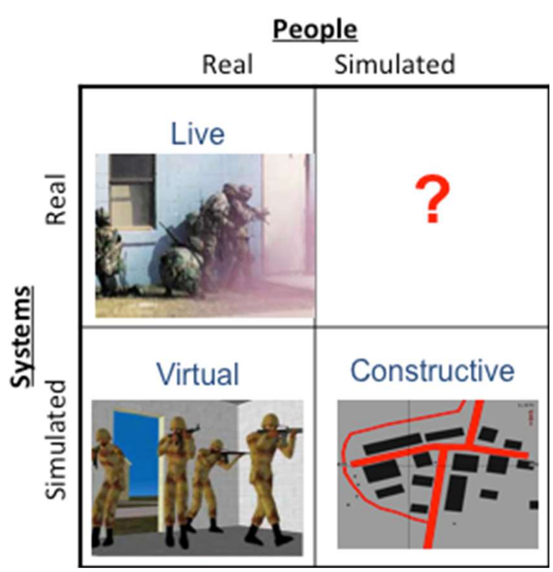

Figure 2. Categorizing simulations by the way humans interact with them [9]

LVC and AI

When the LVC taxonomy was created in 1991, there was no agreed name for the upper right quadrant that suggests simulated people operating real equipment, as there were no examples of this type of interaction. However, technology has advanced to the point where simulated humans in the form of modern AI are operating real systems. For example, the DARPA AlphaDogfight Trial was a competition where AI controlled a simulated F-16 fighter jet in aerial combat against an Air Force pilot flying in a virtual reality simulator (Goecks, et al., 2023). The AI-based system defeated an experienced pilot in a simulated dogfight, demonstrating that an AI-based pilot may surpass human abilities. Although this quadrant remains officially unnamed, it is sometimes called *Autonomy*.

THE CASE FOR EVOLVING THE LVC CONSTRUCT

The military community is traditionally and understandably risk-averse; we must be intelligent about how to design, operate, and train for the new hybrid of human-AI teams. In theater, teams will consist of humans and AI systems with heterogeneous capabilities and specialization. Teams will include human-human, human-AI, and AI-AI dynamics with each being bidirectional. Operations require a capability-oriented form of teaming where hybrid teams make decisions and act on them to produce a synergistic if not unified capability with real-world consequences. All levels of sensing, communicating and coordinating, computing, deciding and executing will need to be explored, and all may take place in the face of uncertainty and severe time constraints as well as being distributed to varying degrees depending on the operation. This introduces a complex and intractable problem space, one for which we believe LVC may offer the best framework for meaningful and comparative study leading to better, intentional design, and well-founded training.

A Proposal for Extending the LVC Paradigm

To incorporate the new dimension of AI into the live, virtual, constructive (LVC) paradigm, Loper and Sitterle (Loper & Sitterle, 2023) propose an expanded model that goes beyond the Autonomy quadrant in Figure 2, leading to the revised framework depicted in Figure 3. Specifically, Loper and Sitterle propose treating AI-enabled systems as a distinct "Actor" within the simulation framework. This categorization separates AI systems from real people and traditional M&S components, which include simulated platform interfaces (used to create a sense of realism in virtual environments) and fully abstracted representations of dynamics (constructive).

