U.S. ARMY COMBAT CAPABILITIES DEVELOPMENT COMMAND ARMY RESEARCH LABORATORY

Assessing Human-Autonomy Teams: How to Verify?

05 OCT 2023

Ralph W. Brewer II

Computer Scientist

Army Research Directorate – Humans in Complex Systems

Approved for public release; distribution is unlimited.

BACKGROUND

- **Unmanned platforms**
  - These platforms will feature reduced crew-system ratios and will be controlled through a mix of manual teleoperation and autonomy.

- **Integrating humans and autonomous systems into teams is complex**
  - Humans and machines think, decide, and act differently
  - Machines operate at dramatically different time scales than humans
  - Humans and machines have different capabilities to adapt to mission demands

- **Assessing human-autonomy teams is challenging.**
  - Current methods focus on scripted evaluation scenarios that are expensive and periodic.

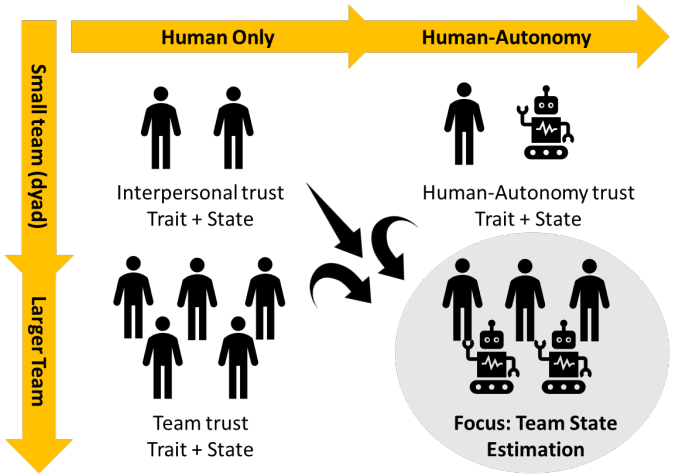
- **Need assessment methods that can accommodate these new team characteristics and dynamics**
  - Heterogeneous teams
  - Continuous assessment
  - Learning and adaptation

TEAM ASSESSMENT TOOL DEVELOPMENT

In response to the need for novel assessment methods, two software tools are being developed:

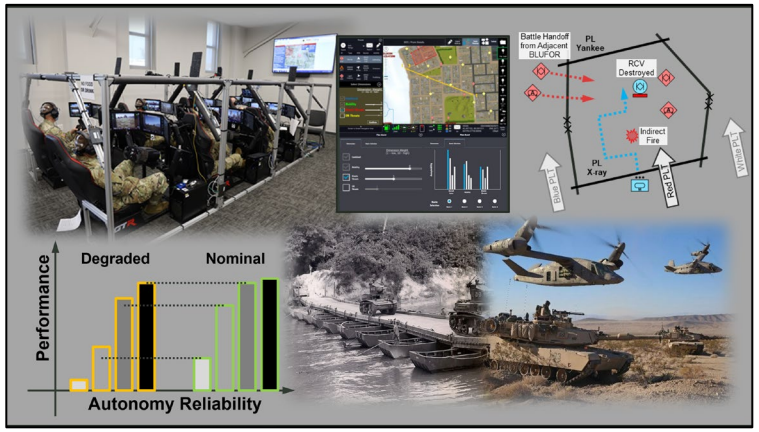
Human-Autonomy Teaming Trust Toolkit (HAT^3)

- Goal: provide a multimodal approach to assess the dynamic nature of human-autonomy team trust in complex environments through real-time state estimation.



Toolkit for Continuous Team Assessment (T-CAT)

- Goal: provide a capability for collaborative, continuous team assessment via a software platform that enables re-streaming and analysis of human-autonomy teams in real/near-real time in order to improve team effectiveness.



Approved for public release; distribution is unlimited.

HUMAN-AUTONOMY TEAMING TRUST TOOLKIT (HAT^3)

Description: Modular software tool for continuous, unobtrusive, real-time, multimodal assessment of team trust and appropriate trust

Rationale: Existing trust measures rely solely on self-report which can be obtrusive and are unable to capture changes in trust and cohesion over time. A multimodal approach is needed to understand what impacts trust over time and identify when breakdowns occur so interventions for proper calibration can be implemented.

