Human Factors in Robots, Drones and Unmanned Systems, Vol. 179, 2025, 73–82
https://doi.org/10.54941/ahfe1006376

Trust in AI and Autonomous Systems

Elizabeth Mezzacappa, Dominic Cheng, Lucas Hess, Nikola Jovanovic, Robert DeMarco, Jose Rodriguez, Madeline Kiel, Kenneth Short, Alexis Cady, Jessika Decker, Mark Germar, Keith Koehler, Nasir Jaffery, and Lawrence D'aries

U.S. Army Combat Capabilities Development Command Armaments Center, Picatinny Arsenal, NJ 07806, USA

ABSTRACT

In 2023, the Office of the Undersecretary of Defense established the Center for Calibrated Trust Measurement and Evaluation (CaTE) aimed at establishing methods for assuring trustworthiness in artificial intelligence (AI) systems with an emphasis on the human-autonomy interaction. As part of the CaTE effort, the DEVCOM Armaments Center's Tactical Behavior Research Laboratory was tasked with developing standards for testing and measuring calibrated trust in AI-enabled armament systems. Qualitative and quantitative measures of trust were collected from 114 Soldiers in table-top, force on force, simulated environments, demonstrations, trainings, and engineering integration events. A survey instrument, configured specifically for assessing trust in AI weapon systems, has been created for this research. Embedding with Soldiers during operational exercises using actual systems, the researchers were able to gather video footage and audio recordings of human systems integration (HSI) issues. Information from live exercises was used to configure a virtual environment experiment using the same terrain, controllers, and systems. This presentation will give an overview of the research program, with the emphasis on novel HSI data collection methods

Keywords: DoD 3000.09, Responsible AI, Autonomous weapons, Survey, Behavioral recording, Trust, Soldier data, Calibrated trust measurement and evaluation

INTRODUCTION

Without a doubt, artificial intelligence (AI) and autonomous armament systems are and will continue to be a critical component of warfighting (Army Rapid Capabilities and Critical Technologies Office, 2024; Office of the Under Secretary of Defense for Policy, 2023; US Army Futures Command, 2022). From the beginning of the US DoD considerations of these types of weapons, Warfighter trust in these systems was identified as a major concern for developers (Defense Science Board, 2012; 2016). Currently, Warfighter trust in automation and artificial intelligence is a focus of many publications (Brill et al., 2016; Hancock et al., 2021; Hancock et al., 2023; Kohn et al., 2021; Lee & See, 2004; Montgomery, 2019; Porter et al., 2020; Sablon, 2025; Smith, 2019; Weltman et al., 2023).

Moreover, the US DoD Responsible Artificial Intelligence Strategy and Implementation Pathway (DoD Department of Defense, 2022; DoD

Responsible AI Working Council, 2022) lists Warfighter Trust as Tenet 2: *Ensure warfighter trust by providing education and training, establishing a test and evaluation and verification and validation framework that integrates real-time monitoring, algorithm confidence metrics, and user feedback to ensure trusted and trustworthy AI capabilities*.

The Center for Calibrated Trust Measurement and Evaluation (CaTE) was established to assist the Department of Defense (DoD) in its efforts to ensure that artificially-intelligent (AI) systems are safe, reliable, and trustworthy before being fielded to government users in critical situations (Carnegie Mellon University News, 2024). CaTE addresses both the dynamics of how systems interact with each other, and especially the interactions between AI and humans – to establish trusted decisions in the real world. CaTE is a collaborative research and development center that works with all branches of military services on areas such as human-machine teaming and measurable trust.

Within the CaTE effort, the US Army Combat Capabilities Development Command Armaments Center (DEVCOM AC) Tactical Behavior Research Laboratory (TBRL) was tasked with developing tools to measure trust in AI-enabled autonomous lethal armaments. There are multiple published trust questionnaires for assessing levels of trust in systems (Jian et al., 2000; Schaefer, 2016; Wojton et al., 2020). However, the general nature of those questionnaire items fails to capture the unique nuances of trust in weapon systems in battle. This gap in knowledge required more operationally-relevant questionnaire items to be developed to provide insight into trust in the battlefield AI autonomous armaments in military scenarios.

