



ARL-TR-10070 • MAR 2025



AI in Command and Control: The Role of Transparency in Human–AI Planning

by Shan G. Lakhmani

DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

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.



AI in Command and Control: The Role of Transparency in Human–AI Planning

Shan G. Lakhmani

DEVCOM Army Research Laboratory

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
March 2025	Technical Report	START DATE 1 May 2024	END DATE 31 December 2024
4. TITLE AND SUBTITLE AI in Command and Control: The Role of Transparency in Human–AI Planning			
5a. CONTRACT NUMBER		5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
5d. PROJECT NUMBER		5e. TASK NUMBER	5f. WORK UNIT NUMBER
6. AUTHOR(S) Shan G. Lakhmani			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) DEVCOM Army Research Laboratory ATTN: FCDD-RLA-FD Orlando, FL 32826			8. PERFORMING ORGANIZATION REPORT NUMBER ARL-TR-10070
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 ORCID ID: Shan G. Lakhmani, 0000-0001-6052-439X			
14. ABSTRACT To support operational performance in the context of command and control (C2), partnerships between humans and AI systems have been proposed. To maximize the benefits of human–AI partnerships, however, AI partners must meet the cognitive and informational needs of their human partners; these AI partners must be transparent. This report describes how transparent systems have led to benefits in important operational outcomes, including performance, trust, and situation awareness. The goal of this report is to advocate for the design and implementation of transparent AI for C2 by identifying other contexts where transparent AI systems have led to positive outcomes and establishing parallels for possible implementation in the military decision-making process and C2.			
15. SUBJECT TERMS Humans in Complex Systems, command and control, transparency in artificial intelligence, human-autonomy teams, performance, trust in artificial intelligence, situation awareness			
16. SECURITY CLASSIFICATION OF:		17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT UNCLASSIFIED	b. ABSTRACT UNCLASSIFIED	c. THIS PAGE UNCLASSIFIED	UU 33
19a. NAME OF RESPONSIBLE PERSON Shan G. Lakhmani			19b. PHONE NUMBER (Include area code) (407) 208-3432

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

Contents

1. Introduction	1
2. The Case for Transparency in C2	3
3. Performance	4
3.1 Sensemaking and Creating Plans	5
3.2 Narrowing Down Options and Making a Choice	7
3.3 Replanning and Adjustment	8
4. Trust	9
4.1 Analytical Processes	9
4.2 Affective Processes	11
5. Situation Awareness	12
5.1 Transparent AI and Shared SA	14
5.2 Maintaining SA of a Transparent AI	15
6. Conclusion	16
7. References	19
List of Symbols, Abbreviations, and Acronyms	27
Distribution List	28

1. Introduction

Existing approaches to command and control (C2) use training, tactics, and procedures to enact a commander's intent across the battlefield at large, using the myriad of forces under their command to accomplish the goals of the mission (Kase et al. 2022; McDowell et al. 2024). However, these battlefields are ill-structured, increasingly complex, and feature interactions across more domains than ever before, and they are only anticipated to become more so in the future (Goecks and Waytowich 2024; McDowell et al. 2024). As a result, complex systems—such as AI, robotic technology, and others—are being developed to maintain overmatch against adversaries that are met on these battlefields (Schaefer et al. 2015; McDowell et al. 2024).

Future AI systems can be used to maintain overmatch in multiple ways. One way is through delegation. Some of AI's strengths include quickly processing large amounts of data, performing repetitive tasks, and multitasking (de Winter and Dodou 2014); therefore, delegating those kinds of tasks to AI systems frees humans to focus on other tasks that better suit their capabilities, such as complex decision-making (Chen and Barnes 2014). Another way future AI systems can be used to maintain overmatch is through the facilitation of teamwork. C2 is characterized by teams of people working together to fulfill a commander's intent (McDowell et al. 2024). Teams use teamwork behaviors (i.e., communication and coordination) (Salas et al. 2015) to pursue their shared goal of fulfilling a commander's intent. AI can support these behaviors by aiding communication and passing relevant information to appropriate team members (Sycara and Sukthankar 2006) or using visual cues to help maintain pertinent mutual knowledge between teammates (Bradshaw et al. 2009) to facilitate coordination. A final way that future AI systems can be used to maintain overmatch is through partnership with humans. Humans and AI each have independent, but often overlapping, breadths of capabilities. By partnering together, humans and AI use their capabilities to support not only themselves but their partners as well; this human–AI interaction results in a joint capability that is greater than the sum of a human's and AI's individual capabilities (Metcalf et al. 2021; Sycara and Sukthankar 2006).

To take best advantage of future AI systems to maintain overmatch, however, these AI systems must be designed to be human-centered, supporting human cognitive needs (Endsley 2023). If AI is designed and implemented without the knowledge of how it can best support its human partners, then, ironically, the addition of AI, which is meant to aid the performance of required tasks, may hamper that performance. Without accounting for the specific information that humans need

during a task, or the modalities of the user interface they are using, the addition of AI can add to the amount of information that must be considered, increasing both task complexity and cognitive load (Lakhmani et al. 2019; Bläsing and Bornewasser 2021). Without appropriate feedback concerning the behavior of the AI, especially in situations in which people are cognitively burdened, they will more readily delegate responsibility to AI systems or accept information from AI systems when inappropriate (Hinds et al. 2004; Lee and See 2004). Finally, not taking human informational needs into account can lead to situations in which the AI pushes humans out of the informational loop, causing them to lose track of the status of the AI (Bolstad et al. 2006). This can result in out-of-the-loop performance problems, that is, a diminished ability to detect system errors and perform tasks manually due to a loss of situation awareness (SA) (Endsley and Kiris 1995; Endsley 2017). These potential shortcomings of AI integration can be mitigated, or even negated, through the design and careful implementation of transparent AI (Chen et al. 2018).

Transparency can refer to a quality of a system—or AI in this instance—that allows it to convey information about its intent, awareness of the environment, reasoning process, and future plans (Chen et al. 2014; Lyons and Havig 2014). A transparent system, one that has the quality of transparency, can yield fewer out-of-the-loop performance problems than their more-opaque counterparts by facilitating human oversight of the system, greater SA of the system and concomitant environment, and a more calibrated level of trust in the performance of the system (Chen and Barnes 2014; Endsley 2023). With the added complexity arising from the human–AI partnership, rather than just implementation of automation, transparency would have to take into account not just the shared capabilities of humans and AIs (Metcalf et al. 2021), but also the common ground shared between human and AI partners (Bradshaw et al. 2011). Models of transparency, such as the dynamic Situation Awareness-based Agent Transparency (SAT) model, can be used to guide the implementation of systems that take into account these overlapping areas of information and responsibilities (Chen et al. 2018). Careful implementation of transparency in human–AI partnership is particularly important when AI systems are facilitating both the monitoring of complex, dynamic environments and the use of that information to make decisions in accordance with a commander’s intent (Chen et al. 2018; McDowell et al. 2024).

