ARL-TR-10322 ● APR 2026

Technically Savvy Soldier (TSAVVY) FY23–FY25 Final Report (Summary Technical Report, Sep 2023–Sep 2025)

by Catherine Neubauer, Heather Roy, Leah Enders, Kimberly Pollard, Stephen Gordon, Bianca Dalangin, William Gerichs, Alex Stauff, Kevin W. King, and Amar Marathe

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

ARL-TR-10322 ● APR 2026

Technically Savvy Soldier (TSAVVY) FY23–FY25 Final Report (Summary Technical Report, Sep 2023–Sep 2025)

Catherine Neubauer, Heather Roy, Leah Enders, Kimberly Pollard, Kevin W. King, and Amar Marathe
DEVCOM Army Research Laboratory

Stephen Gordon, Bianca Dalangin, William Gerichs, and Alex Stauff
DCS Corp, Alexandria, VA

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
		START DATE	END DATE
April 2026	Summary Technical Report	September 2023	September 2025

4. TITLE AND SUBTITLE
Technically Savvy Soldier (TSAVVY) FY23–FY25 Final Report (Summary Technical Report, Sep 2023–Sep 2025)

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Catherine Neubauer, Heather Roy, Leah Enders, Kimberly Pollard, Stephen Gordon, Bianca Dalangin, William Gerichs, Alex Stauff, Kevin W. King, and Amar Marathe

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
DEVCOM Army Research Laboratory
ATTN: FCDD-RLA-FA
Aberdeen Proving Ground, MD 21005

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-10322

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

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

13. SUPPLEMENTARY NOTES
ORCIDs: Catherine Neubauer, 0000-0002-6686-3576; Heather Roy, 0000-0002-7843-2335; Leah Enders, 0000-0002-2205-2245; Kimberly Pollard, 0000-0002-5849-1987; Stephen Gordon, 0000-0003-4350-6899; Bianca Dalangin, 0000-0002-6084-2803; Kevin King, 0000-0003-4888-7491; Amar Marathe, 0000-0002-6253-6105;

14. ABSTRACT
As technology evolves at a rapid pace, it is now more imperative than ever that human skills evolve as well. Advanced, intelligent systems are becoming commonplace in industry, academia, and government sectors, requiring personnel to skillfully work with, guide, and adapt these systems. However, identification of such skills and specific skill-based training has not kept pace with rapidly evolving technology. To meet this demand, the Technically Savvy Soldier (TSAVVY) Program was established to identify specific skills and characteristics that may predict successful human–AI working proficiency, develop testbeds for assessing the degree to which these skills map onto desired performance outcomes, and empirically validate a predictive model of technological fluency (TF). We conceptually define TF as an individual's ability to effectively and creatively use technology, without formal training. The model aims to predict TF based on knowledge, skills, and behaviors (KSBs) and builds on early efforts to define TF and preliminary KSB predictors. Through literature reviews and subject matter expert interviews, researchers refined the model by encapsulating elements of technological proficiency—effectively using technology—and TF—the ability to creatively use, synthesize, or adapt emerging technology. Using cognitive science-based approaches and Future Operating Environment-oriented testbeds, we validated the model to identify KSBs that predict Soldier performance over time. Results indicate that KSBs that fall with the Information Management and Cognitive Ability competencies are critical facets of the TF model. Regarding teams, KSBs relating to interpersonal skills competency, including increased team trust and adaptive exchanges of leadership, are tied to greater performance in TF-related tasks. By gaining a better understanding of the factors contributing to an individual's likelihood of becoming TF, we can better identify training opportunities and TF standards across different military occupations to ensure Soldiers are prepared to adapt and excel within complex environments that use evolving technology.

15. SUBJECT TERMS
Humans in Complex Systems, technological fluency, technological adaptation, human-guided technological adaptation, technological literacy, Summary Technical Report, STR

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT	b. ABSTRACT	C. THIS PAGE		
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UU	66

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
Catherine Neubauer	(520) 691-5046

STANDARD FORM 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

Acknowledgments

The Technically Savvy Soldier (TSAVVY) Team acknowledges the following institutions who contributed to the efforts described in this report: George Washington University, The Institute for Human and Machine Cognition, Nova Southeastern University, The Army Research Institute, and DCS Corp.

Executive Summary

The Technically Savvy Soldier (TSAVVY) Research Program was a 3-year effort (FY23–FY25) supporting Army modernization priorities by defining, modeling, and empirically validating technological fluency (TF) as a critical Soldier competency for operating advanced, adaptive, AI-enabled systems. The program was designed to provide the Army with tools to accelerate Soldier adaptability, inform unit composition and talent management, and improve classification into Military Occupational Specialties (MOSs), and was executed across three phases.

Phase 1 (FY23) established the theoretical foundations of TF through subject matter expert input, detailed job analyses, and a large-scale literature review. Within this phase, a large focus was on developing a working definition of TF, which was later defined as "a competency wherein people's knowledge, skills, and behaviors (KSBs) enable them to guide and operate with novel learning-capable systems toward near-optimal performance, with little-to-no formal training."[1] This work identified a list of promising candidate TF KSBs and organized them into five higher-order categories: 1) Disposition and Motivation; 2) Cognitive Abilities; 3) Social and Teaming Skills; 4) Adaptability and Response to Change; and 5) AI-Relevant Knowledge and Experience. These results yielded a theory-driven TF model with initial tools for measuring predictors and performance criteria.

Phase 2 (FY24) transitioned this framework into an initial validation effort led by the Army Research Institute (ARI) who conducted a series of Soldier studies. These studies included Soldiers from diverse branches and MOSs to test the working TF model by asking Soldiers, who routinely work with a variety of technologies, to rank the importance of each KSB toward achieving not only technology proficiency (TP), (i.e., the ability to use a technology effectively) but also TF. Findings demonstrated that TP is technology-specific, whereas TF generalizes across systems. ARI refined the initial TF model into a higher-resolution framework highlighting seven key facets: 1) Information Management; 2) Dispositions; 3) Interpersonal Skills; 4) Critical Thinking; 5) Knowledge and Experience; 6) Cognitive Ability; and 7) Management.[2] This updated model provided clear targets for training-related improvements and baseline proficiency expectations across MOSs.

[1] Neubauer C et al. Measuring and predicting technical fluency: how knowledge, skills, abilities, and other behaviors can contribute to technological savviness. DEVCOM Army Research Laboratory (US); 2024. Report No.: ARL-TR-9884.

[2] Brown J et al. Preliminary evaluation of the technological fluency competency model. Army Research Institute for the Behavioral and Social Sciences (US); forthcoming 2026.

Phase 3 (FY25) focused on the empirical validation of the integrated TF model, led by the U.S. Army Combat Capabilities Development Command (DEVCOM) Army Research Laboratory (ARL). ARL developed multiple testbed platforms to evaluate participant behavior and performance with advanced AI-enabled systems under conditions where system capabilities were degraded, uncertain, dynamically changing, or artificially enhanced. These experiments were designed to capture both Soldier adaptability and the predictive power of KSBs for TF across complex, dynamic, and technologically driven environments. Within these environments, Soldiers are considered technologically fluent if they can 1) identify deficiencies in AI-based systems, 2) retrain them on the fly, and 3) evaluate system effectiveness. The program used an iterative model development approach in which theoretical models were progressively validated through ARI's Soldier studies and ARL's empirical testbed experimentation. Deliverables included a theory-driven TF framework, an empirically validated integrated model, and prototype assessments linking KSBs to TF performance outcomes. By the end of FY25, the program delivered a fully validated TF model, providing the Army with actionable insights for MOS classification, talent management, and Soldier adaptability to future technologies. Results will transition to ARI for integration into Soldier assessment pipelines. This report describes the program structure, summarizes technology development and validation activities across phases, and provides recommendations for Army decision makers and future research efforts.

1. Introduction

1.1 Mission Description

The U.S. Department of War Office of Strategic Capital has listed AI among the top critical technology areas targeted to help maintain United States national security and are pushing efforts to expand AI capabilities to increase operational effectiveness (OUSW[R&E] 2023). The U.S. Army seeks to take advantage of the recent surge in technological advancements, specifically in AI, to improve strategic capabilities in the field. AI has the potential to significantly decrease our adaptive and planned response time in offensive and defensive maneuvering and provide real-time information identifying surrounding threats, friendlies, obstacles, supplies, and navigation (to name a few). This technology is developing at a rapid pace worldwide with new applications daily, giving rise to the constant battle of keeping abreast of advancements and potential applications while maintaining a sufficient level of user expertise to outpace our adversaries. Within the military sector, and as emerging technologies reshape the modern battlefield, the U.S. Army faces a critical challenge: how to prepare Soldiers to work seamlessly with advanced systems while also adapting to rapid, often unpredictable, changes in technology. Thus, understanding the knowledge, skills, and behaviors (KSBs), or rather the unique combination of skills and characteristics that enable an individual to effectively leverage emerging technologies, can assist the Army in maintaining overmatch against United States adversaries.

The capability to work effectively with advanced, adaptive technologies requires more than prescribed, technical know-how. It demands a more complex, flexible, and adaptive ability to work with and guide that technology. It requires being both *technologically proficient (TP) and technologically fluent (TF)*. We define TF as the ability to use technology not only in proficient ways (i.e., TP), but also in innovative, perhaps unexpected ways and with little prior training (i.e., TF). It reflects an individual's ability to flexibly guide and adapt technology within complex environments. Central to this initiative is the concept of a Technologically Savvy Soldier (TSAVVY), a Soldier who demonstrates both foundational knowledge and the behavioral adaptability required to effectively interact and work with AI and emerging technologies. Of particular interest, is whether some individuals are better equipped for navigating this complex technological landscape, with the hope of understanding what the unique KSBs may be within those technologically savvy groups to instruct, operate, and adapt these technologies on the fly. Specifically, we seek to understand and identify the important facets that facilitate better performance with technology. With this

problem space in mind, work for FY25 focused on understanding whether models of individual and/or team TF can be developed that effectively predict such performance.

1.2 FY23 Work Defining KSBs that Contribute to TF

Acknowledging the unique problem space and motivations for success, work under this program unfolded in several phases. Beginning in FY23, the focus was on clearly outlining, understanding, and describing the construct of TF through subject matter expert (SME) input, detailed job analyses, and in-depth literature reviews. Our team worked collaboratively with academic partners and other government organizations to develop a working definition of TF, which was ultimately defined as "a competency wherein people's knowledge, skills, and behaviors (KSBs) enable them to guide and operate with novel learning-capable systems toward near-optimal performance, with little-to-no formal training" (Neubauer et al. 2024, pg. 1). This effort culminated in a large-scale literature review that laid the foundations for understanding TF definitions and other related terms from the research, as well as ways to measure this competency; however, no work had specifically looked at the term technologically fluency as it related to military-relevant contexts and Soldier skill enhancement. The literature review also documented existing KSBs within related fields such as human computer interaction, digital literacy, and computer literacy to name a few (Neubauer et al. 2024). Additionally, this work yielded a list of promising, yet preliminary, candidate KSBs that may contribute to TF. We further organized these KSBs into five higher-order categories:

1) **Category 1: Disposition and Motivation.** Disposition and motivation refer to relatively stable patterns of thoughts, feelings, drives, and tendencies that define an individual's unique character or viewpoint; however, they can also be influenced by situational factors (some more than others). Relevant KSBs that fall within this category include *extraversion, agreeableness, conscientiousness, neuroticism, openness to experience, stress tolerance, self-efficacy and self-confidence, motivation and (vocational) interest, and self-directed learning (SDL) and proactive personality*. Some studies suggest that certain dispositions or motivational traits may be predictive of TF. For example, a study by Oksanen et al. (2020) found that individuals high in openness were more likely to trust AI robots, which may relate to adaptability and willingness to explore new methods and ideas, aiding in the incorporation of AI and future technologies by fostering innovative approaches and embracing change. Further, a general willingness to work with, perform, or drive changes in technological usage and adaptation relates to motivation and (vocational) interest. According to Green (2005),

motivation to use Internet technologies appears to initiate as a developmental path starting first with extrinsic motivational factors (e.g., duty, and need), which then shifts into an intrinsic/extrinsic mixture of factors for usage (e.g., diversion, entertainment, and membership into a technological culture); however, intrinsic motivational factors may be a more prominent predictor of TF. Finally, SDL skills are also expected to be required when learning with technological systems (Azevedo et al. 2004), as users must search, vet, and integrate information from digital sources (Greene et al. 2014). Here, TF individuals need to learn technology and AI-related information and behaviors on the fly, be proactive, and not wait for someone to teach them.

2) **Category 2: Cognitive Abilities.** Cognitive abilities refer to the mental processes that individuals use to acquire, process, and apply information and may be crucial KSBs that enhance TF in individuals. Relevant KSBs that fall within this category include *general and (logical) reasoning, probabilistic thinking and pattern recognition, spatial skills, graph literacy, numeracy, general cognitive ability, cognitive biases, systems thinking and strategic thinking, working memory, theory of mind, learning efficiency, and metacognition*. In this context, several researchers have developed frameworks that examine how reasoning interacts with AI technologies including the Joint Cognitive Systems (JCS) framework (Woods and Hollnagel 2006). The JCS framework posits that the collaboration between both the human and AI systems is integrated and complementary where humans contribute reasoning capabilities and AI contributes to processing vast amounts of data, pattern recognition, and computation speed. Additionally, some work has demonstrated the efficacy of probabilistic reasoning (i.e., a skill that allows an individual to navigate uncertainty by using probabilities and identifying trends in regular or irregular data) in optimizing resource allocation strategies, and interpretation of AI-generated insights (Silverman et al. 2019). Others have underscored the importance of accurate probabilistic interpretations in bolstering the reliability of AI systems, particularly in critical domains such as medical risk (Fagerlin et al. 2007) and financial forecasting (Chen and Zhu 2020).
3) **Category 3: Social and Teaming Skills.** As AI and advanced technologies continue to permeate industries and organizations, people increasingly find themselves operating in mixed human–autonomy teams, where social skills vital to effective functioning in human–human teams are likely to continue to be valuable in human–machine teams (Lakhmani et al. 2022). KSBs we placed in this category include *cultural awareness, social skills and teamwork, leadership and project management, and effective*

communication. These types of skills generally facilitate effective collaboration and allow TF individuals to collaborate with others. Additionally, Techataweewan and Prasertsin (2018) specifically called out collaboration skills as one of four other factors that are vital for enhancing digital and TF. These authors further outlined that collaboration skills within digital domains consist of using digital technologies in collaboration with others, either as the leader or a member of a team, by working together to achieve team goals.

