REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
06/09/2023	Conference Presentation	

4. TITLE AND SUBTITLE

TRUST ISSUES: Understanding warfighter trust in an airspace deconfliction engine

5a. CONTRACT NUMBER
W56KGU-18-D-0004

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Dar-Wei Chen; Craig Doescher;

5d. PROJECT NUMBER

5e. TASK NUMBER
W56KGU22F0017

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The MITRE Corporation
7515 Colshire Drive
McLean, VA 22102

8. PERFORMING ORGANIZATION REPORT NUMBER
PRS-23-1663

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Defense Advanced Programs Agency Adaptive Capabilities Office

10. SPONSOR/MONITOR'S ACRONYM(S)
DARPA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES
Prepared for Military Operations Research Symposium (MORS).

14. ABSTRACT

Summary: The MITRE Corporation is supporting DARPA on the assessment of ASTARTE, an AI-driven airspace deconfliction engine (COA recommender); Methods presented here are an attempt at understanding, in the domain of airspace deconfliction, operator trust in automation and automation transparency; Initial results presented in this work reflect early operator feedback on a prototype of the ASTARTE engine; These methods have uncovered opportunities to understand trust in AI engines and could be adapted for other use cases as well;

15. SUBJECT TERMS

Airspace; Air Traffic Management; operator trust; airspace conflicts; ASTARTE; Air Space Total Awareness for Rapid Tactical Execution

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE		13	Susan Carpenito
					19b. TELEPHONE NUMBER *(Include area code)* 781-271-7646

TRUST ISSUES:

Understanding warfighter trust in an airspace deconfliction engine

Dar-Wei Chen, Ph.D. (Senior Cognitive Engineer)
Craig Doescher (MITRE ASTARTE Task Lead)
The MITRE Corporation

This technical data deliverable was developed using contract funds under Basic Contract No. W56KGU-18-D-0004.

MITRE | SOLVING PROBLEMS FOR A SAFER WORLD®

Distribution Statement A. Approved for Public Release; Distribution Unlimited. Public Release Case Number 23-1663

ASTARTE:
Air Space Total Awareness for Rapid Tactical Execution

Relevant entities

- DARPA (sponsor)
- MITRE Corporation (advisor)
- Raytheon (performer)

Domain: Airspace deconfliction

- Long-range fire missions in large-scale combat operations (LSCO)
- Operating manned and unmanned aircrafts within and between airspaces
- Supporting the JAGIC at the division level (Joint Air-Ground Integration Center)

DESCRIPTION | An **AI-driven decision engine** that uses algorithms and sensor networks to:

- Provide dynamically-updated **4D airspace pictures**
- Anticipate and **solve airspace conflicts**
- **Recommend COAs** based on varying mission priorities
- **Provide situation awareness** of adversary airspace

the ASTARTE engine will be assessed via some **“traditional” metrics**…

BAA metrics (broad agency announcement)

- **Time** to generate a plan or to re-plan
- **Percentage** of conflicts correctly identified
- **Percentage** of COAs deemed operationally acceptable
- **Number** of entities tracked
- **Number** of missions handled simultaneously

Communicating with key systems: AFATDS, TAIS, JADOCS, etc.

Acronyms
- **AFATDS**: Advanced Field Artillery Tactical Data System
- **TAIS**: Tactical Airspace Integration System
- **JADOCS**: Joint Automated Deep Operations Coordination System

Handling key data types: ATO/ACO, UAP, USMTF, etc.

Acronyms
- **ATO**: Air tasking order
- **ACO**: Airspace control order
- **UAP**: Unit airspace plan
- **USMTF**: United States Message Text Format

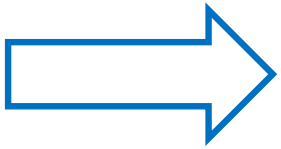
MITRE

…but the **focus of this presentation today** is an initial assessment of **operator trust and automation transparency** of an ASTARTE system prototype

Sociotechnical system: A collection of humans, technology, and work (SOPs, policies, etc.), working to accomplish a goal (Dorton, 2022)

Transparency

Enables operators to understand why a technology is performing a given action



Appropriate trust

Operators calibrating trust in automation to the right degree and in the right situations

