

TRUST AND HUMAN-AUTONOMY TEAMING

A thesis presented to the Faculty of the U.S. Army
Command and General Staff College in partial
fulfillment of the requirements for the
degree

MASTER OF MILITARY ART AND SCIENCE
Information Advantage Scholars

by

CHRISTOPHER KEAN, MAJ, USAF
B.A., University of Washington, Seattle, WA 2009
M.A., University of Oklahoma, Norman, OK, 2014

Fort Leavenworth, Kansas
2022

Approved for public release; distribution is unlimited. Fair use determination or copyright permission has been obtained for the inclusion of pictures, maps, graphics, and any other works incorporated into this manuscript. A work of the United States Government is not subject to copyright, however further publication or sale of copyrighted images is not permissible.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 10-06-2022		2. REPORT TYPE Master's Thesis		3. DATES COVERED (From - To) AUG 2021 – JUN 2022	
4. TITLE AND SUBTITLE Trust and Human-Autonomy Teaming				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Christopher Kean				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Command and General Staff College ATTN: ATZL-SWD-GD Fort Leavenworth, KS 66027-2301				8. PERFORMING ORG REPORT NUMBER	
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 Approved for Public Release; Distribution is Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The purpose of this research paper is to examine what influences human trust as it relates to human-autonomy teaming. The research was conducted as a mixed method study conducted via an online survey experiment. The experiment analyzed several independent variables and a single dependent variable (trust in an Artificial Intelligence (AI) system). The independent variables were an individual's digital literacy, an AI system's level of transparency, and the participant's experience level concerning an AI system's abilities/performance. Participants were given a hypothetical vignette, with random treatments to measure the independent variables, in which they assumed the role of lead planner on an operational team paired with the hypothetical AI system. The results show the importance of underlying baseline levels of trust in AI, regardless of treatment conditions. Finally, the study highlights that AI system transparency to increase trust might be more limited than previously thought.					
15. SUBJECT TERMS Trust, Human-Autonomy Teaming, Artificial Intelligence, Course of Action Development, Operational Planning					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT (U)	b. ABSTRACT (U)	c. THIS PAGE (U)			19b. PHONE NUMBER (include area code)
			(U)	127	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

MASTER OF MILITARY ART AND SCIENCE

THESIS APPROVAL PAGE

Name of Candidate: Christopher R. Kean

Thesis Title: Trust and Human-Autonomy Teaming

Approved by:

_____, Thesis Committee Chair
Trent J. Lythgoe, Ph.D.

_____, Member
Matthew W. Broaddus, Ed.D.

_____, Member
Peter S. Im, MSS, MMAS, MSSI

Accepted this 10th day of June 2022 by:

_____, Assistant Dean of Academics for
Dale F. Spurlin, Ph.D. Degree Programs and Research

The opinions and conclusions expressed herein are those of the student author and do not necessarily represent the views of the U.S. Army Command and General Staff College or any other governmental agency. (References to this study should include the foregoing statement.)

ABSTRACT

TRUST AND HUMAN-AUTONOMY TEAMING, by Major Christopher Kean, 127 pages.

The purpose of this research paper is to examine what influences human trust as it relates to human-autonomy teaming. The research was conducted as a mixed method study conducted via an online survey experiment. The experiment analyzed several independent variables and a single dependent variable (trust in an Artificial Intelligence (AI) system). The independent variables were an individual's digital literacy, an AI system's level of transparency, and the participant's experience level concerning an AI system's abilities/performance. Participants were given a hypothetical vignette, with random treatments to measure the independent variables, in which they assumed the role of lead planner on an operational team paired with the hypothetical AI system. The results show the importance of underlying baseline levels of trust in AI, regardless of treatment conditions. Finally, the study highlights that AI system transparency to increase trust might be more limited than previously thought.

ACKNOWLEDGMENTS

I am indebted to multiple people for helping me complete this thesis. This work could not have been possible without their consistent guidance, insight, and patience.

First, I'd like to thank my committee members. A special thanks to Dr. Robert Arp, who provided initial guidance on this topic, but was unable to remain on the committee due to scheduling conflicts. Thanks to Dr. Mathew Broaddus for joining my committee late and hitting the ground running to provide inputs. Mr. Im, as both a committee member and professor for the Information Advantage Scholars, was instrumental in reminding me to keep the big picture in mind. His continual motivation helped ensure that this thesis was written to add to the larger Information Advantage space and provide value to the joint force. Finally, Dr. Trent Lythgoe, as my thesis chair, provided the most input throughout this entire process. The guidance on quantitative studies and statistical analysis methods was critical in my successful completion of this thesis. His willingness to be available to answer questions, during long nights and weekends, allowed for me take the results of this study further than I could have ever initially hoped for when I started.

Next, I'd like to thank my fellow students and scholars for their help in reading and reviewing my progress throughout the year. I was lucky enough to be a part of two amazing groups during my time at CGSC. The early stages of the CGSC curriculum and the start of my thesis were manageable due to the friendships established with all the students of 11D. Special thanks are also due to the teaching team of 11D, especially Mr. Dwayne Wagner, who constantly pushed for me to strive to achieve more during my year here. Additionally, the other Information Scholars were crucial in helping me cross the

finish line in completing this thesis. Our shared discussions helped refine my thinking and their influence is apparent in these final pages.

Last, but certainly not least, immeasurable levels of gratitude are reserved for my fantastic wife, Meggie. Your constant patience in listening to me think out loud was the only way I made it through this process with my sanity intact. You and our wonderful daughters provided me the outlet I needed to step away from this project, reserve my energy, and maintain my motivation throughout the year. I know this year has been more challenging due to my decision to complete an MMAS and there is no way I could have done it without your support.

TABLE OF CONTENTS

	Page
MASTER OF MILITARY ART AND SCIENCE THESIS APPROVAL PAGE	iii
ABSTRACT.....	iv
ACKNOWLEDGMENTS	v
TABLE OF CONTENTS.....	vii
ACRONYMS.....	x
ILLUSTRATIONS	xi
TABLES	xiii
CHAPTER 1 INTRODUCTION	1
Background	1
Problem Statement.....	5
Purpose of the Study	5
Research Questions.....	6
Secondary Research Questions:	6
Hypotheses.....	6
Assumptions.....	7
Definition of Terms	7
Scope.....	8
Limitations and Delimitations	9
Significance of the Study	10
Summary	10
CHAPTER 2 LITERATURE REVIEW	12
Introduction.....	12
Human-Autonomy Teaming	12
Emerging Role of Human-Autonomy Teaming	15
Trust	17
The Role of Trust in Decision-Making	17
Trust in Human-Human Teams	19
Incremental Theories	20
Non-Incremental Theories	21
Transformational Theory	22
Trust in Human-Autonomy Teams	22
Trust and AI System Transparency.....	24

Integrated Model Trust for Human-Autonomy Teaming	26
Knowledge and Literacy of AI Systems	28
Data Literacy	28
Digital Literacy	30
Summary	34
CHAPTER 3 RESEARCH METHODOLOGY	35
Introduction.....	35
Method	35
Participants.....	36
Procedure and Treatments.....	37
Measures	38
Data Analysis.....	40
Ethical Considerations	40
Summary	41
CHAPTER 4 ANALYSIS	43
Descriptive Statistics.....	43
Hypothesis Findings	49
Hypothesis #1 Findings.....	51
Hypothesis #2 Findings.....	52
Hypothesis #3 Findings.....	52
Overall Findings.....	52
Qualitative Analysis.....	53
Assessment of Vignette Treatments.....	56
Additional Analysis	60
Factor Analysis and Trust in AI.....	62
Possible Explanations for a Participant’s Self-Reported Trust in AI.....	64
Trust in AI and Digital Literacy and Sub-competencies	65
Trust in AI and STEM Degrees	66
Trust in AI and view on AI Capabilities	66
CHAPTER 5 CONCLUSIONS AND RECOMMENDATIONS	68
Recommendations for Joint Force Human-Autonomy Teams	70
Limits of Transparency in System Design.....	70
Increasing Trust in AI	71
Selection of Military Members for Human-Autonomy Teaming.....	71
Implications for Training to Increase Trust in AI.....	71
Limitations	73
Recommendations for Additional Research	74
Explainability and Transparency in AI System Design.....	74
Factors Influencing AI Trust.....	75
Levels of Automation and its Effect on Trust.....	76

Conclusions.....	76
APPENDIX A RESEARCH SURVEY	78
Vignette.....	78
Self-Reported Trust Survey	82
Digital Literacy Survey.....	83
Demographical Questions.....	85
Appendix B RECRUITMENT E-MAIL.....	88
Appendix C CONSENT STATEMENT.....	89
APPENDIX D DEMOGRAPHIC STATISTICS	92
APPENDIX E ADDITIONAL QUANTITATIVE ANALYSIS	98
APPENDIX F QUALITATIVE ANALYSIS.....	105
BIBLIOGRAPHY	110

ACRONYMS

AI	Artificial Intelligence
CGSC	Command and General Staff College
COA	Course of Action
HAT	Human-Autonomy Team
JPP	Joint Planning Process
OPLAN	Operation Plan

ILLUSTRATIONS

	Page
Figure 1. Integrated Model of Trust for HAT	27
Figure 2. Participant COA Selection Responses	44
Figure 3. Mean of Participant's Self-Reported Trust	45
Figure 4. Mean of Participant's Self-Reported Digital Literacy	48
Figure 5. Connotation of Responses to the Open-Ended Question	56
Figure 6. Responses Expressing Desire/Appreciation for Experience with AI System by Experience Treatment Group	57
Figure 7. Responses Expressing Desire/Appreciation for Transparency with AI System by Transparency Treatment Group	58
Figure 8. Responses Expressing Desire for Explainability with AI System by Transparency Treatment Group	59
Figure 9. Scree Plot on Principal Component Factor Analysis of Self-reported Trust Survey	61
Figure 10. Updated Integrated Model of Trust for HAT	69
Figure 11. Demographic Question: Age	97
Figure 12. Frequency Distribution of Self-Reported Trust Survey Question #1	98
Figure 13. Frequency Distribution of Self-Reported Trust Survey Question #2	99
Figure 14. Frequency Distribution of Self-Reported Trust Survey Question #3	99
Figure 15. Frequency Distribution of Self-Reported Trust Survey Question #4	100
Figure 16. Frequency Distribution of Self-Reported Trust Survey Question #5	100
Figure 17. Frequency Distribution of Self-Reported Trust Survey Question #6	101
Figure 18. Frequency Distribution of Self-Reported Trust Survey Question #7	101
Figure 19. Frequency Distribution of Self-Reported Trust Survey Question #8	102
Figure 20. Frequency Distribution of Self-Reported Trust Survey Question #9	102

Figure 21. Frequency Distribution of Self-Reported Trust Survey Question #10	103
Figure 22. Hierarchical Qualitative Coding Results	109

TABLES

	Page
Table 1. Sub-competencies Derived from the European Union’s <i>Digital Competence Framework for Citizens</i>	32
Table 2. USAF Digital Literacy Sub-competencies	33
Table 3. Responses per Survey Treatment.....	43
Table 4. Participant COA Selection Frequencies	44
Table 5. Frequency Distributions on Individual Questions in Self-Reported Trust Survey 47	
Table 6. Ordinal Logistic Regression Results for Transparency and Experience (H ₁ and H ₃ , Respectfully).....	50
Table 7. Ordinal Logistic Regression Results for Digital Literacy (H ₂).....	51
Table 8. Qualitative Codes from the Open-Ended Question	54
Table 9. Rotated Factor Matrix Showing Values above .5	62
Table 10. Ordinal Logistic Regression Results for Mean Trust in AI	64
Table 11. Linear Regression Results of Predictors of Participant’s Trust in AI (Mean)....	65
Table 12. Demographic Question: Gender	92
Table 13. Demographic Question: Race/Ethnicity - Are you Hispanic, Latino, or of Spanish origin?	92
Table 14. Demographic Question: Race/Ethnicity - How would you describe yourself? ..	93
Table 15. Demographic Question: Grade	93
Table 16. Demographic Question: Service	94
Table 17. Demographic Question: Branch.....	95
Table 18. Demographic Question: What is the highest level of education you have completed?	96
Table 19. Demographic Question: Is your degree Science, Technology, Engineering, Mathematics (STEM)	96

Table 20.Descriptive Statistics of Self-Reported Trust Survey Sub-Competencies.....	103
Table 21.Linear Regression Results of Predictors of Participant’s Trust in AI (Mean).. <td>104</td>	104
Table 22.Qualitative Code Descriptions and Examples	105

CHAPTER 1

INTRODUCTION

Background

It is 2028. A joint planning team is beginning the early steps of the Joint Planning Process (JPP) to develop courses of action (COA) for an operation against a great power competitor. The team is augmented by an artificial intelligence (AI) system capable of advanced COA development and COA analysis. During planning the AI provides the team several COAs. However, the complexity of provided COAs dissuades the team from choosing them. Instead, the team selects five COAs that were developed by the staff organically. The team uses the AI system to wargame the COAs. The AI system analyses each COA and recommends the COA with the highest probability of success. However, the AI's recommended COA is only fourth best according to the team's human-led wargame results. Unable to reconcile the AI's reasoning for the COA recommendation, the team presents the COA based on the human-led results instead. The mission ultimately fails.

What went wrong, and why was the human joint planning team unable or unwilling to work successfully with the AI system? While the above situation was hypothetical, it showcases some of the upcoming challenges that the military will face with human-autonomy teaming. Specifically, overcoming the potential challenges of human trust in AI. Trust is well established as having a positive impact on team

performance.¹ Importantly, this impact is also shown to extend to human-autonomy teams (HAT).² Ensuring trust between humans and AI systems will be paramount for any eventual paradigm shift towards AI to be successful. This research study examines what factors influence trust in an AI system as a part of a human-autonomy team during operational planning. Additionally, this study provides senior leaders with possible understanding on the traits that planners might need to possess to be effective in a human-autonomy team environment and how the military can train those planners accordingly.

The *Joint Operating Environment 2035* describes the challenges of the future security environment through the lens of several global trends, among them the innovative use of high and low technologies.³ Specifically, AI is increasingly seen as a major disruptive technology, with the Joint Operating Concept noting that “[e]merging autonomous robotic systems are being increasingly used to augment, rather than simply replace, individuals and platforms.”⁴ As the military begins augmenting organizations with AI systems, there are several implications that will impact how effective human-

¹ Bart A. De Jong, Kurt T. Dirks, and Nicole Gillespie, “Trust and Team Performance: A Meta-Analysis of Main Effects, Moderators, and Covariates,” *Journal of Applied Psychology* 101, no. 8 (August 2016): 1145, <https://doi.org/10.1037/apl0000110>.

² Nathan J. McNeese, Mustafa Demir, Erin K. Chiou, and Nancy J. Cooke, “Trust and Team Performance in Human–Autonomy Teaming,” *International Journal of Electronic Commerce* 25, no. 1 (January 2021): 52, <https://doi.org/10.3389/fpsyg.2021.589585>.

³ Chairman of the Joint Chiefs of Staff (CJCS), *Joint Operating Environment 2035: The Joint Force in a Contested and Disordered World* (Washington, DC: Joint Chiefs of Staff, 2016), ii.

⁴ *Ibid.*, 18.

autonomy teaming will be in this future environment. This study investigates some of those implications through the narrower lens of how the military might use AI systems to augment the decision-making capabilities of planners and commanders.

The power of augmented decision-making has two immediate use cases for the military: planning and execution. Moreover, the level of integration between a human and the AI system will possibly change depending on the time constraints present and which human decision-making system it is attempting to complement. This level of integration can be described across a spectrum of how much human involvement there is towards the final decision.

During tactical engagements, commanders and their subordinate leaders make decisions under time constraints in rapidly changing environments. These tactical decisions are embodied in the principle of mission command and decentralized execution. As Joint Publication 3-0 notes, “[m]ission command is built on subordinate leaders at all echelons who exercise disciplined initiative and act aggressively and independently to accomplish the mission.”⁵ The clear implication being that AI augmented decision-making for commanders would lead more effective decisions and in turn tactical success on the battlefield.