Features: 5 modules

- **Subjective:** self-report questionnaires
- **Communication:** communication flow, network analysis, linguistic analysis
- **Physiological***:* heart rate, heart rate variability, pupillometry (in progress)
- **Behavioral:** eye-tracking, interface interactions (future)
- **Affective:** facial expression (future)

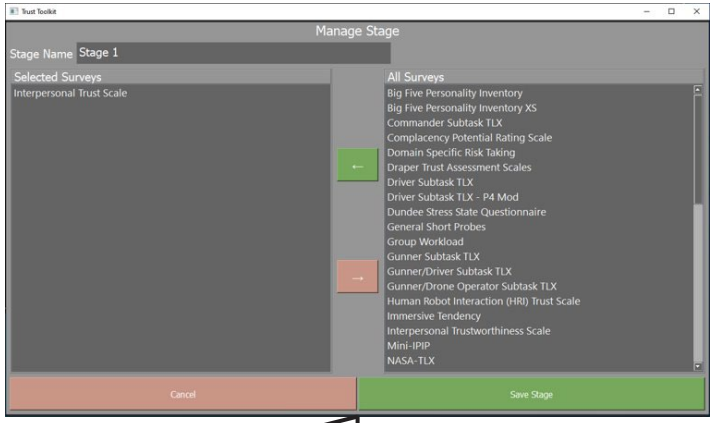
Targeted users: Researchers, analysts, vehicle commanders and commanding officers

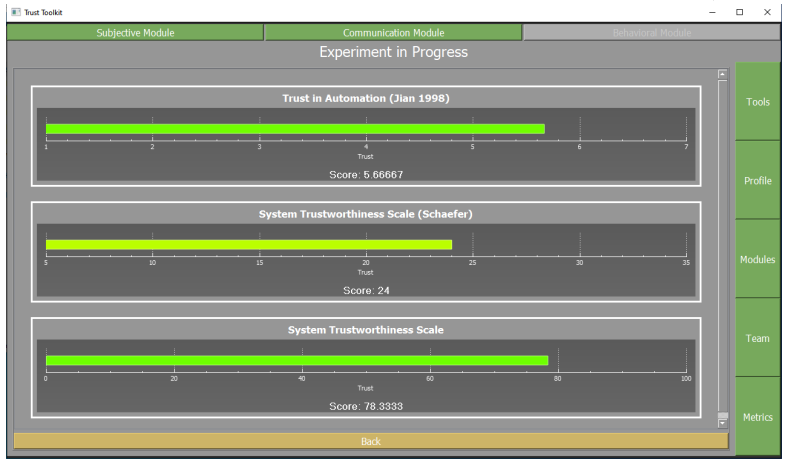
Usage: Laboratory and Field

TRUST TOOLKIT - SUBJECTIVE MODULE

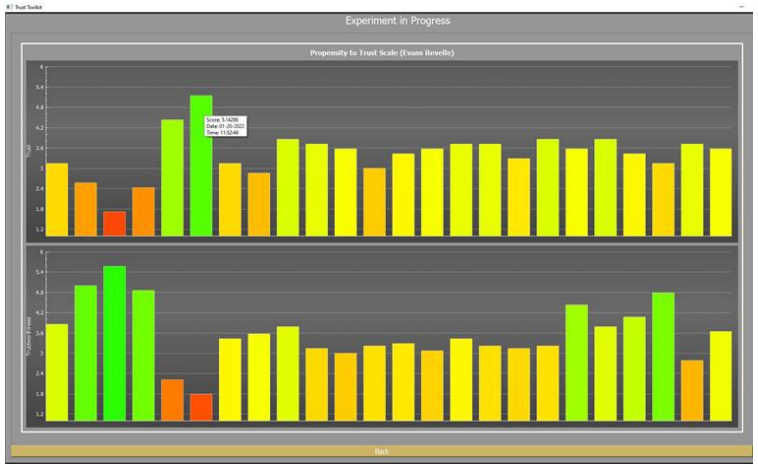
Description:

- 10 self-report measures of trust
  - Trait-based scales
    - Trust propensity
  - State-based scales
    - Interpersonal trust
    - Trust in autonomy
- 30 self-report measures related to trust
- Customizable





Colored bars indicate one crew member's scores on three trust scales: Trust in Automation (top), and two System Trustworthiness Scales (middle and bottom).