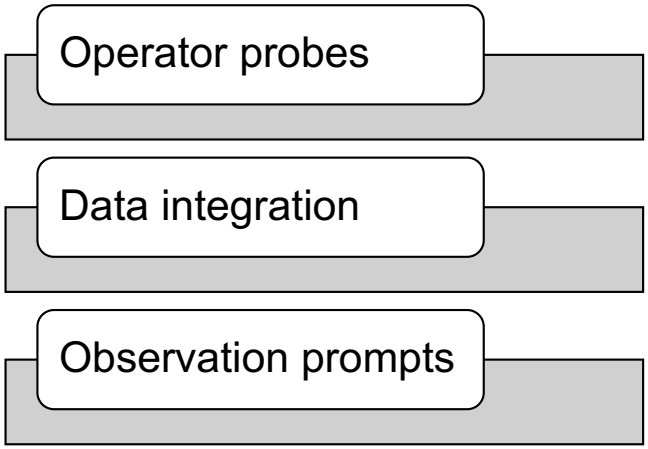
- **"Goldilocks" amount of information** provided to enable proper interpretation of system actions
- **Layman terminology** (e.g., showing code doesn't help most people!)
- **Better operator control** of system usage
- **Uncertainty, probabilities, etc.** boost transparency
- **Manual take-over** is more possible when operators understand how system works

- **Automation: Not always perfect** (transparency enables operators to see that)
- **Over-trusting automation** leads to:
  - Suboptimal decision-making
  - Complacency and SA loss (e.g., driver-less cars)
  - Skill degradation
- **Under-trusting automation** leads to: Unnecessarily high operator workloads

Trust: "The attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (Lee & See, 2004)

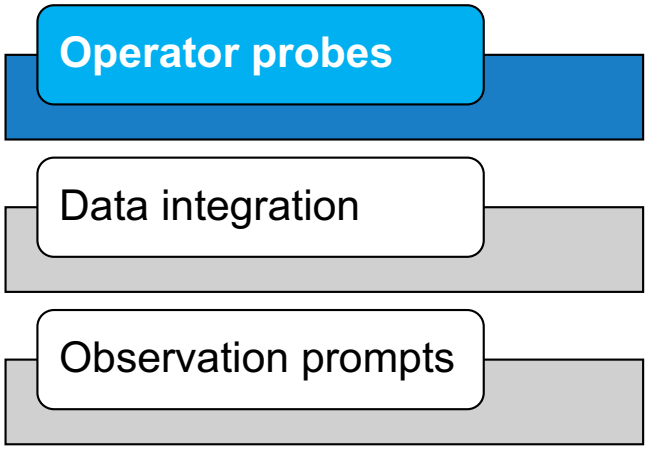
MITRE

Here are initial trust and transparency areas of interest for the ASTARTE engine and airspace deconfliction (adapted from previous literature)



MITRE

Here are initial trust and transparency areas of interest for the ASTARTE engine and airspace deconfliction (adapted from previous literature)



"[ASTARTE can] provide a solution normally given by the operator"

- SME #2

Notes on usage

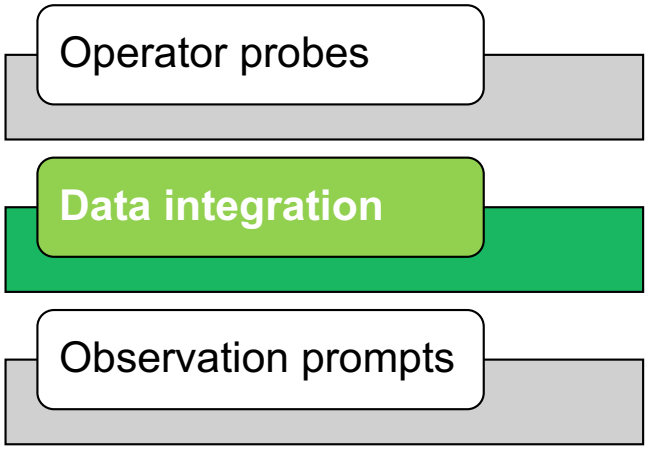
- Response format: 0-10 scale (Likert)
- Administration: As survey questions during or after exercises

Topics to be probed

A. General trust in COA (course of action) recommender engine
B. Provenance: Trust in developers of engine
C. Familiarity of engine compared to similar technologies
D. Integration of engine with rest of existing operations
E. Understanding pros and cons of proposed COAs
F. Rec factors: Confidence in ASTARTE to account for all relevant factors when generating COAs
G. Predictability: Consistency of ASTARTE-generated COAs with current processes' COAs
H. Directability: Degree of feeling in control of COA generation process