C2 is currently a human-dominated process. To deal with the operational challenges anticipated on the battlefield of the future—such as dealing with denied, degraded, intermittent, and limited (DDIL) communications and having to juggle multiple data streams across multiple domains—robust partnerships between humans and AI systems will be needed to maximize the capabilities of both partners and

maintain overmatch against adversaries (Metcalf et al. 2021; McDowell et al. 2024). Collaboration between human and AI partners is expected to yield dynamic interactions and emergent capabilities that can exceed the individual actions and capabilities of the human or AI alone (Metcalf et al. 2021; McDowell et al. 2024). This partnership is meant to lead to more efficient decision-making, faster coordination with less required direct communication, reduced load on working memory, and decreased operating costs (Kilgore and Voshell 2014; Kase et al. 2022; McDowell et al. 2024). However, partnership requires learning from each other, mutual adaptation, and understanding of one's partners, which, in human teams, requires some level of communication (Salas et al. 2015; Cooke et al. 2016). To gain the benefits of a partnership between humans and AIs, the AIs need to be able to communicate information they have in a way conducive to joint work (Bradshaw et al. 2011; Lyons 2013). Essentially, getting the most benefits from human–AI partnerships in C2 requires those AI partners to be transparent.

2. The Case for Transparency in C2

To perform effectively as individuals and as teams (Endsley 2023), people need to have an appropriate understanding of the task they are doing, the context in which they do it, and individual agents (e.g., teammates, AI systems, and enemy actors) acting within the environment (Mathieu et al. 2000; DeChurch and Mesmer-Magnus 2010). AI can facilitate individual performance (Sycara and Sukthankar 2006) and act as partners to their human counterparts (You and Robert 2017); however, to unlock the synergy between human and AI partners—especially in the context of C2 (Metcalf et al. 2021; McDowell et al. 2024)—transparent AI is crucial (Chen et al. 2018). An opaque AI partner can lead to negative outcomes. A lack of transparency in AI can lead an operator to have false expectations of the AI, leading to neglect or disuse of the AI when use is warranted (Ishak and Nathan-Roberts 2015; Lyons et al. 2016). A lack of transparency in AI can also yield out of the loop performance issues, where AI's human partners have degraded SA, reduced speed when detecting problems, and slower intervention when problems *are* detected (Chen and Barnes 2014; Endsley 2017). However, a transparent AI can communicate relevant information to their human partners, which can have positive outcomes. A transparent AI can communicate meta-information about the information being used (e.g., uncertainty, information context, or credibility) and can positively influence decision-making, workload, and performance (Pfautz et al. 2006; Helldin et al. 2013). With knowledge about an AI system's reliability, confidence, the degree to which parameters are met, and other information about the AI itself, humans can more effectively partner with AI (Helldin et al. 2013; Lyons 2013). A transparent AI can also aid its human partner's SA by providing

insight into 1) whether the AI can perform in the current situation and 2) its ability to detect and incorporate relevant situational data into its decision-making processes. This performance reduces the likelihood of being surprised by the AI (Miller 2014; Endsley 2023).

The purpose of this manuscript is to emphasize the importance of transparency in the human–AI partnership and to advocate for its implementation. To do so, I identify how transparency can lead to positive outcomes in some common indicators of success: performance, trust in the AI system, and SA. Tasking in C2 is often centered around decision-making (Headquarters, Department of the Army 2022; Kase et al. 2022), so I discuss how transparent systems have facilitated positive outcomes in different phases of decision-making. I also set expectations based on existing research. Trust in an AI system can modulate how people work with that system (Lee and See 2004); therefore, I discuss how transparency can be used to set suitable expectations so that AI can be used appropriately. A transparent AI can also facilitate an awareness and understanding of the battle space and the actors within it, which can yield faster and more informed decisions (Endsley et al. 2003; Endsley 2023), so I discuss how transparency can be used to support one’s SA. After discussing how transparency can influence these three outcomes, I conclude this report with a discussion on the importance of human-centered design of AI and how this approach can benefit the human–AI partnership in C2.

3. Performance

In the C2 space, performance is oriented around the decision-making process; therefore, when discussing how transparency can affect performance, I primarily focus on the decision-making process. In the current, primarily analog approach to C2, smooth operation is dependent on a diverse team of experts in a command post adhering to well-established processes to make decisions, networking to convey those decisions to the relevant parties, and communicating across the group to establish sufficient understanding to maintain unity of effort (Headquarters, Department of the Army 2022). Introducing AI into the decision-making process is expected to facilitate the speed and quality of decision-making in C2. AI can quickly gather and process information from a wide array of sources in a variety of environments so that plans can be constructed (Kauppe et al. 2010; Endsley 2023). AI can facilitate faster and more informed decisions by mitigating uncertainty and helping maintain shared SA throughout the process (Pfautz et al. 2006; Ososky et al. 2014). However, the goal is not to replace humans with AI but to create effective human–machine partnerships that leverage the strengths and abilities of humans and machines.

Although the presumption of *fixed* human strengths and fixed machine strengths is, at best, an oversimplification, humans and machines *do* have independent strengths and capabilities, some of which overlap (Dekker and Woods 2002; Metcalfe et al. 2021). Although both humans and machines can have the general capability, humans are considered stronger when it comes to integrating new information into the decision-making process, using flexible procedures, and exercising judgment (Dekker and Woods 2002; Bradshaw et al. 2011). Conversely, although both humans and machines can have the general capability, machines are considered stronger when it comes to carrying out long sequences of actions in predetermined ways, deductive reasoning and computation, and responding to control signals with precision (Bradshaw et al. 2011; de Winter and Dodou 2014; Metcalfe et al. 2021). Humans, working with AI, can stay informed of ongoing events and computationally instantiate their models of the world; AI, working with humans, can keep their models aligned to context and adapt to unexpected changes in the world; human-machine partnerships not only leverage their own strengths to accomplish their shared goals, but they are able to use their own strengths to enhance *each other's* strengths (Bradshaw et al. 2011; Metcalfe et al. 2021).

Human-AI partnerships can be more responsive to changes in goals, the working environment, or operational challenges—so any replanning or retasking needed to achieve the desired end-state can be done quicker (Sycara and Sukthankar 2006; Endsley 2023). To take advantage of the shared capabilities, human and AI partners must work interdependently (Metcalfe et al. 2021). This joint work requires that the partners have a compatible (although not necessarily identical) knowledge base, both about the environment and the agents (human *and* AI) within it (Cannon-Bowers and Salas 2001; Bradshaw et al. 2009). However, maintaining this common ground requires transparency on the part of both the human and AI agents (Chen et al. 2018). Next, I discuss how transparency has been used in human-machine interactions to facilitate gathering information and making plans, narrowing down options and making decisions, and adjusting plans and replanning.

3.1 Sensemaking and Creating Plans

Once the Commander's Critical Information Requirements and intent are established, the next step for the C2 staff is to establish an initial understanding of the situation and the problems at hand (corresponding to mission analysis) and identifying potential solutions to those problems (corresponding to course of action development) (Marr 2001). One approach to understanding a problem space and identifying possible solutions is the Playbook approach. This approach uses the metaphor of a sports team's playbook, where a coach uses predefined but easily modifiable templates to convey their goals and specific behaviors that need to be

completed (Miller and Parasuraman 2007; Mercado et al. 2015). A play is a familiar structure to organize information around, it can model a complex task or situation, and it does not require one to understand the entire task model to use it effectively (Miller and Parasuraman 2007; Miller 2014). By using plays, human–AI partners can quickly and flexibly switch between multiple complex sets of predetermined actions that they both understand (Stowers et al. 2016; Handelman et al. 2023). This approach allows for faster decision-making that can be more reactive to the changing needs of a dynamic battlefield. Using a simulated heterogeneous multi-unmanned vehicle planning task, researchers examined a transparent playbook approach with a simulated plan recommendation engine (Mercado et al. 2015; Chen et al. 2018). Rather than solely providing research participants with plan information and a choice, researchers also provided meta-information explicitly addressing the factors that led to a particular recommendation and projected outcomes of each plan (Mercado et al. 2015). They were able to visualize the factors that undergirded plan success (i.e., vehicle speed, sensor coverage, and vehicle capabilities), the paths that unmanned vehicles would take through relevant environmental constraints, and meta-factors that could impact expected performance (e.g., unknown interaction with environmental constraints and potentially out-of-date information). Providing this additional information led to the selection of the better plan more often than when research participants were solely given basic plan information (Mercado et al. 2015; Chen et al. 2018). Having simulated AI systems provide information about the reasoning underlying its recommendation and potential uncertainties that color those recommendations allowed research participants to cross reference that information with what they already knew. This agent transparency led to the opportunity for partnership between human and machines agents, leading to more successful outcomes. In this instance, a transparent system using a playbook approach led to more accurate decision-making; this increase in accuracy is expected to occur in the C2 space, when transparent AI is implemented.