On the other hand, decisions made during planning at the operational level are often complicated by the sheer amount of information available. Although time constraints are more forgiving, data can still overwhelm the planning process and

⁵ Chairman of the Joint Chiefs of Staff (CJCS), Joint Publication (JP) 3-0, *Joint Operations* (Washington, DC: Joint Chiefs of Staff, 2016), II-2.

decisions are often made with incomplete information.⁶ In planning situations, it is clear to see the potential benefits of AI systems, mainly to help analyze the enormous amounts of data in the operational environment. This study focuses on the later situation and examines what influences trust during human-autonomy teaming in an operational planning environment, to maximize the effectiveness of a HAT.

Importantly, the common decision-making processes behind quick decisions that are made in a tactical situation can be different than those made during a longer and more deliberative process, such as planning. The difference between these two processes, the system 1 used to make quick decisions and system 2 used to make more deliberate decisions, was originally proposed by Stanovich and West and popularized by Daniel Kahneman.⁷ It is important to examine how these two human decision-making systems might interact with an AI system. Specifically, to increase the effectiveness of human-autonomy teaming, augmented systems might need to be calibrated depending on which system of human decision-making they are supporting. While not specifically examined in this study, the level of integration of an AI system and which level of system of decision-making it interacts with might have an overall effect on trust.

Finally, it is useful to compare human-autonomy teams to human-human teams to generate insights. Of note, and relevant to this study, are the insights related to the relationship that trust plays in human-human teaming effectiveness. Meta-analysis has

⁶ Chairman of the Joint Chiefs of Staff (CJCS), Joint Publication (JP) 5-0, *Joint Planning* (Washington, DC: Joint Chiefs of Staff, 2016), IV-1.

⁷ Daniel Kahneman, *Thinking, Fast and Slow* (New York: Farrar, Straus, and Giroux, 2011), 20.

shown there are several possible factors that influence trust in automation, including the traits of the humans and the AI systems that comprise a HAT.⁸ However, there are still gaps in the literature concerning the effect of some of the human traits (e.g., cognitive and/or emotive factors) and how influential they might be in generating trust in a human-autonomy team.⁹

Problem Statement

As the military moves towards augmenting human decision-making with AI systems, trust will be a fundamental part of HAT effectiveness. While research has indicated that several factors influence trust, particular characteristics relating to the human and system factors that influence trust need to be better understood.¹⁰ A better understanding of these factors is critical to effective AI system design and enables the joint force to focus human training requirements. Ultimately, successful human-autonomy teaming is predicated on military members trusting the underlying AI systems.

Purpose of the Study

This study examines, through the lens of trust and the application of current decision-making theories, the factors that influence an individual's trust in an AI system.

⁸ Kristin E. Schaefer, Edward R. Straub, Jessie Y.C. Chen, Joe Putney, and A.W. Evans, "A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems," *Human Factors: The Journal of the Human Factors and Ergonomics Society* 58, no. 3 (May 2016): 387, <https://doi.org/10.1016/j.cogsys.2017.02.002>.

⁹ Ibid., 391.

¹⁰ Ibid., 387.

Due to the vast possible applications of different AI systems for military purposes, the research narrows the focus on human-autonomy teaming during planning. While the primary purpose is to analyze the desired characteristics needed to create trust during human-autonomy teaming, the study will also examine if additional training for military members is required to ensure these characteristics are developed.

Research Questions

What influences human trust as it relates to the implementation of human-autonomy teaming during operational planning?

Secondary Research Questions:

- 1) Does the interaction between a person and an AI system in the decision-making process affect a person's trust of that system?
- 2) What human factors influence a person's trust in an AI system?
- 3) What AI system factors influence a person's trust in an AI system?

Hypotheses

H₀: There is no relationship between a person's knowledge of an AI system, their digital literacy, or the transparency of the AI system and a person's trust in that system.

H₁: There is a positive relationship between a person's experience level of an AI system's performance and their trust in the system.

H₂: There is a positive relationship between a person's digital literacy and their trust in an AI system.

H₃: There is a positive relationship between the transparency of an AI system and a person's trust in that system.

Assumptions

Assumptions in this paper are related to the hypothetical nature of AI systems under discussion and how close they will ultimately represent the possible AI systems in the future. First, this paper assumes that an AI system will eventually be able to provide objectively sound advice during COA development and COA analysis. Additionally, this paper assumes that given the hypothetical nature of the AI systems being described, participants' answers will accurately reflect what they would do when interacting with such a system.

Definition of Terms

Digital Literacy – the ability for a person to “[use] technology to identify, critically evaluate, and synthesize data and information; explore, create and manage digital content; and appropriately interact in a virtual environment.”¹¹

Explainability – the desire of an individual to know why an AI system made a particular decision.

Human-Autonomy Teaming – A human-machine pairing in which the machine operates and is viewed as a teammate. Compared to human-automation teaming in which the machine operates and is viewed as a tool.¹²

¹¹ Laura Barron, *Evaluation of Digital Literacy as an Air Force Foundational Competency*, Air Education and Training Command (AETC) Technical Report No. 3 (Randolph Air Force Base, San Antonio, TX: AETC, July 2021), 20, <https://apps.dtic.mil/sti/pdfs/AD1143616.pdf>.

¹² Joseph B. Lyons, Katia Sycara, Michael Lewis, and August Capiola, “Human–Autonomy Teaming: Definitions, Debates, and Directions,” *Frontiers in Psychology* 12 (May 2021): 2, <https://doi.org/10.3389/fpsyg.2021.589585>.

Human-in-the-loop – The lowest end of automation where there is “manual or frequent human interaction.”¹³

Human-on-the-loop – A level of integration where an AI system is semi-autonomous, but a “human monitors and has the ability to intervene.”¹⁴

Human-out-of-the-loop – the extreme end of the level of integration, representing a fully autonomous AI system, with “no human intervention.”¹⁵

Trust – “[A] belief in, and willingness to act on the basis of, the words, actions and deeds of another.”¹⁶

Scope

While there are numerous possible applications for the use human-autonomy teams, the scope of this study limits the application to mission planning. Additionally, there are multiple factors that may influence the effectiveness of human-autonomy teams; however, this study focuses on the factors that influence human trust in AI systems.

¹³ Kristin E. Schaefer, Deborah R. Billings, James L. Szalma, Jeffrey K. Adams, Tracy L. Sanders, Jessie Y. Chen, and Peter A. Hancock, “A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Human-Robot Interaction,” (Army Research Laboratory, Aberdeen Proving Ground, MD, July 2014), 27, <https://apps.dtic.mil/sti/pdfs/ADA607926.pdf>.

¹⁴ Ibid.

¹⁵ Ibid.

¹⁶ Roy Lewicki and Chad Brinsfield, “Framing Trust: Trust as a Heuristic,” in *Framing Matters: Perspectives on Negotiation Research and Practice in Communication*, ed. William A. Donohue, Randall G. Rogan, and Sandra Kaufman (New York: Peter Lang Publishing, 2011), 115.

Limitations and Delimitations

The main limitation of this study is the time available to conduct the research. This limitation is mitigated by narrowing the scope of the problem. While there are multiple implications related to the impact that human-autonomy teaming will have on military operations, the study focuses on just the impact during planning. Additionally, there are various factors that likely influence trust in human-autonomy teams, however, the study focuses on the specific factors of data literacy, AI system transparency, and experience of AI performance.¹⁷ Together this narrower focus ensured the research was completed in the time available.

Additional limitations are also related to the ability to accurately measure some of the variables in the study. There is an assumption that a level of trust will be able to be generalizable across participants, since scoring a specific level of trust is inherently subjective. Additionally, there is a limitation concerning the reliability of measuring digital literacy from a self-reported survey and how well an index can approximate an individual's data literacy. Finally, there are additional limitations in the generalizability of the population under study and how accurately it would reflect a representative sample of an operational planning team.

The major delimitation present regarding the scope of the paper is related to the selection of the dependent variable of trust as it relates to human-autonomy teaming effectiveness. This selection was due to the potential impact the notion of trust has as a possible explanation for the wide variety of outcomes during augmented decision-making

¹⁷ Schaefer et al., "A Meta-Analysis of Factors Influencing the Development of Trust in Automation," 387.

in the planning process. Additionally, there are multiple variables that are likely to have a relationship with a person's trust and those variables might provide key insights to the design of human-autonomy teams.

Significance of the Study

AI has the potential to be an asymmetrical advantage to the United States in future competition and conflict. As the military begins to incorporate human-autonomy teams into its decision-making processes, it is critical that leaders understand the factors that ensure effective human-autonomy teams. This study will help military leaders understand how to maximize effective integration between military planners and AI systems. Additionally, the results of this study will aid design teams in improving AI system design and improve the rollout of systems to enable effective teaming. Finally, this study will provide insights into areas to focus additional training for military members.

Summary

Operational planners are increasingly facing an operating environment that is so complex that decision-making is impacted through information overload. The military has plans to solve this problem by using AI systems to augment human planners. However, researchers have yet to fully examine what makes an effective human-autonomy team and importantly, what influences a human's trust in an AI system. This question is complicated by progressively more automated AI systems where human interaction is more limited. How can the military ensure that these systems are used by planners to their maximum potential?

This study examines the possible factors that influence a person's trust in an AI system. One possible factor is an AI system's level of transparency. However, it is not enough to scrutinize just the design of the AI system, the military also needs to consider what characteristics military planners might need to be more effective teammates with AI systems. The characteristic that is explored during this study is a person's data literacy. Finally, the experience level that a human has with an AI system's performance, is also a factor that is analyzed. Together these possible factors can provide insight for military leaders to better design AI and train military members to enable successful human-autonomy teaming.

However, to further study these possible relationships, there first must be a solid framework regarding how humans make decisions. This framework will also need to examine the role of trust in the decision-making process. The study of human-to-human teams and the role of trust provides insights that carry over to human-autonomy teams. The next chapter will examine the existing literature to begin to answer these questions and establish the conceptual framework regarding trust and decision making that will be used for the rest of the study.

CHAPTER 2

LITERATURE REVIEW

Introduction

This chapter will review the current literature on three topics necessary to understand how humans and AI systems are likely to behave during the operational planning process, and how these topics helped to inform the research questions and hypotheses (i.e., what influences human trust as it relates to the implementation of human-autonomy teaming, specifically during operational planning?).

The first topic is human-autonomy teams. The evolution of AI is a rapidly changing field, so having a common baseline for terminology is critical throughout the remainder of this paper. The second topic is this study's dependent variable, trust. This examination will begin with an overview of the multiple definitions of trust, and how it impacts human decision-making. The role of trust as it relates to human-to-human teams, as well as human-autonomy teams will also be covered. The final topic is digital literacy. Like the field of AI, this is an emerging concept, so a review on the current literature regarding the definition and what constitutes digital literacy, as well as how it can be measured, is warranted. This will include a review of the various subcomponents related to digital literacy, such as data and informational literacy, as well as why digital literacy is hypothesized to influence trust in an AI system.

Human-Autonomy Teaming

Up until this point, there has been a deliberate effort to use a single term to describe the interaction of an AI system and human interacting, specifically, human-

autonomy teaming. For the purposes this study, the definition for human-autonomy teaming will be based on the work of Lyons et al., namely, a human-machine pairing in which the machine operates and is viewed as a teammate.¹⁸ Additionally, due to common usage, the autonomous portion of this team will often be referred to as an AI system. These choices are due in part to the prevailing nature of the term in the literature as well as the apparent acceptance of the term among both the Air Force Research Laboratory as well as the Army Research Laboratory.¹⁹ However, there are still other terms that are used in the literature to describe the various levels of interaction humans have with machines. A partial list of discovered terms in the literature include Human-Machine teaming, Human-Robot Interaction, and Human-Automation, among others.²⁰ However, for the purposes of this paper, most of the distinctions between these terms does not bear further examination. The one exception is between Human-Automation teaming and Human-autonomy teaming.

Within the scope of human-autonomy teaming, there are three levels of integration (automation): human-in-the-loop, human-on-the-loop, and human-out-of-the-

¹⁸ Lyons et al., “Human–Autonomy Teaming,” 2.

¹⁹ Lyons et al., “Human–Autonomy Teaming,” 2.; U.S. Army Combat Capabilities Development Command, Army Research Laboratory (ARL), “Human Autonomy Teaming (HAT),” Open Campus, accessed November 9, 2021, <https://www.arl.army.mil/opencampus/HAT>.

²⁰ John Laird, Charan Ranganath, and Samuel Gershman, “Future Directions in Human Machine Teaming Workshop,” (Basic Research Innovative Collaboration Center, Arlington, VA, 16-17 July 2019), <https://basicresearch.defense.gov/Portals/61/Future%20Directions%20in%20Human%20Machine%20Teaming%20Workshop%20report%20%20%28for%20public%20release%29.pdf>; Lyons et al., “Human–Autonomy Teaming,” 11; Schaefer et al., “A Meta-Analysis of Factors Influencing the Development of Trust in Automation,” 387.

loop.²¹ These levels show a clear shift to distinguish between automation and autonomy, with the later specifically including the role of the machine's "intelligence" in their ability to aid decision making.²² One of the most comprehensive and referenced definitions showing how automation can be thought of as a spectrum, ending in complete autonomy, is by Parasuraman et al., who describe the levels of automation (LOA) across ten levels:

10. The automation decides everything, acts autonomously, ignoring the human.
9. The automation informs the human only if it, the robot, decides to
8. The automation informs the human only if asked
7. The automation executes automatically, then necessarily informs the human
6. The automation allows the human a restricted time to veto before automatic execution
5. The automation executes that suggestion if the human approves
4. The automation suggests one alternative
3. The automation narrows the selection down to a few options
2. The automation offers a complete set of decision/action alternatives
1. The automation offers no assistance: the human must make all decisions and actions²³

²¹ Schaefer et al., "A Meta-Analysis of Factors Influencing the Development of Trust in Automation," 387.

²² Lyons et al., "Human–Autonomy Teaming," 1.

²³ Raja Parasuraman, Thomas B. Sheridan, and Christopher D. Wickens, "A Model for Types and Levels of Human Interaction with Automation," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* 30, no. 3 (May 2000): 287, <https://doi.org/10.1109/3468.844354>.

However, it has also been noted that a scale as detailed, and with a high number of distinct levels, might be too hard to implement. Therefore, the Army Research Lab offered an alternative solution by grouping the Parasuraman et al. categories into three categories: “human-in-the-loop (manual or frequent human interaction), human-on-the-loop (human monitors and has ability to intervene), and autonomous (no human intervention).”²⁴ These definitions have also appeared to have been adopted by senior leaders in the military.²⁵ Therefore, human-in-the-loop, human-on-the-loop, and human-out-of-the-loop will be the definitions used to for the remainder of the paper.

Emerging Role of Human-Autonomy Teaming

As mentioned in chapter 1, there are two broad cases that AI aided decision-making appears to be most beneficial for military operations: planning and execution. Lingel et al. describe several categories of AI technologies: computer vision, natural language processing, expert systems, planners, machine learning (ML), and robotics.²⁶ While these technologies have cross-over into both cases of planning and execution, the focus of planning for this research paper will in turn limit the following discussion of

²⁴ Schaefer et al., “A Meta-Analysis of Factors Influencing the Development of Trust in Automation,” 387.

²⁵ U.S. Congress, Senate, *Statement of General Terrence J. O’Shaughnessy, United States Air Force Commander, United States Northern Command and North American Aerospace Defense Command before the Senate Armed Services Committee*, February 13, 2020, 58.

²⁶ Sherrill Lingel, Jeff Hagen, Eric Hastings, Mary Lee, Matthew Sargent, Matthew Walsh, Li Ang Zhang, and David Blancett, *Joint All-Domain Command and Control for Modern Warfare: An Analytic Framework for Identifying and Developing Artificial Intelligence Applications* (Santa Monica, CA: RAND Corporation, 2020), 18, <https://doi.org/10.7249/RR4408.1>.

emerging human-autonomy teaming to the categories of expert systems, planners, and machine learning.

The benefit of investment in these AI technologies as they apply to operational planning is summarized succinctly by Lingel et al.:

Expert systems could be used to recommend effects to achieve operational and tactical objectives. Planning systems could be used to pair various air, space, and cyber assets against targets and to generate a time-phased scheme of maneuver. Finally, ML could be used in conjunction with other categories of AI to allow C2 systems to learn how to perform tasks when expert knowledge is not available or when optimal tactics, techniques, and procedures (TTPs) are unknown.²⁷

Overall, these AI technologies working in conjunction serve a similar role to a planner doing COA development. Moreover, the ability for a machine learning AI system to operate under uncertain conditions provides for similar result, hypothetically, to the doctrinal definition of operational art.²⁸ In addition to COA development, the other immediate use case in planning is during COA analysis. Through modeling and simulation, AI systems could provide insights into the likelihood of mission success, paired with deeper risk assessments, when analyzing various COAs compared to human wargaming.²⁹ However, one thing that still needs to be discussed is how these possible AI systems will be integrated into human teams and how much humans will trust them.