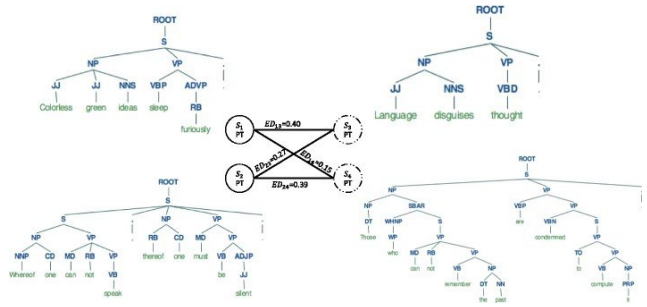
Historical view of one team members' scores over time. Colors indicate a continuous spectrum of subjective trust state which ranges from high (green ranges) to medium (yellow-orange ranges) to low (red ranges).

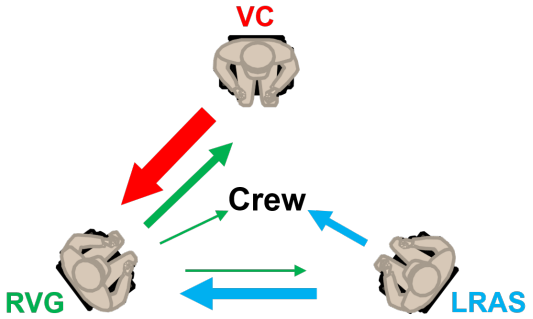
TRUST TOOLKIT – COMMUNICATION MODULE

Description:

- **Three methods for visualizing and analyzing team communication that enable an understanding of team dynamics and trust**
  - **Real-time Event, Flow, and Coordination Tool (REFLECT)**
    - Communication flow, rates
  - **Network Analysis: in depth analysis of communication behaviors**
    - Individual-level analysis
    - Team-level analysis
  - **Bag of Words: linguistic analysis of communication content**
    - Empath text analysis tool

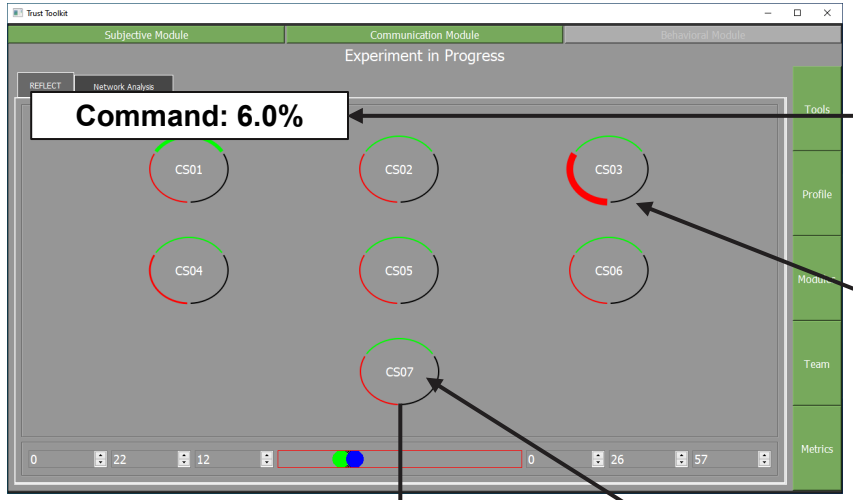
NOTE: Data shown on next slides were collected from a vehicle crew of 7 members during a simulation experiment. Each crew station is labeled CS01 – CS07. Each crew member performed a specific role: Commander, Gunner, Driver.