4) **Category 4: Adaptability and Response to Change.** Technology is advancing at a rapid and accelerating pace, with some advanced forms of AI even capable of updating their models and changing their behavior in real time. This rapid change and unpredictability put tremendous pressure on the human to be flexible in their thinking and strategies, to adapt to new conditions, and to have enough comfort with change and uncertainty to remain focused and effective in their tasks. A person's ability to adapt not just themselves but also to adapt elements of the situation, including the AI, will be vital. Under this category, we place the KSBs of *situational (general) adaptability, employee agility, flexible thinking/cognitive flexibility, and comfort with uncertainty versus need for cognitive closure*. For example, individuals who are flexible in their ways of thinking (i.e., those with a propensity to adapt to new situations with less resistance [Barak and Levenberg 2016]) may be able to adapt to, as well as drive, technological adaptation, and can effectively use new technologies faster than those with more rigid thinking (Barak and Levenberg 2016). This KSB may allow individuals to adapt, re-strategize, or restructure their plan of action to complete a task goal when faced with novel or unexpected situations, both of which may be readily apparent when working with new or evolving technology.
5) **Category 5: AI-Relevant Knowledge and Experience.** In the context of AI, direct experience with AI or knowledge of it can be important in ensuring that individuals use AI technologies effectively and accurately. For example, individuals who have expertise in data analysis can ensure that the data input into the AI system is accurate and reliable. Similarly, individuals who have expertise in the field in which the AI system is being used can ensure that the output generated by the system is relevant and useful. KSBs that fall within this category include *knowledge, general understanding of AI, computational thinking, digital literacy, beliefs about AI, algorithmic thinking, AI literacy, propensity to trust in technology and trust calibration, and video gaming experience*. Someone with these KSBs would have a general knowledge of how AI works, what the processes are, and how to

best use those capabilities. Proficiency enables them to use AI correctly, as intended, and in the right situations. Conversely, a lack of understanding can result in unintentional misuse, whereas high proficiency may lead to creative and beneficial unintended uses and ensuring that users are able to identify where novel technologies break down. Moreover, beliefs about AI are linked to proficiency in technological tasks because individuals who believe in AI's superiority are more motivated to adopt and effectively use AI-driven tools and advice (Von Walter et al. 2021). Moreover, AI literacy and computational thinking exhibit a positive association, wherein computational thinking facilitates the understanding, recognition, and evaluation of AI-based technologies (Celik 2023). This synergy between computational thinking and AI literacy underscores the interconnectedness of cognitive skills and AI proficiency.

This program also provided subsequent tools for measuring each KSB, as well as providing a theory-driven TF model aimed at capturing predictors and performance criteria.

1.3 FY24 Empirical Evaluation of KSBs and Development of TF Competency Model

1.3.1 FY24 Overview and Description of Human-Guided AI Experimental Testbed Development

One of the primary stage gates for this program was to develop a set of experimental testbeds capable of eliciting human-guided, or rather human-driven, AI adaptation behaviors and performance markers based on how well humans trained and worked with an AI. We were tasked with designing these testbeds to be conceptually in line with how we characterize a "TSAVVY" Soldier, which is 1) someone who can detect deficiencies in advanced systems, 2) retrain or adapt them on the fly, and 3) evaluate them once trained. This is also in line with our definition of someone who is TF. Several testbeds capable of eliciting these behaviors and interactions were developed by ARL and DCS Corporation, with multiple studies executed (Institutional Review Board [IRB] nos. ARL 23-054; ARL 24-098; ARL 24-135) and one proposed but halted due to loss of funding (ARL 24-117). With these testbeds, we were able to incorporate into the design several crucial capabilities that included the ability to gather large amounts of data on general population technology users and delivering a task that assessed how a human–AI team could solve a task together while eliciting behaviors and performance markers critical to guiding "adaptive technology." The parameters of the tasks and AI itself could also be changed, with tasks that required human intervention and adaptable/flexible

skillsets. Progress on the platform development allowed our team to begin targeting increased Soldier-centered tasks to explore a wide range of capabilities and to tap into real-world environments and problems Soldiers of the future will face. Finally, the design of our last platform (e.g., DYNAMO) allowed us to move into the teaming space, where multiple humans could work together as with an AI agent. The development of these platforms also allowed us to empirically test and refine our candidate list of KSBs as well as refine and validate the TF competency model (described in further detail in Section 2). Ultimately, outcomes from this work will be delivered to ARI to develop assessment and training procedures for TF. The following sections give a brief overview of each of the testbeds developed.

1.3.2 Description of ARL Testbeds

1) **Heterogenous Teaming Puzzle Experimental Testbed:** The first testbed developed under this effort took the form of a falling blocks puzzle game in which the human participant played the game under various conditions including those with access to a pretrained or user-trained AI to optimize their performance. A within-subjects design was used, where participants were tasked with completing six game conditions playing a variation of the game Tetris including: 1) horizontal play baseline (Figure 1), 2) user-trained AI horizontal condition, 3) vertical play baseline condition, 4) user-trained vertical AI condition, 5) user-trained vertical AI condition (starting with an AI trained for horizontal gameplay), and 6) user-trained vertical condition with basic strategies that were given in the instructions to optimize performance. In this condition, recommended strategies included clearing multiple columns at the same time and strategically maintaining and building space to drop the blocks. The order of conditions four and five were counterbalanced across participants. In the baseline conditions, the AI was turned off and the human played as they would if they played a typical video game. In the user-trained conditions, participants had to train an AI using reinforcement learning to optimize performance in the game. We were able to utilize the online data collection platform Prolific and obtained data from a total of 319 participants including 84 females (mean [M] = 36.13, standard deviation [SD] = 9.97), 231 males (M = 36.67, SD = 9.39), and 4 of unreported age and gender. For a full description of the study procedures see Dalangin et al. (2024).

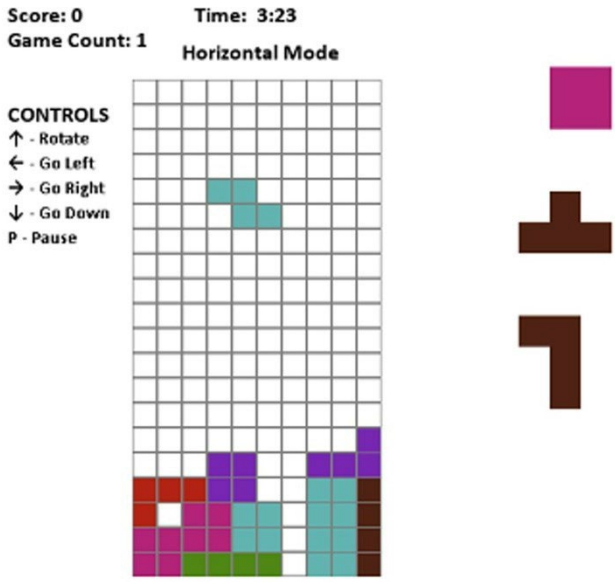

Figure 1. Screenshot of one of the horizontal gameplay baseline conditions. Here, participants are able to change the direction of the falling blocks using keyboard keys and can see the next three blocks.

Findings indicated that increased preference for the user-trained AI was associated with improvements in trust in the AI and increased satisfaction with the AI, but not performance (Dalangin et al. 2024). These results suggest psychological ownership, the degree users perceive an object as "mine," is an important factor for increasing positive interactions with AI systems and tools. We further evaluated the predictive value of various KSBs on the participants' performance with training their AI. This was achieved by examining the game scores of each participants' trained AI models, scored post hoc in a sandbox. The KSBs of interest were selected based on ARL's previous literature review (Neubauer et al. 2024). The main findings of the KSB analysis were that 1) individual users did differ notably in the speed at which they were able to train their AI to strong performance and separated into four clusters that varied in their speed to proficiency, 2) KSBs significantly predicted cluster membership, 3) KSBs significantly predicted final scores *within* speed-to-proficiency clusters, and 4) KSBs significantly predicted final scores in general. Key predictive KSBs that emerged included dispositional traits such as conscientiousness (a subscale of the Big 5) and comfort with training AI. The state-level KSB of intrinsic motivation also predicted speed-to-proficiency cluster assignment. Within clusters, KSBs such as anxiety, depression, age, and recent usage of AI predicted performance, but only with certain clusters. Across all clusters combined, AI literacy subscales and comfort with training AI emerged as the only significant predictors. These results suggest a nuanced picture for the relationship of KSBs to TF. Comparing these findings to the

categories in the ARI TF model, we see support for both Knowledge and Experience and Disposition and Motivation categories as influencing TF, as demonstrated in the heterogenous teaming puzzle task. A journal article with comprehensive details on these results will be forthcoming (Neubauer et al. 2026).

2) **AI for Dynamic Pathfinding in Threatening Scenarios (ADAPTS) Experimental Testbed:** The second testbed, ADAPTS (ARL 24-098; N = 140), tasked participants with navigating a simulated urban environment, while avoiding threats and completing objectives (i.e., collecting resources and avoiding tripwires), under time pressure to make it to a set extraction point. Participants were advised to gain the most points as possible by collecting resources scattered around the environment and by avoiding tripwires (which would detract points), while also making it to the set extraction point in time. Participants were granted access to either a pretrained or self-trained AI tool to aid navigation of the optimal route through the environment.

   To navigate the environment, participants were provided with a mini map and compass bar (Figure 2). They could also see their score and available time remaining displayed on the screen at all times. Participants could elect when to activate the AI for a cost of two points. Once activated, the AI remained active throughout the run. The AI then provided an optimal route to the participant, appearing as a green route running throughout the environment. They could choose to adhere to or deviate from this route as much as they desired. AI usage in ADAPTS was calculated as the percentage of time during the task that the user engaged an AI guide for navigation assistance. Participants completed two runs of each of these conditions (i.e., self-trained vs. pretrained). For the purposes of the current report, we only consider the pretrained AI conditions as these map more readily to the third platform's study paradigm.

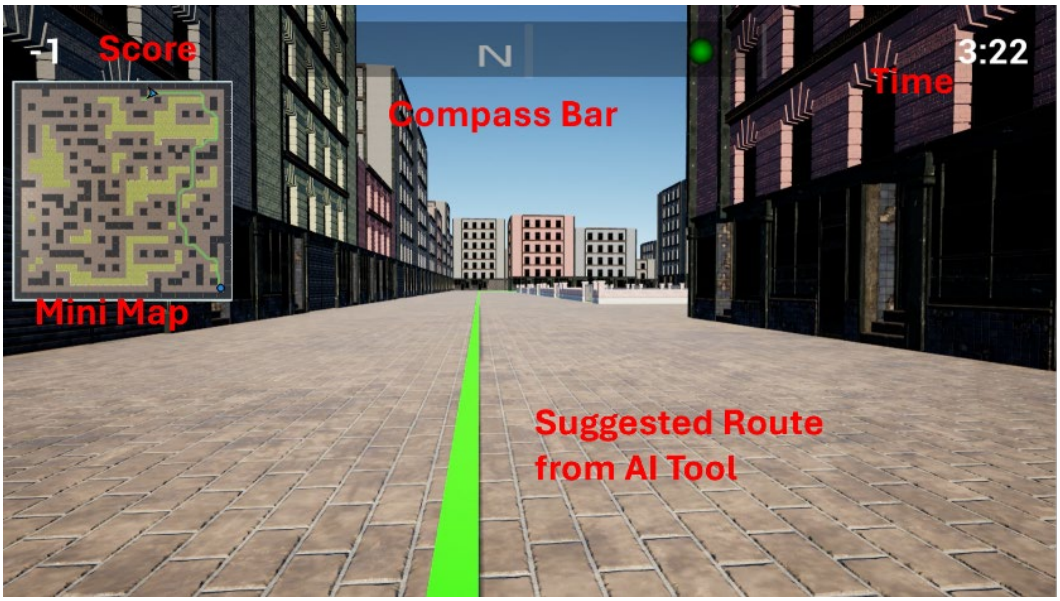

Figure 2. ADAPTS virtual environment and AI capability.

3) **Dyadic Navigation with AI Systems for Mission Operation (DYNAMO) Experimental Testbed:** DYNAMO was specifically designed as a follow-on study to scale up an ADAPTS-like paradigm using AI navigational aids from an individual to a teaming (dyadic) study (Figure 3). While the ADAPTS study examined individual participants, the Army is inherently a teaming organization, where bringing more people into an operation can quickly influence the direction and approach to missions. Leadership and team management has further been identified in ARI's TF model as a factor related to TF. Therefore, it is critical to understand the influence of human teammates not only on our own behavior, but also how teammates can influence our interactions with AI and autonomy. Toward this effort, the DYNAMO testbed was created to investigate how teaming and emerging leadership behaviors inform TF individuals, performance, and group cohesion.

Figure 3. DYNAMO virtual teaming environment with AI capability.

4) **Landing AI-Navigated Drones in Environments of Risk (LANDER):** Due to funding cuts, we could not complete data collection for the fourth testbed: LANDER. We developed the LANDER testbed to explore TF performance in user interaction with an AI-enabled unmanned aerial system (UAS). Participants would train the AI on how to land a UAS safely and would do so under a series of challenging and unexpected conditions. The participant's behavior during the trials would be recorded, as would their success in landing the UAS (Figure 4). Their trained AIs would later be evaluated in sandbox conditions to measure how well the participants trained the AIs.

Figure 4. LANDER testbed environment with AI-drone landing capability.

In addition to measuring KSBs from the participants, we would also explicitly probe the three parts of our conceptualization of TF (see Section 1.3.1): 1) identify deficiencies in AI-based systems, 2) retrain them on the fly, and 3) evaluate if those systems are working effectively. The LANDER software was transitioned to ARI for direct future use in Soldier training and for ARI to use to further their research on TP, TF, MOS designations, and trainable KSBs.

1.3.3 Results in Context with Other ARL Testbeds and Studies

Additional collaborative work with academic partners resulted in the development of two additional TSAVVY testbeds as follows.

1) ARL collaboration with external partners at the Institute for Human Machine Cognition (IHMC) and Florida State University led to the development of a simulated robot testbed in which the robot's human teammate must develop an understanding of the robot's rules and functions to successfully complete puzzles with different hazards and challenges. Recently completed data collection by our partners was used to compare performance on this testbed versus a battery of KSBs suggested and co-developed by ARL, with a results publication forthcoming. The testbed was built within specifically designed experimental environments and specifications. These environments each included multiple empirical outcome variables that measure performance on tasks that require TF.