²⁷ Lingel et al., *Joint All-Domain Command and Control for Modern Warfare*, 18.

²⁸ CJCS, JP 5-0, I-3.

²⁹ Terrence J. O'Shaughnessy, "Decision Superiority Through Joint All-Domain Command and Control," *Joint Forces Quarterly* 99, no. 4 (October 2020): 79, <https://ndupress.ndu.edu/Portals/68/Documents/jfq/jfq-99/jfq-99.pdf>.

Trust

Trust is “a belief in, and willingness to act on the basis of, the words, actions and deeds of another.”³⁰ While there are multiple frameworks for understanding trust, there are there are two dominant conceptualizations, behavioral and psychological.³¹ Since the research topic and question for this paper are related to the role that trust has related to the decision-making process, this review focuses on the psychological approach.

The decision to trust someone is both cognitive and emotional. The cognitive component of trust is more researched with early accounts by Lewis and Weigert noting that, “[people] cognitively choose whom we will trust in which respects and under which circumstances, and we base the choice on what we take to be ‘good reasons, constituting evidence of trustworthiness.’”³² It is this component that is of particular interest to the research question. The “evidence of trustworthiness” when dealing with an AI system, vice another human, is hypothesized to be dependent on several factors. Of interest to this study is whether these factors include a person’s digital literacy or particular characteristics of the AI system—specifically transparency.

The Role of Trust in Decision-Making

Trust plays an important role in decision-making as a cognitive heuristic. However, to better understand how this heuristic works, a review of the theoretical lens

³⁰ Lewicki and Brinsfield, “Framing Trust: Trust as a Heuristic,” 115.

³¹ Ibid., 114.

³² David J. Lewis and Andrew Weigert, “Trust as a Social Reality,” *Social Forces* 63, no. 4 (June 1985): 970, <https://doi.org/10.2307/2578601>.

of human decision-making is required. While there are multiple theories on the process of human decision-making, for the purposes of this paper, we will be using the work of Daniel Kahneman and his two-system approach. As previously mentioned, Kahneman argues that human decision-making is broken into a system 1 that humans use to make quick decisions and system 2 used to make more deliberate decisions.³³ However, it is important to note that even during the deliberate decision-making process using the slower system 2, the cognitive biases of system 1 are still running. It is through system 1 where Kahneman argues that biases and heuristics are used as cognitive short cuts in the decision-making process. These biases and heuristics, while cognitively resourceful, can often be flawed. Of note, this argument is not universally accepted, for example, Gary Klein argues that it is in the quick decision-making process that a naturalistic decision-making process occurs, allowing experts to use their vast experience to make accurate judgments.³⁴

While a worthy discussion, the distinction between the two systems is outside the scope of this paper. However, the approach by Kahneman allows for a useful lens to examine trust under, so it will be the preferred framework for the remainder of the paper. The usefulness of viewing trust as a heuristic is also the main argument of Lewicki and Brinsfield.³⁵ They mention that this approach is useful for many reasons, particularly

³³ Kahneman, *Thinking Fast and Slow*, 20-21.

³⁴ Daniel Kahneman and Gary Klein, "Conditions for Intuitive Expertise: A Failure to Disagree," *American Psychologist* 64, no. 6 (2009): 515, <https://doi.org/10.1037/a0016755>.

³⁵ Lewicki and Brinsfield, "Framing Trust: Trust as a Heuristic," 125.

because by viewing trust as a cognitive frame, it is “subject to the both the positive effects (i.e., enabling the actor to more effortlessly assimilate large amounts of information), and negative effects (i.e., biases, and perceptual distortions), often associated with heuristic use.”³⁶ Inherently it is in these effects that the independent variables are trying to uncover in the study, and to better understand what influences an individual’s framing of trust related to AI systems.

Trust in Human-Human Teams

Although the study of human trust in AI systems is new, the study of trust writ large is not. Examining how humans trust between themselves may provide insights into trust in human-autonomy teams. As noted by McNeese et al., the concept of trust between humans and the subsequent performance of the teams is well researched.³⁷ Critically, the role of trust and its effect on team performance in human-human teaming seems to also hold for human-autonomy teams.³⁸

Costa et al., theorized a multilevel model of trust in teams, which is founded on several theories regarding how trust in teams develops.³⁹ The theories surrounding trust can be grouped into three general categories: incremental, non-incremental, and

³⁶ Lewicki and Brinsfield, “Framing Trust: Trust as a Heuristic,” 114.

³⁷ McNeese et al., “Trust and Team Performance in Human–Autonomy Teaming,” 52.

³⁸ Ibid., 67.

³⁹ Ana Cristina Costa, C. Ashley Fulmer, and Neil R. Anderson, “Trust in Work Teams: An Integrative Review, Multilevel Model, and Future Directions,” *Journal of Organizational Behavior* 39, no. 2 (February 2018): 172, <https://doi.org/10.1002/job.2213>.

transformational.⁴⁰ These theories provide the framework for examining how individual level factors combined with the interaction of a team is integral for team trust.⁴¹

Incremental Theories

Incremental theories describe trust as a psychological construct between members that is increased (or decreased) over time. There are two main approaches to trust under the incremental theory lens, social exchange theory and social information processing theory.⁴² Many of the approaches under social exchange theory begin with the assumption that the initial baseline for trust is zero and then it is increased or decreased through encounters.⁴³ The interaction between cooperation and monitoring, provides reinforcing cycles based on the reciprocity between members, with cooperation increasing trust and violations decreasing it.⁴⁴

Similarly, social information theory describes the process of increasing trust between team members. However, unlike social exchange theory, the social information approach theorizes that the context of the team shapes trust.⁴⁵ For example, studies have shown that trust levels are dependent on several contextual considerations, such as the

⁴⁰ Costa, Fulmer, and Anderson, “Trust in Work Teams,” 172.

⁴¹ Ibid., 173.

⁴² Ibid., 172.

⁴³ Roy J. Lewicki, Edward C. Tomlinson, and Nicole Gillespie, “Models of Interpersonal Trust Development: Theoretical Approaches, Empirical Evidence, and Future Directions,” *Journal of Management* 32, no. 6 (December 2006): 999, <https://doi.org/10.1177/0149206306294405>.

⁴⁴ Costa, Fulmer, and Anderson, “Trust in Work Teams,” 172.

⁴⁵ Ibid.

medium of introduction, time available, and organizational context (e.g., competitive versus cooperative) all played a role in driving trust.⁴⁶

Non-Incremental Theories

The development of the non-incremental theories originated to explain research findings that showed a non-baseline zero level of trust which ran contrary to the incremental theories.⁴⁷ This level of trust can be determined for several reasons, such as the individual's disposition to trust, institution safeguards, and/or cognitive processes to facilitate processing judgements on trustworthiness.⁴⁸ Generally viewed, the non-incremental theories can be grouped as either belonging to the social categorization approach or the swift trust theory.

Social categorization theory postulates that trust develops quicker in teams where members are more homogeneous and have social groupings (e.g., by gender or profession) compared to teams that are diverse.⁴⁹ Comparably, the swift trust theory also argues for a quick establishment in team trust but posits that trust can occur in the absence of social categorizations. This is accomplished through members using

⁴⁶ Sharon N. Hill, Kathryn M. Bartol, Paul E. Tesluk, and Gosia A. Langa, "Organizational Context and Face-to-Face Interaction: Influences on the Development of Trust and Collaborative Behaviors in Computer-Mediated Groups," *Organizational Behavior and Human Decision Processes* 108, no. 2 (March 2009): 195, <https://doi.org/10.1016/j.obhdp.2008.10.002>.

⁴⁷ Costa, Fulmer, and Anderson, "Trust in Work Teams," 172.

⁴⁸ Harrison D. McKnight, Larry L. Cummings, and Norman Chervany, "Initial Trust Formation in New Organizational Relationships," *The Academy of Management Review* 23, no. 3 (July 1998): 476, <https://doi.org/10.5465/amr.1998.926622>.

⁴⁹ Costa, Fulmer, and Anderson, "Trust in Work Teams," 172.

expectations of trust to analogous scenarios and is most often seen in temporary teams with clearly defined tasks.⁵⁰

Transformational Theory

While both the incremental and non-incremental theories view trust from a unidimensional approach with regard to trust (i.e., trust and distrust as polar opposites along a single dimension), the transformational theory suggests that there are different types of trust and it transforms over time.⁵¹ The various stages of trust development are calculus-based (deriving the cost/benefit of keeping/breaking the relationship), knowledge-based (knowing the teammate well enough to predict behavior), and identification-based (understanding the teammate's desires and intentions).⁵² While, the stages are developed over time, they are not mutually exclusive. Relationships might not progress through each stage, with proponents arguing that many relationships stay in the knowledge-based stage indefinitely.⁵³

Trust in Human-Autonomy Teams

One of the challenges in the literature regarding human-autonomy teaming is that most of the relevant AI technologies are not fielded. However, Schaefer et al.'s findings

⁵⁰ Debra Meyerson, Karl E. Weick, and Roderick M. Kramer, "Swift Trust and Temporary Groups," in *Trust in Organizations: Frontiers of Theory and Research*, ed. Roderick M. Kramer and Tom R. Tyler (Thousand Oaks, CA: SAGE Publications, 1996).

⁵¹ Lewicki, Tomlinson, and Gillespie, "Models of Interpersonal Trust Development," 1006.

⁵² Ibid., 1007.

⁵³ Ibid., 1008.

that human trust in automated systems comprises both cognitive and emotional factors suggests that human-machine trust has commonalities with trust between people.⁵⁴ It is important to note that the presence of both cognitive and emotive factors lends credibility to the chosen theoretical lens of viewing trust as a heuristic. This section will look at the current literature as it applies not only to hypothetical human-autonomy teaming, but also teams across the various levels of automation.

Schaefer et al. noted during a meta-analysis, that many of the factors that are considered in examining trust in human-autonomy teaming are directed at the AI system components and not at the factors relevant to the human.⁵⁵ However, their meta-analysis listed understanding the automation, ability to use automation, and expectancy of automation as human cognitive factors. While the human emotive factors consisted of attitudes toward automation, confidence in automation, satisfaction with automation, and comfort with automation.⁵⁶ Finally, Schaefer et al.'s model describes human demographic traits (age, gender, ethnicity/culture, and personality) as possible moderating factors for trust.⁵⁷ For AI systems, the meta-analysis revealed an AI system's capability was the strongest moderating factor.⁵⁸

⁵⁴ Schaefer et al., "A Meta-Analysis of Factors Influencing the Development of Trust in Automation," 10.

⁵⁵ Ibid., 24.

⁵⁶ Ibid., 25.

⁵⁷ Ibid.

⁵⁸ Ibid., 24.

However, Schaefer et al. also noted that, while there is literature to support the theoretical importance of AI features (e.g., levels of automation, communication mode, etc.) towards trust, current studies are unable to show a stable effect.⁵⁹ Likewise, there is a recommendation for more research on human cognitive factors.⁶⁰ The remaining sections will be focused on a closer examination of three specific factors found in Schaefer et al.'s meta-analysis, namely, the AI system factors of communication and feedback (hypothesized to be represented by system transparency) and the human cognitive factor of understanding (as explained by an individual's digital literacy).

Trust and AI System Transparency

A central theme in the literature concerning trust is that importance of understanding intent between the trustor and trustee. Additionally, during interactions, the ability to understand intent can positively influence a person's confidence in decisions and provide the ability to predict a trustee's actions, also increasing trust. To this end, an AI system's level of transparency is the designed way to provide this intent to the human teammate.⁶¹ This section will provide an overview of the role that the transparency of an AI system plays to influence trust.

⁵⁹ Schaefer et al., "A Meta-Analysis of Factors Influencing the Development of Trust in Automation," 24.

⁶⁰ Ibid., 23.

⁶¹ Schaefer et al., "Communicating Intent to Develop Shared Situation Awareness and Engender Trust in Human-Agent Teams," 29.

Roff and Danks highlighted the relationship between a person's confidence in a system and their trust. As system reliability decreases, trust does as well.⁶² However, this relationship is also influenced by an individual's ability to see an AI system as predictable. As Chen and Barnes highlight, one aspect of trust is merely the belief that the recipient of a person's trust will act a certain way.⁶³ In this sense, a person does not need to have the knowledge of why an AI system is behaving why it is, but merely needs to be able to predict what it will do.⁶⁴

However, there are also deeper notions of trust that run counter to the above claim. This is especially true with the definitions of trust that incorporate an aspect of vulnerability from the trustor towards the trustee. As Danks puts it:

In these relationships, trust-relevant beliefs involve understanding, and perhaps even internalizing some of the values, preferences, and beliefs of the trustee. More colloquially, both the trustor and the trustee “share a mental model” of the world; in essence, the trustor knows roughly *what* the trustee will do, and also *why* she pursues that course of action. [Emphasis in the original]⁶⁵

Mercado et al. were able to demonstrate that there was a positive relationship between trust and transparency during an experiment with an AI system being used to augment the decision-making of operating a simulated unmanned vehicle.⁶⁶ The transparency levels

⁶² Heather M. Roff and David Danks, “‘Trust but Verify’: The Difficulty of Trusting Autonomous Weapons Systems,” *Journal of Military Ethics* 17, no. 1 (January 2018): 6, <https://doi.org/10.1080/15027570.2018.1481907>.

⁶³ *Ibid.*, 6-7.

⁶⁴ *Ibid.*, 6.

⁶⁵ *Ibid.*, 6-7.

⁶⁶ Joseph E. Mercado, Michael A. Rupp, Jessie Y.C. Chen, Michael J. Barnes, Daniel Barber, and Katelyn Procci, “Intelligent Agent Transparency in Human-Agent Teaming for Multi-UxV Management.” *Human Factors: The Journal of the Human*

tested ranged from the AI system simply providing a recommendation (level 1), providing a recommendation and reasoning for the recommendation (level 2), and finally a recommendation, reason, and level of uncertainty (level 3).⁶⁷ Critically, Mercado et al.'s results also showed not only an increase in trust, but a higher level of performance by human operators due to transparency level.⁶⁸

The resulting research in the importance of transparency has been used in several recommendations for AI system design.⁶⁹ For example, Lee and See recommended that systems be designed to “show the process and algorithms of the automation by revealing intermediate results in a way that is comprehensible to the operators.”⁷⁰ However, the understanding an individual will have on various AI processes and algorithms will vary. Therefore, it is important to understand the underlying factors that might play a role in that understanding, mainly through examining the concept of digital literacy.

Integrated Model Trust for Human-Autonomy Teaming

Using the understanding of trust in human-human teams, as well as the research into possible influences of trust in automation just discussed, allows for the creation of an

Factors and Ergonomics Society 58, no. 3 (May 2016), 401, <https://doi.org/10.1177/0018720815621206>.

⁶⁷ Ibid., 406.

⁶⁸ Ibid., 408.

⁶⁹ Schaefer et al., “A Meta-Analysis of Factors Influencing the Development of Trust in Automation,” 37.

⁷⁰ John Lee and Katrina See, “Trust in Automation: Designing for Appropriate Reliance,” *Human Factors* 46, no. 1 (2004): 74, https://doi.org/10.1518/hfes.46.1.50_30392.

integrated model of trust for human-autonomy teams shown in figure 1. As seen, this model breaks the factors of trust in automation into three sections: human factors, AI system factors, and environmental factors. Additionally, the experience of the human and demonstrated reliability of the AI system are increased through interaction.