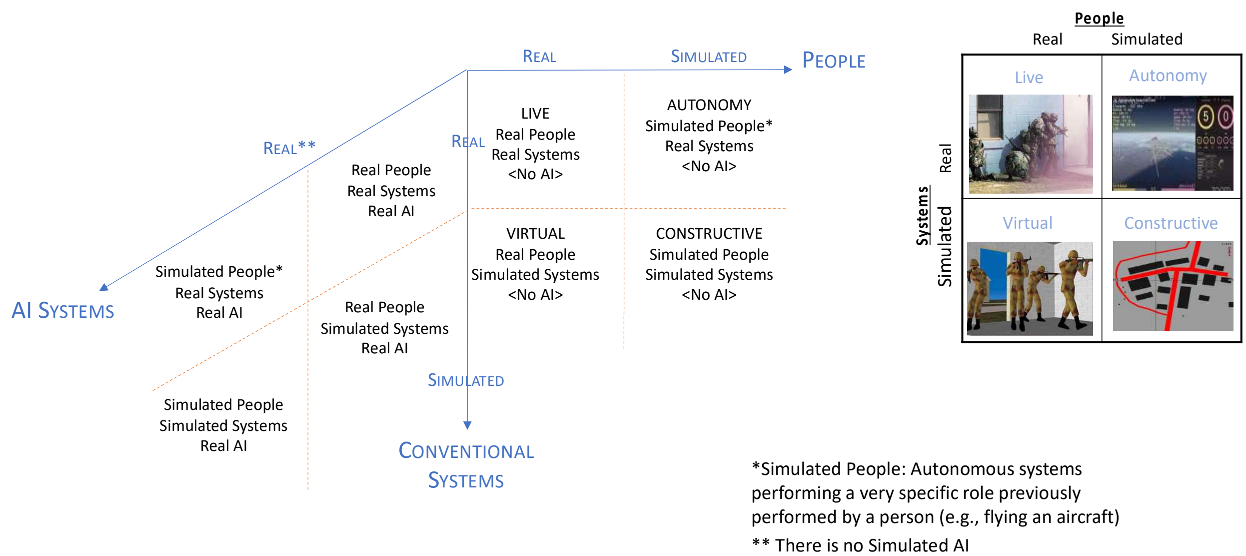

Figure 3. The revised LVC model of Loper and Sitterle [11]

The revised LVC construct aims to define effective human-AI teaming. As shown in Figure 4, we use binary notation to indicate if an LVC entity is real, simulated, or not present, excluding "simulated" AI because lower-fidelity AI can't accurately represent true performance.

	No Adaptive/ Learning-Enabled AI		Real Adaptive/ Learning-Enabled AI	
	Real People	Simulated People	Real People	Simulated People
Real Systems	Live_C (1, 1, 10)	Autonomy_C (0, 1, 10) (*Current blank square*)	Live_E (1, 1, 1)	Autonomy_E (0, 1, 1)
Simulated Systems	Virtual_C (1, 0, 10)	Constructive_C (0, 0, 10)	Virtual_E (1, 0, 1)	Constructive_E (0, 0, 1)
	Conventional LVC		*Extended LVC*	

State:
X (P, S, A)
P – People
S – System (Conventional)
A – Adaptive/ Learning-Enabled AI

Binary representation:
0 – Simulated
1 – Real
10 – Not present

Subscripts:
C – Conventional
E - Extended

Figure 4. The extended LVC model for human-AI teaming [11]

Since AI is software based, using the actual AI software is more practical than modeling a lower-fidelity version. Unlike simulated data used for training, fully implemented AI systems can be trained on such data before integration into the LVC model. The extended LVC framework includes adaptive, learning-enabled AI, while narrow AI and non-learning systems fit within the traditional LVC model. Reflecting AI's presence or absence, the model avoids distinctions between "real" and "not real" AI since such distinctions are unsupported by established literature.

THE FUTURE OF HUMAN SYSTEMS INTEGRATION IN HUMAN-AI TEAMING

Human Systems Integration (HSI) is defined by the Department of Defense (DoD) in DoD Instruction 5000.95 as "the System Engineering process and program management effort that provides integrated and comprehensive analysis, design, and assessment of requirements, concepts, and resources for human factors engineering, manpower, personnel, training, safety and occupational health, force protection and survivability, and habitability (the 'domains' of HSI)" (Office of the Under Secretary of Defense for Research and Engineering, 2022). With the release of DoDI 5000.95 on 1 April 2022, the Office of the Under Secretary of Defense for Research and Engineering (OUSD(R&E)) emphasized the importance of integrating human considerations across system design to optimize performance and minimize life-cycle costs (Office of the Under Secretary of Defense for Research and Engineering, 2022). Similarly, the International Council on Systems Engineering (INCOSE) defines HSI as "an approach that integrates technology, organizations, and people effectively" (International Council on Systems Engineering, 2023).

While HSI has historically focused on systems engineering and traditional human factors, the rise of AI-enabled systems and their integration within LVC environments demands an expanded perspective. Increasing system complexity, coupled with the shift toward AI-driven capabilities, presents new challenges and opportunities for HSI practitioners (Reqi Systems Engineering Articles, 2024). As modern systems operate within intricate networks involving humans, machines, and AI agents, the boundaries between human and machine functions are increasingly blurred. In LVC operations, this interconnectedness is exemplified by the need for seamless system-of-systems integration, where humans and AI collaborate dynamically across time zones and locations to perform complex tasks.