Specifically, the testbed can record data on when and how a human chooses to interact with AI (i.e., behavior variables) as well as the performance of the AI. We used performance variables as empirical indicators of TF. For example, such variables included points earned in a game, time taken to complete a task, and AI usage. Additionally, a battery of subjective questionnaires was administered to measure and assess specific KSBs of interest.

2) ARL collaboration with external partners at Nova Southeastern University (NSU) led to the development of a competency and skills assessment "Cyber Range" testbed. During Phase 1 of this effort, 20 SMEs validated a set of TF and cybersecurity competencies that included 31 Knowledge units, 5 Skills, and 9 Tasks (KSTs). The validated KSTs then were assessed on a pilot group of 26 freshman students that had little or no experience within the Cyber Range platform. The 26 pilot students were then provided with training on Generative Artificial Intelligence (GenAI) prompt development techniques that included both general guidance and specific prompt developments for TF and cybersecurity purposes. Phase 2 involved having the students return again to perform the same set of Cyber Range assessments, now with the use of GenAI as their "assistant." Initial results, from participants who had little to no knowledge with GenAI, demonstrated a significant increase in their overall TF competency assessment (i.e., 33.12% without GenAI, and 50.86% with GenAI, $p < 0.001$). Two publications are in progress, with an additional invited keynote address that was delivered in FY25 (Neubauer 2025).

1.3.4 FY24 ARI Competency Model of TP and TF

In FY24, ARL partners at the Army Research Institute (ARI) used the initial theoretical results to lead their own model validation effort across diverse branches and Military Occupational Specialties (MOSs), helping establish baseline proficiency expectations. Here, models progressed from initially describing factors related to innate TF to later models that identified targets for training-related improvements in TF. Preliminary studies done by ARI partners tested the working model by asking Soldiers, who work with a variety of different technologies, a series of questions to determine how Soldiers ranked each KSB competency for importance toward becoming not just TP but TF with that specific technology (Brown et al. 2026). ARI found that while TP was technology specific, TF was not dependent on the specific technology. Within this new model, ARI highlighted seven key facets that appear to be important predictors of TF: *Information Management, Dispositions, Interpersonal Skills, Critical Thinking, Knowledge and Experience, Cognitive Ability*, and *Management* (Figure 5). Given programmatic

needs and deliverables, these findings helped provide an updated TF model that could be validated with ARL's empirical data collections to help identify KSB factors that may contribute to both aspects of this model.

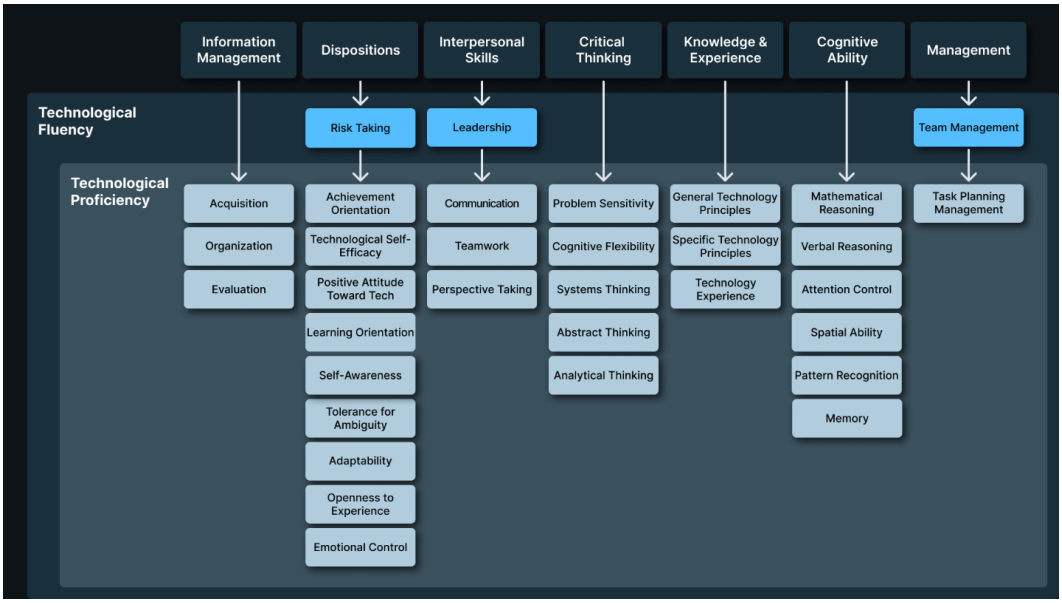

Figure 5. ARI competency model of TF and TP (Brown et al. 2026).

1.4 FY25 Plan to Empirically Validate ARI's Competency Model

In FY25, we primarily focused on the empirical validation of the integrated TP/TF model (Figure 5), given to us by ARI. To achieve this, we used the development of multiple testbeds (described in Section 1.3.2) to identify which KSBs were related to participants' behavior and performance with advanced, adaptive, AI-enabled technologies. The KSBs of interest were selected based on ARL's previous literature review (Neubauer et al. 2024) and task-relevant skills required. Each experimental paradigm provided participants with the opportunity to engage with AI to optimize their task performance. Conditions included system capabilities that were either degraded, uncertain, changed in unanticipated ways, or were artificially enhanced.

ARL launched several large-scale online experimental efforts using these testbed platforms to reach a large cohort of participants. Each experimental effort empirically evaluated a user's performance with AI technology and captured KSBs with the aim to assess the TF model's effectiveness in complex, military-relevant, and evolving scenarios. Each experimental testbed included elements of working with an AI component, but varied in several aspects (e.g., degree of engagement with AI, user feedback mechanisms). Even though TF is thought to be technologically agnostic, characterizing interaction differences and mechanisms of

engagement between the AI technology and user are important considerations when interpreting results. Therefore, testbeds were also characterized by experimenter ratings (from the point of view of the end user) of each technology's unique properties using a set of User Engagement Characteristics (UECs) (Appendix A). These UECs provided a greater context to describe the validation of the TF model as it relates to the specific type of technologies. Additionally, for the purposes of this report and requirements for the model validation analysis effort (discussed in further detail below), we focused specifically on the ADAPTS and DYNAMO testbed data as these two studies provided enough conceptual and data-centric overlap to compare TF-relevant KSBs within the competency model.

Ultimately, the goal for this work is to identify and evaluate methods for measuring an individual's tendency toward TF. Subsequently, this work will be transitioned to ARL's partner, ARI, for whom the primary objective is to improve classification of Soldiers into MOSs using TF-relevant assessments.

1.5 Model Validation Justification, Aims, and Hypothesis

While the preliminary TF model and SME survey data collected by ARI provide a framework of expectations for which KSBs are predictive of TP and TF, these findings are subjective to the rater and their estimation of proficient or fluent performance with specific technologies that the SMEs chose to rate. This first (nonempirical) validation of the TF model, performed by ARI and colleagues, used SME interviews and surveys to draw connections between candidate KSBs, TF, and UECs; however, the model must be empirically tested to determine whether the proposed KSBs predict TF outcomes in real, AI-enabled task contexts. An ideal test of this model would involve multiple different testbed platforms and associated tasks to provide a best estimate of predictive power. Unbiased empirical evidence of how these KSBs predict performance, especially in novel tasks, could be beneficial in testing how well this TF model would operate for the wide range of existing and potential future AI-enabled and adaptive technologies in the Army. Collectively examining outcomes from multiple test beds (instead of a single test bed or multiple independently) enables us to examine which KSBs predict TF more broadly.

Therefore, the aim of the current work is to *empirically* test the TF model's predictions by examining data from multiple testbeds collectively. This allows us to assess performance with complex, adaptive technologies across multiple domains. KSBs measured from all participants serve as independent variables. Statistical analysis will reveal what aspects of the model (e.g., which KSBs) receive empirical support as important features in a TF model and are therefore worth

pursuing for future study, MOS assignment, or other team assignment or training purposes. Future goals for this collaborative effort between ARL and ARI include using the TF model to develop and validate assessment batteries that can predict Soldiers' TF and evaluate Soldiers on TF, while using that information to feed into future classification algorithms and/or decisions. Section 3 outlines the specific model validation methodology that was used for this effort.

2. Methods

2.1 Model Validation Study Participants

The ADAPTS (ARL 24-098) and DYNAMO (ARL 24-135) studies, described in detail above, were conducted online via Prolific and carried out in accordance with ARL's Human Research Protection Program (HRPP) and conducted in compliance with the ARL HRPP and the Declaration of Helsinki. The protocols were approved by the ARL HRPP. All participants reviewed and acknowledged an information sheet detailing each study prior to volunteering. Participants were compensated for their participation following Prolific guidelines at an approximately $12 per hour rate.

2.1.1 ADAPTS Participants and Procedures

In the ADAPTS (ARL 24-098) study (N = 140), participants were recruited through Prolific, an online research platform. Participants were fluent in English and 18 years of age or older. The participant sample included 45 females with an average (M) age of 35.38 years (SD = 11.47, range: 18–58 years), 93 males (M = 31.06, SD = 10.17, range: 18–73 years), and 2 nonreported (M = 28, SD = 0). As the task was visual in nature, participants had normal or corrected-to-normal vision without colorblindness. Since this study required normal visual processing, participants also did not suffer any significant brain trauma or injuries in the last 3 months as that could cause cognitive or visual processing issues. Participants completed the experiment across two sessions with an expected total duration of approximately 2 h (experimental session) and 25 min (pre-screener session). Following Prolific recommendations, participants were compensated at a $12.00/h rate. Participants were compensated based on the expected total time rather than the maximum time allowed by Prolific (double the expected time). Participants were compensated $5.00 for completing the pre-screener and $24.00 for completing the experimental session. Data from 121 participants who completed navigation tasks in both conditions *and* trained their AI tool were used for the reported analysis.

2.1.2 DYNAMO Participants and Procedures

In the DYNAMO (ARL 24-135) study (N = 40 dyads), participants were tasked with navigating a virtual environment with a partner while disarming as many bombs as possible to secure the area and accrue points and avoid enemy actors (i.e., losing points). Disarming bombs required participants to coordinate key presses together. Participants were provided with a mini map, chatlog, and a simple communication wheel (Figure 3) to coordinate with their partner. Both participants had the capability to call a pretrained AI tool that generated an optimal route that identified bombs along the way to defuse and avoided enemy actors. Like ADAPTS, once called, the AI would display as a green navigational route. However, the AI would provide the next segment of the route rather than the whole route through the environment. The team then needed to decide when and if to continue to use the AI. Unlike ADAPTS, there was no cost to calling the AI. AI usage in DYNAMO was calculated as the number of times the AI was called upon during the task. Further discussion of dispositional and trait-level KSBs collected in these two studies is in Section 2.2.1. A*ssessing the KSBs across both studies allowed a large amount of crossover in not only game dynamics but in the assessments included in the studies, which gave us a more robust breadth to compare the TF competency model from the ADAPTS and DYNAMO datasets.*

2.2 Primary Analysis: Tech Proficiency Model

2.2.1 Dispositional and Trait-Level KSBs Collected via Questionnaires

In addition to task performance, the ADAPTS and DYNAMO studies each included the collection of multiple subjective assessments designed to target specific KSBs of individuals, which are included within the TF model. While these studies are independent from one another—each with independent objectives and task requirements—each study included the collection of individual self-assessment questionnaires to capture subjective ratings of KSBs, and each battery had an element of human-guided AI training. With DYNAMO following ADAPTS, we were able to include common measures, allowing us to execute this model validation pipeline across these two studies. Figure 6 provides an overview of the specific KSBs assessed in both ADAPTS and DYNAMO, as well as which KSBs were assessed in both. The second portion of the figure specifically illustrates which of the KSBs we assessed fit into the ARI competency model categories. See Figure 5 and Brown et al. (2026) for further description of categories and competencies.

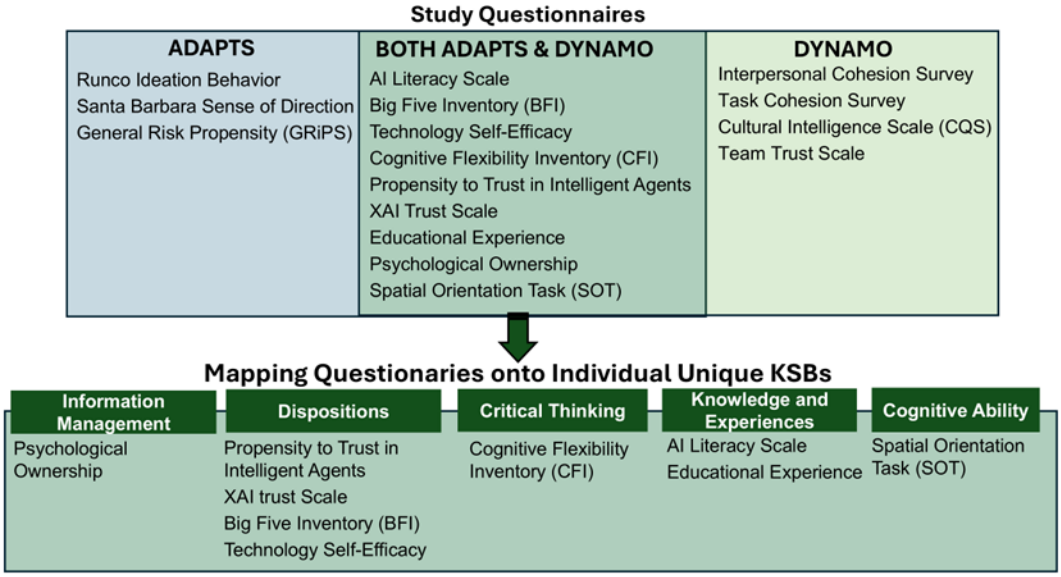

Figure 6. ADAPTS and DYNAMO study questionnaires used to assess KSBs of interest.

2.2.2 Trait-Level KSB Collected via a Test

In addition to self-report questionnaires, an aptitude test was also administered to participants. Due to the importance of spatial navigation in both testbeds (ADAPTS and DYNAMO), spatial ability was considered a relevant KSB that could help predict performance. Participants were therefore asked to complete a Spatial Orientation Task (SOT; Friedman et al. 2020) to measure their spatial skills. The SOT tests a participant's ability to imagine different perspectives or orientations in space. Participants were presented with a picture of an array of objects with an arrow circle (Figure 7) and were instructed to imagine they were standing at one object in the array and facing another. Their task was to then draw a line showing the direction to a third object from that perspective.