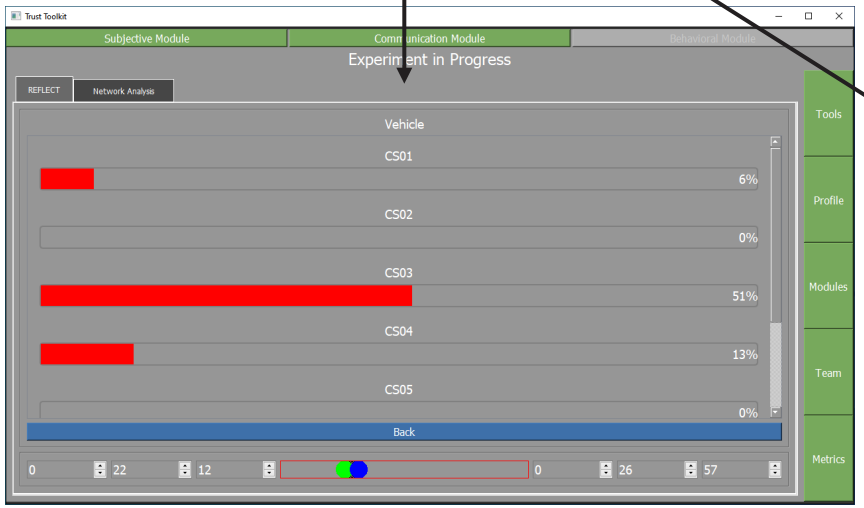




TRUST TOOLKIT – COMMUNICATION MODULE

REFLECT: communication flow, rates





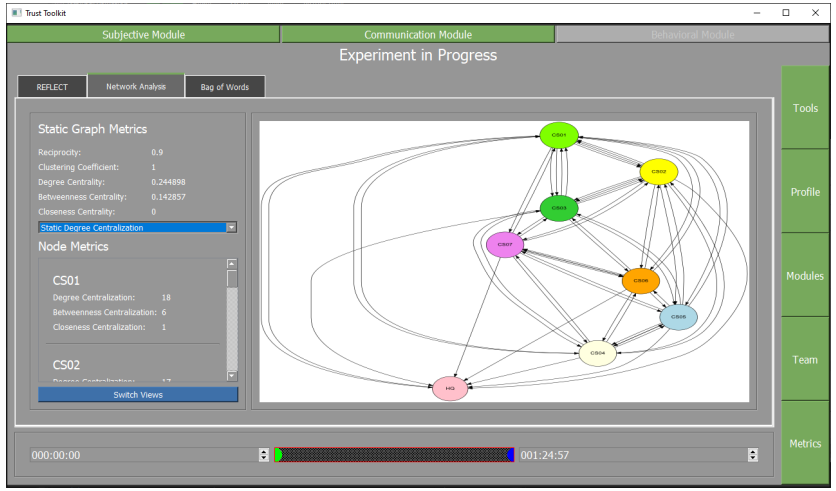
- Hover cursor over a crewstation to view percentage of time participant spent talking
- Line thickness indicates quantity of communication on the command network for crewstations 1 through 7 during a given timeframe
- Clicking on a crewstation allows user to view each participant's time spent speaking on that network.

TRUST TOOLKIT – COMMUNICATION MODULE

Network Analysis: in depth analysis of communication behaviors

- **Individual-level measures**
  - Degree centrality
  - Betweenness centrality
  - Closeness centrality
- **Network-level measures**
  - Reciprocity
  - Clustering coefficient
  - Degree centralization
  - Betweenness centralization
  - Closeness centralization

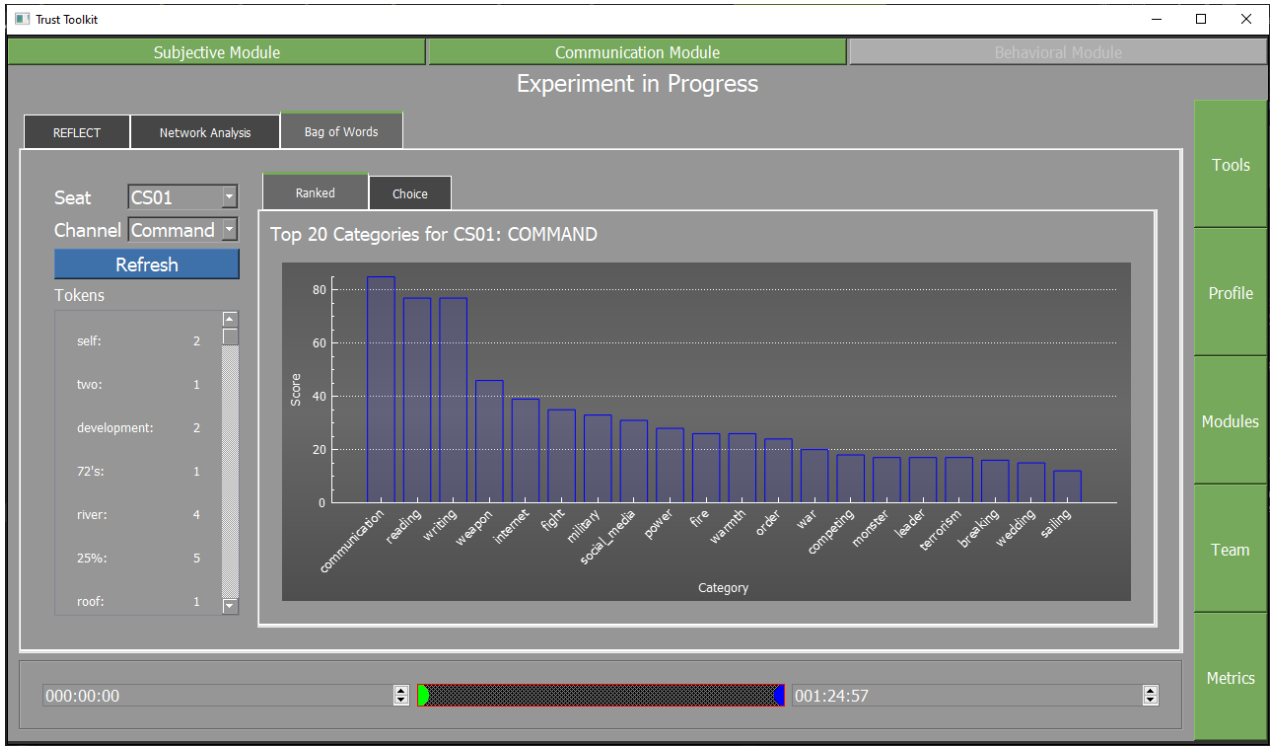
(U)



