



EVALUATING USER TRUST IN LARGE LANGUAGE MODELS

Authors: Samantha K. Berg¹, Nicholas Waytowich², Vinicius G. Goecks², and Andrea Krausman²

¹Oak Ridge Associated Universities, ²U.S Army DEVCOM Army Research Laboratory

Background

- As Large Language Models (LLMs) become integrated into future military command and control (C2) systems for streamlining the Military Decision-Making Process (MDMP), understanding and measuring trust in these AI systems is critical.
- The U.S. Army Combat Capabilities Development Command Army Research Laboratory is developing a research initiative to evaluate trust in LLM-based decision support tools. Initial Soldier touchpoints with COA-GPT included direct observation of Soldier-LLM interactions, structured free-response assessments, and detailed annotation of interaction processes.
- A series of follow-on touchpoints to capture feedback from users and subject matter experts (SMEs), as well as targeted experimentation round out our research program. These interactions will combine informal feedback sessions with structured evaluations, allowing us to gather both spontaneous user insights and systematic data on trust formation.

COA-GPT Touchpoint

- Users completed a simulated mission, using COA GPT as an assistive aid for task identification and risk assessment.
- COA-GPT was used to execute the Mission Analysis.
- COA-GPT generated the first output. Users had the opportunity to choose to provide feedback so the output could be regenerated, or approve it as is, and move on to the next substep.

Figure 1. Sample Mission Path.

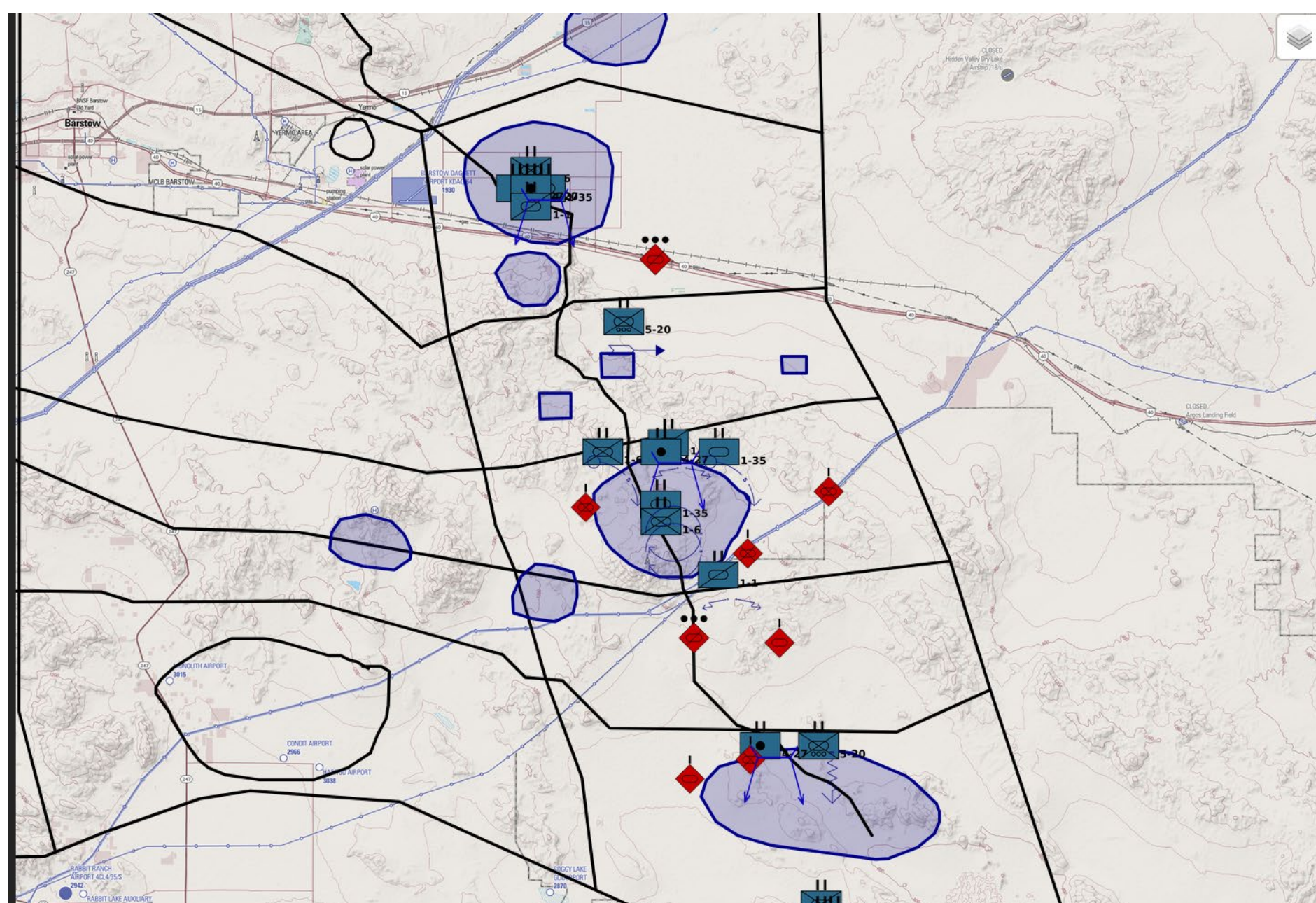


Figure 2. MDMP Substeps.