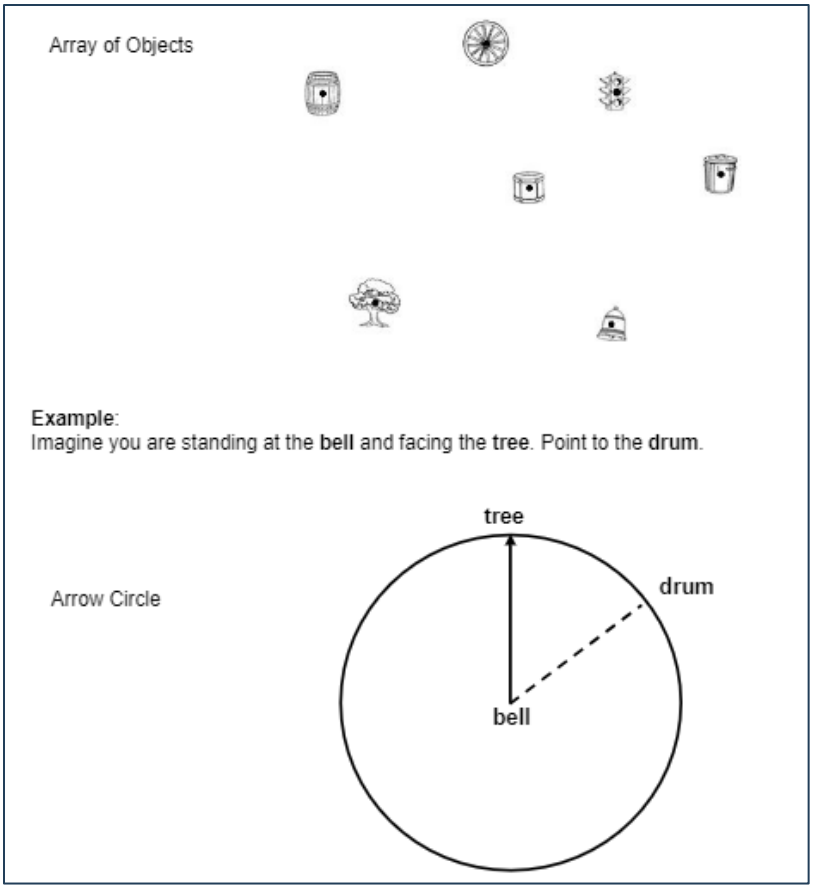

Figure 7. SOT example.

2.2.3 TF Model Validation Plan: Predictor Variables and Performance Outcome

Our primary model validation analysis combined similar variables from the ADAPTS and DYNAMO datasets. To empirically test the TF model, we first measured a series of trait-level KSBs (e.g., dispositional traits, life experiences, and so on; see Figure 6) from each participant using a battery of questionnaires and the SOT test (Figure 7). Descriptions of these measures and citations are provided in Appendix B.

As an initial step to validate the model, we first identified common measures collected using the ADAPTS and DYNAMO testbeds (Figure 6) and then aligned them according to the KSB categories in the ARI TF model. Of the 15 questionnaires included across the studies, there were 8 questionaries and 1 aptitude test in common between both testbeds. Some questionaries only had one global outcome score (e.g., Explainable AI [XAI] Trust Scale), whereas others (e.g., Big Five Inventory [BFI], Trust in AI and Satisfaction, and AI Literacy Scale) had multiple dimensions and thus, multiple outcome scores to include in the model.

SOT score (one variable) was also calculated as the average angular error, in degrees, between the estimated and actual orientation. Variables included in the analysis from both datasets included one score from the XAI Trust Scale, five scores from the BFI, one score from the Psychological Ownership Scale, one score from the Propensity to Trust Intelligent Agents Scale, two scores from the Technology Self-Efficacy Scales, three scores from the Cognitive Flexibility Scale, eight scores from the AI Literacy Scale, and one score from the SOT. Additionally, each of these studies included collection of basic demographic variables that may impact a participant's TF. At this time, only educational level (one variable) was included. Finally, because each of the studies involved incorporating AI, we wanted to capture participants' propensity to engage with the AI element, or AI usage. The ADAPTS and DYNAMO tasks were different in terms of how participants interreacted with the AI element. In ADAPTS, participants could only call the AI once, but once called, it would remain functional for the remainder of that run. Therefore, for ADAPTS, AI usage was defined as the percentage of the trial that they had the AI assisting them navigate the environment. For the DYNAMO study, AI usage was calculated as the number of times the team called the AI for assistance. To allow comparison across the two studies, AI usage was z-scored across each study independently and added as a predictor in the TF model. In all, 24 predictors were included in the empirical TF model.

Overall performance was included as the dependent variable (Figure 8). Scoring for ADAPTS and DYNAMO performance was similar in that both consider a positive (e.g., points accrued) and negative effect (e.g., points lost). For the ADAPTS study, performance score was calculated as the total number of resources gathered minus the number of tripwires set off. For the DYNAMO study, performance score was calculated as the total number of bombs deactivated minus the total number of enemy hits on the team. Given the large range in enemy hits accrued by teams in DYNAMO, the number of hits was capped at 50. DYNAMO performance scores were averaged across the team to create a single performance score for each dyad for each run. Both ADAPTS and DYNAMO performance scores were averaged across the two runs to create a single score for each participant (ADAPTS) or dyad (DYNAMO). Performance scores were z-scored for each study independently prior to including in the model.

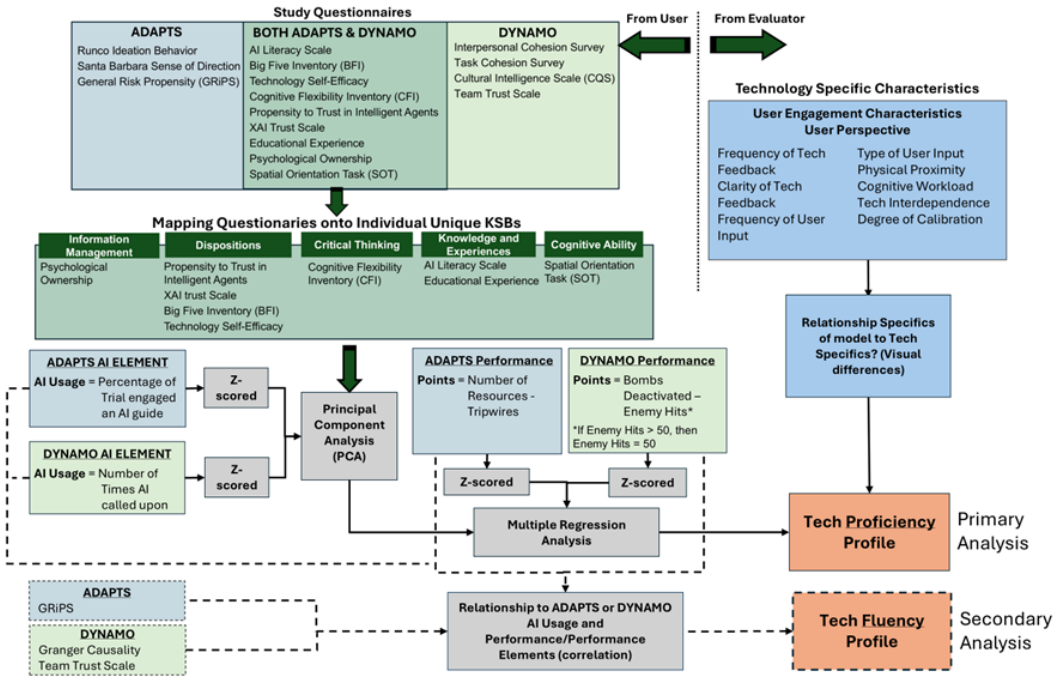

Figure 8. TF model validation pipeline. Primary analysis (solid arrows) investigated how KSBs from ADAPTS and DYNAMO corresponded to performance and were indicative of TF. Secondary analysis (dashed arrows) examined additional predictors of ADAPTS and DYNAMO and how these were related to performance on these testbeds individually and predictors of TF. UECs were considered to provide context.

2.2.4 TF Model Validation: Data Analysis

Prior to combining the datasets along these variables, both datasets were examined independently for outliers and data completeness. Initially, ADAPTS contained 140 observations and DYNAMO 40 teams. Outliers at or greater than 3 standard deviations from the mean were removed, resulting in 34 observations being removed from ADAPTS (1 for performance and 33 for AI usage). No outliers were present in DYNAMO. In ADAPTS, if participants had multiple runs (more than the expected two), only two full runs were kept. This typically occurred if a participant experienced technical difficulty during the first run and had to restart, resulting in a partial first run and two full subsequent runs. Within DYNAMO, if a participant appeared in multiple teams, only complete runs were kept. Like ADAPTS, this typically occurred when there were technical difficulties and the teams needed to be rematched to run again. The final combined dataset included 137 observations.

For the 24 predictor variables, principal component analysis (PCA) was applied to reduce dimensionality of the combined ADAPTS and DYNAMO dataset and identify the primary patterns of variation. Multiple regression analysis examined the relationship between the most encompassing principal components identified by the PCA and performance score (z-scored). Primary (and secondary, detailed

below) analysis was performed using R (R Core Team 2025) and a p-value <0.05 was used as a threshold for significance for all analysis.

2.3 Secondary Analysis of TF Model

2.3.1 Investigating General Risk Propensity and Team Trust

The ARI competency model of TF included factors relating to team trust and riskiness to predict TF. Although ADAPTS and DYNAMO did not overlap in their ability to include questionnaires capturing these KSBs, ADAPTS did collect questionnaire information related to riskiness with the General Risk Propensity Score (GRiPS) questionnaire, while DYNAMO collected questionnaire information related to trust with the Team Trust Scale (see Figure 6). Therefore, a secondary analysis using the ADAPTS dataset looked at how GRiPs correlated to performance, AI usage, time left (i.e., amount of time left when crossing threshold), number of resources gathered, and number of tripwires activated in the ADAPTS task. Another secondary analysis using the DNAMO dataset captured how overall team trust and components of trust, perceived team trust, and cooperation team trust were correlated with performance, AI usage, distance traveled (i.e., averaged between teammates), number of communications (i.e., communications summed between teammates), points (i.e., number of bombs disarmed), and hits (i.e., number of enemy hits on team). Spearman Rank Order correlation methods were used in this secondary analysis because of violation of normality on performance variables for both ADAPTS and DYNAMO.

2.3.2 Evaluating the Impact of Team Dynamics on DYNAMO Performance

To objectively quantify the emergent team dynamics within the DYNAMO testbed, a Granger causality analysis was applied to participant movement data. This technique, previously termed "Granger Leadership" (King et al. 2023), uses a pairwise Granger causality F-test on the X,Y,Z velocity vectors of each dyad to determine if one participant's movement history is predictive of their partner's. A p-value-based classification method was used, designating a participant as the leader if their corresponding p-value was less than 0.05, and also lower than their partner's p-value. This process yielded a continuous, moment-by-moment time series of the leadership state (i.e., Leader, Follower, or No Leader) for each dyad, as illustrated in Figure 9.

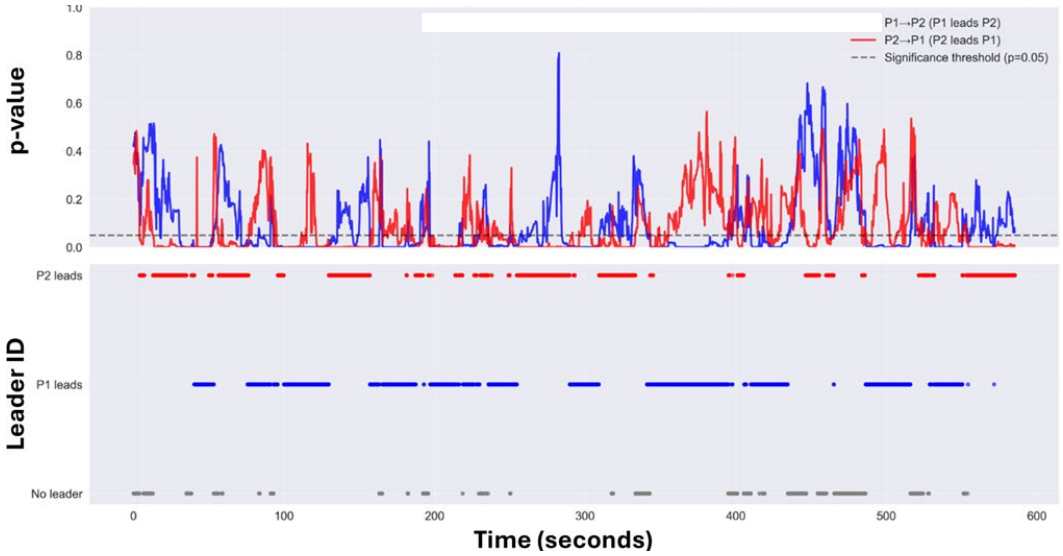

Figure 9. Example of leadership detection over time in DYNAMO from a single dyad.

From this high-resolution time series, a suite of summary metrics was derived to characterize the overall interaction patterns for each dyad in DYNAMO. These metrics included: the overall role designation of Leader versus Follower based on total time leading; a Leadership Dominance Index to quantify imbalance; Switch Rate, Initiation Rate, and Reciprocal Switch Count to measure leadership stability; and the Average Leadership Bout Duration to measure persistence.

A final methodological step involved creating role-centric KSB variables for subsequent modeling. Individual participant KSBs were mapped from their arbitrary P1/P2 designations to their empirically determined Leader or Follower roles. This transformation produced new dyad-level variables (e.g., Leader Age, Follower Age, and Player Age) that enabled a direct analysis of how the traits of the person filling a specific role related to team dynamics and performance. The next section discusses potential differences related to different types of technologies (i.e., User Engagement Characteristics), how these features can be measured, and whether they predict variation in TF.

2.4 UECs

It is important to acknowledge the vast spectrum of technologies that Soldiers may interact with. All available technologies (e.g., AI, autonomous systems, weapons, sensors, and so on) cannot be characterized by a single description. While ARI found that TF was not dependent on the specific technology used, they found TP was technology specific. It is thus necessary to acknowledge the UECs defining technological systems to provide context for interpreting a TF model. UECs capture specific characteristics of the available technology including aspects such as the

frequency of the technological feedback, types of user input, and how clear the technology provides feedback to the user. ARI developed a series of UECs, a full list and description of which can be found in Table 1. Scoring information for the UECs can be found in Appendix A. Additionally, analysis of the UEC data can be found in section 3.4.