- Left panel displays network statistics
- Adjustable sliding window to show how the network is changing over time (e.g., What did the network look like over the last 15 min? Over the last hour?).
- Attributes of individual nodes, such as size or color, can be modified in real time to reflect centrality scores. For example, nodes most important for routing information through the network could be enlarged relative to nodes with lower centrality scores.

TRUST TOOLKIT – COMMUNICATION MODULE

- **Bag of Words: linguistic analysis of communication content**
  - Uses *Empath* text analysis tool (https://github.com/Ejhfast/empath-client)
  - Evaluates number or words that fit different categories (positive or negative sentiment, perception cognition, past tense) as potential indicators of trust.



Approved for public release; distribution is unlimited.

TRUST TOOLKIT PATH FORWARD

- Additional modules will be added:
  - **Physiological** – heart rate, heart rate variability (in-progress)
  - **Behavioral** – eye-tracking, interface interactions (future)
  - **Affective**- facial expression (future)

- Incorporate Human-Autonomy Team Cohesion scale

Table 1. Scale Reliability (Cronbach's alpha)

Subdimension	Item Scale	High Cohesion Scenario (α)	Low Cohesion Scenario (α)
Function	Function	0.93	0.92
	GCQ	0.91	0.85
Exclusivity	Exclusivity	0.90	0.93
	GCQ	0.90	0.91
Pride	Pride	0.97	0.97
	GCQ	0.93	0.91
Complementarity	Complementarity	0.92	0.93
	GCQ	0.91	0.91
Resilience	Resilience	0.95	0.96
	GCQ	0.90	0.87

Table 2. Correlations between the Subdimensions and their corresponding Group Cohesion Questionnaires (GCQs)

Subdimension	Scenario	t	df	p	r
Function	High Cohesion	16.26	122	<.001	0.84
	Low Cohesion	9.87	115	<.001	0.68
Exclusivity	High Cohesion	10.57	121	<.001	0.69
	Low Cohesion	12.12	119	<.001	0.74
Pride	High Cohesion	15.92	122	<.001	0.82
	Low Cohesion	16.77	123	<.001	0.83
Complementarity	High Cohesion	14.03	121	<.001	0.79
	Low Cohesion	10.06	119	<.001	0.68
Resilience	High Cohesion	13.19	108	<.001	0.79
	Low Cohesion	10.72	108	<.001	0.72

- Expand capability beyond trust and cohesion to include system performance (autonomy) and other types of human performance measures and latent states of interest (e.g., situation awareness, workload, dynamic resource allocation, stress etc.)

- Developers working on scripts for ingesting additional data formats

TOOLKIT FOR CONTINUOUS ASSESSMENT OF TEAMS (T-CAT)

Description: Toolkit that consists of mission re-streaming and visualization software that enables the user to view mission data and metrics in real-time and view and manipulate data and metrics post hoc. The Toolkit is an integral part to the After-Action Review (AAR) to increase the Soldiers understanding of the human-autonomy teaming dynamic.