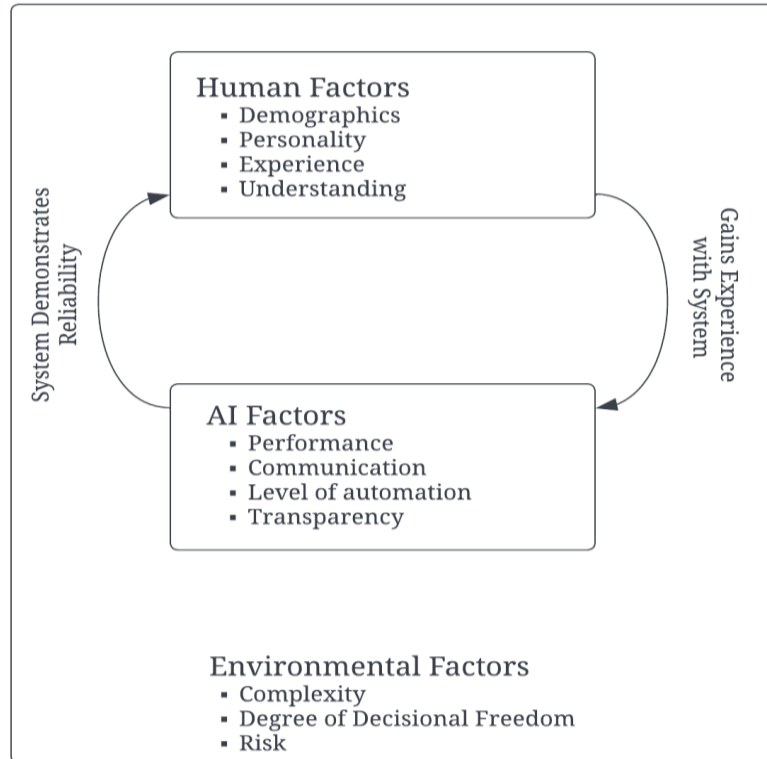


Figure 1. Integrated Model of Trust for HAT

Source: Created by author based on models of Kristin E. Schaefer, Deborah R. Billings, James L. Szalma, Jeffrey K. Adams, Tracy L. Sanders, Jessie Y. Chen, and Peter A. Hancock, “A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Human-Robot Interaction,” (Army Research Laboratory, Aberdeen Proving Ground, MD, July 2014), <https://apps.dtic.mil/sti/pdfs/ADA607926.pdf>; Kevin Hoff and Masooda Bashir, “Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust,” *Human Factors: The Journal of the Human Factors and Ergonomics Society* 57, no. 3 (May 2015): 407-34, <https://doi.org/10.1177/0018720814547570>.

Knowledge and Literacy of AI Systems

One of the key gaps in the literature is that we do not know much about the human factors that might influence trust in an AI system. However, one likely factor is an individual's basic knowledge around AI technologies. This link between knowledge of teammates and trust is aligned with the previously discussed trust frameworks, mainly the non-incremental and transformational theories. There are multiple possible ways to measure an individual's knowledge around AI systems. This section will discuss the current literature related to the defining and studying of two such areas, data literacy and digital literacy. Additionally, instruments used to measure a person's digital literacy will be evaluated for potential use for data collection.

Data Literacy

The importance of data is increasing becoming relevant in both the public and private sectors, and the military is no exception. However, just as the concept of “big data” is maturing, so is the idea of a person's data literacy. Several definitions exist on what constitutes data literacy. For example, Ridsdale et al. defined it as “the ability to collect, manage, evaluate and apply data, in a critical manner.”⁷¹ Meanwhile, Wolff et al., take a more nuance and detailed approach claiming that data literacy should be framed as a set of competencies, which are ultimately dependent how the level of data handling

⁷¹ Chantel Ridsdale, James Rothwell, Michael Smit, Hossam Ali-Hassan, Michael Bliemel, Dean Irvine, Daniel Kelley, Stan Matwin, and Bradley Wuetherick, “Strategies and Best Practices for Data Literacy Education,” (Knowledge Synthesis Report, Dalhousie University, Nova Scotia, Canada, 2015), 8, <https://dalspace.library.dal.ca/handle/10222/64578>.

needed to solve problems.⁷² Finally, as Frank et al., noted, any attempt to define data literacy in such a formative stage “is as much prescriptive as descriptive.”⁷³

While defining data literacy is important to ensure a common framework, the more important review of the literature is related to how to measure data literacy. As noted by Smolnikova, the assessment of data literacy is still in its infancy. Many assessments are directed at either the data literacy of a corporation or focused on how to teach data literacy.⁷⁴ Ultimately, one of the most promising instruments was developed by Pratamaet et al. and has passed preliminary study to validate the measurement to test initial levels of data literacy.⁷⁵ However, for the purposes of this study, administering such a comprehensive exam would be excessive and would also negatively impact participant recruitment. Therefore, it will be required to use an instrument that measures a participant's data literacy through a self-assessment.

⁷² Annika Wolff, Daniel Gooch, Jose J. Cavero Montaner, Umar Rashid, and Gerd Kortuem, “Creating an Understanding of Data Literacy for a Data-Driven Society,” *The Journal of Community Informatics* 12, no. 3 (August 9, 2016), 18, <https://doi.org/10.15353/joci.v12i3.3275>.

⁷³ Mark Frank, Johanna Walker, Judie Attard, and Alan Tygel, “Data Literacy - What Is It and How Can We Make It Happen?” *The Journal of Community Informatics* 12, no. 3 (September 19, 2016): 6, <https://doi.org/10.15353/joci.v12i3.3274>.

⁷⁴ Marketa Smolnikova, “Next Step: Data Literacy Measurement,” in *Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management* (Budapest, Hungary: SCITEPRESS - Science and Technology Publications, 2020), 4, <https://doi.org/10.5220/0010146402340240>.

⁷⁵ M. A. Pratamaet, Supahar Pahar, Diah P. Lestari, Widia K. Sari, Tri S. Y. Putri, and V.A.K. Adiatmah, “Data Literacy Assessment Instrument for Preparing 21 Cs Literacy: Preliminary Study,” *Journal of Physics: Conference Series* 1440 (January 2020), <https://doi.org/10.1088/1742-6596/1440/1/012085>.

Bonikowska et al. summarize the currently available tools for measuring data literacy, with many of them being offered by private sector companies or not-for-profit organizations whose goals are to train data literacy.⁷⁶ Many of these instruments are based on an individual's self-assessment and are based on previously discussed competencies. For example, the organization "Data to the People" based their framework on the competency map developed by Ridsdale et al., while the software company Qlik uses a propriety framework.⁷⁷

Digital Literacy

Digital literacy is the ability for a person to "[use] technology to identify, critically evaluate, and synthesize data and information; explore, create and manage digital content; and appropriately interact in a virtual environment."⁷⁸ This definition was created for the United States Air Force in conjunction with the identification of digital literacy as a foundational competency and an assessment instrument.⁷⁹ The definition and competency framework were adapted from the European Union's Digital Competence Framework for Citizens.

⁷⁶ Aneta Bonikowska, Claudia Sanmartin, and Marc Frenette, "Data Literacy: What It Is and How to Measure It in the Public Service," (Analytical Studies Branch, Statistics Canada, Ottawa, Ontario, Canada, August 2019, 10, http://epe.lac-bac.gc.ca/100/201/301/weekly_acquisitions_list-ef/2019/19-33/publications.gc.ca/collections/collection_2019/statcan/11-633-x/11-633-x2019003-eng.pdf).

⁷⁷ Ibid., 12.

⁷⁸ Barron, *Evaluation of Digital Literacy as an Air Force Foundational Competency*, 20.

⁷⁹ Ibid., 4.

The Digital Competence Framework for Citizens, identified five dimensions of digital competence, each with several sub-competences: information and data literacy, communication and collaboration, digital content creation, safety, and problem solving.⁸⁰ This framework was further refined by Skov to four domains of digital competence: information, communication, production, and safety.⁸¹ These domains are defined as:

Information: “Ability to identify, locate, retrieve, store, organize, and analyze digital information and evaluate relevance and purpose”

Communication: “Ability to communicate, collaborate, interact with, and participate in virtual teams, and networks as well as make use of appropriate media, tone, and behavior”

Production: “Ability to create, configure, and edit digital content, solve digital problems, and explore new ways to take advantage of technology”

Safety: “ability to use digital technology safely and sustainably in relation to data, identity, and work injuries and to pay attention to legal consequences, rights, and duties”⁸²

Furthermore, Skov, identified several sub-competencies that fall under the four domains. Their definitions are seen in table 1.

⁸⁰ Lieve Van Den Brande, Stephanie Carretero, Riina Vuorikari, and Yves Punie, *DigComp 2.0: The Digital Competence Framework for Citizens. Update Phase 1: The Conceptual Reference Model* (Luxembourg: Luxembourg Publication Office of the European Union, 2016), 8-9, <https://data.europa.eu/doi/10.2791/11517>.

⁸¹ Anders Skov, “What is Digital Competence,” Center for Digital Dannelsen, accessed February 17, 2022, <https://digital-competence.eu/front/what-is-digital-competence/>.

⁸² Barron, *Evaluation of Digital Literacy as an Air Force Foundational Competency*, 5.

Table 1. Sub-competencies Derived from the European Union's
Digital Competence Framework for Citizens

Sub-competency	Definition
INFORMATION SEARCH	Ability to search and find digital information, navigate between many online resources and sort through irrelevant information
CRITICAL EVALUATION OF INFORMATION	Ability to process, understand and critically evaluate digital information when sent and received
INFORMATION SELF-SERVICE	Ability to seek out and benefit from self-service solutions online
INFORMATION STORAGE	The ability to format, organize and store digital material while keeping both safety and accessibility in mind
ACTIVE INTERNET PARTICIPATION	Ability to make use of, express opinions, or otherwise contribute actively in digital environments
DIGITAL COLLABORATION	Ability to use technologies and media for teamwork, coordination, and collaboration processes
DIGITAL SOCIAL AWARENESS	Ability to reconcile behavior, tone, language, and technology with regard to context and social relations
DIGITAL MEDIA CHOICE	Ability to interact through a wide range of digital platforms and to be able to choose the best media for communicating with a specific recipient or group
DIGITAL PRODUCTION AND SHARING	Ability to create, assemble parts and modify content in many different formats., e.g. images, text, video or sound
DIGITAL EXPLORATION	Ability and willingness to stay updated on the technological developments and explore new digital opportunities
AUTOMATION	Ability to modify or create digital solutions that can fully or partially automate and perform a task
DIGITAL CONFIGURATION	Ability to adjust applications and devices to their own personal preferences, as well as to solve technical problems or tasks
DIGITAL LAW	Knowledge about current laws and licenses for digital behavior, information, and content
ONLINE IDENTITY MANAGEMENT	Ability to monitor/protect personal info online & know the consequences of personal digital footprints
DATA PROTECTION	Ability to identify and protect sensitive data and understand related risks
COMPUTING HEALTH	Ability to care for both physical and mental health in an everyday life surrounded by technology and media

Source: Anders Skov, "What is Digital Competence," Center for Digital Dannelsen, accessed February 17, 2022, <https://digital-competence.eu/front/what-is-digital-competence/>.

Ultimately, the Air Force adapted this model to create an assessment on digital literacy based on eight sub-competencies: information literacy, communication medium choice, technological exploration, appropriate social media use, information protection, software proficiency, active internet use, general computer skills.⁸³ The included sub-competency definitions can be seen in table 2.

Table 2. USAF Digital Literacy Sub-competencies

Sub-competency	Definition
Information Literacy	Recognizes when information is needed and locates, evaluates, and uses the needed information to achieve goals
Communication Medium Choice	Recognizes how technology may help or hinder communication; selects the most appropriate medium (e.g., e-mail, phone, video telecon) for communication based on situational needs
Technological Exploration	Seeks out, learns, and adopts new technologies; pioneers the use of new technologies within one's organization or unit
Appropriate Social Media Use	Maintains awareness of others' access to information posted in social media or other online forums, and manages one's online identity and reputation appropriately
Information Protection	Identifies and safeguards sensitive electronic data, mitigating potential risks
Software Proficiency	Creates and modifies digital content (e.g., text, graphics, numeric data) using commonly available software
Active Internet Use	Seeks out and benefits from potential self-service solutions online
General Computer Skills	Uses computers to organize information; adjusts settings and applications as needed to troubleshoot common computer problems

Source: Laura Barron, *Evaluation of Digital Literacy as an Air Force Foundational Competency* (Air Education and Training Command (AETC) Technical Report No. 3, Randolph Air Force Base, San Antonio, TX: AETC, July 2021), Table 10, 23, <https://apps.dtic.mil/sti/pdfs/AD1143616.pdf>.

⁸³ Barron, *Evaluation of Digital Literacy as an Air Force Foundational Competency*, 23.

From these definitions it is clear to see the overlap between the concepts of data literacy and digital literacy. Indeed, digital literacy is a broad enough term to also contain the idea of data literacy. It is for these reasons, combined with the availability of the Air Force assessment, that digital literacy will be used as the concept to measure an individual's possible knowledge of AI systems for this study.

Summary

This chapter reviewed the literature on human-autonomy teaming, trust, and digital literacy. Additionally, the literature review provided justification of the research question under study. That is, what influences human trust as it relates to the implementation of human-autonomy teaming? Finally, the literature review also provided insights into where gaps existed, further highlighting the need to examine what additional factors might moderate the trust a person has when teamed with an AI system.

CHAPTER 3

RESEARCH METHODOLOGY

Introduction

This chapter outlines the research design and methods that are used to analyze what influences human trust as it relates to the implementation of human-autonomy teaming during operational planning. The research was accomplished using an original survey instrument, composed of an experimental vignette, two self-assessment surveys, and demographical questions.

Method

The research was conducted as a cross-sectional, mixed-method study conducted via a survey experiment. Specially, the survey experiment manipulated two treatments, transparency, and experience, through a hypothetical vignette. The dependent variable, trust in AI, was evaluated through a participant's willingness to use a hypothetical AI recommendation. Additionally, trust in AI was measured through a self-assessment survey. This methodology was chosen due to the nature of the research question and the large amount of previous research already showing the impact of trust and its importance towards team effectiveness. This method allowed for a more precisely focused set of questions related to possible variables and their relationship to trust. Finally, these potential variables were easily measured using existing and validated instruments and/or common demographical questions.

Participants

The sample for this study was U.S. volunteers from the resident student body of the Command and General Staff College (CGSC). The target population for this study is U.S. military operational planners. Although Command and General Staff Officer Course (CGSOC) students are a convenience sample they are also reasonably representative of the target population: operational planners on a joint staff doing operational planning. In other words, many operational planners on joint staffs (at least those in the Army) are likely graduates of the resident CGSOC course. Furthermore, based on the timing of data collection (post JPME core courses), the population had recently completed courses and exercises on the Joint Planning Process and Military Decision-Making Process to mirror operational planning experience. This is important as it suggests that all participants should be familiar with the steps in COA development and analysis that the hypothetical AI system will be aiding with during the vignettes. These factors are helpful in strengthening the findings of the research as it relates to generalizability towards operational planners across the joint force.

Recruitment was conducted via email to CGSOC students sent through the CGSC Quality Assurance Office (QAO). The initial recruitment email (Appendix B) was sent to 986 students on 28 March 2022, with 164 students randomly assigned to each survey treatment. A follow-on email (the same as the initial e-mail) was sent out once half-way through the survey window to the students that had not completed the survey. The survey ended on 31 March 2022.

Procedure and Treatments

The data was collected using a web-based survey (Appendix A) created in Verint and administered to U.S. students at the Army's CGSOC through the QAO. The QAO randomly assigned U.S. resident students into 1 of 6 groups and sent them the recruitment email (Appendix B) to the appropriate survey treatment. In each treatment condition, participants were asked to read a hypothetical vignette in which they assumed the role of the lead planner on a joint operational team with a hypothetical AI system.

Two factors were manipulated in the treatment vignettes. The first factor was experience (new/experienced). The second factor was AI system transparency (low/medium/high). The result was a 3 by 2 factorial design with six total treatments. The treatment statements are provided in brackets throughout the vignette statement (Appendix A) to show clarity in changes between the six vignettes.

The transparency treatment was manipulated by providing statements that increased in detail (from low to high), providing the reader more information on the sources of data that the AI system used for its calculations. Additionally, the transparency treatments increased the level of explanation that the AI system provided during the wargaming results, going from simply providing a recommendation (low transparency) to showing a recommendation, the system's evaluation factors, and the system methodology (high transparency). The experience treatment was manipulated through a binary statement during the vignette (i.e., either providing to the participant that they have previous knowledge of the system's performance or the opposite).