Another way that AI can facilitate the decision-making process is by summarizing information it has processed using a large language model (LLM). Similar to a human communicating information to a partner to build a shared understanding of the current situation, ChatGPT—a popular LLM—can summarize literature, which allows researchers to gain an initial understanding of a wide field of information relatively quickly (Cooke et al. 2017; van Dis et al. 2023). ChatGPT can be trained on a specific literature, such as military doctrine, and produce output based on that specific literature (Min et al. 2023; Goecks and Waytowich 2024). In game-based, military-relevant scenarios, COA-GPT—an LLM designed to act as a military commander’s assistant—was able to generate actionable COAs that could be successfully updated in response to human feedback. COA-GPT is expected to lead

to more rapid and informed decision-making than manual COA generation alone. Although LLMs are powerful tools that can enhance functionality, they are complex systems whose inner working mechanisms can be opaque, which can yield negative repercussions (Gunning and Aha 2019; Bünger 2024; Zhao et al. 2024). An LLM that takes into account its human partner must take a more transparent approach to human–AI interaction. An LLM that can trace its inferential steps from initial prompt to COA and present that process in human-understandable terms may yield the benefits of increased interpretability (Gunning and Aha 2019; Israelsen and Ahmed 2019). This can include a greater understanding of the LLM’s capabilities and limitations, easier identification of hallucinations or inappropriate content, as well as insight into biases, risks, and areas for improvement (Zhao et al. 2024). Focusing on how different actions support different priorities (Mercado et al. 2015) and providing surface-level information about why a plan was chosen with the option of in-depth information (Wright et al. 2020) are potential strategies for implementing transparency into this kind of AI. LLMs with greater traceability can lead to more successful outcomes by allowing faster decision-making while mitigating potential drawbacks of hallucinations. Understanding where the LLM gets its information can also help human partners build an understanding of relevant areas in the physical environment and help inform their trust in the AI.

3.2 Narrowing Down Options and Making a Choice

Once a problem has been defined and several potential solutions have been put together, one has to narrow down these options and decide which of these solutions best address the problem. This aspect of decision-making corresponds to COA comparison (Marr 2001).

Choosing from several options can be difficult, but AI systems can be used to make that process easier. AI can be used to recommend one of a series of choices, rank plans in terms of likelihood of success, or bin options in terms of important factors or pre-established metrics. These decisions can be automated through the use of conditional logic, where the system can take particular actions in response to specific circumstances that trigger them (Parasuraman et al. 2000; Ur et al. 2014). As the number of trigger-action rules increase, the humans making the rules will have a harder time envisioning the outcomes; using projected outcome transparency (SAT Level 3) (Chen et al. 2014), AI can provide projected, expected outcomes, reducing the number of items that a human would have to keep in mind and concomitantly reducing their cognitive burden. Presenting projected outcomes not only facilitates human trust in AI, but it directly benefits SA (Endsley 2017; Metcalfe et al. 2021).

When there is an abundance of information, recommender systems, such as Amazon’s book recommender or Pandora’s personalized radio station, are another way that AI can make decision-making easier. Recommender systems reduce decision paralysis by giving recommendations based on user preferences, similarity between users, or user history (Cramer et al. 2008). Transparency in recommender systems requires users to understand why a particular recommendation was made (Kim and Hinds 2006), so describing the factors that led to the recommendation is expected to increase the likelihood that the recommendation is adopted. For example, an AI can recommend different uncrewed vehicles to assist with a search and rescue task; a transparent system can match each vehicle with the factors that led to its recommendation (Mercado et al. 2015). A transparent AI recommender might suggest two vehicles: one is chosen due to its speed, whereas the other is chosen for its offensive capabilities. With the information provided by the transparent AI, the human partners would be able to use their judgment and choose a faster vehicle, if they anticipate a search and rescue task. Transparent AI, in this example, allowed human partners to make a faster choice (because they were between two, rather than more, vehicles) and a more well-informed choice (because they were able to see the factors that led to the vehicle recommendations).

3.3 Replanning and Adjustment

In an ideal setting, once a solution is implemented, it may be completed successfully. However, battlefield environments are hardly ideal, and, as such, rapid replanning is often necessary. This aspect of decision-making corresponds to COA execution and assessment, to some extent, although it also corresponds to the larger continual process that must be done when the situation demands a reassessment, which, given the dynamic nature of the battlefield and the vast swaths of ever-changing information that needs to be considered, happens often (Marr 2001).

The first part of rapid replanning is recognizing that the plan needs to be adjusted in the first place. Maintaining a constantly updating understanding of the situation (SA) allows one to detect when a change in plan may be needed, but human and AI partners, across hierarchical and peer groups, are needed to avoid the challenges of maintaining it (e.g., attentional narrowing, data overload, spurious models of the situation) (Endsley and Connors 2008; Endsley 2023). Furthermore, disseminating this understanding to other staff members, leaders, and other echelons is important to maintain SA among a larger group (Endsley 2023), but the act of communication can have an “overhead” cost, especially if done so inefficiently (MacMillan et al. 2004). AI can be used to help maintain an accurate and up-to-date SA, as well as communicating this awareness among team members (Sycara and Sukthankar

2006). The maintenance and distribution of SA, using transparent AI, is discussed in greater depth in Sections 5, 5.1, and 5.2.

Minor changes in the environment can result in minor changes to a plan. However, given the complex nature of the battlefield, small changes on the battlefield can rapidly escalate, leading to disruption in mission plans, communication, and even to the overall battle effort (Kasdaglis et al. 2014). AI can use sensors to detect changes in the battlefield and quickly respond to them (Bradshaw et al. 2011; Kilgore and Voshell 2014). Transparent AI can make their human partners aware that its change in behavior is a result of a detected change in the environment (e.g., the physical field, the equipment being used, Blue Force [BLUFOR] positioning, and opposing force [OPFOR] activity). Without transparency, the AI's change in behavior would seem wildly inappropriate, until its human partner notices the stimulus that elicited the change in AI behavior (Stubbs et al. 2007). Conversely, humans may notice a change in the environment that the AI has not (Bradshaw et al. 2011). In this case, the human must introduce that change to the AI; for example, humans using COA-GPT were able to examine the COA that the AI output, and then they were able to provide additional information that the AI did not have to update the COA to something more appropriate (Goecks and Waytowich 2024). Transparency helps prevent situations in which AI machines do unexpected, difficult-to-explain things, which helps prevent confusion and helps human-machine partners make more well-informed decisions (Miller 2014).

