**AI Challenge for Armaments Systems Engineering – A Preliminary Perspective on Human System Integration**

David Du and Nathan Lau

Virginia Tech

This team (still in progress toward full formation) will explore Human Systems Integration (HSI) or the human elements in Systems Engineering (SE) in the context of armaments equipped with Artificial Intelligence (AI) for operations in clearing mines. This paper presents the team’s preliminary perspective on trust in AI and some potential focus areas that would be essential to develop trustworthy AI and trusted-AI operations.

*Perspective on Trust in AI*. Trust is commonly defined as reliance or confidence in the dependability of a someone or something (APA Dictionary of Psychology)[[1]](#footnote-1). Trust or reliance between humans is essential for efficient social cooperations. For cooperation without time pressure, productivity and positive outcomes can be achieved without formal contracts and endless negotiation on acceptable behaviors for an infinite number of possible situations. In fact, even legal contracts include highly subjective terms, like “in good faith”, to connote trust between parties. For cooperation under high time pressure, trust moderates[[2]](#footnote-2) the types and amount of overt communication on idiosyncratic details across situations that compete for cognitive resources and response time. However, *trust is not a free lunch*. First, trust is refined in accuracy, precision (also known as resolution), and specificity over interactions across different situations, which take time and effort. For example, a soldier might trust each other with their life in the battlefield but not necessarily in the kitchen after a meal. Second, trust can be mis-calibrated (under-/over-trust) with insufficient and ineffective interactions between parties, leading to unexpected, negative consequences. In light of these insights, the team asserts that fostering human-AI trust requires a more comprehensive, iterative, and human-centric approach to system design.

Based on the general nature of human-human trust, we derive the following premise[[3]](#footnote-3) as the basis in advancing HSI and SE in the context of the mine clearing case study:

1. Trust is essential for cooperation because system complexities and situational idiosyncrasies preclude computationally exhaustive means to predict behaviors and/or outcomes. That is, unanticipated events will demand human and AI to cooperate in adaptive problem solving.
2. Trust is an emergent property from the interactions between human and AI. That is, trust can never be fully attributed to AI design (given premise 1).
3. AI could possess design characteristics (and humans could have traits) that accelerate the emergence of trust in the system.
4. The accuracy/calibration, precision/resolution, and specificity of trust are dependent on *quantity* and *quality* of interactions between human and AI.
5. Human and AI must be equipped with interaction mechanisms to achieve optimal calibration, resolution, and specificity of trust on the system as a whole.

Assuming the premise, albeit incomplete, to be true, the team contends that SE must provision appropriate interactions between stakeholders/users at different stages of the AI product life cycle for optimal calibration, resolution, and specificity of trust.

*Provisioning Human-AI Interactions.* In this AI Challengefor Armaments SE, the team will investigate when and how to provision “user interactions”, loosely defined, in the different stages of AI life cycle to optimize trust and performance during the Rapid Safe Passage (RSP) mission. Figure 1 presents an extension to the well-established Department of Defense (DoD) Systems Engineering Process (Defense Acquisition University, 2022)[[4]](#footnote-4) within the context of mine-clearing operations from the perspective of the human operator at Command and Control (C2). Comprehensive and iterative in nature, the process diagram depicts the top-down engineering approach that the operator employs to translate initial mission requirements into a series of informed decisions and actionable outputs.

A diagram of a system

Description automatically generated

Figure 1. Preliminary Extension of the DoD SE Process to Account for Human Factors and Trust

Based on the SysML model of the existing system provided for this AI Challenge, four areas within the DoD SE process are identified to be potentially meaningful for provisioning interactions between the human and AI (Figure 2). Specifically, the team plans to investigate provisioning interactions in the requirements loop, design loop, system analysis & control, and verification process for the next phase of the challenge.