In addition to the vignette, participants also answered two self-assessment surveys and a standard demographical question battery. The first self-assessment survey was

conducted to measure the participants trust in a variety of situations, to get a baseline trust score. The second self-assessment measured a participant's digital literacy. The demographical questions collected the participant's gender, race/ethnicity, age, grade, branch of service and occupational specialty, and educational background.

Measures

The survey (Appendix A) focused on analyzing several independent variables and a single dependent variable. The secondary research questions and hypotheses, broadly group the independent variables into three categories: human factors impacting trust, AI system attributes that affect trust, and the nature of the human-autonomy team's relationship and how that might impact trust. Specifically, the independent variables under consideration are an individual's digital literacy, an AI system's level of transparency, and the individual's knowledge of previous system performance (i.e., their experience with the system).

The dependent variable measured a person's trust in a hypothetical AI system. Trust was measured in three ways. First, the participants were given a behavioral question on the survey that measured, on an ordinal scale, whether to follow an AI system's recommendation. Secondly, there was an open-ended qualitative question, asking the participants to describe their reasoning for making the selection. Finally, there was a self-assessment survey that measures trust, on a Likert-type scale, in various situations (e.g., trust in co-workers, trust in strangers, trust in AI systems, etc.).

The dependent variable was measured after participants receive the treatment. After reading the vignette participants were asked to choose which COA to recommend to the commander. The COA options are to accept the AI's recommendation, select a

middle ground COA that compromises between the human and AI recommendations (i.e., not the preferred choice of either, but the second choice of both), or reject the AI recommended COA and use the COA recommended due to the human team's wargaming results. These selections provide an ordinal measurement of a person's trust in the AI system from high to low. The participants then had the opportunity to explain their reasoning for their selection with an open-ended question. Finally, trust was measured using a Likert scaled self-reported survey. This survey also served as a baseline for a participant's general level of trust, as it contains trust questions relating to other humans, in addition to automated systems.

The final independent variable (individual's digital literacy) was measured using the Air Force Digital Literacy self-assessment described in chapter 2. Finally, the demographical information collected (gender, race/ethnicity, age, grade, service, branch, level of education and degree major) allowed for a more comprehensive examination of digital literacy outside of self-assessment. For example, Science, Technology, Engineering, and Mathematics (STEM) majors and more technically aligned branches served to operationalize digital literacy as an additional measurement beyond the self-assessment. Additionally, demographical information was used to examine any other possible human factors that might influence trust outside of digital literacy. Finally, demographical information permitted the research's analysis to examine any demographic deviations between the target population of military planners and the participants.

Data Analysis

Data analysis used SPSS to show any statistically significant relationships between the independent and dependent variables. Since the dependent variable was measured on an ordinal scale (i.e., low, medium, and high), an ordinal logistic regression estimated the relationship between the dependent variable and the three primary independent variables. Additional ordinal logistic regressions were estimated an interactional model to control for the collected demographic variables and to see if any potential relationships exist that were not covered in the original hypotheses. Finally, additional analysis was conducted to do a principle component factor analysis for the self-reported trust survey.

The qualitative analysis was based on the single open-ended question where participants were asked to provide their reasoning for why they selected a particular COA after reading the vignette. The qualitative data was manually coded using a grounded theory approach. The coding took place in three cycles. First a cycle of open coding was performed resulting in 27 initial codes. The second cycle was an axial coding stage that refined the codes down to 23 unique codes. The resulting codes were then grouped hierarchically, by category, in the final stage of coding. Finally, the codes were identified as to whether they presented a viewpoint with a positive, negative, or neutral connotation regarding the AI system as a teammate.

Ethical Considerations

There were various safeguards used throughout data collection to ensure participant confidentiality. First, the survey was administered online through the CGSC QAO allowing for anonymous submission. Before beginning the online survey, potential

participants were provided a consent statement (Appendix C) that informed them of their rights. Participants were not allowed to continue until verifying that they agreed with the statement. QAO reviewed the open-ended responses and ensured that no personally identifiable information (PII) was included. Finally, all collected information was aggregated during data analysis and presentation, to ensure that no participants were identified through aggregated demographical data.

To ensure data storage compliance all survey and data storage follows best practices. All data collected is stored on the primary researcher's official military O365 OneDrive. Additionally, all data is stored in a restricted folder that is accessible only to the researcher and the committee members (and to the IRB for auditing purposes if required). Finally, files that contain unaggregated data is password protected, with the password being known only by the researcher and committee members. These considerations help ensure that data remains secure and is saved for a minimum of three years.

Summary

This research study examined any potential relationships between several independent variables and an individual's trust in a hypothetical AI system. Specifically, the three hypotheses investigated trust as it relates to a person's data literacy, an AI system's transparency, and the level of experience with an AI system's performance. This mixed methods study was accomplished by data collection via an online survey that was provided to a convenience sample of resident CGSC students. While this convenience sample will cause limitations on the generalizability of the results, ultimately students at

CGSC are well positioned to represent notional operational level planners on the joint staff.

CHAPTER 4

ANALYSIS

The survey was completed by 155 students out of 986 e-mailed for a response rate of approximately 16%. A breakdown of the response rates per experimental treatment is provided in table 3. The full breakdown of the participant's demographical information can be found in Appendix D.

Table 3. Responses per Survey Treatment

Treatment	# of participants
High Transparency/New (No Experience)	27
High Transparency/Experienced	22
Medium Transparency/New (No Experience)	29
Medium Transparency/Experienced	28
Low Transparency/New (No Experience)	27
Low Transparency/Experienced	22
Total	155

Source: Created by author.

Descriptive Statistics

Among the 155 responses, 154 were valid (i.e., one participant did not make any selection before submitting the survey). The responses to the question measuring the dependent variable, a participant's level of trust in AI, are provided in figure 2 and table 4. The participant's trust is ordinally measured in the hypothetical AI system as represented by their COA selection presented after the vignette. Selecting the human preferred COA (choice #1) indicates low trust in AI, while selecting the AI recommended COA (choice #3) indicates high trust. There is also a balanced COA (choice #2) the

indicates a medium level of trust in AI. As seen, low trust was the most preferred choice across all treatments, with 55.84% of participants selecting to use the human wargaming results instead of the AI system’s recommendation.

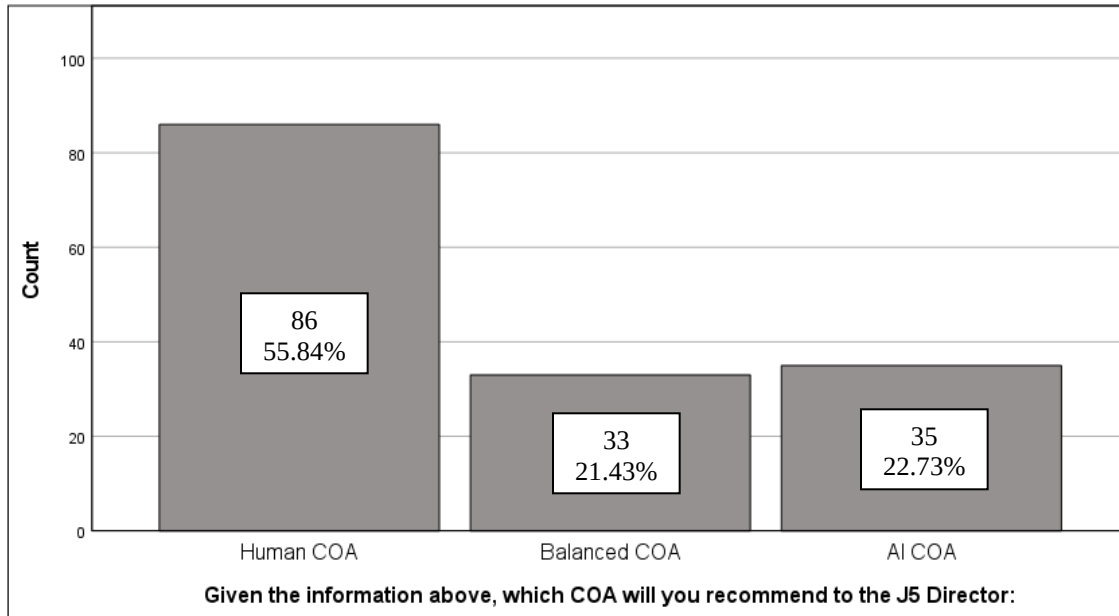


Figure 2. Participant COA Selection Responses

Source: Created by author.

Table 4. Participant COA Selection Frequencies

COA selection	Frequency	Percent	Valid Percent
Human COA	86	55.5	55.8
Balanced COA	33	21.3	21.4
AI COA	35	22.6	22.7
No Response	1	.6	
Total	155	100.0	100.0

Source: Created by author.

In addition to the COA selection question designed to measure trust, the participants were also asked a series of ten questions designed to measure trust in a variety of different teammates. All the questions were along a 7-point Likert-type scale, with 1 representing low trust and 7 representing high trust. The scores were then averaged to provide a baseline trust score. The mean scores are presented in figure 3. The mean was an average score of 4.7, with a standard deviation of .55, showing that across various teammates, participants had a baseline trust that was slightly higher than neutral.

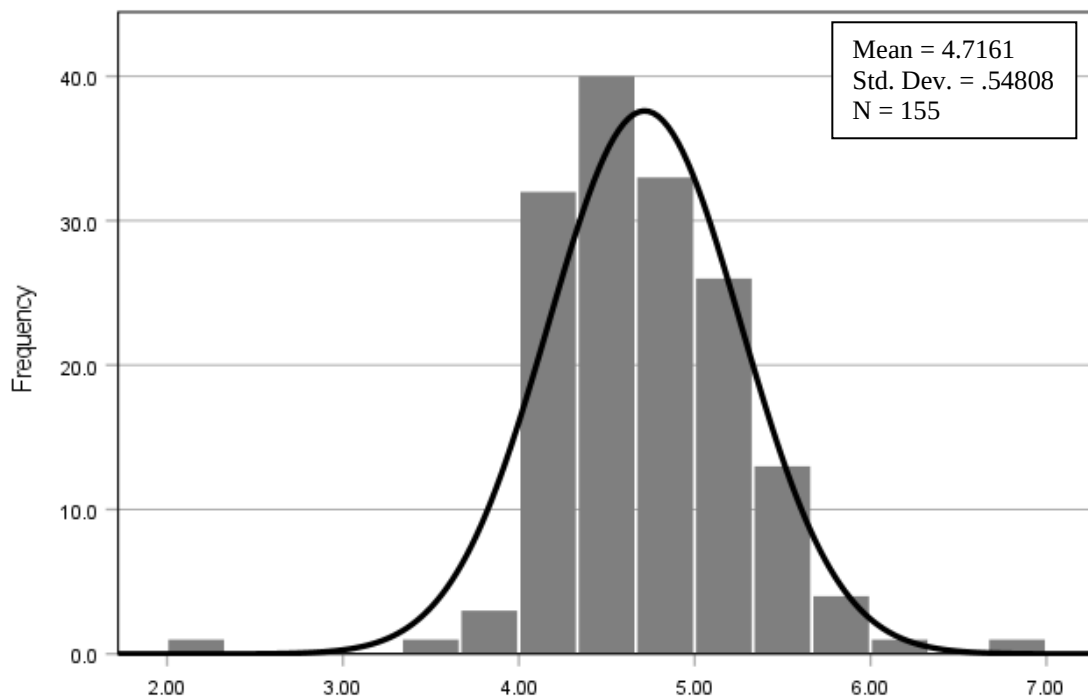


Figure 3. Mean of Participant's Self-Reported Trust

Source: Created by author.

A further breakdown of the means and standard deviations related to the specific questions is seen in table 5, with the full histograms of each question in Appendix E. As expected, based on the mean trust score, most of the individual questions also had a mean above 4, indicating that most participants had above a neutral level of trust for the various teammates presented in the questions. The obvious exception was the incompetent teammate, who had the lowest mean trust score.

Another interesting result can be seen by looking at the last three questions relating to various levels of automation. The questions were designed to map to the levels of automation discussed in chapter 2, namely human-in-the-loop, human-on-the-loop, and human-out-of-the-loop. While the means for the human-in-the-loop and human-on-the-loop questions are roughly equivalent, there is a noticeable drop in trust when it comes to a human-out-of-the-loop level of automation. The incompetent teammate and the human-out-of-the-loop AI system were the only teammates that had a mean trust score below neutral. A possible indication on the limits of trust during Human-Autonomy Teaming with consideration to the level of automation.

Table 5. Frequency Distributions on Individual Questions
in Self-Reported Trust Survey

Question:	N	Range	Mean	Std. Dev.
A new member of your team (i.e., you do not have insight into their background or history of performance)	155	[1, 7]	4.35	1.01
A team-member of the same service as you (e.g., if you are in the Army, the team member is also in the Army)	154	[2, 7]	4.71	.89
A team-member of a different service as you (e.g., if you are in the Army, the team member is in the Air Force, Navy, etc.)	155	[1, 7]	4.36	.94
A team-member you have worked with before and found incompetent	155	[1, 7]	2.52	1.09
A team-member you have worked with before and found competent	155	[1, 7]	5.43	.85
A team-member you have worked with before and found exceptional	155	[2, 7]	6.09	.95
A team-member you have a long history of working with successfully	155	[2, 7]	5.66	.94
A team-member augmented with a computer decision-making aid	152	[2, 7]	5.16	.95
A team-member augmented by an Artificial Intelligence System	154	[2, 7]	5.14	1.08
An Artificial Intelligence System (providing a recommendation without a human team-member's input)	154	[1, 7]	3.74	1.19

Source: Created by author.

The final self-reported survey measured the participant’s digital literacy through the Air Force’s Digital Literacy self-assessment. The Score is along a 4-point Likert scale with 1 being low digital literacy and 4 being high digital literacy. Additionally, the participants were able to indicate an “Unknown/No Opportunity to Perform” response to each question. For the purposes of the mean score calculations, these responses were ignored. The mean scores of the participants are seen in figure 4. The mean score was 3.36 with a standard deviation of .31. Additionally, the Air Force Digital Literacy contains eight sub-competencies. The full breakdown of the sub-competency statistics can be found in Appendix E.

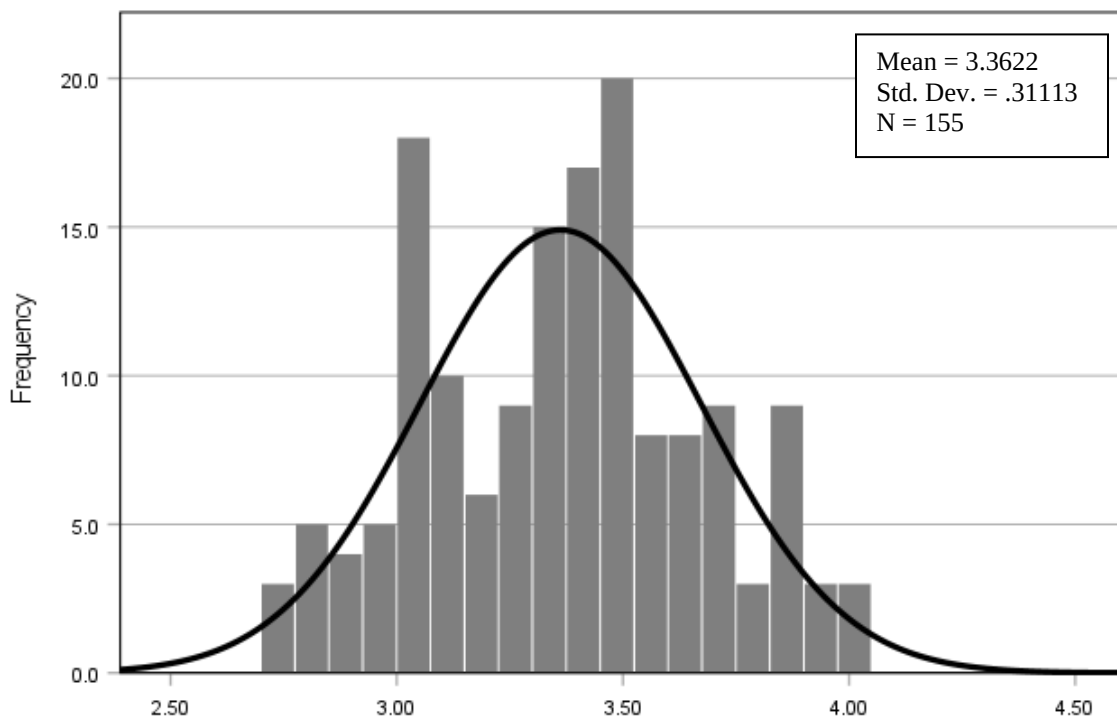


Figure 4. Mean of Participant’s Self-Reported Digital Literacy

Source: Created by author.

