



**SYNTHETIC TASK ENVIRONMENT
FOR HUMAN-AI TRAINING**

EXPLORING THE TRADESPACE

James E. McCarthy, Lillian K.E. Asiala, LeeAnn Maryeski, & Nyla Warren

Sonalysts, Inc.



BLUF

- The research team conducted a survey of researchers and military subject matter experts to identify important features of a synthetic task environment (STE) designed to support research focused on Human-AI teaming, especially the training of human members of those teams.
- Thematic analyses of survey responses allowed the team to identify:
 - Important STE features
 - Desirable attributes of the STE architecture
 - Desirable attributes of the task domain
 - Important aspects of scenario authoring
 - Important data collection and performance assessment capabilities
 - Important data analysis and visualization capabilities
 - Important communication capabilities
 - Important agent capabilities
- The team will explore these capabilities in more detail as our project continues.



Agenda

- Background
- Methods
- Results
- Next Steps





Project Background



Need

- Significant Cross-service interest in hybrid teaming
 - Human-AI Teaming
 - Manned-Unmanned Teaming (MUM-T)
 - Human-Autonomy Teaming (HAT)
- Various developments are likely to stimulate this interest
 - Increasing capabilities of AI-powered agents
 - The need to further compress decision cycles
 - The proliferation of data streams and courses of action (COAs) made possible by emergent concepts such as Joint All-Domain Command and Control (JADC2)



National Academies of Sciences, Engineering, and Medicine (NASEM) Consensus Report

- In 2021, the NASEM produced a consensus report that defined the primary research objectives for Human-AI teaming
- The objectives were distributed over nine focus areas
 - Training Human-AI Teams
 - AI Transparency and Explainability
 - Trusting AI Teammates
 - Human-AI Team Interaction
 - Human-AI Teaming Processes and Effectiveness
 - Human-AI Teaming Methods and Models
 - Situation Awareness in Human-AI Teams
 - Identification and Mitigation of Bias in Human-AI Teams
 - HSI Processes and Measures of Human-AI Team Collaboration and Performance



National Academies of Sciences, Engineering, and Medicine (NASEM) Consensus Report

- In 2021, the NASEM produced a consensus report that defined the primary research objectives for Human-AI teaming
- The objectives were distributed over nine focus areas
 - **Training Human-AI Teams**
 - AI Transparency and Explainability
 - Trusting AI Teammates
 - Human-AI Team Interaction
 - Human-AI Teaming Processes and Effectiveness
 - Human-AI Teaming Methods and Models
 - Situation Awareness in Human-AI Teams
 - Identification and Mitigation of Bias in Human-AI Teams
 - HSI Processes and Measures of Human-AI Team Collaboration and Performance



Step 1: Develop Research Plan

Tasks	Outcomes	Next Steps
<ul style="list-style-type: none">• Review literature pertaining to:<ul style="list-style-type: none">▪ Training of human-AI teams▪ Adjacent research areas<ul style="list-style-type: none">○ Competencies required for all-human teams○ Methods to develop those competencies○ Selected topics in agent performance (e.g., trust and trust repair)	<ul style="list-style-type: none">• Summary of foundational research• Research premises• Research hypotheses• Preliminary STE definition	<ul style="list-style-type: none">• Refine STE definition by surveying the research community, SMEs, and various stakeholders (e.g., Government PIs) to identify the attributes that would make the STE valuable/accepted.



Methods



Development Method

- We decided to separately survey researchers (academics and Government PIs) and military SMEs
- For each group, we used the research plan to identify elicitation objectives for the corresponding survey
- Define survey probes for each elicitation objective
 - Likert Items – Assessment of features/capabilities identified in the research plan
 - Open-ended – Queried respondents for features/capabilities that were not included in the plan
- Conducted internal QA
- Authored in Survey Monkey
- Piloted with one member of each community
- Refined the surveys based on the results of the pilot study



Delivery Method

- Each respondent was assigned a unique access link
 - Enable progress tracking
 - Enabled the use of different computers/browsers to complete the survey
 - Allowed respondents to review/refine answers to previous questions
- The links were provided via emails generated by Survey Monkey
- Reminder emails were distributed, as necessary
 - Survey Monkey
 - Personal emails



Analysis Method

- Open-Ended Items
 - The two lead researchers conducted independent thematic analyses
 - The lead researchers then compared their lists to develop a consensus set of themes
 - Mapped individual responses to the consensus list to indicate the level of overlap across the responses
- Likert Items
 - Calculated various descriptive statistics