4. Trust

Although trust is a weighty concept, with numerous factors and contexts, it can generally be defined as the attitude that an agent will help individuals achieve their goals in a situation characterized by uncertainty and vulnerability (Mayer et al. 1995; Lee and See 2004). Trust is based on three factors: 1) ability, the competency of the trustee to influence the domain; 2) integrity, the degree to which the trustee adheres to an acceptable set of principles; and 3) benevolence, the extent to which the intent of the trustee aligns with that of the truster (Mayer et al. 1995; Lee and See 2004). This emergent property, generated by the interdependence and interaction between entities, is influenced by analytical and affective processes (Lee and See 2004; Ososky et al. 2014). As such, transparency influences trust through facilitating these processes.

4.1 Analytical Processes

One aspect of trust is expectation of capability. Although an AI would ideally perform perfectly at all times, that is rarely the case, especially when it is acting in

a dynamic, complex environment. The extent to which one should trust an AI, and consequently the extent to which one should rely upon its output, should be calibrated (Lyons 2013). Trust calibration refers to how closely a human's level of trust corresponds to the capabilities of an AI; when a person's trust in the AI matches its capabilities, it is referred to as calibrated trust (Lee and See 2004). For example, the facial recognition feature on your phone used to unlock it is a simple AI. If you know that it works best under bright lights, and thus you only use it in those conditions, then you have calibrated trust with your phone's face recognition AI. On the other hand, trust miscalibration leads to over trust and under trust (i.e., distrust), both of which can impact appropriate system use.

Over trust refers to when a person's trust in the AI exceeds its capacity, which can result in misuse or using the system beyond its capabilities (Lee and See 2004). One example of an AI that tends to trust a little too much is the autocomplete on one's phone. This AI can correct spelling errors and allow for faster textual communication, but one still must review the output or else risk an unwanted or inappropriate message. Distrust refers to when a person's trust in the AI falls short of its capabilities, which can result in disuse of the systems when the system is capable. One example of this in everyday life is the AI in cars that assists with parking. For many, parallel parking is a trial, so the use of parking assistance—either camera-based parking guidance systems or more advanced systems that use multiple sensors and the car's own steering system to park itself with minimal help from the driver (Hearst Autos n.d.)—can be a boon. However, if drivers do not know or do not trust this system, then they will lose time and damage their car attempting to parallel park, even when a technological solution to this problem exists.

In the real world, performance and errors are not randomly distributed. In a decision-making capacity, AI will make different kinds of errors depending on its perceptual capabilities, its classification algorithms, and its other capabilities (de Visser et al. 2014). A transparent AI, when it detects that an error exists, can indicate that fact to their human partner. If it misses its own errors, however, it can still indicate contexts and circumstances where its strengths are limited. A key challenge when using AI is perceptual limitations that undermine its performance (Endsley 2023). For example, an AI may have difficulty detecting flying targets when it rains; if a target flies by during a rainstorm, the AI would miss it and not say anything. However, if the issue of perceptual difficulties during rainstorms is a known issue, then a transparent AI can indicate this weakness, alert its human partner to inclement weather, and a human partner can supplement the AI while it rains. Continuing with this example, if an AI can detect flying targets in the rain, but the rain reduces the likelihood of a successful categorization, then a transparent

AI could use visual indicators of confidence or uncertainty, giving its human partners information they can use to make a judgment that the AI could not (Endsley et al. 2003; Pfautz et al. 2006). Transparency allows human partners to identify when the AI needs help and allows human and AI partners to complement each other's strengths and weaknesses (Stubbs et al. 2007; Metcalfe et al. 2021).

In sum, a transparent AI conveys information to its human partner that lets them establish a more accurate understanding of how it works in different contexts and what expectations humans should have concerning the kinds of outputs it can provide (Onal et al. 2014; Ososky et al. 2014). This understanding allows human partners to calibrate their trust in the AI in general, as well as in different contexts, increasing the likelihood that human partners will take advantage of the AI's capabilities appropriately (Lyons 2013).

4.2 Affective Processes

Although there is a strong analytic process involved in trusting an AI, affect represents the core influence of trust on behavior (Lee and See 2004). In situations in which people do not have the historical precedent to analyze when making the judgment to trust an AI, they are much more dependent on affective signals (Karimov et al. 2011). Therefore, an AI has to provide these kinds of signals in order to seem trustworthy. Some of these can be more superficial features, such as appearance or mode of communication (Schaefer et al. 2016). For embodied AI, physical appearance (Robinette et al. 2016) or physical proximity to a human teammate (Henkel et al. 2014) can affect feelings of trust in an AI. Modes of communication influence trust; therefore, given that music and prosody may better influence affective response, using sounds (e.g., beeps or clicks) or verbal responses can influence the feeling of trust (Lee and See 2004; Schaefer et al. 2016).

However, using these signals—that can influence a human's feelings of trust—does not necessarily *make* a system more trustworthy. For example, a slot machine can have an exciting and interactive interface, with smiling cartoon characters, catchy music, and slick infographics about the high odds of a payout. These features hide the slot machine's underlying purpose, that is, parting the slot players from their money, behind a veil of faux trustworthiness. An AI's features must be tied into its trustworthiness, or else its human partners may feel betrayed when the AI does not live up to the standards it set with its trust indicators. This feeling of betrayal can hamper trust in the system in the long term (Yagoda and Gillan 2012).

Using text-to-speech or human voice audio as an output can yield greater feelings of trust in the system using them (Karimov et al. 2011). This human-like interaction with machines can engender human-like expectations (Perzanowski et al. 2001). If

the machine cannot meet these human-like expectations, however, humans will notice these shortcomings and exhibit a marked decrease in trust (Lee and See 2004; Merritt et al. 2013). If an AI's indicators of trust do not match its trustworthiness, then people will get frustrated with it and eventually stop using it (Ishak and Nathan-Roberts 2015). Transparency can link affective trust with the system's trustworthiness by confirming one of the bases of trust: benevolence, which is defined as the extent to which the intents and motivations of the truster and trustee align (Lee and See 2004; Lyons 2013; Feitosa et al. 2020).

Humans readily attribute intent to the actions of nonsentient actors (Ososky et al. 2014); therefore, it is important for the design of the AI to clearly convey the intended functionality behind its creation (Lyons 2013). The presentation of the AI has to reflect the purpose behind it. An AI meant for social interactions would benefit from a human-like avatar and naturalistic voice responses, to not only reflect that social purpose but to engender human-like treatment (de Visser et al. 2014; Lakhmani et al. 2016). AI can communicate that its intent matches that of its human teammates, meeting the standard of benevolence and facilitating affective trust (Lee and See 2004; Lyons 2013). An AI, by associating the specific tasks it is pursuing with the goals it is trying to accomplish, can facilitate trust by making sure the human and AI are on the same page (Lyons 2013; Lyons et al. 2016); if they are not on the same page, then the human, seeing the AI goals mismatching the human goals, can adjust the AI's goals, restructuring the AI's tasks so they meet the desired needs (Chen et al. 2018). Without this extra transparency information about AI capabilities, goals, and explanations of failures, expectation violations can feel like a betrayal and hamper one's willingness to trust in the system as one once did (Lyons 2013; Krausman et al. 2022). With the extra transparency information, the AI seems more trustworthy and thus, when cognitive resources are limited, the AI is more likely to engender affective trust (Hoff and Bashir 2015).