Generate Products							
MDMP Step 1: Receipt of Mission							
Actions	Human Feedback	Generate	MDMP Substep	Prompts	Generated Products	Status	
Approve	Human Feedback	Generate	3. Update Running Estimates	View Prompts	Reasoning	View Estimates	Not Reviewed
Approve	Human Feedback	Generate	4. Conduct Initial Assessment	View Prompts	Reasoning	Main Output	Not Reviewed
Approve	Human Feedback	Generate	5. Issue Suggested Commander's Initial Guidance	View Prompts	Reasoning	Main Output	Not Reviewed
Approve	Human Feedback	Generate	6. Issue Initial Warning Order #1	View Prompts	Reasoning	Main Output	Not Reviewed
					Download Product		
MDMP Step 2: Mission Analysis							
Approve	Human Feedback	Generate	1. Analyze Higher Headquarters' Plan or Order	View Prompts	Reasoning	Main Output	Not Reviewed
Approve	Human Feedback	Generate	2. Perform Initial Intelligence Preparation of the Battlefield	View Prompts	Reasoning	Main Output	Not Reviewed
Approve	Human Feedback	Generate	3. Determine Specified, Implied, and Essential Tasks	View Prompts	Reasoning	Main Output	Not Reviewed
Approve	Human Feedback	Generate	4. Review Available Assets and Identify Resource Shortfalls	View Prompts	Reasoning	Main Output	Not Reviewed
Approve	Human Feedback	Generate	5. Determine Constraints	View Prompts	Reasoning	Main Output	Not Reviewed
Approve	Human Feedback	Generate	6. Identify Facts and Develop Assumptions	View Prompts	Reasoning	Main Output	Not Reviewed
Approve	Human Feedback	Generate	7. Begin Risk Assessment and Management	View Prompts	Reasoning	Main Output	Not Reviewed
Approve	Human Feedback	Generate	8. Develop Initial CCB and EEP	View Prompts	Reasoning	Main Output	Not Reviewed
Approve	Human Feedback	Generate	10. Develop Initial Information Collection Plan	View Prompts	Reasoning	Main Output	Not Reviewed
Approve	Human Feedback	Generate	11. Update Plan for the Use of Available Time	View Prompts	Reasoning	Main Output	Not Reviewed
Approve	Human Feedback	Generate	12. Develop a Proposed Problem Statement	View Prompts	Reasoning	Main Output	Not Reviewed
Approve	Human Feedback	Generate	13. Develop a Proposed Mission Statement	View Prompts	Reasoning	Main Output	Not Reviewed

Feedback

- Once the user completed a substep, they provided a written evaluation of the output generated and gave it a confidence score using a Likert scale (1 to 5). Follow-up surveys were completed.
- Questions assessed dynamics such as COA-GPT's ability to identify tasks and issue accurate statements. User trust and confidence in COA-GPT was also evaluated.

Table 1. Sample Feedback Survey Questions.

- How well did COA GPT identify Specified Tasks in the Higher HQ order?
 - This question related to task identification. Answers were rated on a Likert Scale of the following: [Not at all, Not Very Well, Well, Very Well, and Extremely Well].
- Did you prompt COA GPT multiple times during Mission Analysis? If yes, how many times?
 - This question included a quantitative free response metric and was evaluated using Yes/No.
- How much did you trust the information COA GPT provided?
 - This question aimed to evaluate trust, and the following Likert rating scale was used: [Not at all/ Slightly/ Moderately/ Mostly/ Completely].

Preliminary Results

- COA-GPT can speed up much of the manual processing work that needs to be done during the MDMP execution.
- Initial results of our assessment:
 - COA-GPT can generate proper Warning Orders (WARNORDs) with the required content and format.
 - COA-GPT can find conflicting information in the provided Operations Order (OPORD) and other scenario documents.
 - COA-GPT can readily adapt to user feedback and update the generated products according to the user's preference.

Future Directions

As LLM capabilities mature, we will systematically examine trust dynamics across increasingly complex operational contexts. The next step is to complete a Second COA-GPT Assessment where military SMEs evaluate the quality of the MDMP products generated by COA-GPT. A Third COA-GPT Assessment is also planned focusing on generating the geospatial products of the MDMP. Finally, we intend to complete a COA-GPT Demonstration to allow personnel to experience and utilize COA-GPT as well as provide additional feedback. Our research will engage with Soldiers across various specializations, with initial focus on basic interaction patterns before progressing to more specialized operational scenarios.

FOR FURTHER INFORMATION:

U.S. ARMY COMBAT CAPABILITIES
DEVELOPMENT COMMAND
ARMY RESEARCH LABORATORY:
WWW.ARL.DEVCOM.ARMY.MIL

POINT OF CONTACT:

Samantha Berg
Samantha.k.berg2.ctr@army.mil
407-590-7701