Rationale: Current methods used for assessing human-autonomy teams focus on simple use-case interactions in highly structured environments. T-CAT will enable continuous assessment of Soldier crews, autonomy, and their interactions in realistic, unscripted, and dynamic scenarios.

Features:

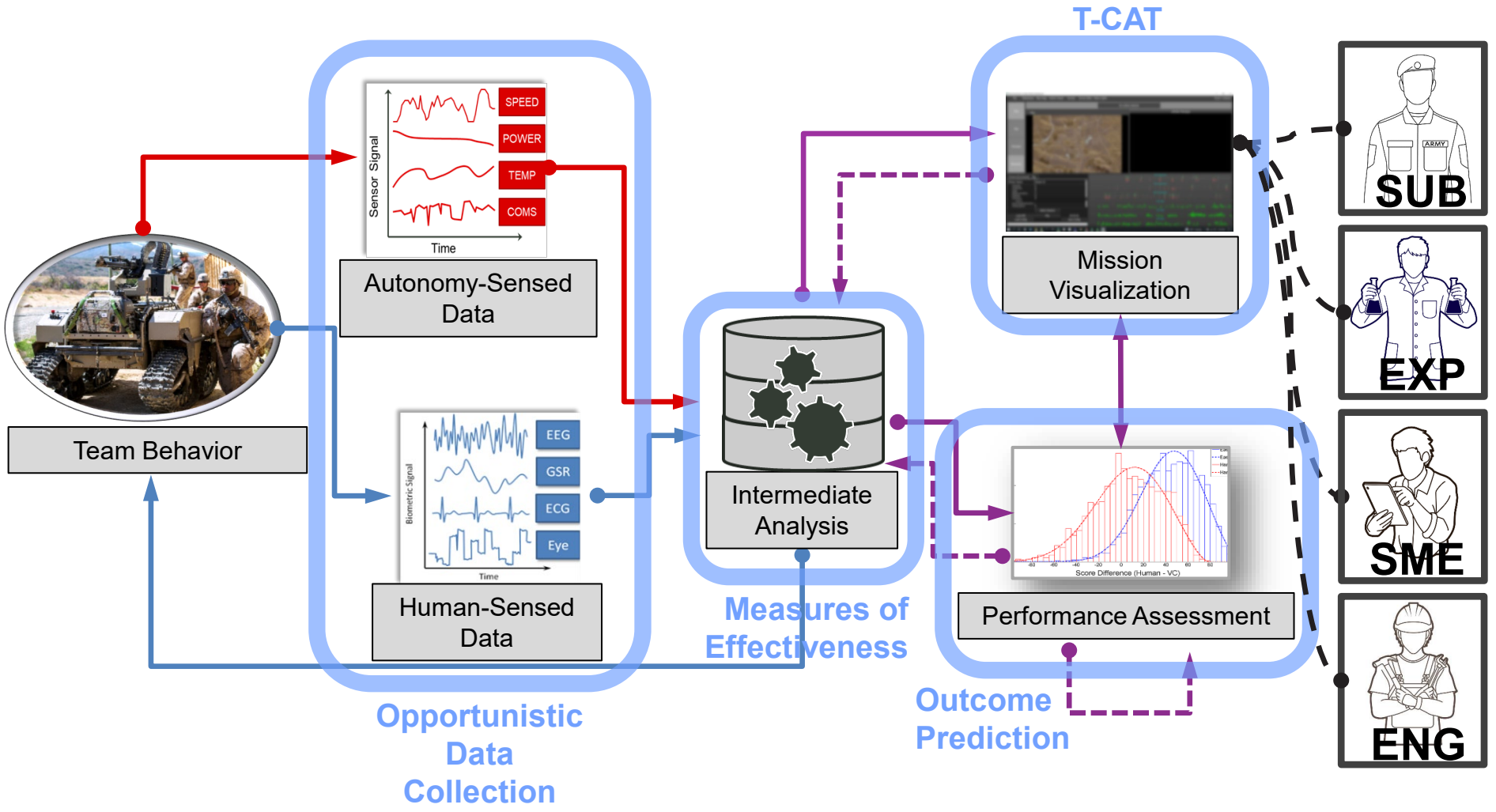
- Near-real time & post hoc data processing, synchronization, visualization, and analysis
- Near-real time & post hoc data annotation which enables users to log events as they occur to preserve critical data
- Embedded algorithms and predictive models estimate team's overall effectiveness