5. Situation Awareness

SA is defined as “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future” (Endsley and Jones 1997). According to Endsley's (1995) popular model of SA, this diagnosis is made up of three levels: 1) sensory perception of cues within the environment, 2) comprehension of how perceived cues will impact one's goals and objectives, and 3) projection of how this information will affect future states of the operating environment (Bolstad et al. 2006). Using these three levels, each individual establishes a situation model, that is, a specific understanding of the current situation (Cooke et al. 2000). Each person's individual situation model is different, but if team members wish to

collaborate, each team member's disparate situation model needs to have some level of commonality (Bradshaw et al. 2004). To establish this level of commonality, team members can use processes, such as communication or coordination, to share unique information and build a shared understanding (Endsley 2015; Salas et al. 2015). The degree to which team members have the same SA on shared SA requirements is known as shared SA (Endsley 2015). In human-AI partnerships, AI partners need to be able to establish a shared SA with their human partners, such as in human-human teams, to collaborate effectively.

Although AI are more robust against real-world complexities than they once were (Gunning and Aha 2019), brittleness in the face of new types of situations is still a challenge (Endsley 2023). Human partnerships with an AI can shore up that brittleness problem (Bradshaw et al. 2011), but challenges arise when humans lose track of the AI machine's activities, fail to realize when AI machines act incorrectly, and misapprehend why the AI is acting the way it does (Endsley 2017). When humans are divorced from the task, or they fail to spot when an AI is working on that task has erred, then they can fall out of the loop. The out-of-the-loop performance problem refers to an issue in which a human working with automated systems (which can include AI) can lose track of the elements they are responsible for but are under the system's control (Bolstad et al. 2006; Endsley 2017). This can lead to negative outcomes, such as increased error potential, higher cognitive load, and lower SA (Stubbs et al. 2007; Grote et al. 2014; Kilgore and Voshell 2014).

Transparent AI can be used to help mitigate this out-of-the-loop problem by helping humans maintain an active awareness of the AI and the unique information that it can hold (Stubbs et al. 2007). One approach to transparency is the SAT model, which is specifically designed to help shape human-AI interactions that keep humans in the loop (Chen et al. 2014). The SAT model identifies what information is needed to support a human's SA of their AI collaborator (Chen et al. 2014). Corresponding to Endsley's (1995) three levels of SA, the SAT model identifies three levels of information: 1) the AI's current state and goals (SAT level 1); 2) the AI's reasoning process behind its actions and the constraints under consideration (SAT level 2); and 3) the AI's projection of the future state, including expected outcomes, probabilities of success, and limitations (SAT level 3). Having accurate SA of an AI's actions, reasoning, and its effect on the decision space helps keep humans in the loop and can prevent the negative effects that can occur if they fall off (Chen et al. 2018).

AI is good at learning from extremely large amount amounts of data (e.g., movie streaming services analyzing ratings, streaming data, and movie meta data to feed into a personalized recommendation system) (Israelsen and Ahmed 2019), and, if properly designed, it can help keep humans informed of ongoing events (e.g., blind

spot detection systems in cars that continuously scan the areas alongside and behind one's vehicle, alert drivers when an object enters that space, and identify vehicles and objects in one's blind spots) (Bradshaw et al. 2011). Keeping humans informed—and thus in the loop—requires two kinds of information. First, humans working with an AI partner need to maintain shared SA regarding the goals and environment they all share: this environment includes the entirety of the battlefield representation, including physical and digital space, BLUFOR and OPFOR status and location, and resources. Second, humans need information about the AI's reasoning process and expected outcomes that result from its recommendations or actions (Chen et al. 2014). Transparent AI can help humans maintain 1) shared SA with their AI partners, 2) an awareness of what the AI is doing, and 3) an understanding of why AI produced the output that it produced.

5.1 Transparent AI and Shared SA

One of the benefits of working with an AI is that, by delegating work to it, humans can save time and effort (Miller 2014). However, doing the work manually has its own set of benefits that one no longer gets if one delegates the task, partially or totally, to an AI. For C2, manually putting together a COA allows one to build a rich understanding of the environment (as previously denoted), logistics, and meta-information (e.g., sensor reliability, historical trends, expectations of lag or latency) (Pfautz et al. 2006). Delegating some of the tasks needed to put together a COA to an AI saves human partners time and effort, but human partners would not build the level of understanding that they would have built if they did the task manually. Therefore, AI needs to provide humans information about the environment relevant to the human–AI partnership as a whole to ensure that human partners stay in the loop.

As previously discussed, humans develop an individualized understanding of situations, and that understanding does not include information processed by AI, unless that information is explicitly shared with them (Lyons 2013; Endsley 2015). An AI sharing all information with its human partner would yield cognitive overload, defeating the purpose of having AI in the first place (Miller 2014); therefore, a transparent AI would have to adapt to its human partners' informational needs (Karsenty and Botherel 2005), providing the data necessary to meet the human and AI's shared SA requirements (Endsley 2015). Meeting shared SA requirements has been predictive of team performance (Endsley 2023).

AI can provide this necessary information in several different ways. If the human needs to build an initial understanding quickly, an AI can use an LLM to summarize relevant information (Goecks and Waytowich 2024). If a human is maintaining

awareness of an environment (e.g., logistics, battlefield representations, and BLUFOR and OPFOR information) but has not built a deep understanding of the environment, the sensors, and all the intricacies that are built with experience, then an AI can supplement that understanding by bringing certain meta-information (e.g., interaction effects between particular actors, information context, source reliability or credibility, relevance to mission goals or broader operational context, and temporal qualifiers) into the person's awareness (Pfautz et al. 2006).

5.2 Maintaining SA of a Transparent AI

AI can assist the decision-making process by filtering, managing, or processing data used to make decisions (Pfautz et al. 2006), narrowing down potential options (Parasuraman et al. 2000), or providing explicit recommendations for actions (Helldin et al. 2013). AI makes these decisions using a complex series of rules that, depending on the AI's design, may not match the process that a human uses to make decisions (Sukthakar et al. 2013; Felzmann et al. 2020). Depending on one's role in the decision-making process, however, one may need to understand how the AI made its decisions.

For humans who are directly collaborating with AI partners, they need a transparent AI that describes its processes to its human partner. The SAT model is one popular way to map out the kinds of information that would be needed (Chen et al. 2014). This would include information about an AI's current activities (e.g., is it filtering raw data, is it creating plans using data run through decision-making algorithms, and is it winnowing down existing plans to provide human partners with the three "best" choices), the processes that it used to produce those activities (e.g., filtering criteria for raw data, what data it used to inform plan-making and how those algorithms used that information to make plans, and how it determined what the best choices were), and how current actions affect future expectations (e.g., uncertainty regarding current and future filtered data, the scope of data used to create plans, and levels of certainty regarding selected plans' success). For collaborators, transparency does not require a full accounting of every assumption for every decision (Miller 2014). Too much information can lead to information overload (Helldin et al. 2013). Human collaborators with AI partners need enough information about AI systems to allow them to make accurate mental models of the AI and its future decisions (Sycara and Sukthakar 2006; Bradshaw et al. 2011). Mutual predictability has been important for the development of skilled human teams (Bradshaw et al. 2009) and for the maintenance of unity of effort in military teams (McDowell et al. 2024); therefore, the ability to be predictable to its teammates is a key capability for AI partnering with humans (Endsley 2023).

