

AFRL SUNY Trusted AI Challenge Series

Topic #2: Human-Artificial Intelligence Performance Optimization: Trust and Joint Action for Digital Data Analysis

Sponsor: AFRL

Objective: This topic seeks to optimize human-AI performance and efficiency through the lens of shared joint action for digital data analysis and reporting. Achieving this objective requires new methods and greater precision for human-AI analytic joint action that consider trust calibration and require system learning and adaptation, with preference for human-AI collaboration that have real-world applications.

Description: Current research on trust in AI is nascent; however, the state of the literature suggests that the trust construct for AI must be evaluated from the side of the human trustee and the AI trustor, and should consider shared context or situation awareness to achieve joint action. In addition, these results also suggest that the traditional trust model for automation continues to be viable for human-AI collaborations, with ability, integrity, and benevolence as the primary bases of trust,¹ with the caveat that these indicators generally require intent to cooperate versus coordinate.

To achieve this objective, trust must be appropriately calibrated between the human operator and its AI partner to effectively and efficiently deliver analytic products. Such products will vary across timelines, when the human-AI partnership have disparate goals beyond the required collaboration (e.g. scheduling, push-pull demands) to produce required scheduled and on-demand analyses. In these cases, human analysts must trust the data pushed and pulled from the AI, and the AI must learn to adapt to the preferences, topics of interest, expertise, and scheduling demands for one or more human analysts to calibrate trust. The topic sponsor may provide a subject-matter expertise and evaluation criteria by which to evaluate the effectiveness of the approach. Experts may suggest different capabilities, whether in-house or external, for potential testing scenarios and provide input and scope for shared analytic joint action tasks of interest to current and future lab programs.

¹ Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20, 709–734.

Guidance: Success for this topic would entail demonstrable improvement of human-AI performance and efficiency in joint analytic tasks that involve analysis and reporting. A desired solution would include measures or mechanisms of trust calibration and agent learning and/or adaptation to one or more human analysts. Selected efforts will be for a short-term effort, such as a seedling effort, and the anticipation is that the final solution could be further developed. All solutions would include suggested measures of performance and efficiency that consider trust, learning, and adaptation. Solutions may also be selected for extension and additional funding. Proposers should provide a research extension plan at the effort's conclusion to be considered for additional research funding and extension.

Summary: This topic seeks ideas that consider optimizing performance and efficiency in analytic joint action tasks. Potential solutions should address operator-trustee trust calibration during said action, with a preference for research, development, and testing using real-world or simulated real-world joint action analytic contexts, which could range from human resource or training data and reporting requirements to tasks where analyses using big data, or large amounts of structured and/or unstructured data in one or more media formats, could benefit from human-AI collaboration.