HSI and AI-Enabled Systems in LVC Environments

The DoD HSI Guidebook (Office of the Under Secretary of Defense for Research and Engineering, 2022), while not a formal standard, provides best practices for addressing HSI challenges in system design. However, AI-enabled systems introduce unique considerations, particularly in the context of human-AI teaming. The guidebook implicitly acknowledges the importance of adapting HSI practices to address collaboration between humans and AI, emphasizing the need to mitigate risks such as *cognitive overload* and safety hazards in operations involving complex, time-sensitive decision-making. For example, system capability documents should clearly define "human-in-the-loop" requirements, specifying scenarios for "human in control," "manual override," or "fully autonomous operations". LVC environments provide an essential testing ground for exploring these dynamics.

By simulating real-world scenarios, LVC platforms enable system designers to evaluate how human-AI teams interact, identify potential failure points, and refine system requirements. Incorporating HSI principles into these environments ensures that AI systems are designed to complement human capabilities, fostering trust, adaptability, and shared situational awareness. This is particularly important as AI agents become increasingly autonomous, requiring clear interaction protocols and transparent decision-making processes to build confidence among human operators.

To address the evolving landscape of AI-enabled systems, the HSI framework must expand to include consistent representations of systems where humans and AI are treated as collaborating agents (Boy, 2024). Emerging research suggests that multi-agent system models developed in AI are closely aligned with the system-of-systems approaches used in systems engineering. In this context, an "agent" can be viewed as an entity capable of perceiving situations, deciding, and executing actions—whether human or AI-enabled. This cross-disciplinary alignment offers opportunities for "cross-fertilization" between AI and systems engineering, enabling the development of robust multi-agent representations tailored to human-AI teaming.

The increasing interconnectivity and complexity of these systems demand that HSI practitioners account for not only human factors but also the interplay between AI agents and human users. Agent-oriented system modeling provides a valuable tool for understanding these relationships, allowing for the development of adaptive, resilient systems that can dynamically allocate tasks based on real-time conditions.

HSI Standards and Best Practices for LVC and AI Integration

Standards and best practices are critical for ensuring the successful integration of HSI principles into AI-enabled LVC environments. SAE Standard SAE6906A, "Standard Practice for Human Systems Integration," offers a valuable starting point for addressing HSI challenges in AI-enabled systems (G-45 Human Systems Integration, 2023) . This standard outlines HSI processes throughout the system life cycle, from design and development to decommissioning, and emphasizes the importance of addressing HSI considerations proactively in procurement and acquisition programs.

While originally developed for the automotive industry, SAE6906A provides a framework that can be adapted to military applications, particularly in the context of LVC operations. By incorporating lessons learned from the automotive sector's advancements in AI-enabled human-machine interfaces, the DoD can develop tailored standards that address the unique demands of military systems. These standards should account for the dynamic nature of human-AI teaming, emphasizing scalability, interoperability, and human-centered design.

AN EXAMPLE USE CASE FOR HUMAN-AI TEAMING IN LVC

To showcase the value of LVC training with both human and AI participants, consider the following mission rehearsal. This LVC event evaluates the integration and performance of a Human-AI team in a dynamic scenario, testing Intelligence, Surveillance, and Reconnaissance (ISR) and Search and Rescue (SAR) operations. The focus is on assessing Human-AI collaboration, particularly in handling uncertainty and adapting to changing missions. Human analysts and intelligent UAVs work as a team, exchanging information and opinions to enhance situational awareness and operational efficiency, thereby improving their response to mission-critical tasks in a realistic training environment.

Use Case Narrative

The mission begins with a directive to locate a high-value target. A swarm of intelligent AI-enabled UAVs autonomously deploy to gather intelligence, processing data at the edge. This processed intelligence is shared with human analysts, who request further details from the UAVs. The UAVs adapt to evolving mission objectives, responding to analyst requests and seeking human interpretation of collected intelligence (e.g., images, audio). Analysts, working individually and collaboratively, use the information to generate and evaluate hypotheses about the target's location, directing UAVs to address information gaps.

When an unexpected helicopter crash occurs, the UAV swarm splits: one sub-swarm continues ISR activities, and the other focuses on search and rescue at the crash site. Human analysts also split, with one group focusing on the high-value target and the other on the crash site. In the search and rescue effort, analysts interpret crash site intelligence, identify critical gaps, and task the UAV sub-swarm. UAVs provide processed data and request human interpretation. This iterative exchange highlights the dynamic human-AI teaming, balancing competing mission priorities for an effective response.