For humans in leadership positions or at higher echelons—who synthesize information from multiple sources to make larger decisions—transparent AI partners need be inspectable (Feltzmann et al. 2020). Human–AI partnerships can result in a wide range of capabilities (Metcalf et al. 2021), but final responsibility falls to human partners. As such, humans in these positions need to be able to audit AI decision-making, walking through the technical logic that led a decision (Feltzmann et al. 2020) and trace AI errors and misbehaviors (Wortham et al. 2017). Traceability in transparent AI allows users to debug possible AI errors (Wortham et al. 2017), adjust the AI’s activities or priorities (Chen et al. 2018), and accurately attribute accountability (Lyons 2013). Being able to do all of this allows leaders to do their due diligence and pursue the commander’s intent.

6. Conclusion

Establishing human–AI partnerships is being proposed to overcome anticipated challenges in future battle spaces (Kase et al. 2022). Designing AI with transparency in mind is expected to allow humans and AI to make the most of their respective capabilities and maximize the emergent capabilities that result from human–AI interaction (Metcalf et al. 2021; Kase et al. 2022). Transparency is anticipated to facilitate these outcomes by allowing human partners to “see into” the systems they use, allowing them to maintain appropriate expectations of their AI partners and accurately predict the AI’s actions (Karsenty and Botherel 2005; Endsley 2023; van de Merwe et al. 2024). As discussed in the report, transparency—in contexts adjacent to but different than C2—has led to improvements in performance, trust, and SA. These three outcomes are critical for successful human–machine partnership, and I have argued that transparent AI are expected to yield these same improvements in C2.

Performance—defined here as speed and accuracy of decision-making—is a desired outcome of human–AI partnership in general, and C2 specifically (McDowell et al. 2024; Steyvers and Kumar 2024). AI can facilitate decision-making by providing information about the environment and recommending COAs based off of that information. A transparent AI can lead to faster decision-making by providing meta-information to reduce uncertainty, which circumvents uncertainty mitigation strategies, such as forestalling, weighing pros and cons, and collecting more information (Lipshitz and Strauss 1997; Pfautz et al. 2006). Transparent AI can further yield faster, more accurate decision-making by providing reasoning information about recommended COAs; transparent AI can provide qualifications or explanations of its recommendations, so its human partner spends less time justifying an acceptance or finding a reason to reject the recommendation (Cramer et al. 2008). Therefore, transparent AI that takes human

informational needs into account is preferable to more opaque AI because it can lead to faster, more efficient decision-making.

Trust in these AI systems not only affects users' feelings towards the AI with which they work, but it also influences the frequency and appropriateness of that use (Krausman et al. 2022). Transparent AI makes it easier for users to align their use of the AI's outputs (e.g., acceptance of recommendations, unalloyed use of summarized data) with the conditions that best suit the AI's capabilities, by providing information that helps them understand the system's reliability in different contexts (Lee and See 2004; Lyons 2013). Given that human-centric features—such as multiform communication, group ideation, and repetition—have evolved to build trust in human partners and maintain unity of effort, AI-centric efforts, such as transparency, can do the same (McDowell et al. 2024).

SA allows one to stay responsive—not only to the environmental factors needed to create and update COAs, but also to the partners with whom they create and update COAs—with improved SA enhancing the decision-making process, teamwork, and coordination (Kase et al. 2022; Endsley 2023). By providing information that lets humans “see into” their AI partners' decision-making process, transparent AI can externalize the capabilities, statuses, goals, and actions that humans would normally have to intuit or observe; humans can more easily maintain shared SA if the AI provides some of the requisite information to them directly (Stubbs et al. 2007; Endsley 2015, 2023). Transparent AI can help maintain unity of effort under DDIL conditions by allowing mutual awareness between human–AI partners and reducing the amount of explicit communication needed to maintain SA (MacMillan et al. 2004; McDowell et al. 2024).

One of the strengths of the modern C2 approach has been the way it centers human needs to ensure that the prerequisites for situational understanding are sufficiently met (McDowell et al. 2024). An AI system designed to be transparent centers human cognitive needs, providing the information their human partners need to make accurate decisions in a timely manner (Lyons 2013; Endsley 2023). Transparent AI can result in faster interventions (Helldin et al. 2013), more accurate expectations (Schaefer et al. 2016), and reduced cognitive load (You and Robert 2017), all of which is expected to allow human–AI partnerships to work at a faster pace and take advantage of fleeting windows of opportunity. These advantages are expected to help the Army of the future maintain transformational overmatch, even as operational challenges abound and the informational complexity of the battlefield surges.

Although transparency is hardly a panacea for challenges to human–AI partnerships, it is an approach for examining and designing human–AI interaction

for the purpose of meeting human informational needs. Future discussions of transparency in human–AI partnerships will focus on specific approaches to implementing transparency and how these approaches can be applied to different levels of the military decision-making process in the larger C2 paradigm.

7. References

- Bläsing D, Bornewasser M. Influence of increasing task complexity and use of informational assistance systems on mental workload. *Brain Sciences*. 2021;11(1):102.
- Bolstad CA, Costello AM, Endsley MR. Bad situation awareness designs: what went wrong and why. *Proceedings of the 16th World Congress of International Ergonomics Association*; 2006 July 10–14; Maastricht, Netherlands. c2006. p. 1–6.
- Bradshaw JM et al. Dimensions of adjustable autonomy and mixed-initiative interaction. In: *Agents and computational autonomy: potential, risks, and solutions 1*. Springer Berlin Heidelberg. c2004. p. 17–39.
- Bradshaw JM et al. From tools to teammates: joint activity in human-agent-robot teams. *Human Centered Design: First International Conference*; 2009 July 19–24; San Diego, CA. *Human Centered Design*; c2009. p. 935–944.
- Bradshaw JM, Feltovich PJ, Johnson M. Human-agent interaction. In: Boy G, editor. *Handbook of human-machine interaction*. CRC Press; c2011. p. 293–302.
- Bünger MA. Exploring user trust in natural language processing systems: a survey study on ChatGPT users [master's thesis]. Uppsala University; 2024. <https://uu.diva-portal.org/smash/get/diva2:1873121/FULLTEXT01.pdf>
- Cannon-Bowers JA, Salas E. Reflections on shared cognition. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*. 2001;22(2):195–202.
- Chen JY, Barnes MJ. Human-agent teaming for multirobot control: a review of human factors issues. *IEEE Transactions on Human-Machine Systems*. 2014;44(1):13–29.
- Chen JY et al. Situation awareness-based agent transparency. *Army Research Laboratory (US)*; 2014 Apr. Report No.: ARL-TR-6905.
- Chen JY et al. Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical Issues in Ergonomics Science*. 2018;19(3):259–282.
- Cooke NJ, Demir M, McNeese N. Synthetic teammates as team players: coordination of human and synthetic teammates. *Cognitive Engineering Research Institute Mesa (US)*; 2016 May. Report No.: RE2016844 01.