Targeted users: Scientists, Engineers, SMEs, and Study Participants

Usage: Laboratory and Field

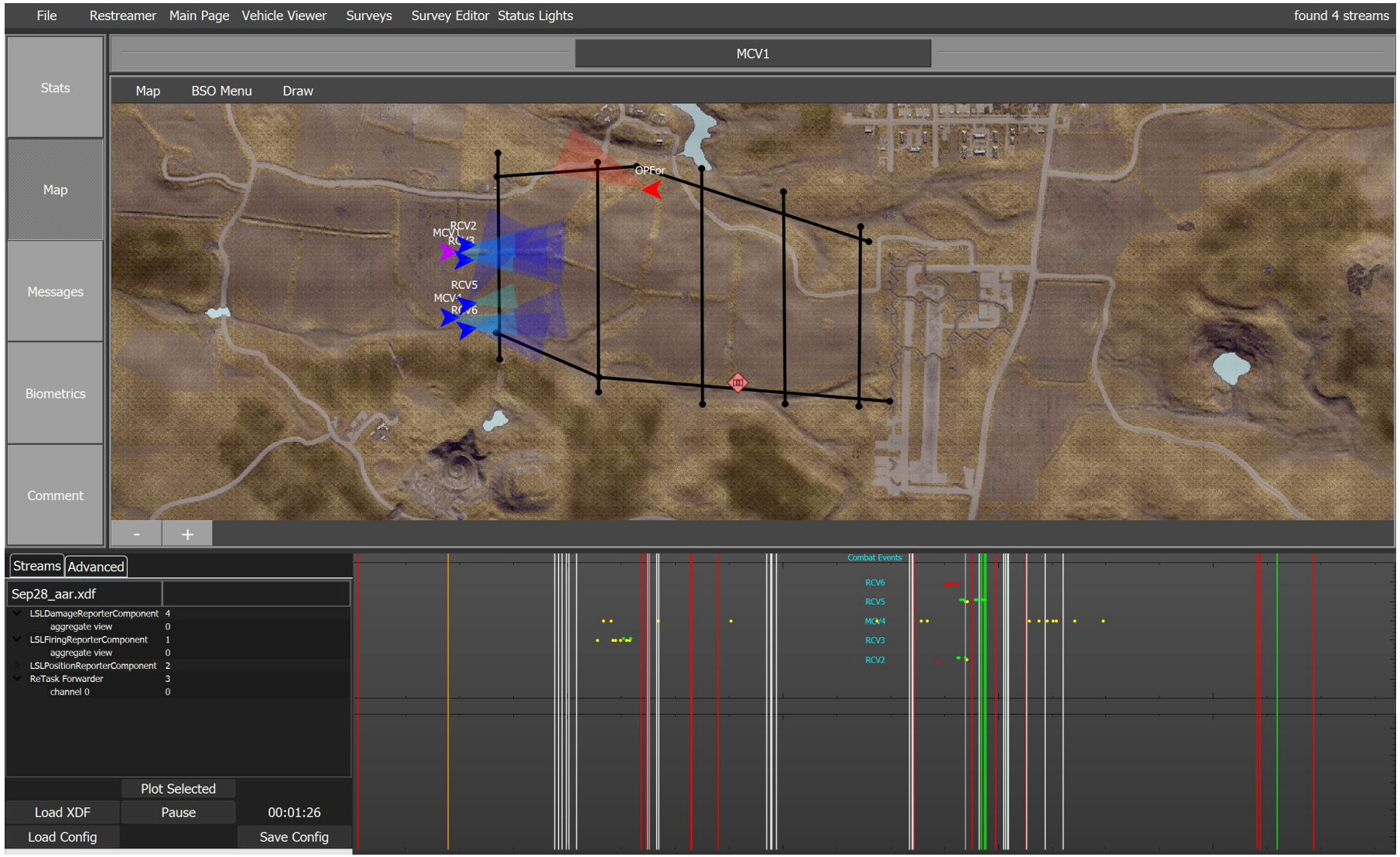
Approved for public release; distribution is unlimited.

T-CAT ARCHITECTURE



Approved for public release; distribution is unlimited.