MITRE

Here are initial trust and transparency areas of interest for the ASTARTE engine and airspace deconfliction (adapted from previous literature)



Notes on usage

- Response format: 0-10 scale (Likert), select all integration issues that apply
- Administration: As survey questions during or after exercises

Operators are asked about the following data types and integration issues:

Data types

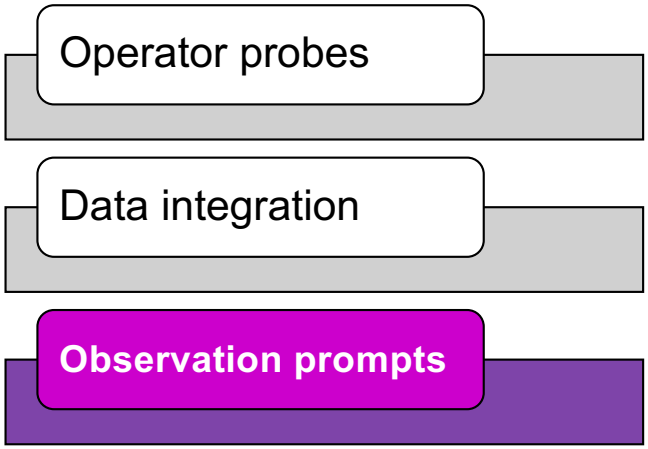
I. AFATDS data builds: System preferences
II. AFATDS data builds: Geometries
III. ASTARTE delivery: System integration
IV. ASTARTE planning: Counterfire units identified
V. ASTARTE planning: Permissions and authorities
VI. ASTARTE planning: Data paths
VII. ASTARTE planning: System integration requirements

Code letters for integration issues

A. The data was not located where I wanted or expected
B. The data was not presented in the numerical or stylistic format that I wanted or expected
C. The data was incomplete and I was expecting more
D. The data was not correct
E. The data was missing altogether
F. The element was not timely in its updates and in its interactions with other systems
G. It was not obvious whether or which ingested data fields were accounted for in the generation of COAs

MITRE

Here are initial trust and transparency areas of interest for the ASTARTE engine and airspace deconfliction (adapted from previous literature)

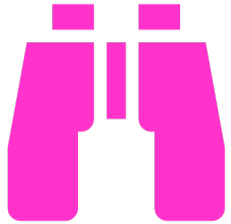


Notes on usage

- Subject matter experts (SMEs) use these prompts to take notes on how the operators and ASTARTE are performing
- Notes can be taken during and after exercises

Topics covered in prompts

- Operators manually performing in place of ASTARTE (distrust)
- Operator surprise or confusion regarding recommended COAs
- Can operators identify “questionable” generated COAs, and do operators stick with them anyway? (over-trust)
- Are risks readily apparent to operators?



Highlights from preliminary SME feedback:
Operator probes

Current trust-related feedback

Trust in the ASTARTE engine is largely dependent on the quality of the **info provided by C2 systems**

Need to see more of the **variables that affect COA generation**

Experienced operators have "healthy distrust" of new systems and will need **time to "see it for themselves"**

the ASTARTE engine largely feels and performs like other relevant technologies (AFATDS, TAIS)

Organizations and processes that developed the ASTARTE engine are generally considered trustworthy (could use more joint fires SMEs, though)

the ASTARTE engine will be more well-integrated and trusted **as a microservice** for other systems, as opposed to a separate system and interface

MITRE

Highlights from preliminary SME feedback:
Observation prompts

It's unclear…

- …why asset A was unavailable
- …why target X was prioritized by the ASTARTE engine over target Y
- …how many criteria need to be designated red or yellow before a COA is unusable
- …why the "accept" button for a COA is still available if the COA was considered unusable

Observations

1. Operators trust the ASTARTE prototype less when mission locations and times are disparate (recommendations are less reliable)
2. The ASTARTE prototype does not make explicit when data needed for a fully-informed recommendation is unavailable

Future work

1. Use the questions and prompts at ASTARTE exercises
2. Wisdom of crowds: Collect trust- and transparency-related feedback in other ways
3. Improve the ASTARTE engine's transparency based on observations and operators' responses → Operators can better calibrate trust in automation