- Cooke NJ, Salas E, Cannon-Bowers JA, Stout RJ. Measuring team knowledge. *Human Factors*. 2000;42(1):151–173.
- Cooke NJ, Stout RJ, Salas E. A knowledge elicitation approach to the measurement of team situation awareness. *Situational Awareness*. Routledge; c2017. p. 157–182.
- Cramer H et al. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-adapted Interaction*. 2008;18:455–496.
- DeChurch LA, Mesmer-Magnus JR. The cognitive underpinnings of effective teamwork: a meta-analysis. *Journal of Applied Psychology*. 2010;95(1):32.
- Dekker SW, Woods DD. MABA-MABA or abracadabra? Progress on human–automation co-ordination. *Cognition, Technology & Work*. 2002;4:240–244.
- de Visser EJ, Cohen M, Freedy A, Parasuraman, R. A design methodology for trust cue calibration in cognitive agents. *VAMR 2014, Held as Part of HCI International. Proceedings of the Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments: 6th International Conference*. 2014 June 22–27; Heraklion, Crete, Greece. Springer International Publishing; c2014. Part I 6; p. 251–262.
- de Winter JC, Dodou D. Why the Fitts list has persisted throughout the history of function allocation. *Cognition, Technology & Work*. 2014;16:1–11.
- Endsley MR. Measurement of situation awareness in dynamic systems. *Human Factors*. 1995;37(1):65–84.
- Endsley MR. Situation awareness misconceptions and misunderstandings. *Journal of Cognitive Engineering and Decision Making*. 2015;9(1):4–32.
- Endsley MR. From here to autonomy: lessons learned from human–automation research. *Human Factors*. 2017;59(1):5–27.
- Endsley MR. Supporting human-AI teams: transparency, explainability, and situation awareness. *Computers in Human Behavior*. 2023;140:107574.
- Endsley MR, Bolté B, Jones DG. *Designing for situation awareness: an approach to user-centered design*. CRC Press; c2003.
- Endsley MR, Connors ES. Situation awareness: state of the art. 2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century; 2008 July 20–24; Pittsburgh, PA. p. 1–4.

- Endsley MR, Jones WM. Situation awareness, information dominance, and information warfare. Air Force Armstrong Laboratory (US); 1997. Report No.: AL/CF-TR-1997-0156.
- Endsley MR, Kiris EO. The out-of-the-loop performance problem and level of control in automation. *Human Factors*. 1995;37(2):381–394.
- Feitosa J, Grossman R, Kramer WS, Salas E. Measuring team trust: a critical and meta-analytical review. *Journal of Organizational Behavior*. 2020;41(5):479–501.
- Felzmann H, Fosch-Villaronga E, Lutz C, Tamò-Larrieux A. Towards transparency by design for artificial intelligence. *Science and Engineering Ethics*. 2020;26(6):3333–3361.
- Goecks VG, Waytowich N. COA-GPT: generative pre-trained transformers for accelerated course of action development in military operations. 2024 International Conference on Military Communication and Information Systems (ICMCIS); 2024 Apr 23–24; Koblenz, Germany. IEEE. p. 01–10.
- Grote G, Weyer J, Stanton NA. Beyond human-centred automation—concepts for human–machine interaction in multi-layered networks. *Ergonomics*. 2014;57(3):289–294.
- Gunning D, Aha D. DARPA’s explainable artificial intelligence (XAI) program. *AI Magazine*. 2019;40(2):44–58.
- Handelman DA et al. Multi-agent playbook for human-robot teaming. *Proceedings of the Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications V*; 2023 June; SPIE; c2023. Vol. 12538, p. 123–135.
- Headquarters, Department of the Army. Commander and staff organization and operations. Headquarters, Department of the Army; 2022 May. Field Manual No.: FM 6-0.
- Hearst Autos. What is park assist? Quick guide. (n.d). <https://www.caranddriver.com/research/a32814141/park-assist/>
- Helldin T et al. Transparency of military threat evaluation through visualizing uncertainty and system rationale. EPCE 2013, Held as Part of HCI International; *Proceedings of the Engineering Psychology and Cognitive Ergonomics. Applications and Services. 10th International Conference*; 2013 July 21–26; Las Vegas, NV. Springer Berlin Heidelberg; c2013. Part II 10. p. 263–272.

- Henkel Z, Bethel CL, Murphy RR, Srinivasan V. Evaluation of proxemic scaling functions for social robotics. *IEEE Transactions on Human-Machine Systems*. 2014;44(3):374–385.
- Hinds PJ, Roberts TL, Jones H. Whose job is it anyway? A study of human-robot interaction in a collaborative task. *Human-Computer Interaction*. 2004;19(1-2):151–181.
- Hoff KA, Bashir M. Trust in automation: integrating empirical evidence on factors that influence trust. *Human Factors*. 2015;57(3):407–434.
- Ishak D, Nathan-Roberts D. Analysis of elderly human-robot team trust models. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*; 2015 Sep; Los Angeles, CA: SAGE Publications; c2015. Vol. 59, No. 1, p. 65–69.
- Israelsen BW, Ahmed NR. “Dave... I can assure you... that it’s going to be all right...” A definition, case for, and survey of algorithmic assurances in human-autonomy trust relationships. *ACM Computing Surveys*. 2019;51(6):1–37.
- Karimov FP, Brengman M, Van Hove L. The effect of website design dimensions on initial trust: a synthesis of the empirical literature. *Journal of Electronic Commerce Research*. 2011;12(4).
- Karsenty L, Botherel V. Transparency strategies to help users handle system errors. *Speech Communication*. 2005;45(3):305–324.
- Kasdaglis N, Newton O, Lakhmani S. System state awareness: a human centered design approach to awareness in a complex world. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*; 2014 Sep; Los Angeles, CA: SAGE Publications; Vol. 58, No. 1, p. 305–309.
- Kase SE et al. The future of collaborative human-artificial intelligence decision-making for mission planning. *Frontiers in Psychology*. 2022;13:850628.
- Kaupp T, Makarenko A, Durrant-Whyte H. Human-robot communication for collaborative decision making—a probabilistic approach. *Robotics and Autonomous Systems*. 2010;58(5):444–456.
- Kilgore R, Voshell M. Increasing the transparency of unmanned systems: Applications of ecological interface design. *VAMR 2014, Held as Part of HCI International. Proceedings of the Virtual, Augmented and Mixed Reality. Applications of Virtual and Augmented Reality: 6th International Conference*. 2014 June 22–27; Heraklion, Crete, Greece. Springer International Publishing; c2014. Part II 6; p. 378–389.

- Kim T, Hinds P. Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction. *Proceedings of the ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*; 2006 Sep 6–8; Hatfield, UK. IEEE; c2006. p. 80–85.
- Krausman A et al. Trust measurement in human-autonomy teams: development of a conceptual toolkit. *ACM Transactions on Human-Robot Interaction*. 2022;11(3):1–58.
- Lakhmani S, Abich J, Barber D, Chen J. A proposed approach for determining the influence of multimodal robot-of-human transparency information on human-agent teams. *AC 2016, Held as Part of HCI International. Proceedings of the Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience: 10th International Conference*; 2016 July 17–22; Toronto, Ontario, Canada. Springer International Publishing; c2016. Part II 10. p. 296–307.
- Lakhmani SG, Wright JL, Schwartz MR, Barber D. Exploring the effect of communication patterns and transparency on performance in a human-robot team. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*; 2019 Nov. Sage CA: Los Angeles, CA. SAGE Publications; c.2019. Vol. 63, No. 1, p. 160–164.
- Lee JD, See KA. Trust in automation: designing for appropriate reliance. *Human Factors*. 2004;46(1):50–80.
- Lipshitz R, Strauss O. Coping with uncertainty: a naturalistic decision-making analysis. *Organizational Behavior and Human Decision Processes*. 1997;69(2):149–163.
- Lyons JB. Being transparent about transparency: a model for human-robot interaction. *2013 AAAI Spring Symposium Series*; 2013 Mar 25–27. Palo Alto, CA.
- Lyons JB et al. Shaping trust through transparent design: theoretical and experimental guidelines. *Advances in Human Factors in Robots and Unmanned Systems. Proceedings of the AHFE 2016 International Conference on Human Factors in Robots and Unmanned Systems*; 2016 July 27–31; Walt Disney World; c2016. p. 127–136.
- Lyons JB, Havig PR. Transparency in a human-machine context: approaches for fostering shared awareness/intent. *VAMR 2014, Held as Part of HCI International. Proceedings of the Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments: 6th*