Results



Sample

SMEs

- n=12
 - *All were Sonalysts' partners

Researchers

- Academic Institutions
 - n=3
- Government Labs
 - n=8



Important STE Features

- System Architecture
 - Open source
 - Modularity
 - Flexibility
 - Consider using an existing STE
- Teaming
 - Flexibility
 - Multiple roles/stations for each member type
 - 6-12 Teammates
- Task Domain
 - Military focus
 - Sufficient complexity and fidelity
 - Significant interdependency
- Data Collection and Analysis
 - Instrumentation
 - Analytics
 - Authorability
- Autonomy
 - Include actual and/or scripted autonomy
- Ease of Use
 - Intuitive displays
 - Game play is easy to learn
 - Scenarios easy to build/modify



Desirable Attributes of the STE Architecture

Cloud-based Approaches

- Better configuration control
- Better supports interaction across labs
- Faster implementation of updates

Local Network Approaches

- Allows labs to function independently
- May be less expensive
- May be more secure
- Works if connectivity is lost



Important Aspects of Scenario Authoring

- Valued Capabilities
 - Flexibility/Simplicity/Usability
 - Ability to simulate real-world interactions
 - Manage linkages among competencies, events, and performance assessments
- Current Challenges
 - Creating scenarios that are technically feasible and aligned with research question
- Simplicity Vs. Power
 - Simplicity
 - Power
 - Both



Important data analysis and visualization capabilities

- Value of Data Analysis
 - Data export is critical
 - Data analysis is nice to have, but may not be necessary
 - Provide robust analysis capabilities
- Value of Visualization
 - As above, data export was viewed as more important than visualization
 - Most favored “quick look” displays that showed single variables and relationships



Research Focus Themes

- Environment/Setting
 - Clear preference for military tasks
 - Authenticity was preferred over fictionalized settings
 - Support for both classified and unclassified scenarios
- Task
 - C4ISR
 - Offensive/defensive operations
 - Military or military analogs (e.g., SAR, bomb disposal)
- Participants
 - Mostly military personnel
 - Some mention of college students
- Measures
 - Team Measures
 - Outcomes
 - Processes
 - States
 - Individual Measures
 - Processes
 - Individual differences
 - Perceptions/opinions
- Other
 - Focus on use-case rather than general purpose solution
 - Importance of robust instrumentation



Important Data Collection and Performance Assessment Capabilities

SMEs

- Areas of emphasis
 - Measures of communication and coordination
 - Course of action assessment
 - Speed, accuracy, and effectiveness
 - Individual states & “co-variants”
- Key assessment dimensions
 - Speed & accuracy
- Noted that assessments tended to focus on outcomes and expert evaluation

Researchers

- Areas of emphasis
 - Measures of communication and coordination
 - Workload measurement
 - Assessment of attention/engagement
- Key assessment dimensions
 - Speed & accuracy
- Noted that authorable surveys would be a nice feature to include



Desirable Attributes of the Task Domain

SMEs

- More willing to entertain military analogs
 - Incident response
 - Capture the flag/paintball
- Points of emphasis
 - Matrix organization
 - Data fusion/filtering and decision-making
 - Communication/coordination
 - Ensure time pressure, complexity, and uncertainty
 - Include surprise, random occurrences, and/or hostile acts

Researchers

- More focused on specific military use cases
- Tasks
 - Mandate interdependency across roles
 - Require variable levels of teamwork



Important communication capabilities

SMEs

- Significant Modalities
 - Face-to-face,
 - Verbal/spoken (Radio, IP, SATCOM, *etc.*)
 - Written (email, chat, messages, *etc.*)
 - System data (COP, system parameters, *etc.*)

Researchers

- Support a wide variety of communication modalities
- Interest in assessing communication processes and content
 - Quality
 - Quantity
 - Pathways
- Significant interest in automatic transcription support
 - Especially when crosstalk is likely



Important agent capabilities

SMEs

- General Skepticism of Agents
- Factors Promoting Trust
 - Good mental model (how the system was developed, works, etc.)
 - Familiarity and positive track record
- Features Harming Trust
 - Poor Performance
 - Actions that seem contraindicated or erratic
 - Surprise
 - Led astray with faulty information
 - Biases
 - Actions taken without “human interface”
- Desired Features
 - Supporting human understanding
 - Acting as good teammates
 - Having good performance
 - Processing/reacting to dense data streams

Researchers

- Support a Range of Adjustable Agent Characteristics
- Act as a Teammate
 - Fill a recognized roll
 - Coordinate with team members
 - Take effective actions
- Act as a Tool
 - Helping humans make the right decisions
 - Take effective actions
 - Get the right information
- Built in vs. External



Next Steps



Continued Progress