Table 1. UECs and definitions.

UEC	Definition
Frequency of technology feedback	The degree to which the technology provides feedback to the user
Clarity of technology feedback	Feedback provided to the user is explicit/clearly defined vs. implicit/ambiguous
Frequency of user input	How often the user must provide input to the technology
Type of user input	How input is communicated to the technology (e.g., number of inputs, sensitivity)
Physical proximity	Location of the technology relative to user (near vs. far), how difficult it is to call it, or how close it is in controlling the test
Cognitive workload	Degree of mental resources required of a person at any one time to interact with the technology
Technological interdependence	Degree to which the technology is a standalone element vs. part of an integrated system of technologies
Degree of collaboration	Degree to which the technology can be operated by a single individual vs. requires team input/coordination

3. Results

3.1 Primary Analysis: TP Model

Principal component analysis on the 24 variables was performed to reduce dimensionality in the dataset and identify the key KSBs that contributed to the highest degree of variance within the KSB dataset. The PCA Scree Plot (Figure 10) was examined and the five initial principal components located prior to the flattening of the curve, the elbow, were retained. In total, these five components explained 61.24% of the total variance (Table 2).

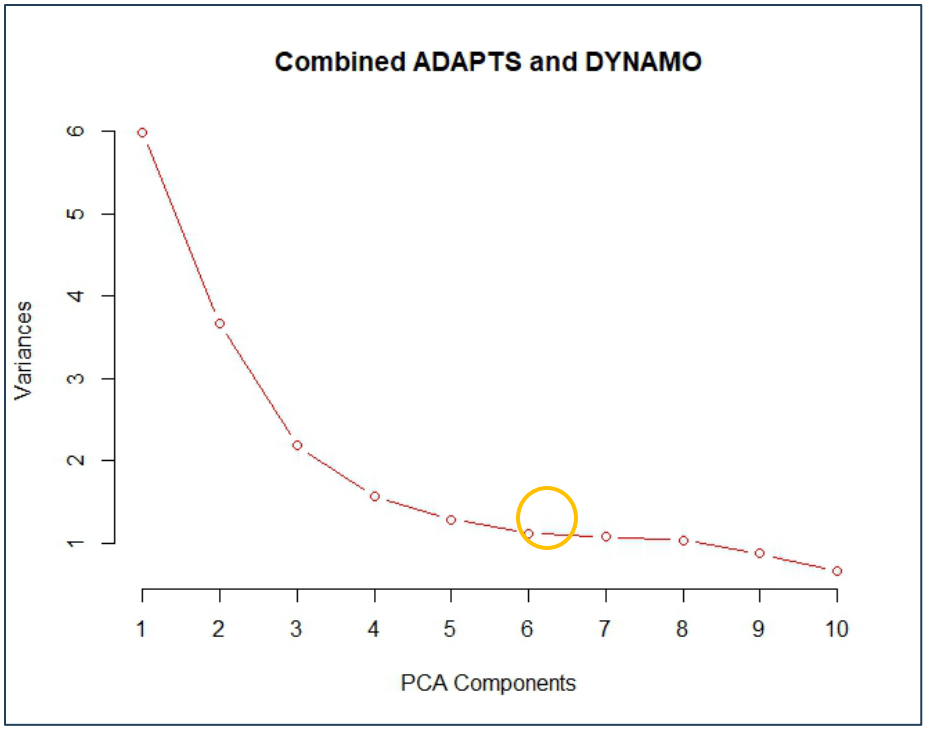

Figure 10. Scree plot: principal components for the combined ADAPTS/DYNAMO dataset in predicting performance score (z-scored).

Table 2. Variance in KSBs explained by retained components PC1–PC5.

Component	Proportion of variance	Cumulative proportion
PC1	0.2495	0.2495
PC2	0.1529	0.4024
PC3	0.0908	0.4932
PC4	0.0652	0.5584
PC5	0.0540	0.6124

In terms of significant loadings of KSB variables onto the first five principal components, the XAI Trust Scale Score (e.g., average Explainable AI Trust) loads strongly onto PC4, with a loading of 0.53. These KSB scores relating to trust are most strongly linked to the model competency category *Disposition*. PC5 was primarily driven by variables related to AI usage (0.57 loading) and SOT performance (0.51 loading), which map onto the TF model's competencies relating to *Information Management* and *Cognitive Ability*. Table C-1, displaying the full PCA loadings for PCA 1–5, is located in Appendix C. These findings suggest that *Disposition, Information Management, and Cognitive Ability* contribute heavily to the variation in the combined ADAPTS and DYNAMO KSB data. Although there were no individual KSBs loaded onto PC1 with a loading over 0.5, AI literacy

contributed notably, with all AI literacy subscales having loadings in the range of 0.24 to 0.36. BFI personality and cognitive flexibility contributed most to PC2, with loadings in the range of 0.23 to 0.42 for all subscales except openness to experience.

Next, a multiple regression analysis examined the relationship between the first five principal components (PC1–PC5) derived from the PCA and our performance variable. The model explained approximately 8% of the variance in the z-scored performance variable ($R^2 = 0.076$, $F(5, 131) = 2.16$, $p = 0.062$). The adjusted R^2 was 0.041. Only PC5 demonstrated a statistically significant positive relationship with performance ($b = 0.233$, $SE = 0.074$, $t(131) = 3.16$, $p = 0.002$). The coefficients for PC1 ($b = –0.006$, $SE = 0.034$, $p = 0.853$), PC2 ($b = –0.038$, $SE = 0.044$, $p = 0.383$), PC3 ($b = –0.010$, $SE = 0.057$, $p = 0.855$), and PC4 ($b = 0.000$, $SE = 0.067$, $p = 0.997$) were not statistically significant. These results provide evidence that *Information Management* and *Cognitive Ability* may be the most critical features impacting performance and thus indicating higher TF. Figure 11 visualizes the correlations between the significant KSB contributors in PC5 and overall performance in ADAPTS and DYNAMO, showing that as trust and AI usage increases, so does performance.

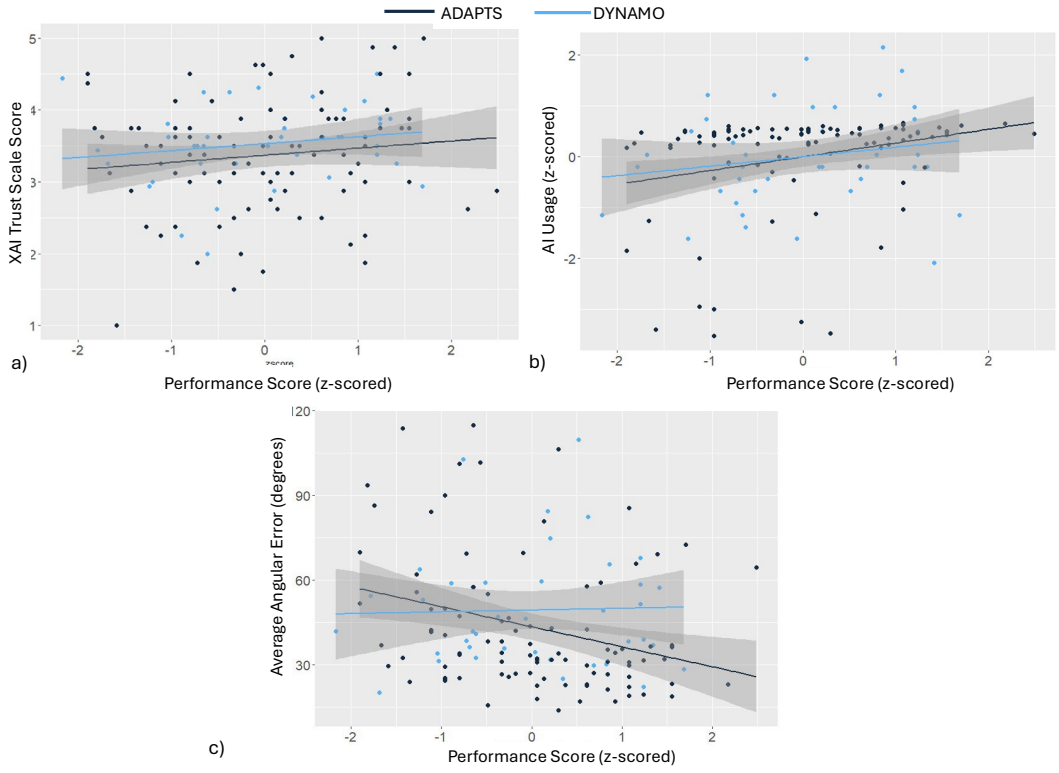

Figure 11. (a) XAI trust score (b), AI usage, and (c) average angular error (SOT) had the highest loadings on principal components derived from the DYNAMO and ADAPTS KSB datasets. Scatterplots of these KSBs are shown here vs. performance score.

3.2 Secondary Analysis: Impact of GRiPS and Team Trust on Performance

Using the ADAPTS dataset, a Spearman correlation found that GRiPS was not significantly correlated with overall Performance (R^2 = –0.146, p = 0.150), AI Usage (R^2 = –0.036, p = 0.565), Time Left (R^2 = 0.097, p = 0.341), Number of Resources (R^2 = –0.016, p = 0.122), or Tripwires activated (R^2 = –0.036, p = 0.723). With the DYNAMO dataset, a higher Overall Team Trust score was significantly correlated with a greater number of Bombs Deactivated (R^2 = 0.414, p = 0.009) and correlated with further average team Distance Traveled (R^2 = 0.523, p = 0.001) during the allotted time in the environment. A greater Perceived Trust component score was also significantly correlated with a greater number of Bombs Deactivated (R^2 = 0.335, p = 0.037) and further average team Distance Traveled (R^2 = 0.473, p = 0.002). For the Cooperative Trust component (Figure 12), a greater score was correlated with greater Bombs Deactivated (R^2 = 0.389, p = 0.015), greater AI usage (R^2 = 0.345, p = 0.029), and greater Distance Traveled (R^2 = 0.463, p = 0.003).

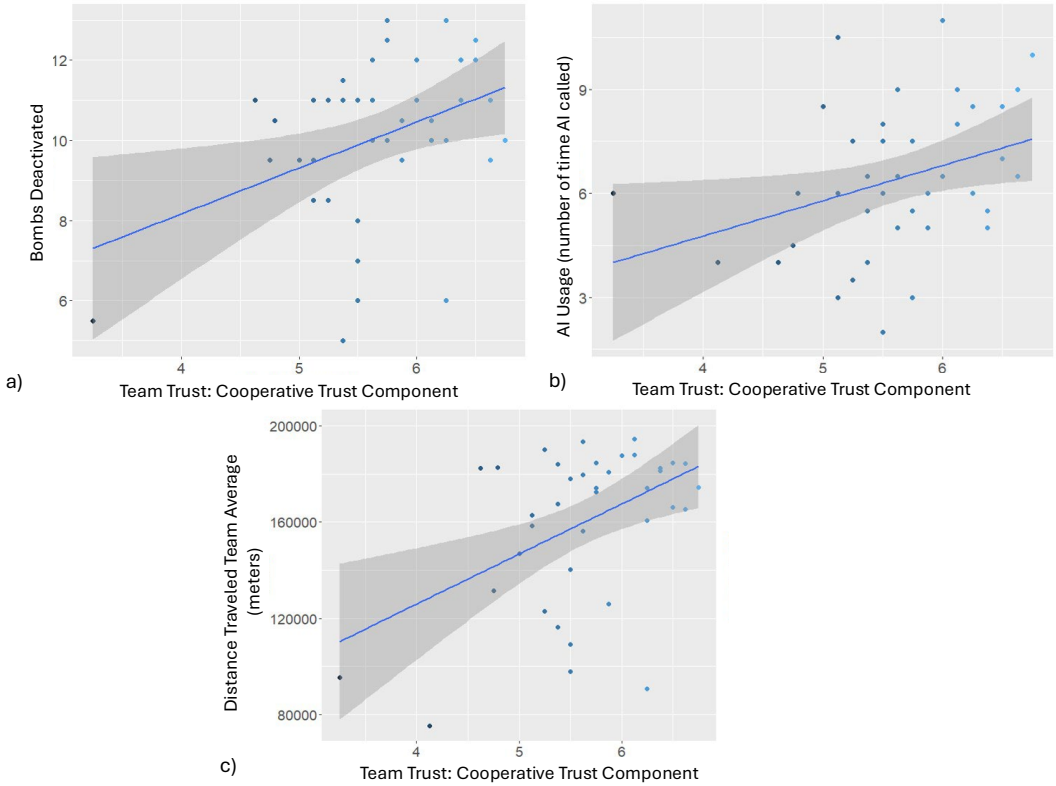

Figure 12. Increased response on the cooperative trust component of the team trust questionnaire was significantly correlated to (a) a greater number of bombs deactivated, (b) greater AI usage, and (c) greater distance traveled on average in the DYNAMO testbed.

3.3 Secondary Analysis: Impact of Emergent Leadership Dynamics in DYNAMO Performance

Granger analysis methodology was used to examine the relationship between emergent leadership dynamics, individual KSBs, and team performance on the DYNAMO testbed. A high-level summary of the observed leadership dynamics is presented in Figure 13. On average, teams spent a significant portion of their time without a clear leader (31.6%), with the designated "Leader" actively leading for 43.7% of the time and the "Follower" for 24.7%. The distributions of leadership volatility (switch rate) and imbalance (dominance) were consistent across most teams, indicating no extreme outliers in behavioral patterns.

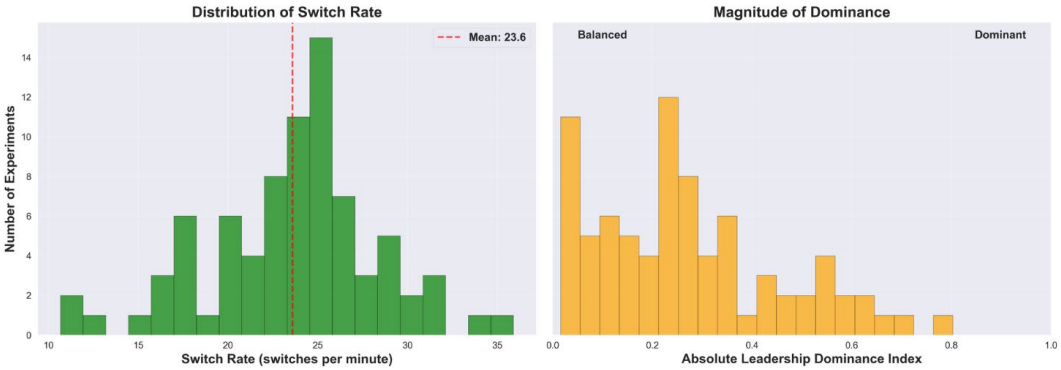

Figure 13. Summary of observed leadership dynamics across dyads.