A diagram of a system

Description automatically generated

Figure 2. Overview of Focus Areas in Human Operator/C2 Process Model

1. *Participatory AI design in the Requirements Loop*

Inside the requirements loop, analysis leading to requirements for AI and other systems design characteristics are specified with major implications on interactions between human and AI and ultimately trust of the overall system. Given that AI is immensely data driven (at least partially model-free), the SE process should integrate participatory design so that operators or users will provide early inputs and acquire early understanding on the competence and limitations of the AI. This includes data curation that define the boundary of competence, limitations, and edge cases of the AI that operators must be cognizant of during operations. From a Bayesian perspective, participatory design enables operators to have a more accurate ‘priors’ before system deployment. In the absence of it, these priors are likely to be influenced by rumors, which generates uninformed expectations and misunderstanding of the system's capabilities and performance. Participatory design could also lead to early requirements on how operators could potentially like to interact with the AI for adapting their decision making synergistically. Together, the emergence of trust could improve because the priors could be closer and interaction mechanisms be more efficient in reach to the best level of trust after deployment.

1. *Information Presentation inside the Design Loop*

Inside the design loop, the transformation from functional to physical architecture is where physical interfaces are defined and subsequently refined. These user interfaces will be designed to support direct interactions between the human operator and AI. The quality of the user interface or interaction designs will dictate how trust will emerge in the system. Some critical and unresolved research questions includes what types and amount of information is essential for human operators for collaborative command and control during operations. Interface/interaction design directly influence the levels of resolution and specificity of trust. Thus, the SE process should explicitly provision for user interface/interaction design during operations for trust to emerge.

1. *Fault/Failure Recovery in Systems Analysis and Control*

Embedded in system analysis and control, the reactive measure of fault/failure recovery is just as vital as the proactive measure of risk identification and mitigation in maintaining human-AI trust. For cooperations under high time pressure, risk, and consequence, strategies and system flexibilities for recovering from failures could significantly determine trust in the system. Rigid systems that do not permit human and/or AI to recover from their errors, particularly with first failure, will erode trust significantly. The team will investigate how the SE process should account for the integration of “user interactions” concerning recovery from failures with adaptive interfaces, user-centric feedback loop, scenario-based simulations for promoting resilience and reliability.

1. *Integrated Systems Validation in Verification*

While verification & validation (V&V) are not traditionally embedded within the SE process employed, the team proposes the inclusion of validation to facilitate integrated systems validation to incorporate human operators as part of the testing and evaluation for assurance of minimal operational performance. This approach has been established for most safety critical systems, particularly in civilian aviation and nuclear power generation. Numerous severe accidents, such as Three Miles Island, Air France 477 and Boeing 737 Max 8, illustrate that the idiosyncratic interactions between human and machine results in unanticipated chain of events that resulted in the accidents. The SE process should explicitly test and evaluate the uncertainty associated with human-AI interactions for beyond design basis scenarios to determine the known and unknown interaction risks for operations that would dictate the actual levels of trust.

In summary, this position paper outlines a preliminary perspective on Human-Systems Integration in developing trustworthy AI for armament SE. By emphasizing the importance of trust as an emergent property of human-AI interactions, the team argues that the SE process should strategically provision user interactions with AI throughout the system life cycle. The proposed focus areas of participatory AI design, information presentation, fault/failure recovery, and integrated systems validation highlight critical points where human-AI collaboration can enhance SE for a more reliable and trustworthy system.

1. American Psychological Association. Trust. In APA Dictionary of Psychology. https://dictionary.apa.org/trust [↑](#footnote-ref-1)
2. Trust does not always reduce communication. High trust can positively influence personnel to questions their peers or superiors. [↑](#footnote-ref-2)
3. The team omitted other fundamental concepts about trust in AI or automation for simplification. [↑](#footnote-ref-3)
4. Defense Acquisition University. (2022). Systems Engineering Guidebook (February 2022 ed.). U.S. Department of Defense. https://www.dau.edu/acquipedia-article/systems-engineering-process [↑](#footnote-ref-4)