Hypothesis Findings

The dependent variable of trust, as measured by the chosen COA selection, is analyzed with the independent variables via an ordinal logistic regression. The independent variable of AI system transparency is measured through three vignette treatments (low, medium, or high), while the independent variable of knowledge of the AI system's performance is measured through two treatments (new or experienced). The independent variable of digital literacy is measured by the mean score on responses to the Air Force Digital Literacy self-assessment that was previously discussed in chapter 2. The summary the ordinal logistic regression testing H_1 and H_3 can be seen in table 6. table 7 shows the ordinal logistic regression testing H_2 .

Table 6. Ordinal Logistic Regression Results for Transparency and Experience (H₁ and H₃, Respectfully)

Variables	Base: Low Transparency/No Experience	
	Interaction	No interaction
Experienced	-0.315 (0.596)	0.020 (0.313)
Medium Transparency	0.273 (0.531)	0.522 (0.389)
High Transparency	0.359 (0.545)	0.533 (0.400)
Experienced*Medium Transparency	0.534 (0.783)	
Experienced*High Transparency	0.385 (0.803)	
/cut1	0.461 (0.394)	0.611 (0.330)
/cut2	1.464*** (0.411)	1.612*** (0.353)

Standard errors in parentheses, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Source: Created by author.

Table 7. Ordinal Logistic Regression Results for Digital Literacy (H₂)

Variables	Base: Low Transparency/No Experience	
	Base Model	Treatment Controls
Digital Literacy	0.488 (0.497)	0.417 (0.518)
Medium Transparency		0.462 (0.397)
High Transparency		0.507 (0.401)
Experienced		0.077 (0.321)
/cut1	1.877 (1.681)	2.012 (1.773)
/cut2	2.871 (1.690)	3.016 (1.782)
Observations	154	154

Standard errors in parentheses, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Source: Created by author.

Hypothesis #1 Findings

The results show that a participant's experience level does not influence trust in an AI system. Using the estimated logistic regression model seen in table 4, the interaction model shows that the experience coefficient does not have a significant ($p < .05$) effect on a participant's choice to trust the AI system. Likewise, there is also no significant ($p < .05$) effect for experience in the non-interaction model. Therefore, the first hypothesis, that there is a positive relationship between a person's experience level and their trust in an AI system, is not supported.

Hypothesis #2 Findings

The results show that a participant's digital literacy does not influence trust in an AI system. Using the estimated logistic regression model seen in table 5, the digital literacy coefficient does not have a significant ($p < .05$) effect on a participant's choice to trust the AI system. Digital literacy continues to show no significant ($p < .05$) effect for trust in AI when transparency and experience are controlled for, as seen in table 5 under the treatment controls column. Therefore, the second hypothesis, that there is a positive relationship between a person's digital literacy and their trust in an AI system, is not supported.

Hypothesis #3 Findings

The results show that the AI system's level of transparency does not influence trust in an AI system. Using the estimated logistic regression model seen in table 4, the interaction model shows that the transparency coefficient does not have a significant ($p < .05$) effect on a participant's choice to trust the AI system. Likewise, there is also no significant ($p < .05$) effect for transparency in the non-interaction model. Therefore, the third hypothesis, that there is a positive relationship between an AI system's level of transparency and a person's trust in that system, is not supported.

Overall Findings

Given the above results, none of the hypotheses were supported at the statistically significant level ($p < .05$). Therefore, the null hypothesis, that there is no relationship between a person's knowledge of an AI system, their digital literacy, or the transparency of the AI system and a person's trust in that system, cannot be rejected.

Qualitative Analysis

Overall, the qualitative analysis provided insights into the possible limitations of transparency as a system design method to increase trust. There appears to be a distinction between transparency and explainability that is not well examined. The desire to know “how” and “why” an AI system makes its conclusions indicates that there are possibly other human factors that moderate the effectiveness of transparency’s influence on trust. This overall interaction between human factors and an AI’s system design, specifically transparency, might be more complicated than currently understood.

The qualitative analysis revealed that participant’s views on Human-Autonomy teaming fell into three major categories: an emphasize on the wargaming results, factors related to humans, or factors related to AI. The human and AI categories were further broken into sub-categories explaining what specific factors relating to humans and AI were important. For example, some participants described AI factors as important by specifically referring to the capability of the AI system (either positively or negatively) while others described the AI’s programming or their confidence in the AI system. Finally, each code is identified as having a positive, negative, or neutral connotation regarding AI. The full breakdown of codes, categories, and the associated connotation rating can be seen in table 8. The full hierarchical breakdown of the codes and categories, along with code descriptions and examples, can be found in Appendix F.

Table 8. Qualitative Codes from the Open-Ended Question

Code	Category Sub-Category	Connotation of Human-Autonomy Teaming
Evaluation Factors Driven (AI Wargaming Method)	Wargaming Results	Positive
Evaluation Criteria Driven (Human Wargaming Method)	Wargaming Results	Negative
Deliberate Compromise	Wargaming Results	Neutral
Mismatch Between Wargaming Evaluation Methods	Wargaming Results	Neutral
Human Bias/Subjectivity	Human Related Factors	Positive
Human Experience	Human Related Factors Human “Intangibles”	Negative
“Art” of War	Human Related Factors Human “Intangibles”	Negative
AI Capabilities Mismatched with Military Operations	Human Related Factors & AI Related Factors Human “Intangibles” & AI Capability	Negative
Unable to Replicate Complex Situations	AI Related Factors AI Capability	Negative
Calculation of Large/Complex Data	AI Related Factors AI Capability	Positive
Lack of Historical Performance	AI Related Factors Confidence in AI	Negative
Lack of Personal Experience with AI System	AI Related Factors Confidence in AI	Negative
Positive Historical Performance	AI Related Factors Confidence in AI	Positive
General Lack of Knowledge Regarding AI	AI Related Factors Knowledge of AI System	Negative
Explainability	AI Related Factors Knowledge of AI System – Programming	Negative

Code	Category Sub-Category	Connotation of Human- Autonomy Teaming
Lack of Knowledge – Calculation Methods	AI Related Factors Knowledge of AI System – Programming	Negative
Lack of Knowledge – Data Inputs	AI Related Factors Knowledge of AI System – Programming	Negative
AI Bias	AI Related Factors Knowledge of AI System – Programming	Negative
Knowledge of AI Methodology	AI Related Factors Knowledge of AI System – Programming	Positive
Misread/Misunderstood Wargaming Results	Invalid Response	Neutral
Response Too Vague/Not Helpful	Invalid Response	Neutral
Mismatch Between Response and COA Selection	Invalid Response	Neutral
No Response	Invalid Response	Neutral

Source: Created by author.

There is a significant association between the connotation of the open-ended response and the participant's COA selection, $\chi^2 (4) = 121.48, p < .001$. This result indicates that the coding methodology accurately matches the participant's reasoning and intent as expressed through their open-ended responses. Among the open-ended responses collected there were a total of 124 valid responses. Each of the valid response codes were given a general indication of whether they represented a positive, negative, or neutral connotation relating to Human-Autonomy Teaming. Each participant's response codes were aggregated to provide an overall connotation. For example, if a participant

provided a response that contained both a negatively coded response and a positively coded response, their overall response would indicate a neutral viewpoint of Human-Autonomy Teams. The overall frequencies of negative, neutral, and positive responses can be seen in figure 5.

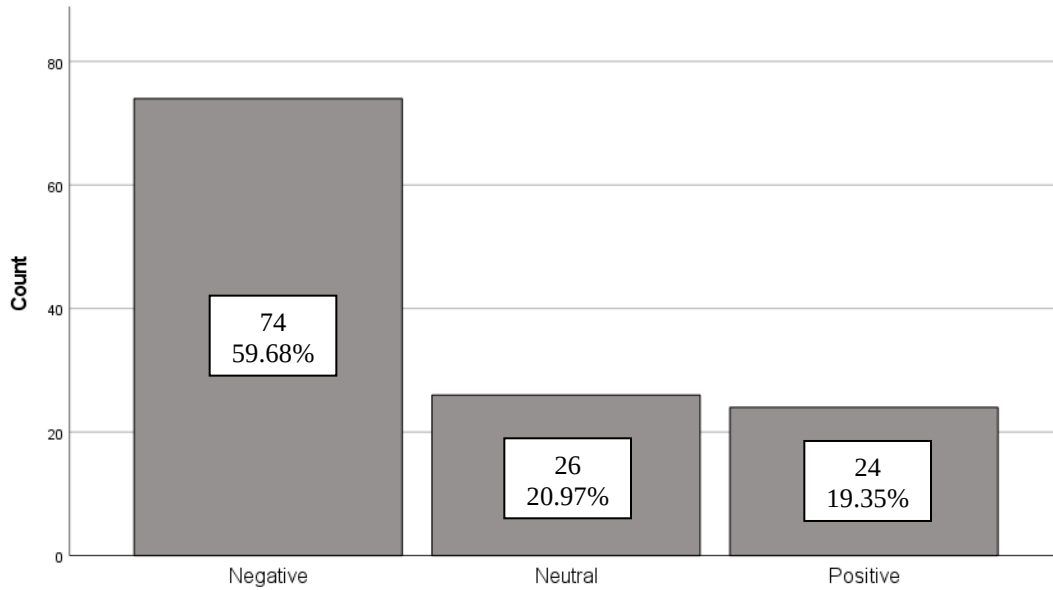


Figure 5. Connotation of Responses to the Open-Ended Question

Source: Created by author.

Assessment of Vignette Treatments

Further analysis with the qualitative coding allows insight into the effectiveness of the vignette treatment conditions. Overall, there were mixed results for the experience treatment. Sixteen of the open-ended responses expressed their confidence in the AI system as either being reduced or increased due to the historical performance of the system. Figure 6 shows how those responses broke down into the two experience-based treatment groups. Unsurprisingly, participants belonging to the new experience group

expressed a desire for more experience with the AI system. However, the experienced treatment group had three responses desiring more experience along with three responses expressing an appreciation for their experience with the AI's performance. This mixed result might indicate that a simple experience statement in the vignette is not sufficient to convey experience to a participant.

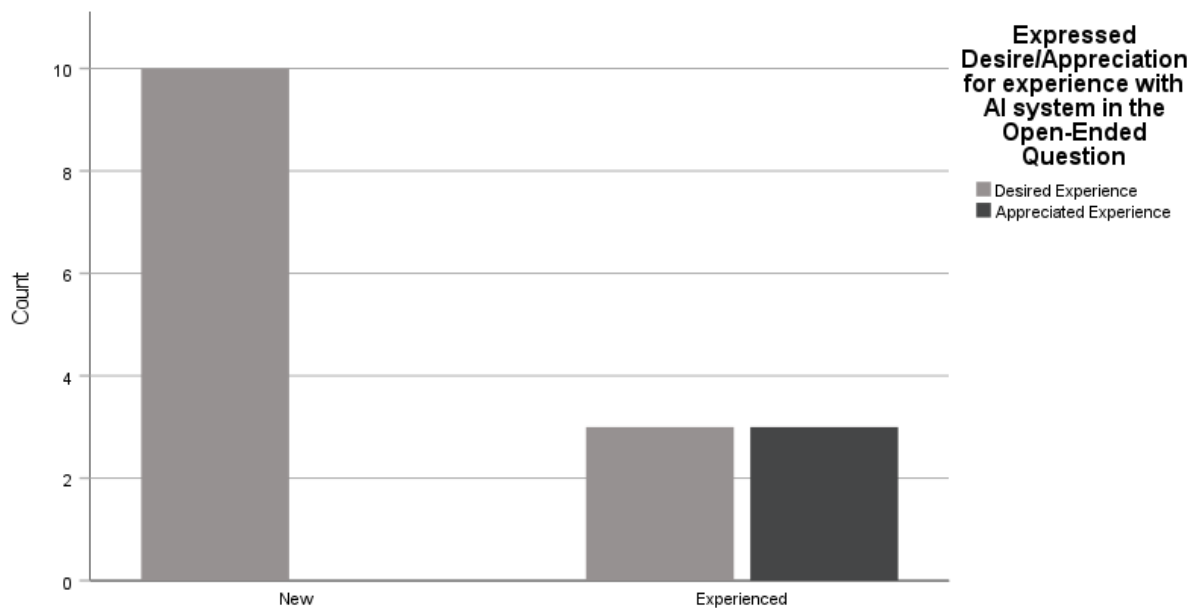


Figure 6. Responses Expressing Desire/Appreciation for Experience with AI System by Experience Treatment Group

Source: Created by author.

This analysis is also applied to the transparency treatments. Coding identified responses that indicated a desire for more transparency into the AI system and those that showed an appreciation with the transparency level of the system. Figure 7 shows the response along with the treatment groups for AI transparency. There were 38 total responses indicating a desire for more transparency and one response expressing

appreciation for the level of transparency. As anticipated, most responses (15 of 38) desiring more transparency belonged to the low transparency treatment group. However, the second most responses (13 of 38) desiring transparency belonged to the high transparency group.

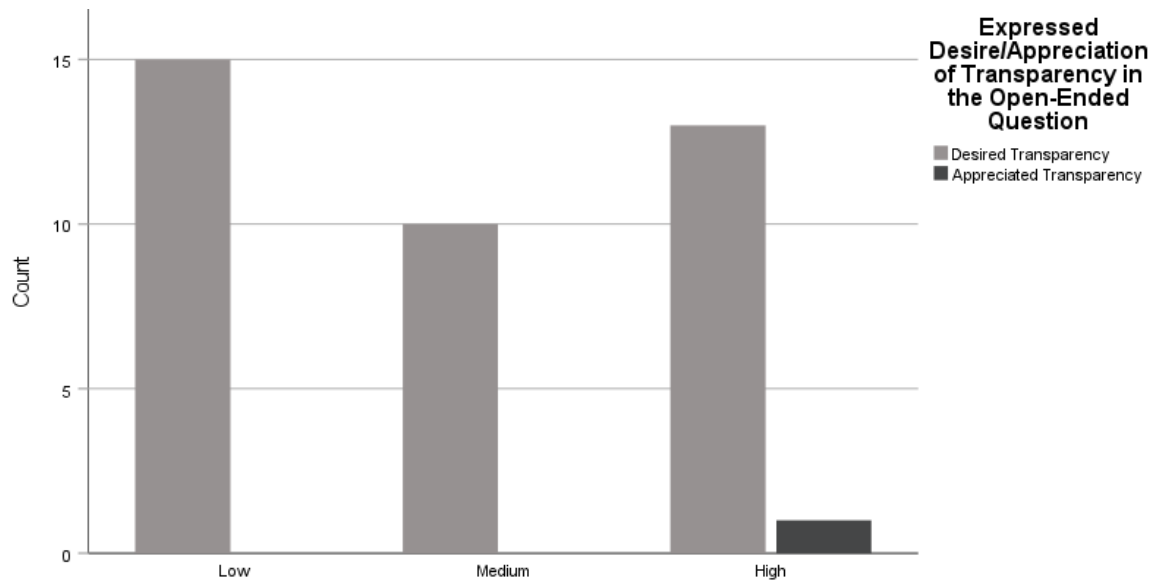


Figure 7. Responses Expressing Desire/Appreciation for Transparency with AI System by Transparency Treatment Group

Source: Created by author.

The counter intuitiveness of this result might be explained in several ways. Similar to the experience treatment, this might indicate that transparency is difficult to accurately convey in a short vignette treatment. However, given the robustness in changes between the transparency treatment groups, especially compared to the simple experience treatment change, this is a less likely explanation. Instead, a possible explanation is that an increase in transparency opens the AI system's methodology to

more questions. This explanation is further supported when isolating the “explainability” responses across the transparency treatment groups as seen in figure 8. Explainability responses were much more present in the low and medium transparency groups indicating a distinction between explainability and transparency.