Question	Avg. rating (1-5 scale; 5 = strongly agree)
Understand presented COAs?	**3.29**
Confident in presented COAs?	**3.43**

Collected at Mission Command Battle Lab assessment (Ft. Leavenworth)

MITRE

Summary

- The MITRE Corporation is supporting DARPA on the assessment of ASTARTE, an AI-driven airspace deconfliction engine (COA recommender)
- Methods presented here are an attempt at understanding, in the domain of airspace deconfliction, operator trust in automation and automation transparency
- Initial results presented in this work reflect early operator feedback on a prototype of the ASTARTE engine
- These methods have uncovered opportunities to understand trust in AI engines and could be adapted for other use cases as well

MITRE

References

Beck, H., Dzindolet, M., & Pierce, L. (2007). Automation usage decisions: Controlling intent and appraisal errors in a target detection task. *Human Factors*, *49*, 429–437.

Chen, J. Y., & Barnes, M. J. (2014). Human–agent teaming for multirobot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems*, *44* (1), 13-29.

Dorton, S.L. (2022). Supradyadic trust in artificial intelligence. *Artificial Intelligence and Social Computing, 28*, 92-100. https://doi.org/10.54941/ahfe1001451

G.B. Duggan, S. Banbury, A. Howes, J. Patrick, and S.M. Waldron. Too Much, Too Little or Just Right: Designing Data Fusion for Situation Awareness. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48(3):528–532, 2004.

Endsley, M. R. (1988, October). Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society annual meeting* (Vol. 32, No. 2, pp. 97-101). Sage CA: Los Angeles, CA: Sage Publications.

Hoffman, R. R., et al. (2006). A Method for Eliciting, Preserving, and Sharing the Knowledge of Forecasters. *Weather and Forecasting*, *21* (3), 416–428.

Körber, M. (2018, March 7). Theoretical considerations and development of a questionnaire to measure trust in automation. https://doi.org/10.31234/osf.io/nfc45

Lee, J. D. (2012). Trust, trustworthiness, and trustability. Presented at the *Workshop on Human-Machine Trust Robust Autonomous Systems*, Ocala, FL, USA.

Lee, J. D. & Moray, N. (1992). Trust, control strategies and allocation of function in human–machine systems. *Ergonomics*, *35*, 1243–1270.

Lee, J.D. & See, K. A. (2004). Trust in technology: Designing for appropriate reliance. *Human Factors*, *46*, 50–80.

Linegang, M. (2006). Human-automation collaboration in dynamic mission planning: A challenge requiring an ecological approach. In *Proceedings of the Human Factors Ergonomics Society Annual Meeting*, 2482–2486.

Madsen, M., & Gregor, S. (2000, December). Measuring human-computer trust. In *11th australasian conference on information systems* (Vol. 53, pp. 6-8). Australasian Association for Information System.

Manzey, D., Reichenbach, J., & Onnasch, L. (2012). Human performance consequences of automated decision aids: The impact of degree of automation and system experience. *Journal of Cognitive Engineering and Decision Making, 6*, 57–87.

McBride, S., Rogers, W., & Fisk, A. D. (2011). Understanding the effect of workload on automation use for younger and older adults" *Human Factors, 53*, 672–686.

Mueller, S. T. & Klein, G. A. (2011). Improving Users' Mental Models of Intelligent Software Tools. *IEEE Intelligent Systems*, pp. 77–83.

Paradis, S., Benaskeur, A., Oxenham, M., & Cutler, P. (2005). Threat evaluation and weapons allocation in network-centric warfare. In Proceedings of the 8th International Conference on Information Fusion, volume 2, pages 1078–1085. IEEE.

Tenhundfeld, N. L., de Visser, E. J., Ries, A. J., Finomore, V. S., & Tossell, C. C. (2020). Trust and distrust of automated parking in a Tesla Model X. *Human factors*, *62*(2), 194-210.

van Dongen, K. & vanMaanen, P. (2013). A framework for explaining reliance on decision aids. *International Journal of Human–Computer Studies*, *71* (4), 410–424.

Wojton, H. M., Porter, D., T. Lane, S., Bieber, C., & Madhavan, P. (2020). Initial validation of the trust of automated systems test (TOAST). *The Journal of social psychology*, *160*(6), 735-750.