To identify potential predictors of team success, a comprehensive correlation analysis was performed between the team performance score and all Granger-related behavioral metrics derived and KSBs. The analysis revealed no single strong predictor; all correlations were weak to moderate. The strongest negative trend was observed with initiation rate ($r = -0.328$), a measure of leadership instability. The most consistent positive trend was related to the average team trust of the leader ($r = +0.237$). The top positive and negative correlates are visualized in Figure 14.

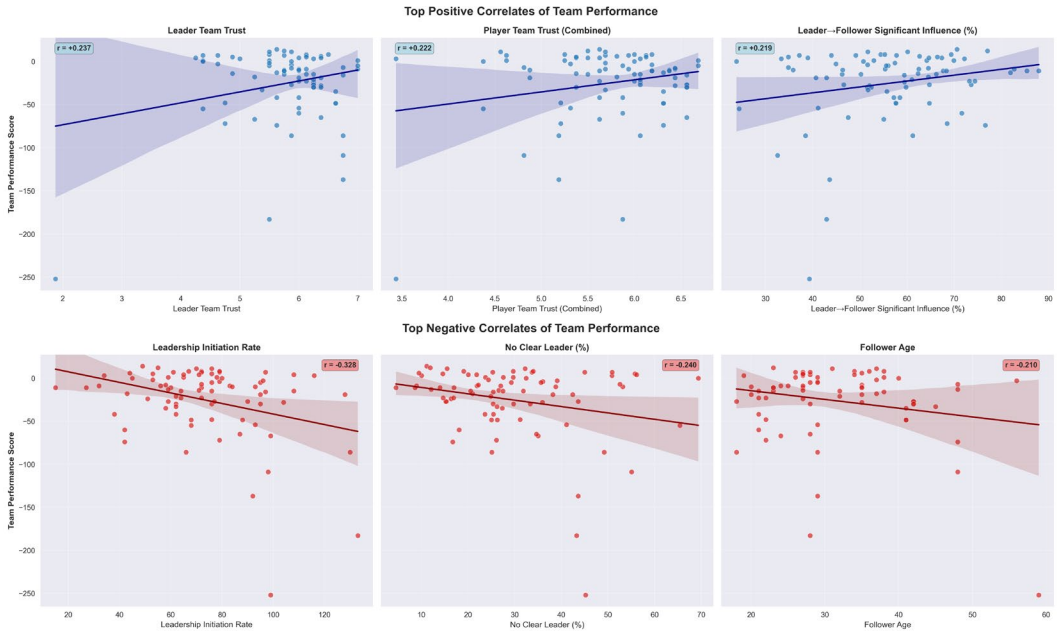

Figure 14. Visualizations of top positive and top negative correlates of team performance.

To synthesize these findings and create the most parsimonious predictive model, a multiple linear regression was performed using a backward elimination approach. Starting with a set of the top eight predictors, the least significant variable was iteratively removed until only statistically significant predictors remained. The final overall model was highly statistically significant ($F(3, 70) = 5.16$, $p = 0.003$) and successfully explained 14.6% of the variance in team performance (adjusted $R^2 = 0.146$). The results of this final model are detailed in Table 3.

Table 3. Final multiple regression model predicting team performance.

Predictor	Coefficient (B)	Std. error	P-value
(Intercept)	–78.03	45.19	0.089
Leader's team trust	+13.29	6.02	0.031
Switch rate (per min)	–2.61	1.04	0.014
Reciprocal switch count	–0.43	0.16	0.007

Note: Model Adj. $R^2 = 0.146$, $F(3, 70) = 5.16$, $p = 0.003$

The final model revealed a nuanced profile of effective teams. Three factors emerged as significant predictors of performance. First, the leader's trust in their teammate was a significant positive predictor. Second, the overall leadership volatility (i.e., switch rate) was a significant negative predictor, suggesting that chaotic and frequent leadership changes are detrimental. The most statistically significant predictor was the reciprocal switch count, which had a strong positive relationship with performance. Reciprocal switching occurs when leadership is passed directly from one person to the other, rather than leadership "terminating" in between, and then a person "initiating" leading. While initiating requires there

to be no leader, followed by another person becoming a leader, reciprocal switching of leadership occurs when there is no lapse in leadership in between the transition from one person to another.

3.4 UEC Analysis

Section 2.4 outlined the use of the UEC for this model validation effort, with further details in this current section. The UECs of the ADAPTS and DYNAMO testbeds were each evaluated by four expert user raters. The researchers and software engineers who designed and implemented the testbeds served as expert users to provide ratings. Each rating of the ADAPTS (Table 4) and DYNAMO (Table 5) technologies was provided by the raters from the perspective of the user of the technology. The average rating and percent agreement are also provided. Average ratings are extracted with scoring labels for easy visual comparison.

Table 4. ADAPTS UEC ratings.

Variable	Rater 1	Rater 2	Rater 3	Rater 4	Average	Agreement (%)
Frequency of tech feedback	5	5	5	5	5	100
Clarity of tech feedback	4	4	4	5	4.25	50
Frequency of user input	5	5	5	5	5	100
Type of user input	2	2	2	2	2	100
Physical proximity	2	5	2	2	2.75	50
Cognitive workload	3	3	3	3	3	100
Tech interdependence	2	5	2	3	3	17
Degree of collaboration	1	1	2	1	1.25	50

Table 5. DYNAMO UEC ratings.

Variable	Rater 1	Rater 2	Rater 3	Rater 4	Average	Agreement (%)
Frequency of tech feedback	4	5	4	4	4.25	50
Clarity of tech feedback	5	5	4	5	4.75	33
Frequency of user input	4	4	4	4	4	100
Type of user input	1	1	1	1	1	100
Physical proximity	1	2	1	2	1.5	33
Cognitive workload	2	3	1	1	1.75	17
Tech interdependence	2	4	1	2	2.25	17
Degree of collaboration	2	4	1	1	2	17

Percent agreement is a simple way to measure inter-rater reliability; however, it only captures whether there was complete or no agreement between raters and not whether scores provided by different raters were dissimilar but close. Therefore, an intraclass correlation coefficient (ICC) inter-rater analysis was selected to provide a nuanced assessment of inter-rater reliability. An excellent degree of reliability in

terms of consistency was found across the raters of the ADAPTS technology. The average measure ICC was 0.926 with a 95% confidence interval from 0.779 to 0.983 ($F(7,21) = 13.440$, $p < 0.001$), indicating raters were consistent with ratings relevant to one another. An excellent degree of reliability in terms of consistency was also found across the raters of the DYNAMO technology. The average measure ICC was 0.954 with a 95% confidence interval from 0.863 to 0.990 ($F(7,21) = 21.667$, $p < 0.001$), indicating raters were consistent with rating relative to each other.

From the UEC average ratings (Table 6), we can see that the two technologies are fairly similar (Figure 15). This makes sense given they are both AI navigation technologies designed to provide optimal routes through an environment to a user. They are also both presented visually to the user in similar ways (i.e., green route line presented in the environment). They differ in terms of how and when you can use the tools throughout the tasks (e.g., ADAPTS you can call the tool once and it persists, DYNAMO you can call the tool frequently, but it does not persist). Considering the UECs, we see that both studies included technology that were most similar in characteristics of clarity, frequency of technology feedback, technological interdependence, and degree of collaboration. The largest difference in UECs from raters between these two technologies was in the physical proximity and cognitive workload ratings. Given the game-like experience of the testbeds and the AI tools provided, conceptualizing physical proximity was difficult and this could partially account for the differences in ratings.

Table 6. Average ADAPTS and DYNAMO UEC ratings.

Variable	ADAPTS average	ADAPTS classification	DYNAMO average	DYNAMO classification	Absolute difference
Frequency of tech feedback	5	Very frequent	4.25	Frequent	0.75
Clarity of tech feedback	4.25	Clear	4.75	Very clear	0.5
Frequency of user input	5	Very frequent	4	Frequent	1
Type of user input	2	Simple	1	Very simple	1
Physical proximity	2.75	Moderately distal	1.5	Proximal	1.25
Cognitive workload	3	Moderate effort	1.75	Low effort	1.25
Tech interdependence	3	Moderately interdependent	2.25	Standalone	0.75
Degree of collaboration	1.25	Very individual	2	Individual	0.75

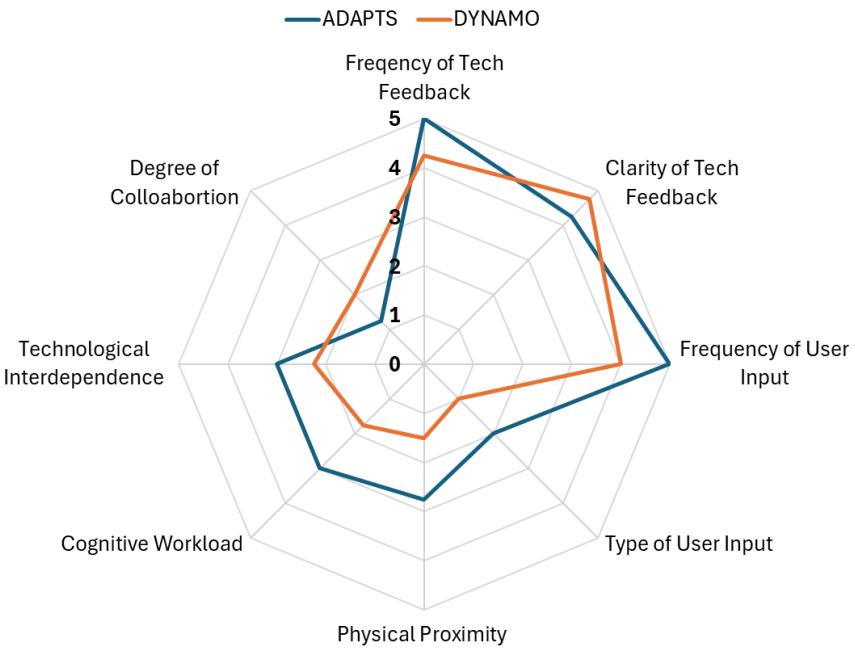

Figure 15. Visualization of ADAPTS and DYNAMO UECs.

Specifically, both the ADAPTS and DYNAMO AI technologies were characterized with frequent to very frequent and clear to very clear technology feedback. Users could use the technologies independently (individual-very individual) and users frequently to very frequently provided simple to very simple input to the AI. In terms of different categorizations, ADAPTS was viewed as requiring moderate cognitive effort and as a moderately interdependent technology, while DYNAMO was seen as a standalone technology requiring low cognitive effort. In terms of proximity to the technology, ADAPTS was viewed as moderately distal (e.g., within visual range) while DYNAMO was viewed as proximal (within arm's reach).

Given that the AI technologies in the ARL testbeds were not drastically different in technological characteristics from one another, it was not appropriate to directly include the UECs ratings themselves in the TF model validation algorithms. Instead, we use these UEC characterizations to conceptualize the technological context in which we interpret the findings relating KSBs to performance using these types of technologies.

4. Discussion

The TSAVVY research program was tasked with observing and conceptualizing, assessing, and maximizing TF in Soldier units and teams. In FY23–25 we accomplished this by 1) identifying candidate KSBs that predict TF performance,

2) developing multiple research-grade testbeds capable of eliciting the types of future Soldier interactions thought to be critical in the Future Operating Environment, 3) developing a working conceptual model of TF, and finally 4) empirically validating said model to understand which KSBs help predict TF performance. Each of these efforts were carried out across three phases.

In Phase 1 (FY23), we established the theoretical foundations of TF through SME input, detailed job analysis, and a large-scale literature review (Neubauer et al. 2024). Within this phase, one effort focused on developing a working conceptual definition of TF, which we defined as a person's ability to creatively adapt a new technology, such as AI, with little or no formal training. In a world where working with advanced technologies such as AI is becoming increasingly important, predicting levels of TF is valuable for personnel selection, team composition considerations, and training assignments. This work also helped our team to identify a preliminary list of candidate KSB predictors, which we organized into five higher-order categories: 1) Disposition and Motivation; 2) Cognitive Abilities; 3) Social and Teaming Skills; 4) Adaptability and Response to Change; and 5) AI-Relevant Knowledge and Experience. These results yielded a theory-driven TF model with initial tools for measuring predictors and performance criteria.

The FY23 effort also focused on examining the influence of a variety of KSBs on performance, working with a trainable AI in a heterogenous teaming puzzle game. Here, participants trained reinforcement-learning AI models to achieve high scores in the heterogenous teaming puzzle game, and their partially trained and fully trained AI models were extracted at different time points and evaluated in a sandbox to assess each AI's performance. We measured participants' KSBs with subjective report questionnaires and aptitude tests. In this effort, we were also able to successfully generate a large dataset across several experimental conditions, which allowed us to examine behavioral and performance variability and potential indicators of TF proficiency. Notably, significant group differences were observed in horizontal conditions, where some participants demonstrated comparatively greater effectiveness in training the AI, while others did not (Neubauer et al. 2025). This was the first "sanity check" to discover whether some individuals were better suited to these types of human–AI tasks, with results indeed confirming this.

Under the previously mentioned FY23 effort, several key KSBs emerged as important for understanding individual performance when training AI systems. Large group differences were shown in some conditions of the heterogenous teaming puzzle task, with additional analyses uncovering four distinct clusters that differed in how rapidly each participant was able to train their AI to proficiency. KSBs were found to predict cluster membership, to predict performance within clusters, and to predict performance across all participants. However, the

relationships were nuanced. Clusters differed in their levels of conscientiousness, intrinsic motivation, and comfort with training AI, with the fastest performers tending to have higher levels of these KSBs. KSBs like anxiety, age, and recent usage of AI also predicted TF, but only within certain clusters. Across all clusters combined, AI literacy subscales and comfort with training AI were the only variables that predicted performance. Combined, these results suggest that KSBs are valuable predictors of TF, but that the influence of different KSBs varies among different clusters of performers.