T-CAT INTERFACE



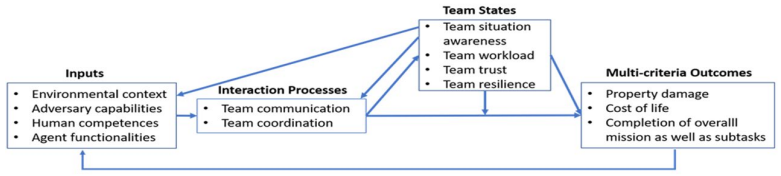
PAYOFF AND PATH FORWARD

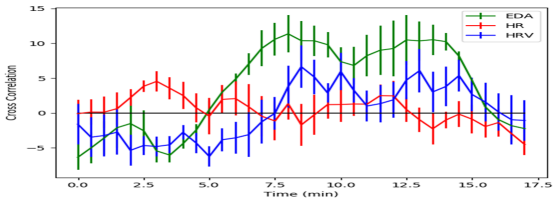
HAT[3] and T-CAT assessment tools provide a capability for continuous assessment of the complex interactions that occur within human-autonomy teams in the NGCV context.

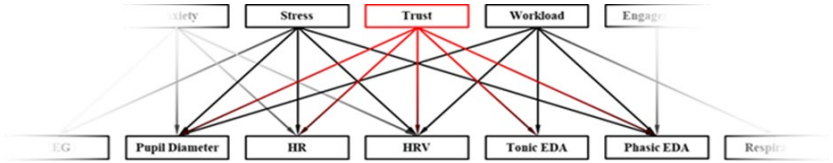
- Enables users to identify opportunities for training, adaptation, or trust interventions and ultimately improve human-autonomy team effectiveness.

Final Capabilities:

1. *Team performance management system that tracks and adapts team performance to maintain, intervene, and change behavior.*
2. *Assessment toolkit that enables S&T, Acquisitions, and T&E communities to characterize Soldier-Autonomy teams.*







OUR RESEARCH TEAM

DEVCOM ARMY RESEARCH LABORATORY

DEVCOM GROUND VEHICLE SYSTEMS CENTER

dcs corp

Georgia Tech School of Psychology College of Sciences

KENNESAW STATE UNIVERSITY

UCF

University of Central Florida

United States Military Academy West Point

ASU Arizona State University

Approved for public release; distribution is unlimited.

CONTACT AND REFERENCES

For more information:

- Mr. Ralph Brewer, ralph.w.brewer.civ@army.mil
- Dr. Andrea Krausman, andrea.s.krausman.civ@army.mil
- Dr. Catherine Neubauer, catherine.e.neubauer2.civ@army.mil

References:

Krausman, A., Neubauer, C., Forster, D., Lakhmani, S., Baker, A.L., Fitzhugh, S.M., Gremillion, G.M., Wright, J.L., Metcalfe, J.S., & Schaefer, K.E. (2022). Trust Measurement in Human-Autonomy Teams: Development of a Conceptual Toolkit. Transactions in Human-Robot Interaction (THRI), 11, 3, https://doi.org/10.1145/3530874.

Lakhmani, S., Neubauer, C....Schaefer, K.E. (2022). Cohesion in human–autonomy teams: an approach for future research, Theoretical Issues in Ergonomics Science, DOI: 10.1080/1463922X.2022.2033876

Neubauer, C., Fitzhugh, S.M., Baker, A.L., Kellihan, B., Jagielski, J., Krausman, A.S. Human-Autonomy Teaming Trust Toolkit (HAT3) Software Development Documentation and User Guide. DEVCOM Army Research Laboratory (US); 2023. Report No.:ARL-TR-9623

Neubauer, C., Forster, D.E., Blackman, J., Lakhmani, S., Fitzhugh, S.M., Krausman, A., Schaefer, K.E., Berg, S., and Rovira, E. (2021) Developing a new human-autonomy team cohesion scale. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Baltimore, MD: HFES.

Baker, A. L., Fitzhugh, S. M., Forster, D. E., & Schaefer, K. E. (2021). Communication metrics for human-autonomy teaming: Lessons learned from us army gunnery field experiments. *Proceedings of the Human Factors and Ergonomics Society, 65*(1), 1157-1161. https://doi.org/10.1177/1071181321651159

Approved for public release; distribution is unlimited.

THANK YOU.


U.S. ARMY