In addition to traditional survey instruments, the TBRL adopted other methodologies consistent with grounded theory (Leedy & Ormrod, 2016). That is, an effort was made to develop some understanding of how the construct of trust manifests specifically in dismounted Soldiers' behaviors and specifically in interactions with AI and autonomous systems on the battlefield. Therefore, a multi-modal research program of qualitative and quantitative data collection was conducted. The methods included survey, focus groups, observation, and controlled virtual environment testing. There are several strengths of the program 1) data are primarily gathered from Soldiers with relevant experience and training during events where these they are using these types of autonomous assets, 2) engineers who design and build the systems support the creation of research test beds, 3) data are collected in high fidelity settings (e.g., virtual and force-on-force exercises).

This article gives a high-level overview of the multi-level methods executed in the CaTE data collection on trust in AI and autonomous weapon systems. Data are still being collected, processed, and analyzed.

Data Collection Venues

There were 6 Soldier data collection events. Soldiers were primarily 11B or 11A (or equivalent) military occupational specialty (MOS)

Table Top Exercise (TTX), n = 13

Participants were 13 Military Role Playing (MRP) Soldiers who were led by a retired general in a table top exercise structured as an action-reaction-counteraction wargaming exercise. Two vignettes were exercised - “Conduct an Attack” and “Seize a Foothold”. For each of these vignettes, three different sets of assets were assigned for use, a) a base case, b) assets projected to be available in 2030, and c) assets projected to be available in 2040. The first was a base case, in which there were no robotic assets. The 2030 assets consisted of a small set of robotic devices, including the focus of the CaTE work, 4 weaponized quadrupeds and 5 lethal UASs. The 2040 assets consisted of a larger set of robotic devices, including 8 weaponized quadrupeds and 10 lethal UASs. Trust questionnaires and focus groups were administered after each of the scenarios. Audio recordings of focus group discussions documented insights into Soldiers’ thoughts on these novel assets.

Live Drop of Unmanned Aerial Systems (UAS) Demonstration, n = 10

Four different UAS capabilities were observed. TBRL researchers instrumented and surveyed 10 Soldiers who have had extensive experience with autonomous and AI-enabled systems. The demonstrations included in-depth explanations about the technologies, hands-on assembly and disassembly, witnessing live fire, and explanations of failures. Trust questionnaires were administered after each of the four demonstrations over two days.

Virtual Environment (SIMX), n = 26

Trust was assessed in Soldiers after they engaged in a virtual environment using a lethal UAS and a lethal quadruped. The Soldiers tested representations of the novel weapon systems within the virtual environments where their performance and reliance on the equipment were measured. Soldiers were also asked to complete trust surveys and participate in a focus group. The goal of these activities was to guide the development of weapon systems and to gauge how their trust in these systems would vary depending on the direction/level of autonomy in the system.

Force on Force Exercises (FFE), n = 18

Every year the Army conducts an exercise that provides Soldiers a chance to train on, work with, and evaluate systems from different vendors. Soldiers are trained on the systems and are then given a chance to operate systems during platoon level and company level force-on-force exercises. The exercises used simulated munitions and were executed over a 3-week period during day and night conditions. Following this experience, the Soldiers provided their feedback to developers and decision-makers. A subset of the 18 participants were 11 members of a platoon who trained and deployed the lethal UAS and provided trust measurement throughout the exercises. Trust questionnaires were administered before training on the UAS, after training, after the platoon level missions, and after the company level missions.

Engineering Integration Event (EIE), n = 23

This event afforded engineers the opportunity to connect novel sensors, weapons, transports, and communication assets through a common network.

This systems-of-systems was then used in platoon level exercises. Trust in the quadruped and AI-enabled mission planning systems were assessed. 23 Soldiers completed questionnaires before and after training, and after completion of the exercises.

Live Fire Event (LFE) for a lethal UAS and a weaponized quadruped, n = 18 UAS/6 quadruped.