The team includes expert human analysts and UAVs. Human analysts specialize in ISR missions, hypothesis generation, and identifying critical information for accurate evaluations. UAVs, equipped with advanced capabilities like route planning, image capture, pattern recognition, and on-edge processing, use AI-driven analysis to recognize features such as caves, debris, and signs of human life, prioritizing anomalies for human review.

The scenario unfolds as follows:

1) **Mission Initiation**: The team receives a commander's intent to locate a terrorist leader.
2) **Analysis Setup**: Human intelligence analysts generate hypotheses about the target's location in the simulated environment.
3) **Swarm Deployment and Intelligence Gathering**: UAVs are deployed, using search and navigation algorithms to capture and analyze data (images, audio, signals) on-board and transmit relevant information.
4) **Pattern Recognition**: A UAV detects possible clothing at a cave entrance and requests verification from an analyst. If confirmed, the UAV continues searching the area; if not, the swarm moves to the next location.

5) **Hypothesis Evaluation**: Analysts generate and evaluate hypotheses, guiding UAVs based on received intelligence. They look for evidence to support or disconfirm hypotheses and direct UAVs accordingly.
6) **Human-AI Coordination**: An analyst confirms whether detected items are relevant (e.g., clothing), influencing the UAV's search pattern.
7) **Focused Searches**: Analysts request specific flyovers to explore areas of interest, such as behind rocks.
8) **Reasoning and Retasking**: Analysts use reasoning about missing information ("Value of Information") to request UAVs to gather specific data to refine hypotheses.
9) **Unexpected Event**: A helicopter crash is simulated, and the information is relayed to the UAVs.
10) **Swarm Adaptation and Analyst Split:**
    a. ISR Sub-swarm and Sub Analyst group: Continues the search for terrorist indicators and signs of human habitation in teaming with the ISR Sub Analysts
    b. Search and Rescue Sub-swarm and Sub Analyst group: Focus on the crash site, using updated pattern recognition to identify wreckage and survivors, working collaboratively.

The Importance of the Human-AI Teaming Use Case

A use case such as the LVC scenario just described provides valuable insights into the performance and coordination of human-AI teams in complex operational environments. It evaluates team performance, identifies skill gaps for both human and AI participants, and informs the development of protocols for communication, data sharing, and decision-making. In a training and mission rehearsal this could include:

1) Team Performance Assessment: Evaluation of human-AI team performance in achieving mission objectives.
2) Skill Gap Analysis: Identification of critical skills required by both human operators and AI systems for optimal team effectiveness.
3) Protocol Development: Development of standardized protocols for communication, data sharing, and decision-making within the human-AI team.

Furthermore, such a use case can be utilized for *Parameter and Configuration Testing* in order to evaluate various team configurations and operational parameters, and to conduct *What-if Analysis* to assess the impact of different decisions, resource allocations, and skill sets on mission outcomes.

Beyond training applications, the scenario facilitates critical testing and analysis, including evaluating different team configurations, operational parameters, and decision-making strategies. Such insights would prove instrumental in establishing standards for LVC simulations to effectively support human-AI teams, in this case, ensuring they meet the demands of modern ISR and search and rescue missions; a step toward advancing human-AI collaboration in mission-critical environments.

The previous is just an example of a use case that would emphasize human-AI teaming and would be addressable by LVC. Many others could be considered as well within the LVC framework such as the examples of AI defense applications discussed by Aaron Frank (Frank, 2022) which include:

1) Attention Management (Live)
   AI monitoring of information flows and anomaly detection.
2) Information Exploitation and Model Validation (Virtual)
   AI optimizing military operations based on specific problems
3) Exploratory Analysis (Constructive)
   AI providing multiple explanations for intelligence, enhancing understanding of adversarial behavior.
4) Autonomy and Principal–Agent Relations (Autonomy)
   Autonomous systems supporting logistics and C4ISR in battlespace operations.