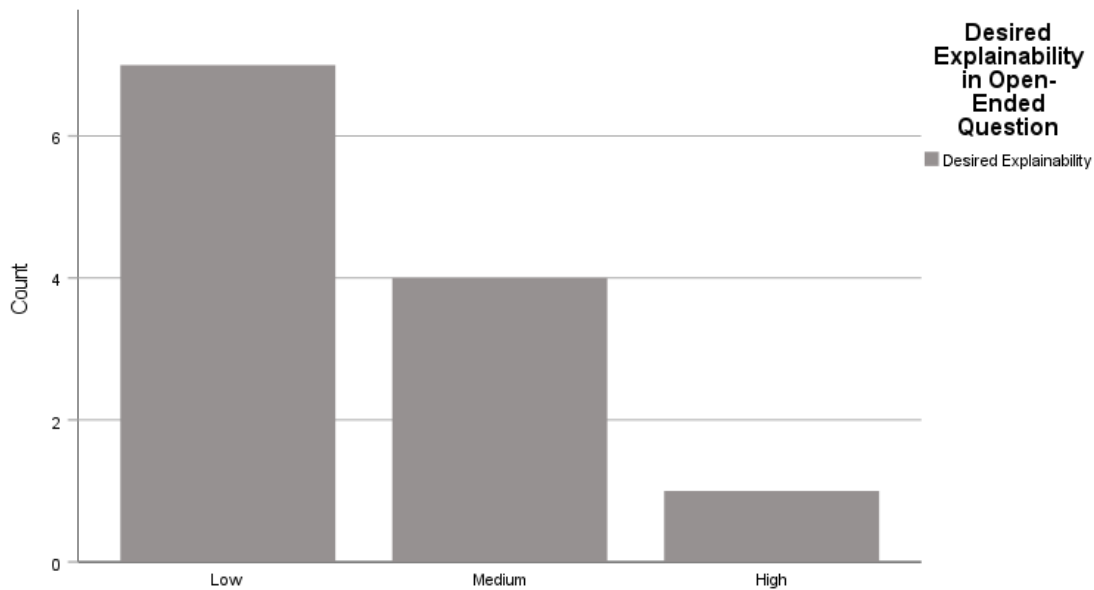


Figure 8. Responses Expressing Desire for Explainability with AI System by Transparency Treatment Group

Source: Created by author.

This distinction between the transparency and explainability might indicate that the type of transparency is what matters in establishing trust. With the low transparency treatment group, Explainability represented a desire to know “why” the system made a particular decision. However, the other negatively coded responses for transparency all expressed a desire to know “how” a system made that decision. For example, once a participant is provided with a confidence interval (as seen in the high transparency group)

they might wonder what calculations the AI performed to arrive at those results. This indicates that it is not sufficient to merely increase the transparency of an AI system, it needs to be done with certain factors in mind.

Additional Analysis

While the null hypothesis could not be rejected (i.e., none of the three other hypotheses were supported) there is additional insight that can be derived from the analysis of self-reported trust survey. Specifically, the self-reported trust survey indicates that participants have a baseline trust regarding AI, and it has a significant ($p < .001$) correlation to their eventual COA selection, even while controlling for experience, digital literacy, and transparency.

Conducting a principal component factor analysis of the ten questions, shows that there are three underlying factors.⁸⁴ The initial analysis was run to obtain eigenvalues greater than 1, with the three qualifying factors explaining 69.4 percent of the variance. Figure 9 shows the scree plot with the inflection point at three factors, while table 9 shows the factor loading after a Varimax rotation, omitting all values below .5.

⁸⁴ Note: the question “A team-member you have worked with before and found incompetent,” was inverted for the factor analysis to ensure directionality of the response was consistent with the other questions.

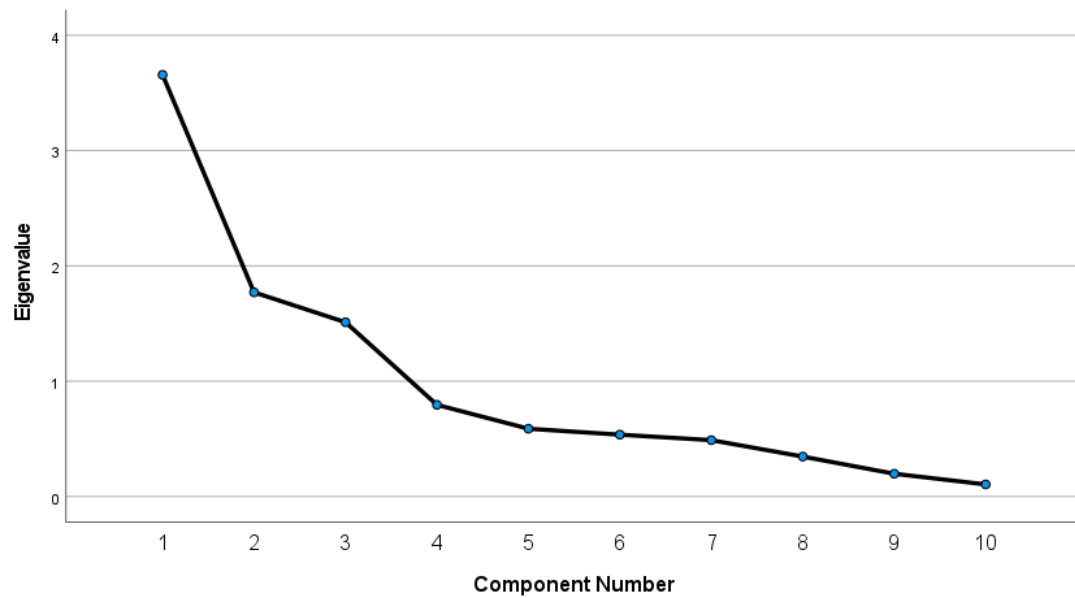


Figure 9. Scree Plot on Principal Component Factor Analysis of Self-reported Trust Survey

Source: Created by author.

Table 9. Rotated Factor Matrix Showing Values above .5

Question:	Component		
	1	2	3
A team-member you have worked with before and found exceptional	.867		
A team-member you have worked with before and found competent	.820		
A team-member you have a long history of working with successfully	.762		
A team-member of a different service as you (e.g., if you are in the Army, the team member is in the Air Force, Navy, etc.)		.808	
A new member of your team (i.e., you do not have insight into their background or history of performance)		.803	
Inverse of team-member Incompetent answer	.553	-.603	
A team-member of the same service as you (e.g., if you are in the Army, the team member is also in the Army)		.547	
A team-member augmented by an Artificial Intelligence System			.855
A team-member augmented with a computer decision-making aid			.816
An Artificial Intelligence System (providing a recommendation without a human team-member's input)			.682

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.^a

a. Rotation converged in 5 iterations.

Source: Created by author.

Factor Analysis and Trust in AI

Using the existing theories of trust discussed in chapter 2, the questions are analyzed to identify potential theoretical explanations for the commonalities between several questions and their factor loadings. Accordingly, factor 1 most likely represents an individual's preference to have experience with the teammate to trust, aligning to the

incremental approaches to trust, specifically social exchange theory. Similarly, factor 2 can be explained as a representation of the characteristics of a team member, such as being part of a homogeneous group, and their impact on a person's trust. This would align with a non-incremental approach to trust, specifically social categorization theory. Finally, factor 3 represents a generalized trust in AI across all the various integration levels (i.e., human-in-the-loop, human-on-the-loop, and human-out-of-the-loop).

Using these factors allows for additional explanations for why participants either trust or do not trust AI (as determined by their COA selection). The responses to the questions related to the factors were averaged giving participant's three mean scores related to the identified sub-categories of trust: experienced based trust (questions 4, 5, 6, and 7), characteristic based trust (questions 1, 2, 3, and 4), and trust in AI (questions 8, 9, and 10). The resulting logistic regression between these variables and COA selection can be seen in table 10. Critically, a participant's trust in AI, as measured by their mean response to the last three questions the self-reported trust survey, is shown to be statistically significant ($p < .001$) to a person's COA selection, across all treatment controls.

Table 10. Ordinal Logistic Regression Results for Mean Trust in AI

Variables	Base: Low Transparency/No Experience			
	Base Model	Treatment Controls	Treatment Controls	Treatment Controls
AI Trust (Mean)	0.701*** (0.201)	0.756*** (0.208)	0.720*** (0.213)	0.860*** (0.229)
Medium Transparency		0.631 (0.403)	0.622 (0.403)	0.659 (0.410)
High Transparency		0.725 (0.416)	0.681 (0.421)	0.687 (0.430)
Experienced		-0.120 (0.324)	-0.154 (0.328)	-0.136 (0.333)
Experience Based Trust (Mean)			0.169 (0.250)	0.277 (0.254)
Characteristic Based Trust (Mean)				-0.662** (0.245)
/cut1	3.475*** (0.957)	4.145*** (1.052)	4.903** (1.550)	3.231 (1.670)
/cut2	4.535*** (0.985)	5.227*** (1.083)	5.988*** (1.576)	4.367** (1.688)
Observations	154	154	154	154

Standard errors in parentheses, *** p<0.001, ** p<0.01, * p<0.05

Source: Created by author.

Possible Explanations for a Participant's Self-Reported Trust in AI

Further analysis is conducted to examine possible explanations for this underlying trust in AI. Using a participant's mean AI trust score as a dependent variable, analysis is run on three possible independent variables: a participant's digital literacy (including a breakdown of the eight sub-competencies), whether a participant has a STEM degree, and whether a participant indicated a positive viewpoint of an AI system's capabilities related to the complexity in an operational environment (as measured by responses to the open-ended question). While, neither digital literacy or STEM were shown to have a

significant ($p < .05$) effect on self-reported trust, there was a significant ($p < .05$) relationship between self-reported trust and holding a positive view on AI capabilities.

Trust in AI and Digital Literacy and Sub-competencies

Digital literacy scores as well as the mean scores for the eight sub-competencies (appropriate social media use, information literacy, communication medium choice, technological exploration, information protection, software proficiency, active internet use, and general computer skills) are ran as the independent variables in a simple linear regression with mean AI trust as the dependent variables. The full analysis of all the digital literacy and sub-competencies results is found in Appendix E. However, a selection of variables (digital literacy, information literacy, technological exploration, and general computer skills) is found in table 11. These variables were selected due to the strength of their theoretical explanations for a possible relationship to trust in AI.

Table 11. Linear Regression Results of Predictors of Participant's Trust in AI (Mean)

Variable	<i>B</i>	<i>SE</i>	<i>p</i>	95% CI	R^2
Digital Literacy	.349	.209	.095	[-.051, .760]	.015
Information Literacy	.266	.157	.088	[-.027, .578]	.022
Technological Exploration	.146	.096	.123	[-.047, .354]	.014
General Computer Skills	.208	.130	.102	[-.043, .496]	.021

95% bias corrected and accelerated confidence intervals and standard errors based on 1000 bootstrap samples

Source: Created by author.

However, as seen, none of the variables are statistically significant ($p < .005$) predictors of a participant's mean trust in AI. However, compared to the other sub-competencies, they offer potential insight into what might cause a higher trust in AI. For example, information literacy is defined in the Air Force's digital literacy self-assessment as being when an individual "Recognizes when information is needed and locates, evaluates, and uses the needed information to achieve goals."⁸⁵ This sub-competency might explain an individual's trust in AI, especially if they have an idea of the source of data inputs into the AI. However, the collected data did not have enough positively coded responses to test this notion.

Trust in AI and STEM Degrees

The relationship between a participant's mean trust in AI and STEM degree was examined using an independent t-test. On average, participants that had a STEM degree had a higher trust of AI ($M = 4.73$, $SE = .097$), than those without a STEM degree ($M = 4.59$, $SE = .105$). This difference, $-.14$, BCa 95% CI $[-.430, .143]$, was not significant $t(144) = -.939$, $p = .329$; additionally, this represented a small effect size, $r = .08$.

Trust in AI and view on AI Capabilities

The relationship between a participant's mean trust in AI and their expressed viewpoint on the capabilities of AI to handle complexity (as measured through the open-ended responses) was also measured using an independent t-test. On average, participants that had a positive view on AI capabilities had a higher trust of AI ($M = 5.13$, $SE = .163$),

⁸⁵ Barron, *Evaluation of Digital Literacy as an Air Force Foundational Competency*, 23.

than those with a negative view ($M = 4.40$, $SE = .213$). This difference, $-.723$, BCa 95% CI $[-1.252, -.232]$, was significant $t(32) = -2.70$, $p = .008$; additionally, this represented a medium effect size, $r = .43$.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

This study began with a hypothetical scenario where a J5 team failed to use an AI system to augment their decision-making during operational planning. The goal of the study was to investigate the possible factors that might influence an individual's trust in AI systems to prevent such failures from occurring in the near future as Human-Autonomy Teams become widely fielded and are no longer regulated to science fiction. Three fundamental questions emerged. Does the interaction between a person and an AI system affect a person's trust of that system? What human factors influence a person's trust in an AI system? What AI system factors influence a person's trust in an AI system?

Based on the above questions and the existing body of research into Human-Autonomy Teaming and trust, a survey experiment was conducted to examine three hypotheses. First, that there is a positive relationship between a person's level of understanding of an AI system's ability and their trust in the system. Second, there is a positive relationship between a person's digital literacy and their trust in an AI system. Third, there is a positive relationship between the transparency of an AI system and a person's trust in that system. Therefore, the resulting null hypothesis was that no interaction between a person's knowledge of an AI system, their digital literacy, or in an AI system's transparency would have an influence on trust during Human-Autonomy Teaming.

Overall, none of the three hypotheses were supported and the null hypothesis could not be rejected. Instead, additional analysis revealed that people have a baseline level of trust in AI and the experiment treatments were unable to influence it.

Furthermore, it appears that digital literacy is not a sufficient indicator of this baseline trust in AI, as hypothesized, however, qualitative analysis does show that a positive view of AI capabilities does have a significant ($p = .008$) effect. Finally, the qualitative analysis adds to the existing literature by showing possible limitations to transparency as an AI system design method to increase trust. Together, these results allow for an update to the integrated trust model introduced in chapter 2, seen in figure 10. Finally, findings have important implications for the Joint Force as AI systems are being designed to augment, or even replace, human decision-making.

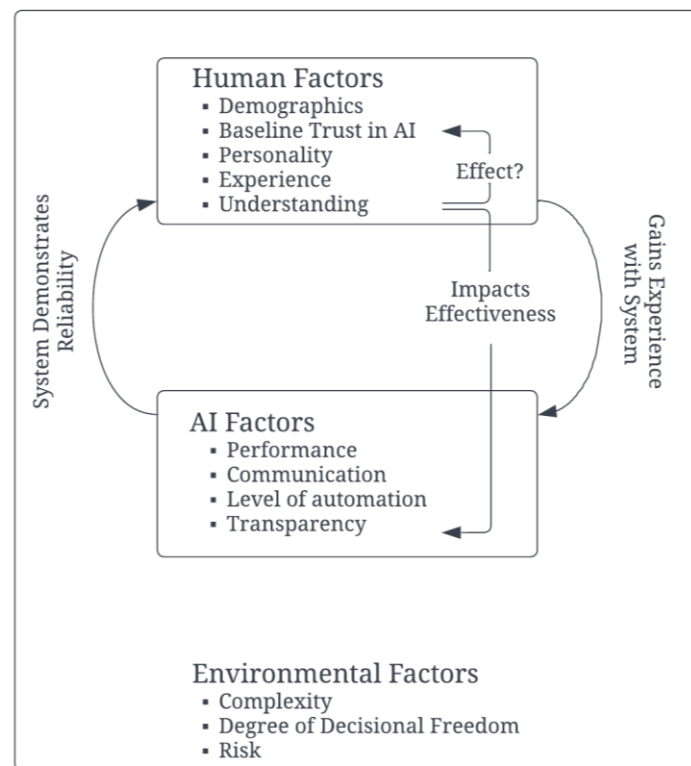


Figure 10. Updated Integrated Model of Trust for HAT

Source: Created by author.

Recommendations for Joint Force Human-Autonomy Teams

Clearly there are implications related to AI system design choices that are revealed through this study, including the importance that people place on transparency and explainability, as seen in the qualitative analysis. However, more importantly, this study adds to the body of research by suggesting that design choices themselves might not be sufficient to impact an individual's level of trust in an AI. Instead, there are still unexplored factors that might explain whether someone chooses to trust an AI system. Of interest to the military is whether these unexplored factors can be measured, and if so, can they be increased through training.