For these live fire events the UAS dropped training rounds and the quadruped shot live rounds. Eighteen Soldiers trained with and used the lethal UAS in scenarios that required the dropping of training rounds over static targets. Soldiers were assigned to one of three groups that used the system with varying levels of successful performance. The exercise terminated when all three UAS devices were broken, providing an important data point on Soldier trust of this weapon system. Six different Soldiers trained with and used the weaponized quadruped against practice targets. The intent was to also shoot live fire at static targets; however, the device malfunctioned and only 4 of the 6 Soldiers were able to zero the weapon and complete a round of practice. The Soldiers completed trust questionnaires before and after training, then after hands-on experience with the lethal UAS and weaponized quadruped.

Research Program Sequence and Integration

The evolution from TTX to LFEs demonstrates a natural progression from low fidelity to high fidelity in terms of situation and lethality capabilities. In addition, information and experience from prior events shaped data collection in later events. That was especially true in the early part of the research program where data from the TTX and FFE events were used to create the SIMX virtual environment. To create a high-fidelity environment, terrain data and the test bed scenarios were derived from the those presented in the TTX. Moreover, the challenges, difficulties, and errors that were committed by operators and systems in the TTX and live FFE were integrated into the SIMX environment. Also, throughout the research program, data collection methods were refined and tailored to the devices that Soldiers would be operating.

Primary Efforts

With these six Soldier data collection events, there were two primary efforts. The first was construction of a valid survey instrument to assess Soldier trust in specific lethal systems with potential autonomous capabilities. The second was the gathering of audio and video recordings of Soldiers observing live fires and planning, executing missions, and conducting after action reports using AI and autonomous weapon systems. The following sections describe the creation of the Soldier trust instrument and the gathering of audio and video recordings.

Trust Survey Construction

TBRL's research goal under the CaTE effort was to develop trust measures and metrics and investigate trust calibration processes for Soldiers with respect to artificially intelligent lethal autonomous weapon systems (LAWS). The research approach was to include both standardized, validated questionnaires from the open literature and to develop future questionnaire items that were specifically appropriate to Soldiers during warfighting operations using LAWS.

There are a number of validated scales evaluating trust in automation and robotics (Jian et al., 2000; Jian et al., 1998; Porter & Fealing, 2022; Schaefer, 2016; Wojton et al., 2020). These questionnaires assessing trust have been validated for uses in automation, however the data are derived from primarily civilian community members or young military personnel with little experience in Soldier operations. One could argue that the stresses associated with battles exacerbate the questions of trust. In addition, the extreme consequences of lethal autonomous systems add another dimension to the concepts of Soldier trust during the operation of these systems. The TBRL therefore seeks to develop a trust survey instrument that is sensitive to the factors uniquely contributing to Soldier interactions with lethal autonomous weapon systems in live military operations.

Development of the Soldier trust questionnaire started with the TTX. In the course of formulating the questionnaire for the TTX process, the incorporation of subject matter expert opinions was imperative to gain a comprehensive understanding of Soldier trust concerning AI and autonomous weapon systems. Construction of the questionnaires required the selection of a credible point system that accurately reflects an individual's trust in artificial intelligence and autonomy on the battlefield. Subject matter experts from West Point Military Academy, particularly from the Future Applied Systems Team (FAST), were consulted. A Cadet interning with TBRL facilitated contact with a MAJ and a SFC, both possessing extensive experience across various military schools, making them highly qualified to assist in the development of this questionnaire.

The questionnaire's design was motivated by the desire to afford Soldiers the opportunity to express their preferences between AI and a "battle buddy" and to elaborate on each response. This emphasis was specifically placed on actions in battle that could result in casualties, recognizing the critical nature of decisions related to life and death on the battlefield. Such actions encompassed maintaining concealability, mitigating fratricide, providing proper covering fire, ensuring flank coverage, and executing kill orders without the need for micromanagement. The questionnaires were deliberately structured with a combination of 1–5 Likert scale questions, true or false questions, and short answer questions to facilitate a comprehensive assessment.