THE ROLE OF MODELING AND SIMULATION ORGANIZATIONS IN OVERCOMING LIMITATIONS

AI-enabled teams have the potential to significantly enhance the adaptive capacity of military forces by combining human judgment and expertise with AI's ability to process vast amounts of data and uncover patterns beyond human capability. Such teams can focus decision-makers on critical challenges, conduct robust exploration to identify novel solutions, and optimize decision-making at strategic, operational, and tactical levels. By integrating human and AI strengths, such teams have the potential to fundamentally transform the landscape of military operations, achieving together what neither AI nor humans can achieve alone. Modeling and Simulation (M&S) organizations are key to this transformation, offering a safe, scalable, and flexible platform for testing and refining human-AI collaboration within Live-Virtual-Constructive (LVC) environments.

To unlock the full potential of human-AI teams, LVC experiments must realistically mirror the complexities of real-world scenarios. AI has already shown how it can improve the design, execution, and assessment of software and systems; this should be applicable to LVC applications as well by helping to develop better models and enabling more realistic analyses of team performance under various conditions. But more importantly, LVC environments provide the means to explore crucial aspects of human-AI teaming, like trust, communication, shared decision-making, and adaptability to changing missions. The challenge is to effectively integrate human-AI teams in AI-enhanced environments, but doing so will require overcoming several key challenges and limitations of current LVC frameworks in the areas of implementation, validation, and operations.

Implementation Challenges:

- **Interoperability:** Integrating diverse systems linking live operators, virtual simulators, and software-driven constructive elements demands smooth interaction across different data models and communication protocols (e.g., HLA, DIS), requiring complex and costly translators.
- **Cybersecurity:** The interconnected LVC environment presents a large attack surface, where breaches can compromise training integrity and have real-world impacts.
- **Hardware and Software:** Limited resources can restrict the scale and fidelity of LVC environments, reducing the number of entities and complexity of behaviors.
- **Data Management:** The large volume of data generated during LVC events poses a significant challenge for after-action reviews.

Validation Risks:

- **Verification, Validation, and Accreditation (VV&A):** Rigorous VV&A is essential for trustworthy LVC platforms, but this process is challenging. Inadequate validation can lead to negative training outcomes.
- **Consistent Fidelity:** Ensuring consistent quality across live, virtual, and constructive components is crucial for maintaining user immersion and trust.
- **Cognitive Load:** Participants may struggle with distinguishing between live, virtual, and constructive entities, leading to confusion and reduced situational awareness.

Operational Barriers:

- **Cost:** LVC systems are expensive and require skilled personnel for maintenance and operation, representing a long-term investment.
- **Scalability:** Performance can decline with increasing numbers of participants and entities.
- **Scenario Creation:** Designing realistic, adaptive, and relevant training scenarios is complex and resource intensive.
- **Cultural Barriers:** Resistance to trusting simulated outcomes over traditional live-only exercises can impede LVC adoption and integration into training programs.

Future Directions and Recommendations

To address these challenges, Modeling and Simulation leadership organizations can play a strong role in providing the necessary governance, technical guidelines, and interoperability frameworks to support the design, development, integration, and execution of synthetic training environments. To advance LVC environments in support of Human-

AI Teaming, we propose that stakeholders in the M&S community undertake several key initiatives. These efforts would address emerging challenges associated with integrating adaptive, learning-enabled AI into LVC environments and optimizing human-AI collaboration for military and other complex operational scenarios. To this end, we recommend the following roles for the Modeling and Simulation community:

1. **Governance:** M&S organizations should lead the development of clear governance structures by defining cybersecurity protocols and developing ethical guidelines for human-AI collaboration. These structures should be in place to ensure compliance and ethical behavior in LVC scenarios, promoting secure and ethical AI integration.
2. **Technical Guidelines:** M&S organizations should seek to create and disseminate technical guidelines for AI-specific interoperability and human-AI teaming representation. These guidelines, including best practices for adaptive AI and standards for integrating advanced AI capabilities, would ensure seamless collaboration and effective representation within LVC environments.
3. **Interoperability Frameworks:** To promote interoperability, M&S organizations must define frameworks that include protocols for AI interaction and standards for integrating diverse AI models. Extending the LVC paradigm for AI roles, these frameworks support consistent interaction and collaboration among varied simulation environments.
4. **Performance Evaluation:** Developing enhanced evaluation metrics is crucial for assessing Human-AI team performance. M&S organizations should collaborate to establish metrics for trust, situational awareness, and mission effectiveness, providing benchmarks to evaluate AI reliability, predictability, and team resilience under dynamic conditions.
5. **Collaboration and Training:** Facilitating collaboration across stakeholders, M&S organizations can lead by organizing working groups and aligning with international standards bodies like NATO and IEEE. By creating shared repositories and promoting best practices, such efforts can enhance multi-domain and international collaboration for human-AI teaming in LVC environments.