In Phase 2 (FY24), we transitioned the conceptual TF framework into a validation effort led by our collaborators at ARI. ARI partners conducted a series of Soldier SME studies, which asked Soldiers (who routinely work with technology) to rank the importance of each KSB toward achieving TP as well as TF. Findings demonstrated that TP is technology specific, whereas TF generalizes across systems. ARI refined the initial TF model into a higher-resolution framework highlighting seven key facets: 1) Information Management; 2) Dispositions; 3) Interpersonal Skills; 4) Critical Thinking; 5) Knowledge and Experience; 6) Cognitive Ability; and 7) Management (Brown et al. 2026). The program utilized an iterative model development approach in which theoretical models were progressively validated through ARI's Soldier studies and ARL's empirical testbed experimentation. This updated model provided clear targets for training-related improvements and baseline proficiency expectations across MOSs and was used for our empirical model validation efforts in FY25.

The Phase 3 (FY25) goal of the Technically Savvy Soldier research program was to develop a comprehensive, empirically validated model of team and individual TF to inform Army initiatives in talent selection, unit composition, and readiness planning. As digital tools become increasingly integrated into operations, it has become critical to identify and maximize individual's TF to optimize their effectiveness, creativity, and ability to adapt to technology. To this end, ARL developed experimental testbeds to identify KSBs related to TF and performance, which ARI is leveraging to create assessment batteries and training concepts.

This work represents a collaborative effort to expand an initial TF model (informed by ARL) to include TP. Here, ARL developed multiple testbed to evaluate participant behavior and performance with advanced AI-enabled systems under conditions where system capabilities were degraded, uncertain, dynamically changing, or artificially enhanced. These experiments were designed to capture both Soldier adaptability and the predictive power of KSBs for TF across complex, dynamic, and technologically driven environments. Within these environments, Soldiers were considered technologically fluent if they could 1) identify deficiencies in AI-based systems, 2) retrain them on the fly, and 3) evaluate system

effectiveness. We validated the expanded ARI model using two ARL testbeds—ADAPTS and DYNAMO—both characterized by UECs emphasizing frequent, clear feedback, simple input, proximal interaction, low cognitive load, and moderate interdependence. It is crucial to interpret our findings within this specific technological context; while ARI found TF independent of specific technology, they found that TP *was* technology specific.

Principal Component Analysis (PCA) of KSB variables from both testbeds yielded an initial five-component model explaining over 60% of the variance. Only components PC4 (*Disposition* – trust) and PC5 (*Information Management and Cognitive Ability* – AI Usage and Spatial Orientation Error) exhibited individual KSB loadings greater than 0.5. Multiple regression analyses revealed that PC5—encompassing *Information Management* and *Cognitive Ability*—significantly and positively predicted task performance. This suggests that, within these AI-driven environments, effectively leveraging AI and possessing strong spatial reasoning skills are critical for success, reflecting the navigational nature of the tasks.

Examining each testbed separately revealed nuanced insights. In ADAPTS, risk-taking was not significantly correlated with performance or AI usage, suggesting this facet may be less relevant in this technological context. DYNAMO, focused on small teams, demonstrated that higher levels of trust—overall, perceived, and cooperative—correlated with increased performance and AI utilization, reinforcing the importance of *Interpersonal Skills* within the TF model. Further, Granger analysis of DYNAMO data highlighted the role of emergent leadership. While leader trust decreased switch rate and increased reciprocal switch count, all indicated more effective teams, *reciprocal switch count* was the strongest predictor of performance. Reciprocal switching of leadership implies leadership passes directly and smoothly from one person to another without terminating first. This suggests high-performing teams exhibit "trusted adaptability"—a foundation of interpersonal trust enabling flexible, fluid leadership transitions based on situational demands.

These testbeds more closely mirror the realities of human–machine interaction in operational environments. Our results indicate that *Information Management, Cognitive Ability,* and *Interpersonal Skills* are critical facets of TF. Future work should prioritize examining these constructs in assessments and training programs, particularly those focused on AI utilization and navigational tasks. While these results are promising, future research is needed to determine the generalizability of these findings to other types of technologies and with Soldier populations.

Implications for Selection and Training

Identifying TF or "TSAVVY" individuals raises important questions for future selection and training protocols. Although some KSBs were found to be predictive in some conditions, and using different testbeds and tasks, the next question to ask is whether individuals with moderate or no proficiency can be elevated to comparable levels via training. Additionally, understanding whether these KSBs of interest are malleable and/or trainable versus which are relatively stable dispositions warrants further investigation. Such work is essential for informing evidence-based training and enhancement strategies and protocols. Preliminary evidence suggests that some skills relevant to AI training may be trainable (Neubauer et al. 2025). Future research should examine whether dispositional factors such as intrinsic motivation or conscientiousness can be developed through training, or whether interventions should primarily target task-specific strategies and practice or experiential opportunities, for example. Finally, not all candidate KSBs identified throughout the program's duration were predictive of TF performance, which warrants consideration. It is possible that certain KSBs are only relevant in particular task environments, or that their influence only emerges when performing in different technological contexts. Future research should aim to refine the understanding of which KSBs most consistently predict successful human–AI interaction across the board. Moreover, as technological capabilities evolve, it is critical to assess whether predictors of proficiency with earlier, first-generation technologies remain valid, or whether new skill sets are required for next-generation AI systems. Expanding research to encompass multiple technology domains will be essential for evaluating the generalizability of KSB predictors and to help guide future team and personnel assignment efforts, help inform training programs, and inspire future research. Finally, programmatic outcomes can help identify personnel who would be well suited to working with AI and personnel who may need different levels or types of training.

5. Conclusion

The current effort had several major accomplishments including leveraging scientific literature and subjective/qualitative feedback from SME focus group interviews to enhance our understanding of the TF construct, unique organization and individual needs, and future Soldier tasks. Based on this work, and in conjunction with our partners, we were able to develop initial conceptual models of TF. These models have since been validated in multiple human-guided AI interaction experiments, where we were able to link possibly predictive KSBs to TF as empirical data became available. Finally, we were able to develop and initiate a successful model validation pipeline methodology, which has been documented

here, with the goal of helping future efforts build predictive TP and TF profiles. Ultimately, this effort will be a critical enabler for talent management and maximization of TF Soldier units and teams.

6. References

Azevedo R, Cromley JG, Seibert D. Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? Contemp Educ Psychol. 2004;29(3):344–370. https://doi.org/10.1016/j.cedpsych.2003.09.002

Barak M, Levenberg A. Flexible thinking in learning: an individual differences measure for learning in technology-enhanced environments. Comput Educ. 2016;99:39–52. https://doi.org/10.1016/j.compedu.2016.04.003

Brown J et al. Preliminary evaluation of the technological fluency competency model. Army Research Institute for the Behavioral and Social Sciences (US); forthcoming 2026.

Celik I. Towards intelligent-TPACK: an empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. Comput Hum Behav. 2023;138:107468. https://doi.org/10.1016/j.chb.2022.107468.

Chen L, Zhu Z. Probabilistic neural networks: foundations and applications. IEEE Trans Neural Netw Learn Syst. 2020;31(8):2827–2841.

Dalangin B, Gordon S, Roy H. Positive interactions with intelligent technology through psychological ownership: a human-in-the-loop approach. Artif Intell Soc Comput. 2024;122:62–74. https://doi.org/10.54941/ahfe1004642

Fagerlin A et al. Measuring numeracy without a math test: development of the subjective numeracy scale. Med Decis Making. 2007;27(5):672–680. https://doi.org/10.1177/0272989x07304449

Friedman A, Kohler B, Gunalp P, Boone AP, Hegarty M. A computerized spatial orientation test. Behav Res Meth. 2020;52(2):799–812. https://psycnet.apa.org/doi/10.3758/s13428-019-01277-3

Green RM. Predictors of digital fluency [PhD dissertation]. Dissertation Abstracts International Section A: Humanities and Social Sciences; 2005.

Greene JA, Yu SB, Copeland DZ. Measuring critical components of digital literacy and their relationships with learning. Comput Educ. 2014;76:55–69. https://doi.org/10.1016/j.compedu.2014.03.008

King KW, Gordon SM, Rabin A. 2023 Oct. Granger leadership in a novel dyadic search paradigm. 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC); 2023 Oct 1–4; Honolulu, HI. IEEE; c2023. p. 4889–4894.

Lakhmani S et al. Cohesion in human–autonomy teams: an approach for future research, Theor Issues Ergon Sci. 2022. https://doi.org/10.1080/1463922X.2022.2033876

Neubauer C et al. Measuring and predicting technical fluency: how knowledge, skills, abilities, and other behaviors can contribute to technological savviness. DEVCOM Army Research Laboratory (US); 2024. Report No.: ARL-TR-9884.

Neubauer C. Future skillsets for digital competencies, cybersecurity, and human AI partnerships. 2025 International Conference on Knowledge Management, Cybersecurity, Learning, and Information Technology; 2025 June 25–28; Siena, Italy.

Neubauer C, Pollard KA, Roy H, Dalangin B, Gordon S. Falling bricks and novel twists: identifying predictors of who can train an AI. Forthcoming 2026.

Oksanen A, Savela N, Latikka R, Koivula A. Trust toward robots and artificial intelligence: an experimental approach to human–technology interactions online. Front Psychol. 2020;11:568256. https://doi.org/10.3389/fpsyg.2020.568256

[OUSW(R&E)] Office of the Under Secretary of War, Research and Engineering. Department of War critical technology areas. Department of War Office of Strategic Capital (US); 2023. https://www.cto.mil/osc/critical-technologies/

R Core Team. R: a language and environment for statistical computing. R Project for Statistical Computing; Vienna, Austria. 2025 [accessed 2026 Feb 17]. https://www.R-project.org/

Silverman J, Purshouse RC, Fleming PJ. Probabilistic model predictive control: theory and applications. IEEE Trans Autom Control. 2019;64(10):4129–4151.

Techataweewan W, Prasertsin U. Development of digital literacy indicators for Thai undergraduate students using mixed method research. Kasetsart J Soc Sci. 2018;39(2):215–221. https://doi.org/10.1016/j.kjss.2017.07.001

Von Walter B, Kremmel D, Jäger B. The impact of lay beliefs about AI on adoption of algorithmic advice. Mark Lett. 2021;33(1):143–155. https://doi.org/10.1007/s11002-021-09589-1

Woods DD, Hollnagel E. Joint cognitive systems: patterns in cognitive systems engineering. CRC Press; 2006.

Appendix A. User Engagement Characteristics Scoring Guide

1. Frequency of Technology Feedback: The degree to which the technology provides feedback to the user				
1	**2**	**3**	**4**	**5**
Very Infrequent On a timescale of months	**Infrequent** On a timescale of days	**Moderately Frequent** On a timescale of hours	**Frequent** On a timescale of minutes	**Very Frequent** Real-time or near real-time
2. Clarity of Technology Feedback: Feedback provided to the user is explicit/clearly defined vs. implicit/ambiguous				
1	**2**	**3**	**4**	**5**
Very Ambiguous Unclear, highly interpretive, or easily misunderstood	**Ambiguous** Partially clear, but includes many ambiguous or confusing signals	**Moderately Clear** Mix of clear and ambiguous feedback; Requires some interpretation	**Clear** Largely clear, but can leave room for interpretation	**Very Clear** Clear and explicit; Specific, quantitative, reasonable, relevant, and actionable
3. Frequency of User Input: How often the user must provide input to the technology				
1	**2**	**3**	**4**	**5**
Very Infrequent On a timescale of months	**Infrequent** On a timescale of days	**Moderately Frequent** On a timescale of hours	**Frequent** On a timescale of minutes	**Very Frequent** Real-time or near real-time
4. Type of User Input: How input is communicated to the technology (e.g., number of inputs, sensitivity)				
1	**2**	**3**	**4**	**5**
Very Simple Basic single-bit input (e.g., a single button or switch)	**Simple** Limited inputs (e.g., a few buttons or straightforward voice commands)	**Moderately Complex** A non-trivial set of simple inputs (e.g., keyboard and mouse, an array of buttons, a touch screen)	**Complex** Multifaceted input methods (e.g., combinations of gestures, voice, and manual controls)	**Very Complex** Highly advanced, sensitive, or intricate input methods

5. Physical proximity: Location of the technology relative to user (near vs. far)				
1	**2**	**3**	**4**	**5**
Very Proximal In user's hand or connected to user's body	**Proximal** Within arm's reach	**Moderately Distal** Within the building OR within visual/auditory range	**Distal** Within a 5 km radius	**Very Distal** Outside 5 miles OR location is irrelevant (e.g., on the cloud)
6. Cognitive Workload: Degree of mental resources required of a person at any one time to interact with the technology				
1	**2**	**3**	**4**	**5**
Very Low Effort Minimal mental effort; automatic or intuitive; requires little or no attention	**Low Effort** Simple to use; A person can multitask while using it	**Moderate Effort** Requires attention and active decision-making	**High Effort** Requires significant attention or problem solving	**Very High Effort** Requires full attention or constant cognitive engagement
7. Technological Interdependence: Degree to which the technology is a standalone element vs. part of an integrated system of technologies				
1	**2**	**3**	**4**	**5**
Very Standalone Functions independently; does not require other technologies to function	**Standalone** Benefits from a platform or backend architecture (e.g., an operating system)	**Moderately Interdependent** Requires other instances of the same technology to be useful	**Interdependent** Requires other instances of the same technology to function at all	**Very Interdependent** Part of an integrated system
8. Degree of Collaboration: Degree to which the technology can be operated by a single individual vs. requires team input/coordination				
1	**2**	**3**	**4**	**5**
Very Individual Operated by a single individual	**Individual** Can be operated by a single individual, but can be more effective with collaboration	**Moderately Cooperative** Equally effective with individual or collaborative use	**Cooperative** Requires cooperation of individuals doing the same thing	**Very Cooperative** Requires coordination of teams of individuals with different roles

Appendix B. Questionnaires from ADAPTS and DYNAMO used for Model Validation Analyses

The following questionnaires are from AI for Dynamic Pathfinding in Threatening Scenarios (ADAPTS) and Dyadic Navigation with AI Systems for Mission Operation (DYNAMO) and are used for model validation analyses.

B.1 Explainable AI (XAI) Trust Scale[1]

Items are along a 1–5 Likert Scale (1 – I agree strongly, 2 – I agree somewhat, 3 – I'm neutral about it, 4 – I disagree somewhat, 5 – I disagree strongly)

1) I am confident in the AI. I feel that it works well.
2) The outputs of the AI are very predictable.
3) The AI is very reliable. I can count on it to be correct all the time.
4) I feel safe that when I rely on the AI I will get the right answers.
5) The AI is efficient in that it works very quickly.
6) I am wary of the AI.
   (adopted from the Jian et al. Scale[2] and the Wang et al. Scale[3]).
7) The AI can perform the task better than a novice human user.
   (adopted from the Schaefer Scale[4])
8) I like using the AI for decision making.
   (adapted from the Madsen–Gregor Scale[5])

Note: The language in these scales may be modified from "tool" or "system" to "AI" to fit the study.