In the next stage of trust survey development, a factor analysis of the responses to the questionnaire was performed using the FFE, EIE, and SIMX data sets. Principal axis factoring was used to extract four factors. Factor 1 represents following mission plan and functionality. Factor 2 represents negligence or danger of the device. Factor 3 represents how the device compares to a Soldier. Factor 4 was an odd outcome; it negatively loads into the model but has positive questions about trust. Consequently, these results suggest that this version of the questionnaire may not be a clear measure of trust as other factors that account for variance in the model are not lining up as one might expect with a trust measure. Additionally, this may be due to a few confusing or double-barreled questions that existed in this survey. Insights into the weaknesses of the questionnaire and the factor structures will be used to improve future iterations of the Soldier trust survey.

Behavioral Recording

Focus Group Recordings

Data for qualitative analytical methods were collected during focus groups through written notes and audio recordings. Because the TTX focus groups were conducted in a classified space, transcriptions of the discussions were screened for sensitive information. Planned analyses include semantic network analyses.

Soldier Recordings

Following ground theory approaches, initial attempts focused on developing an understanding what trust in an autonomous weapon system means to the Soldier. Therefore, a large part of the data collections are recordings, audio and video, of Soldiers talking about, training, interacting, planning, and evaluating these systems. To gain insights relevant to Soldier trust calibration in AI and automated armaments, researchers embedded with Soldiers/-operators of new technologies and decision-makers for a month, from before they receive New Equipment Training (NET), during, and after force-on-force exercises with AI and automated systems and armaments. Researchers recorded Soldier behaviors (verbal and visual) during the planning for simulated battle, conducting operations against an adversary, assessing battle damage, and analyzing performance.

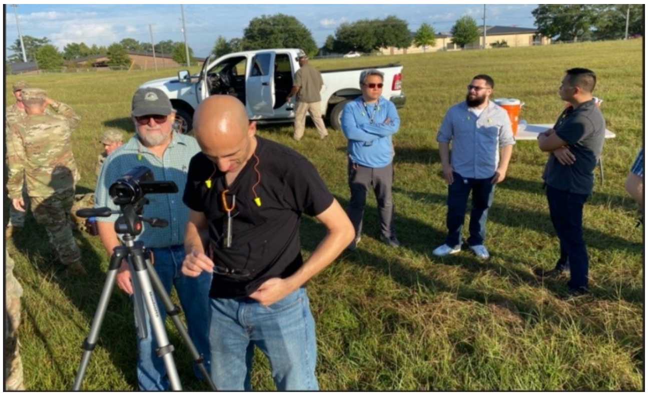

Figure 1: Setting up cameras for live fire recordings.

Instrumentation included audio recorders attached to the uniforms or packs of Commanders and Operators. Body cams were also clipped on the front of operators' uniforms to record interactions with the systems. In addition to the instrumentation of the Soldiers, researchers strategically positioned cameras on tripods to capture interactions. Furthermore, researchers following the platoons and companies used audio and video cameras to capture the wider context of the operations. Video and audio information taken during these exercises are still being processed and analyzed using state-of-the art qualitative analyses to identify trust processes and themes in relation to AI and autonomous armaments. Figures below show data collections.

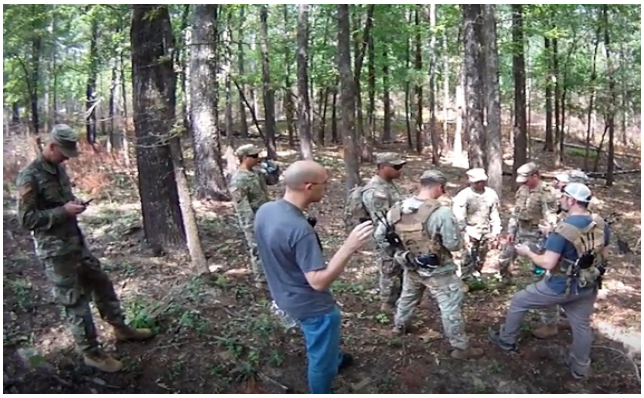

Figure 2: Researcher embedding with squad during force-on-force maneuvers.