- International Conference. 2014 June 22–27; Heraklion, Crete, Greece. Springer International Publishing; c2014; Part I 6. p. 181–190
- MacMillan J, Entin EE, Serfaty D. Communication overhead: the hidden cost of team cognition. In: Salas E, Fiore SM, editors. Team cognition: understanding the factors that drive process and performance. American Psychological Association; c2004. p. 61–82. <https://doi.org/10.1037/10690-004>
- Marr JJ. The military decision making process: making better decisions versus making decisions better. School of Advanced Military Studies, US Army Command and General Staff College; 2001.
- Mathieu JE et al. The influence of shared mental models on team process and performance. *Journal of Applied Psychology*. 2000;85(2):273.
- Mayer RC, Davis JH, Schoorman FD. An integrative model of organizational trust. *Academy of Management Review*. 1995;20(3).
- McDowell K et al. Re-envisioning command and control. 2024 International Conference on Military Communication and Information Systems (ICMCIS); 2024 Apr. IEEE. p. 1–7.
- Mercado JE et al. Effects of agent transparency on multi-robot management effectiveness. Army Research Laboratory (US); 2015. Report No.: ARL-TR-7466.
- Merritt SM, Heimbaugh H, LaChapell J, Lee D. I trust it, but I don't know why: effects of implicit attitudes toward automation on trust in an automated system. *Human Factors*. 2013;55(3):520–534.
- Metcalf JS, Perelman BS, Boothe DL, McDowell K. Systemic oversimplification limits the potential for human-AI partnership. *IEEE Access*. 2021;9:70242–70260.
- Miller CA. Delegation and transparency: coordinating interactions so information exchange is no surprise. VAMR 2014, Held as Part of HCI International. Proceedings of the Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments: 6th International Conference. 2014 June 22–27; Heraklion, Crete, Greece. Springer International Publishing; c2014. Part I 6. p. 191–202.
- Miller CA, Parasuraman R. Designing for flexible interaction between humans and automation: delegation interfaces for supervisory control. *Human Factors*. 2007;49(1):57–75.

- Min B et al. Recent advances in natural language processing via large pre-trained language models: a survey. *ACM Computing Surveys*. 2023;56(2):1–40.
- Onal E et al. Trust and consequences: a visual perspective. *VAMR 2014, Held as Part of HCI International. Proceedings of the Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments: 6th International Conference*. 2014 June 22–27; Heraklion, Crete, Greece. Springer International Publishing; c2014. Part I 6. p. 203–214.
- Ososky S et al. Determinants of system transparency and its influence on trust in and reliance on unmanned robotic systems. *Proceedings of SPIE 9084, Unmanned Systems Technology XVI*; 2014 June; c2014. p. 112–123.
- Parasuraman R, Sheridan TB, Wickens CD. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*. 2000;30(3):286–297.
- Perzanowski D et al. Building a multimodal human-robot interface. *IEEE Intelligent Systems*. 2001;16(1):16–21.
- Pfautz J et al. The role of meta-information in C2 decision-support systems. *Proceedings of Command and Control Research and Technology Symposium*; 2006 June 20–26; San Diego, CA.
- Robinette P et al. Overtrust of robots in emergency evacuation scenarios. 2016 11th *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*; 2016 Mar 7–10; Christchurch, New Zealand. IEEE. p. 101–108.
- Salas E et al. Understanding and improving teamwork in organizations: a scientifically based practical guide. *Human Resource Management*. 2015;54(4):599–622.
- Schaefer KE, Chen JY, Szalma JL, Hancock PA. A meta-analysis of factors influencing the development of trust in automation: implications for understanding autonomy in future systems. *Human Factors*. 2016;58(3):377–400.
- Schaefer KE, Evans III AW, Hill SG. Command and control in network-centric operations: trust and robot autonomy. *Proceedings of the International Command and Control Research Technology Symposium (ICCRTS)*; 2015.
- Steyvers M, Kumar A. Three challenges for AI-assisted decision-making. *Perspectives on Psychological Science*. 2024;19(5):722–734.
- Stowers K et al. Insights into human-agent teaming: intelligent agent transparency and uncertainty. *Advances in Human Factors in Robots and Unmanned*

- Systems. Proceedings of the AHFE 2016 International Conference on Human Factors in Robots and Unmanned Systems; 2016 July 27–31; Walt Disney World. Springer International Publishing; c2016. p. 149–160.
- Stubbs K, Hinds PJ, Wettergreen D. Autonomy and common ground in human-robot interaction: a field study. *IEEE Intelligent Systems*. 2007;22(2):42–50.
- Sukthankar G, Shumaker R, Lewis M. Intelligent agents as teammates. *Theories of team cognition*. Routledge; c2013. p. 313–343.
- Sycara K, Sukthankar G. Literature review of teamwork models. Robotics Institute, Carnegie Mellon University. 2006;31(31):1–31.
- Ur B, McManus E, Pak Yong Ho M, Littman ML. Practical trigger-action programming in the smart home. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2014 Apr. p. 803–812.
- van de Merwe K, Mallam S, Nazir S. Agent transparency, situation awareness, mental workload, and operator performance: a systematic literature review. *Human Factors*. 2024;66(1):180–208.
- van Dis EAM et al. ChatGPT: five priorities for research. *Nature*. 2023; 614(7947):224–226. <https://www.nature.com/articles/d41586-023-00288-7>
- Wortham RH, Theodorou A, Bryson JJ. Robot transparency: improving understanding of intelligent behaviour for designers and users. Proceedings of the 18th Annual Conference on Towards Autonomous Robotic Systems, TAROS 2017. 2017 July 19–21; Guildford, UK. Springer International Publishing; c2017. p. 274–289.
- Wright JL, Chen JY, Lakhmani SG, Selkowitz AR. Agent transparency for an autonomous squad member: depth of reasoning and reliability. Army Research Laboratory (US); 2020 May. Report No.: ARL-TR-8956.
- Yagoda RE, Gillan DJ. You want me to trust a ROBOT? The development of a human–robot interaction trust scale. *International Journal of Social Robotics*. 2012;4(3):235–248.
- You S, Robert L. Teaming up with robots: an IMOI (inputs-mediators-outputs-inputs) framework of human-robot teamwork. *International Journal of Robotic Engineering*. 2017;2(3).
- Zhao H et al. Explainability for large language models: a survey. *ACM Transactions on Intelligent Systems and Technology*. 2024;15(2):1–38.

List of Symbols, Abbreviations, and Acronyms

AI	artificial intelligence
BLUFOR	Blue Force
C2	command and control
DDIL	denied, degraded, intermittent, and limited
LLM	large language model
OPFOR	opposing force
SA	situation awareness
SAT	Situation Awareness-based Agent Transparency

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

1 DEVCOM ARL
(PDF) FCDD RLB CI
TECH LIB