The M&S community is uniquely positioned to lead the advancement of human-AI teaming through its traditional perspective that fuses mathematical principles and human insights. This perspective has historically created powerful models and simulation techniques that have transformed various fields while managing associated risks. As AI continues to evolve, it offers new modes of collaboration that enhance human capabilities within a burgeoning human-AI ecosystem. The flexibility of adaptive, learning-enabled AI systems introduces emergent properties, making it crucial to design and operate these systems safely and effectively, particularly in collaborative roles with warfighters. The LVC paradigm provides a robust platform for refining human-AI teaming by allowing researchers and practitioners to explore, test, and optimize human-AI interactions in realistic, controlled environments.

By embracing the five roles identified in this paper—governance, technical guidelines, interoperability frameworks, performance evaluation, and collaboration and training—the modeling and simulation community can lead LVC-related initiatives to support human-AI teaming. These efforts will not only address immediate challenges but also enhance current LVC methods, solidifying the community's role as a principal enabler for future military and civilian applications of human-AI collaboration.

REFERENCES

Allen, G. W., Lutz, R., & Richbourg, R. (2010). Live, virtual, constructive, architecture roadmap implementation and net-centric environment implications. *ITEA, 31*(3), 355-364.

Boy, G. A. (2024). Human Systems Integration of Human-AI Teaming. *2024 IEEE 4th International Conference on Human-Machine Systems (ICHMS).* Toronto.

Frank, A. B. (2022). Gaming AI without AI. *The Journal of Defense Modeling and Simulation*.

G-45 Human Systems Integration. (2023). *Standard Practice for Human Systems Integration SAE6906A.* SAE International.

Goecks, V. G., Waytowich, N., Asher, D. E., Park, S. J., Mittrick, M., Richardson, J., . . . Narayanan, P. (2023). On Games and Simulators as a Platform for Development of Artificial Intelligence for Command and Control. *The Journal of Defense Modeling and Simulation, 20*(4), 495-508.

Gorman, P. (1991). The Future of Tactical Engagement Simulation. *The Proceedings of the 1991 Summer Computer Simulation Conference.* San Diego.

Gorman, P. F. (2011). Learning to Learn: Reminiscences and Anticipation. *Proceedings, Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).* Orlando.

International Council on Systems Engineering. (2023). *Systems Engineering Handbook, INCOSE-TP-2003-002-05.* Hoboken: John Wiley & Sons, Ltd.

Interservice/Industry Training, S. a. (2013). Fundamentals of Modeling and Simulation Tutorial. *Interservice/Industry Training, Simulation and Education Conference (I/ITSEC).* Orlando.

Loper, M., & Sitterle, V. (2023). Evolving LVC To Include Evaluation Of Human-Ai Teaming Dynamics. *Winter Simulation Conference.* IEEE.

National Academies of Sciences, Engineering, and Medicine. (2021). *Human-AI Teaming: State-of-the-Art and Research Needs.* Washington DC: The National Academies Press.

Office of the Director of Operational Test and Evaluation. (2024). *DoD Manual 5000.101 Operational Test and Evaluation and Live Fire Test and Evaluation of Artificial Intelligence-Enabled and Autonomous Systems.* US Department of Defense.

Office of the Under Secretary of Defense for Research and Engineering. (2022). *DoD Instruction 5000.95 Human Systems Integration in Defense Acquisition.* Washington DC: Department of Defense.

Office of the Under Secretary of Defense for Research and Engineering. (2022). *Human Systems Integration Guidebook.* Washington DC: Department of Defense.

O'Neill, T. A., Flathmann, C., McNeese, N. J., & Salas, E. (2023). Human-autonomy Teaming: Need for a guiding team-based framework? *Computers in Human Behavior*.

Reqi Systems Engineering Articles. (2024, November). *Human Systems Integration: A Comprehensive Guide and Future Trends*. (Reqi) Retrieved from https://reqi.io/articles/human-systems-integration-future-trends

Reyes, D., Dinh, J., & Salas, E. (2019). What Makes a Good Team Leader? *ournal of Character and Leadership Development, 6*(1).

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors, 50*(3), 540-547.