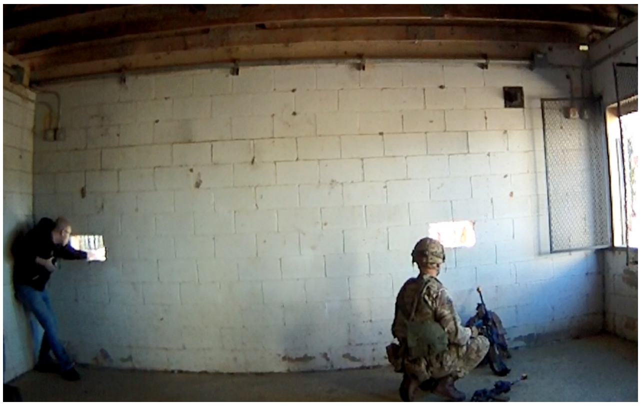

Figure 3: Videoing of UAS missions during exercise.

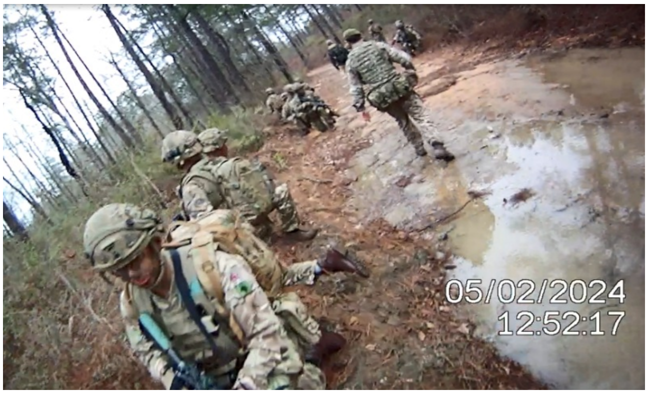

Figure 4: Recording of force-on-force exercises.

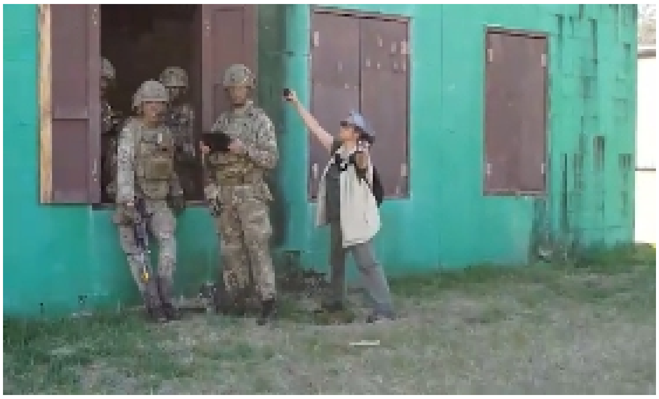

Figure 5: Recording of UAS operators.

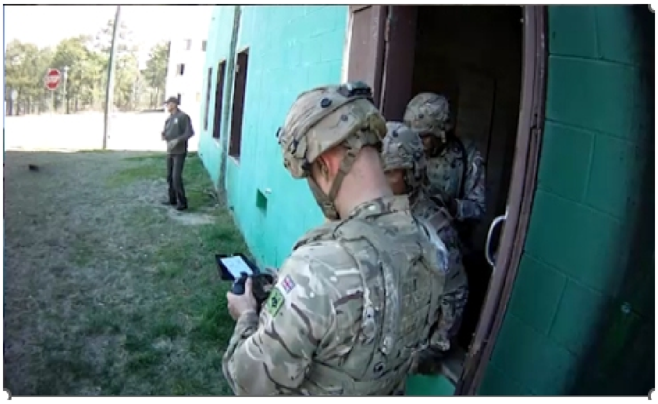

Figure 6: UAS operators.

CONCLUSION

This multi-modal method of assessing trust from Soldier data collection, spanning six separate Soldier Touch Points, yielded a rich archive for analyses. Future funding for the CaTE effort will support ensuring that the data are fully interpreted for insights into Soldier trust in AI and autonomous weapon systems.

ACKNOWLEDGMENT

The authors would like to acknowledge all Soldiers for their service to the country. We would like to thank the DEVCOM Armaments Center engineers and trainers involved with TBRL studies, including the event coordinators, USMA FAST Team, Bryan Almeida, and engineers Matthew Brauer and Bhavanjot Singh. In particular, we would like to thank our lab mates who contributed to leadership, data collection, analyses, and assistance for the project including Gladstone Reid, Joanne Arias, Anthony Robinson, Hillary Rogers, Darnell Simon, and Austin Hall.