[1] Hoffman R, Mueller S, Klein G, Litman J. Measuring trust in the XAI context. PsyArXiv Preprints. 2021. http://doi.org/10.31234/osf.io/e3kv9. Retrieved from: https://digitalcommons.mtu.edu/michigantech-p/15804.
[2] Jian JY, Bisantz AM, Drury CG. Foundations for an empirically determined scale of trust in automated systems. Int J Cogn Ergon. 2000;4:53–71. https://doi.org/10.1207/S15327566IJCE0401_04
[3] Wang et al. A validation of the human-generative artificial intelligence trust scale. Int J Hum–Comput Int. 2025;1–14. https://doi.org/10.1080/10447318.2025.2542881
[4] Schaefer KE. Measuring trust in human robot interactions: development of the "trust perception scale-RHI." In: Mittu R, Sofge D, Wagner A, Lawless WF, editors. Robust Intelligence and Trust in Autonomous Systems. Springer; 2016. p. 191–218.
[5] Madsen M, Gregor S. Measuring human–computer trust. In: Proceedings of the 11th Australasian Conference on Information Systems; 2000 Dec 6–8, Brisbane, Australia. 2000;53:6–8.

B.2 Psychological Ownership

Originally created by: Van Dyne and Pierce (2004).[6]

Adapted by: Delgosha and Hajiheydari (2021).[7]

All measurement items were anchored on a 7-point Likert scale ranging from strongly disagree (1) to strongly agree (7).

1) The AI I trained/The pre-trained AI is my AI.
2) I feel a very high degree of personal ownership for the AI I trained/the pre-trained AI.
3) I sense that I own the AI I trained/the pre-trained AI.
4) The AI I trained/The pre-trained AI incorporates a part of myself.

B.3 The Cognitive Flexibility Inventory[8]

All measurement items were anchored on a 5-point Likert scale ranging from strongly disagree (1) to strongly agree (5).

Instructions: Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age.

1) I am good at "sizing up" situations.
2) I have a hard time making decisions when faced with difficult situations.
3) I consider multiple options before making a decision.
4) When I encounter difficult situations, I feel like I am losing control.
5) I like to look at difficult situations from many different angles.
6) I seek additional information not immediately available before attributing causes to behavior.
7) When encountering difficult situations, I become so stressed that I cannot think of a way to resolve the situation.
8) I try to think about things from another person's point of view.
9) I find it troublesome that there are so many different ways to deal with difficult situations.

[6] Van Dyne L, Pierce JL. Psychological ownership and feelings of possession: three field studies predicting employee attitudes and organizational citizenship behavior. J Organ Behav. 2004;25(4):439–459. https://doi.org/10.1002/job.249

[7] Delgosha MS, Hajiheydari N. How human users engage with consumer robots? A dual model of psychological ownership and trust to explain post-adoption behaviours. Comput Hum Behav. 2021;117:106660. https://doi.org/10.1016/j.chb.2020.106660

[8] Dennis JP, Vander Wal JS. The cognitive flexibility inventory: instrument development and estimates of reliability and validity. Cogn Ther Res. 2010;34(3):241–253. https://doi.org/10.1007/s10608-009-9276-4

10) I am good at putting myself in others' shoes.
11) When I encounter difficult situations, I just don't know what to do.
12) It is important to look at difficult situations from many angles.
13) When in difficult situations, I consider multiple options before deciding how to behave.
14) I often look at a situation from different viewpoints.
15) I am capable of overcoming the difficulties in life that I face.
16) I consider all the available facts and information when attributing causes to behavior.
17) I feel I have no power to change things in difficult situations.
18) When I encounter difficult situations, I stop and try to think of several ways to resolve it.
19) I can think of more than one way to resolve a difficult situation I'm confronted with.
20) I consider multiple options before responding to difficult situations.

B.4 Technology Self-Efficacy Scales

The Tech Self-Efficacy Scale[9] 1 was adapted by ARI from this original scale:

Technology Self-Efficacy Scale 1 [10 items]					
Instructions: Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age.					
I believe I could use a new technology…	Very Inaccurate	Moderately Inaccurate	Neither Accurate nor Inaccurate	Moderately Accurate	Very Accurate
if there was no one around to tell me what to do.	1	2	3	4	5
if I had never used anything like it.	1	2	3	4	5
if I only had the product manual for reference.	1	2	3	4	5
if I had seen someone else using it before trying it.	1	2	3	4	5
if there was no one around to help if I got stuck.	1	2	3	4	5
without someone else helping me get started.	1	2	3	4	5
if there was not a lot of time to complete the task for which the equipment was provided.	1	2	3	4	5
if I only had a built-in help system for assistance.	1	2	3	4	5
if someone showed me how to do it first.	1	2	3	4	5
if I had used similar products like this one to do the same job.	1	2	3	4	5

[9] Laver K, George S, Ratcliffe J, Crotty M. Measuring technology self efficacy: reliability and construct validity of a modified computer self efficacy scale in a clinical rehabilitation setting. Disabil Rehabil. 2012;34(3):220–227.

Technology Self-Efficacy Scale 2					
Instructions: Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age.					
When using technology, I…	Very Inaccurate	Moderately Inaccurate	Neither Accurate nor Inaccurate	Moderately Accurate	Very Accurate
explore capabilities of technologies when I have spare time.	1	2	3	4	5
read manuals.	1	2	3	4	5
like to connect new technologies to my existing technologies.	1	2	3	4	5
like to personalize my technologies/interfaces.	1	2	3	4	5
am the person in the in the group in charge of maintaining the technologies (passwords, system updates, etc.).	1	2	3	4	5

Note: Scale 2 was created by ARI researchers.

B.5 Big Five Inventory[10]

Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who *likes to spend time with others*? Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement.

1 Disagree Strongly	2 Disagree a little	3 Neither agree nor disagree	4 Agree a little	5 Agree strongly

I am someone who…

1. ___Is talkative
2. ___Tends to find fault with others
3. ___Does a thorough job
4. ___Is depressed, blue
5. ___Is original, comes up with new ideas
6. ___Is reserved
7. ___Is helpful and unselfish with others
8. ___Can be somewhat careless
9. ___Is relaxed, handles stress well.
10. ___Is curious about many different things
11. ___Is full of energy
12. ___Starts quarrels with others
13. ___Is a reliable worker
14. ___Can be tense
15. ___Is ingenious, a deep thinker
16. ___Generates a lot of enthusiasm
17. ___Has a forgiving nature
18. ___Tends to be disorganized
19. ___Worries a lot
20. ___Has an active imagination
21. ___Tends to be quiet
22. ___Is generally trusting
23. ___Tends to be lazy
24. ___Is emotionally stable, not easily upset
25. ___Is inventive
26. ___Has an assertive personality
27. ___Can be cold and aloof
28. ___Perseveres until the task is finished
29. ___Can be moody
30. ___Values artistic, aesthetic experiences
31. ___Is sometimes shy, inhibited
32. ___Is considerate and kind to almost everyone
33. ___Does things efficiently
34. ___Remains calm in tense situations
35. ___Prefers work that is routine
36. ___Is outgoing, sociable
37. ___Is sometimes rude to others
38. ___Makes plans and follows through with them
39. ___Gets nervous easily
40. ___Likes to reflect, play with ideas
41. ___Has few artistic interests
42. ___Likes to cooperate with others
43. ___Is easily distracted
44. ___Is sophisticated in art, music, or literature

[10] John OP, Donahue EM, Kentle RL. Big Five Inventory (BFI) [database record]. APA PsycTests. 1991. https://doi.org/10.1037/t07550-000

B.6 Literacy Scale[11]

Responses are provided along a 7-point Likert scale (strongly disagree to strongly agree). The first 12 questions are used to explain the last three questions, with the last three items (AI literacy (overall items)) considered as a "short scale." To score, items are averaged.

1) I have knowledge of the types of technology that AI is built on.
2) I have knowledge of how AI technology and non-AI technology are distinct.
3) I have knowledge of use cases for AI technology.
4) I have knowledge of which human actors beyond programmers are involved to enable human-AI collaboration.
5) I have knowledge of the aspects human actors handle worse than AI.
6) I have knowledge of the aspects human actors handle better than AI.
7) I have knowledge of the input data requirements for AI.
8) I have knowledge of AI processing methods and models.
9) I have knowledge of using AI output and interpreting it.
10) I have experience in interaction with different types of AI, like chatbots, visual recognition agents, etc.
11) I have experience in the usage of AI through frequent interactions in my everyday life.
12) I have experience in designing AI models, for example, a neural network.
13) I have experience in development of AI products.
14) In general, I know the unique facets of AI and humans and their potential roles in human-AI collaboration.
15) I am knowledgeable about the steps involved in AI decision-making.
16) Considering all my experience, I am relatively proficient in the field of AI.

[11] Pinski M, Benlian A. AI literacy - towards measuring human competency in artificial intelligence. Hawaii International Conference on System Sciences 2023 (HICSS-56); 2023. https://aisel.aisnet.org/hicss-56/cl/ai_and_future_work/3

B.7 Propensity to Trust Intelligent Agents[12, 13]

1) One should be very cautious with strangers.
2) Most experts tell the truth about the limits of their knowledge.
3) Most people can be counted on to do what they say they will do.
4) These days, you must be alert or someone is likely to take advantage of you.
5) Most salespeople are honest in describing their products.
6) Most repair people will not overcharge people who are ignorant of their specialty.
7) Most people answer public opinion polls honestly.
8) Most adults are competent at their jobs.

5-point Likert scale from 1 – strongly disagree to 5 – strongly agree

B.8 Educational Experience Demographic Question

During the study, you will be asked a series of questions about yourself. It is very important that you be completely honest, your accuracy in reporting is important for correctly interpreting your data. All information captured will be kept strictly confidential.

Education level:

- Less than high school
- High school graduate
- Associate's degree
- Bachelor's degree
- Master's degree
- Ph.D./Professional degree

[12] Schneider TR, Jessup S, Stokes C, Lohani M, McCoy M. The influence of trust propensity on trust behaviors. Poster presented at APS; 2017 May; Boston, MA.
[13] Jessup S, Schneider T, Alarcon G, Ryan T, Capiola A. The measurement of the propensity to trust automation. Springer International Publishing; 2019. p. 476–489. https://doi.org/10.1007/978-3-030-21565-1_32

Appendix C. Principal Component Analysis (PCA) Loadings on the Individual Knowledge, Skills, and Behaviors (KSBs)

Each color represents strength of loadings (absolute value) on each PCA with dark green, bold lettering denoting loadings that were considered the strongest contributors to variance. Table C-1 displays the full PCA loadings for PCA1–PCA5.

Table C-1. PCA Loadings

Loading contributors	PCA1	PCA2	PCA3	PCA4	PCA5
AI Usage (z-scored)	0.034	0.044	0.078	0.101	**0.568**
Education	0.121	0.083	0.054	0.170	0.100
XAI Trust Scale	0.030	0.081	0.289	**0.532**	0.154
Psychological Ownership	0.061	0.027	0.451	0.249	0.168
Propensity to Trust in Intelligent Agents	0.129	0.117	0.265	0.224	0.260
Tech Self Efficacy Scale 1	0.102	0.108	0.350	0.382	0.066
Tech Self Efficacy Scale 2	0.215	0.106	0.022	0.242	0.262
BFI Openness	0.138	0.143	0.234	0.154	0.146
BFI Conscientiousness	0.075	0.354	0.117	0.068	0.198
BFI Extraversion	0.140	0.234	0.253	0.300	0.058
BFI Agreeableness	0.006	0.259	0.121	0.071	0.280
BFI Neuroticism	0.102	0.346	0.247	0.247	0.001
Cognitive Flexibility Inventory Total	0.154	0.416	0.244	0.011	0.115
Cognitive Flexibility Inventory Alternatives	0.137	0.280	0.372	0.150	0.123
Cognitive Flexibility Inventory Control	0.115	0.413	0.012	0.186	0.064
AI Literacy Scale Technical Knowledge	0.351	0.105	0.049	0.144	0.055
AI Literacy Scale Human Action	0.338	0.087	0.011	0.167	0.078
AI Literacy Scale Input	0.327	0.143	0.024	0.033	0.098
AI Literacy Scale Processing	0.319	0.140	0.002	0.028	0.145
AI Literacy Scale Output	0.293	0.129	0.031	0.038	0.027
AI Literacy Scale Usage	0.241	0.133	0.125	0.124	0.035
AI Literacy Scale Design	0.255	0.143	0.235	0.033	0.003
AI Literacy Scale Overall	0.369	0.095	0.088	0.021	0.063
SOT Average Angular Error	0.071	0.130	0.150	0.225	**0.505**

Notes: AI = Artificial Intelligence; XAI = Explainable AI; BFI = Big Five Inventory; SOT = Spatial Orientation Task

List of Symbols, Abbreviations, and Acronyms

ADAPTS	AI for Dynamic Pathfinding in Threatening Scenarios
AI	Artificial Intelligence
ARI	Army Research Institute
ARL	Army Research Laboratory
BFI	Big Five Inventory
DEVCOM	U.S. Army Combat Capabilities Development Command
DYNAMO	Dyadic Navigation with AI Systems for Mission Operation
FOE	Future Operating Environment
FY	Fiscal Year
GenAI	Generative Artificial Intelligence
GRiPS	General Risk Propensity Score
GWU	George Washington University
HRPP	Human Research Protection Program
ICC	Intraclass Correlation Coefficient
IHMC	Institute for Human and Machine Cognition
IRB	Institutional Review Board
JCS	Joint Cognitive Systems
KSB	Knowledge, Skills, and Behaviors
KST	Knowledge, Skills, and Tasks
LANDER	Landing AI-Navigated Drones in Environments of Risk
M	Mean (Average)
MOS	Military Occupational Specialties
NSU	Nova Southeastern University
PCA	Principal Component Analysis
SD	Standard Deviation
SDL	Self-Directed Learning
SME	Subject Matter Expert
SOT	Spatial Orientation Task

TF	Technological Fluency
TP	Technological Proficiency
TSAVVY	Technically Savvy Soldier
UAS	Unmanned Aerial System
UEC	User Engagement Characteristic
XAI	Explainable AI

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
1 (PDF)	DEVCOM ARL FCDD RLB CB TECH LIB