REFERENCES

Army Rapid Capabilities and Critical Technologies Office. (2024). Human Machine Integrated Formations (HMIF) Overview & Incremental Roadmap Human Machine Integrated Formation Summit IV, Bryan, TX.

Brill, J. C., Bliss, J. P., Hancock, P. A., Manzey, D., Meyer, J., & Vredenburgh, A. (2016). Matters of Ethics, Trust, and Potential Liability for Autonomous Systems Human Factors and Ergonomics Society 2016 Annual Meeting.

Carnegie Mellon University News. (2024, April 30, 2024). SEI and DOD Center To Ensure Trustworthiness in AI Systems. https://www.cmu.edu/news/stories/archives/2024/april/sei-and-dod-center-to-ensure-trustworthiness-in-ai-systems

Defense Science Board. (2012). The Role of Autonomy in DoD Systems. Washington, DC: Defense Science Board.

Defense Science Board. (2016). Report of the Defense Science Board Summer Study on Autonomy. Washington, DC: Defense Science Board.

DoD Department of Defense. (2022). Responsible artificial intelligence strategy and implementation pathway.

DoD Responsible AI Working Council. (2022). U. S. Department of Defense Responsible Artificial Intelligence Strategy and Implementation Pathway.

Hancock, P. A., Kessler, T. T., Kaplan, A. D., Brill, J. C., & Szalma, J. L. (2021). Evolving trust in robots: Specification through sequential and comparative meta-analyses. Human factor, 63(7), 1196–1229.

Hancock, P. A., Kessler, T. T., Kaplan, A. D., Stowers, K., Brill, J. C., Billings, D. R., Schaefer, K. E., & Szalma, J. L. (2023). How and why humans trust: A meta-analysis and elaborated model. Frontiers in psychology, 14.

Jian, J., Bisantz, A., Drury, C., & Llinas, J. (2000). Foundations for an empirically determined scale of trust in automated systems. International Journal of Cognitive Ergonomics, 4, 53–71.

Jian, J., Bisantz, A. M., Drury, C. G., & Llinas, J. (1998). Foundations for an empirically determined scale of trust in automated systems.

Kohn, S. C., de Visser, E. J., Weise, E., Lee, Y., & Shaw, T. H. (2021). A measurement of trust in autonomation: A narrative review and reference guide. Frontiers in psychology, 12.

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. Human Factors, 46(1), 50–80.

Leedy, P., & Ormrod, J. (2016). Practical Research: Planning and Design, 11th Edition. Pearson.

Montgomery, C. S. (2019). Trust in the Machine: AI, Autonomy, and Military Decision Making with Lethal Consequences.

Office of the Under Secretary of Defense for Policy. (2023). DoD Directive 3000.09 Autonomy in Weapon Systems.

Porter, D., & Fealing, C. (2022). Predicting Trust in Automated Systems: Validation of the Trust of Automated Systems Test (TOAST).

Porter, D., McAnally, M., Bieber, C., Wojton, H., & Medlin, R. (2020). Trustworthy Autonomy: A Roadmap to Assurance Part I: System Effectiveness.

Sablon, K. (2025). DoD Joint Robotics and Autonomous Systems Strategy.

Schaefer, K. E. (2016). Measuring Trust in Human Robot Interactions: Development of the "Trust Perception Scale-HRI. In Robust Intelligence and Trust in Autonomous Systems. Springer.

Smith, C. J. (2019). Designing Trustworthy AI: A Human-Machine Teaming Framework to Guide Development.

US Army Futures Command. (2022). Future Operational Environment: Forging the Future in an Uncertain World. (AFC PAM 525-2).

Weltman, G., Kohn, S., Cohen, M., Johnson, A., & Terman, M. (2023). Calibration of ethical trust for lethal autonomous weapons systems (CETLAWS).

Wojton, H., Porter, D., Lane, S., Bieber, C., & Madhavan, P. (2020). Initial validation of the trust of automated systems test (TOAST). Journal of Social Psychology, 160, 1–16.