Limits of Transparency in System Design

One of the key insights of the study, adding to the body of knowledge regarding AI system design, is that increasing system transparency is not enough to increase trust. This study's qualitative analysis shows that the demands by users to understand an AI system's design are not necessarily satisfied as an AI system's transparency increases. A logical explanation for this occurrence is that certain levels of transparency generate further questions into a system's design. These newly generated questions regarding the system's design eventually translate to a negative connotation surrounding the AI system and lead to lower trust. This explanation is supported through the open-ended responses expressing a desire for more transparency being spread evenly across the treatment groups compared to the desire for explainability which decreased as transparency increased. Admittedly, the small sample size of respondents discussing transparency and the limited conditions of the transparency treatment in the experiment mean that this explanation cannot be generalized at this time.

Increasing Trust in AI

The other key insight that emerged from the study is the finding that a person's trust in AI was determined prior to using the system and neither experience with the system nor system design features were significant predictors of increased trust. In other words, if a person entered the experiment with a lower level of trust for AI, then increasing their experience or transparency did not change their trust. This result has several significant consequences for the military regarding possible screening criteria and training opportunities.

Selection of Military Members for Human-Autonomy Teaming

One of the most novel consequences is that the factor analysis of the self-reported trust survey showed that measuring trust in AI can be measured through a survey instrument and it is significantly coordinated to a person's willingness to be on a Human-Autonomy Team. Given a fielded AI system in the future that needs willing human teammates, the military could use a self-reported trust survey as a screener for selection. If the goal is to merely select military members that are going to integrate with fielded AI systems, and the literature suggests that this increases team effectiveness, then the self-reported trust survey, with modifications and further testing, would be a quick and inexpensive solution that could be implemented almost immediately.

Implications for Training to Increase Trust in AI

However, selection is just one of the problems that faces the military eventual fielding of AI systems. This study was not large enough to indicate what a minimum threshold in AI trust should be to warrant selection, and certainly did not survey enough

members to know if the military even has enough members that meet such a criterion. Instead, a more interesting use for understanding this underlying trust in AI is in its application to develop targeted training for military members.

While it was hypothesized that an individual's digital literacy would be one of predictors for trust in AI, this was not shown to be a statistically significant ($p < .05$), although it did begin to approach significance ($p = .095$). Similarly, examining the sub-competencies provides an indication of what factors might influence trust. For example, the sub-competency of information literacy approached ($p = .088$) but did not reach a statistical significance ($p < .05$). Furthermore, both digital literacy and informational literacy had positive coefficients, indicating that increases in literacy would have a positive effect on trust.

In addition to measuring digital trust by design as one of the experiment's independent variables, the collection of demographic and qualitative responses allowed for several other possible predictors of an individual's trust in AI to be explored. As outlined in chapter 4, two specific variables were considered, whether a participant had a STEM degree and whether they had a positive view relating to the capabilities of AI systems. While having a STEM degree was not shown to be significantly ($p < .05$) correlated with AI trust, a positive view on AI system capabilities was significantly ($p=.008$) correlated. One possible explanation is that a positive view in an AI system's capabilities is due to specific knowledge of AI, unlike the more generalized measures of digital or information literacies. If this is the case, then it would imply that increasing AI education efforts could lead to more effective Human-Autonomy Teams overall. However, the limits in qualitative responses and the fact that no dedicated measure to

study AI knowledge was used, means that this notion cannot be supported with this study alone and further research is needed.

Limitations

With a total response rate of roughly 16%, there are implications for the statistical strength of the results. As previously highlighted, the range of respondents for each treatment group was from 22 to 29. With this number of respondents there is the potential for Type II errors (i.e., there is a significant effect that was not able to be detected). This limitation was mitigated with the use of a qualitative open-ended question that allowed additional insight into participant's thinking outside of the quantitative measurements of trust. However, additional research into the areas that are supported by the mixed methods approach is still warranted.

The other limitation that arose during analysis was the possible weakness of the vignette treatments. The qualitative analysis revealed an inconsistency in responses to the experience treatment in particular. While this finding is limited due to the number of responses to the open-ended question specifically mentioning experience as a driving factor in COA selection, it does indicate that experience is harder to convey in a vignette than originally anticipated. This is compounded by the fact that the experience treatment was limited to a single line. Compared to the transparency treatment which had multiple lines and a robust change in graphical representations throughout the different treatment levels. Therefore, experience should continue to be examined as a possible influence into trust.

Recommendations for Additional Research

The recommendations for the application of this study's results for the Joint Force can be strengthened with additional research based on the limits of the study's experimental design, population under study, and the response rate. The recommended areas for improvement aim to bolster this study's findings with more targeted experiments to better understand transparency limitations in system design or widen the data collection to other possible predictors of an individual's baseline trust in AI. Additionally, another possible influence relating to the interaction between human and AI system should be examined, specifically relating to the level of interaction that the system is designed to operate (i.e., human-in-the-loop, human-on-the-loop, or human-out-of-the-loop).

Explainability and Transparency in AI System Design

As discussed in chapter 4, there appears to be a clear distinction between the concepts of explainability and transparency. Explainability in AI is an emerging field and due to the complexity of AI systems, one that has not been adequately solved. Further research is clearly needed and is being appropriately advocated for at multiple level, both in the government as well as in the private sector.

However, research also seems to indicate that AI systems might have limitations regarding the usefulness of transparency. This study was limited in presenting various levels of transparency through vignette treatments, however, more focused research on system design around transparency would provide useful insights. One emerging research question is whether there are any limits to when transparency stops adding value and instead invites more questions from a human teammate. Theoretically, an individual's

knowledge level would impact where this limit would occur, however, there were not enough responses during this study to examine any potential relationship in detail. Finally, there is also an avenue of study regarding the type of information an AI system presents to teammates and the effect that has on trust.

Factors Influencing AI Trust

While the additional analysis conducted during this study provided possible explanations for underlying factors that might influence an individual's baseline trust in AI, more research is needed. The most promising approaches are those that were indicated during this study, but not found to be statistically significant ($p < .05$). Other measures of various literacies, such as information literacy, as defined in the Air Force digital literacy self-assessment, or data literacy as discussed in chapter 2 provide obvious starting points. The challenge, as it was in the development of this study, is in finding an instrument that can adequately measure the literacy in question without hurting participation rates. The more specific a particular literacy might get as a predictor into trust in AI, the more refined the measurement must become, often to the point where instruments become time-consuming for participants.

However, another interesting explanation that emerged through the qualitative responses is a more specific level of knowledge related directly to AI. This theme was shown to be significant in its relationship with trust in AI, however, it was analyzed using only a small sample of participants based on those that responded either positively or negatively regarding AI capabilities. A knowledge test of AI technologies should be possible to create and field and would give a clearer indication as to whether specific knowledge regarding AI is a significant predictor for trust.

Levels of Automation and its Effect on Trust

The final theme that appeared to emerge but was not a planned part of this study's design, was the influence of the level of automation had on trust. As seen in the self-reported trust survey, there was a distinct drop off in trust as automation moved from human-on-the-loop towards human-out-of-the-loop systems. However, the results from this study are insufficient to generalize since each measure of trust was based on a single question that was self-reported. The results can be strengthened with a deliberate design methodology, with levels of automation serves as an independent variable. If the level of automation is a sufficient predictor of trust, then the AI system design implications are clear, so studying this potential relationship is a worthwhile endeavor for future research.

Conclusions

With the advent of AI system being fielded to augment decision-making at the operational levels, it is critical for the military to understand the factors that influence trust in Human-Autonomy Teams. This study's significance was to examine that problem as it related to AI system design and military member education. While the hypothesized factors (level of knowledge, digital literacy, and AI system transparency) were not shown to be significant predictors in trust, the research did show that self-reported trust in AI was a significant predictor in following AI recommendations in a hypothetical vignette experiment. This was true even when controlling for experience, digital literacy, and AI transparency, implying that a person's trust in AI is less malleable to system design choices than originally thought.

Further analysis provides additional insights into the possible causal factors for this trust. While digital literacy was not shown to be significantly significant as a

predictor towards trust, both it and the sub-competency of information literacy began to approach significance ($p = .095$ and $p = .088$, respectfully). Finally, participant's responses indicating a positive view of an AI's capabilities were shown to be significantly correlated with an increase in trust in AI.

If the Joint Force needs to build Human-Autonomy Teams to be successful in future conflicts, this research indicates that it can screen candidates to make those teams as effective as possible by selecting members that have a high trust in AI and are willing to receive inputs and recommendations from an autonomous teammate. However, the research also provides explanations for what might cause a higher trust in AI. By further studying these underlying causes to trust in AI, the military can better provide education opportunities to train the appropriate force for future operational demands.

APPENDIX A

RESEARCH SURVEY

Vignette

Treatment conditions are provided in the brackets. There are three transparency treatments (low, medium, and high) and two knowledge treatments (previous experience or no previous experience) for a 3 x 2 factorial design and 6 different treatments.

Consider the following scenario:

It is 2028 and you are the lead planner on a EUCOM J5 Mission Planning Team (MPT) focused on OPLAN development. Your team has been tasked to update a portion of an OPLAN and present your courses of action (COAs) to the J5 director for a decision. Your team is augmented by an AI system that analyzes the operating environment [Transparency treatment: low - omitted / medium – “using information from multiple Intelligence Community databases, military databases, and open-source databases” / high – “using information from multiple Intelligence Community databases, military databases, and open-source databases and then running thousands of simulations”] to develop possible COAs. The AI system can also wargame COAs and provide an objective recommendation on which COA is best. [Knowledge treatment: No previous knowledge – “This is the first time you or your team have used this system.” / Previous knowledge – “You and your team have used this system in the past for smaller scale plans and have found its performance reliable.”]

You decide to break your team up into two groups (Team A and Team B) to develop COAs 1 and 2, while the AI system develops COA 3. After several days, your teams reconvene and wargame all three COAs based on the following evaluation criteria:

Surprise, Risk, Tempo, Mass, and Flexibility. Simultaneously, the AI system wargames all three COAs. Upon completion, the AI system recommends the COA that it created (COA 3), followed by Team B's COA (COA 2), and lastly Team A's COA (COA 1). [Transparency treatment: Low - omitted / medium – “The AI system provides a detailed list of the various factors it used to arrive at this conclusion.” / high –“The AI system provides a detailed list of the various factors it used to arrive at this conclusion and the methods used for its calculations.”] The AI's recommendations conflict with the results of your team's wargame, which found that Team A's COA (COA 1) is best, followed by Team B's COA (COA 2), and finally the AI's COA (COA 3). The wargaming results are displayed below:

[Low Transparency]

COA Ranking	AI Wargaming Results	J5 Team Wargaming Results
1 (Recommended)	COA 3 (AI COA)	COA 1 (Human Team A COA) <u>COA Evaluation Criteria</u> • Surprise + • Risk - • Tempo + • Mass + • Flexibility -
2	COA 2 (Human Team B COA)	COA 2 (Human Team B COA) <u>COA Evaluation Criteria</u> • Surprise - • Risk + • Tempo - • Mass - • Flexibility -

3	COA 1 (Human Team A COA)	COA 3 (AI COA) <u>COA Evaluation Criteria</u> • Surprise - • Risk - • Tempo - • Mass - • Flexibility +
---	--------------------------	---

[Medium Transparency]

COA Ranking	AI Wargaming Results	J5 Team Wargaming Results
1 (Recommended)	COA 3 (AI COA) <u>COA Evaluation Factors</u> • Calculated probability of mission success: 94%	COA 1 (Human Team A COA) <u>COA Evaluation Criteria</u> • Surprise + • Risk - • Tempo + • Mass + • Flexibility -
2	COA 2 (Human Team B COA) <u>COA Evaluation Factors</u> • Calculated probability of mission success: 87%	COA 2 (Human Team B COA) <u>COA Evaluation Criteria</u> • Surprise - • Risk + • Tempo - • Mass - • Flexibility -
3	COA 1 (Human Team A COA) <u>COA Evaluation Factors</u> • Calculated probability of mission success: 76%	COA 3 (AI COA) <u>COA Evaluation Criteria</u> • Surprise - • Risk - • Tempo - • Mass - • Flexibility +

[High Transparency]

COA Ranking	AI Wargaming Results	J5 Team Wargaming Results
1 (Recommended)	COA 3 (AI COA) <u>COA Evaluation Factors</u> - Calculated probability of mission success: 94% - Calculated loss to forces: 12% <u>Methodology</u> - Simulations Ran: 10,000 - Confidence Interval on Results: +/- 4%	COA 1 (Human Team A COA) <u>COA Evaluation Criteria</u> - Surprise + - Risk - - Tempo + - Mass + - Flexibility -
2	COA 2 (Human Team B COA) <u>COA Evaluation Factors</u> - Calculated probability of mission success: 87% - Calculated loss to forces: 17% <u>Methodology</u> - Simulations Ran: 10,000 - Confidence Interval on Results: +/- 3%	COA 2 (Human Team B COA) <u>COA Evaluation Criteria</u> - Surprise - - Risk + - Tempo - - Mass - - Flexibility -
3	COA 1 (Human Team A COA) <u>COA Evaluation Factors</u> - Calculated probability of mission success: 76% - Calculated loss to forces: 21% <u>Methodology</u> - Simulations Ran: 10,000 - Confidence Interval on Results: +/- 5%	COA 3 (AI COA) <u>COA Evaluation Criteria</u> - Surprise - - Risk - - Tempo - - Mass - - Flexibility +

Given the information above, which COA will you recommend to the J5 Director:

- a) COA 1 (The J5 team's recommended COA and the least favored COA of the AI recommendations)
- b) COA 2 (Neither the J5 team nor the AI's recommended COA, but the second choice of both)
- c) COA 3 (The AI's recommended COA and the least favored COA of the J5 team)

Briefly explain what factors influenced your decision:

Self-Reported Trust Survey

Now consider the following scenario. You are the Director of the EUCOM J5 and must select a COA recommendation from your team. All the possible COAs have passed screening criteria (i.e., they are suitable, feasible, acceptable, distinguishable, and complete). Given your knowledge of the Joint Planning Process (specifically the steps of COA development, COA analysis and wargaming, and COA comparison), what is the probability that you will select the recommended COA, given the team lead is...

Extremely Unlikely	Very Unlikely	Unlikely	Neutral	Likely	Very Likely	Extremely Likely
-----------------------	------------------	----------	---------	--------	----------------	---------------------

- 1) A new member of your team (i.e., you do not have insight into their background or history of performance)
- 2) A team-member of the same service as you (e.g., if you are in the Army, the team member is also in the Army)
- 3) A team-member of a different service as you (e.g., if you are in the Army, the team member is in the Air Force, Navy, etc.)

- 4) A team-member you have worked with before and found incompetent
- 5) A team-member you have worked with before and found competent
- 6) A team-member you have worked with before and found exceptional
- 7) A team-member you have a long history of working with successfully
- 8) A team-member augmented with a computer decision-making aid
- 9) A team-member augmented by an Artificial Intelligence System
- 10) An Artificial Intelligence System (providing a recommendation without a human team-member's input)

Digital Literacy Survey

Rate your level of agreement with each statement: I consistently and effectively...

Strongly Disagree	Disagree	Agree	Strongly Agree	Unknown/No Opportunity to Perform
-------------------	----------	-------	----------------	-----------------------------------

1. Avoid sharing videos and photos on social networking sites that would harm my reputation
2. Block requests from unknown/unwanted people on social networking sites or other online forums
3. Protect personal information shared on social networking sites and other online forums
4. Identify points of agreement and disagreement among Internet sources
5. Evaluate Internet sources
6. Determine the authoritativeness and reliability of Internet sources

7. Choose an appropriate communication medium based on the need to convey nonverbal cues (e.g., in-person meeting or video teleconference for tactfully addressing a disputed issue)
8. Choose an appropriate communication medium based on the need for immediate feedback (e.g., phone call rather than e-mail to address time-sensitive issue)
9. Choose an appropriate communication medium based on the need for careful editing of the message (e.g., e-mail rather than phone to convey complex factual information)
10. Acquire new digital technology before it becomes widespread
11. Keep up with the latest digital technological developments
12. Figure out new high-tech digital products and services without help from others
13. Securely store computer files
14. Lock electronic devices when not in use
15. Exercise caution when transmitting electronic information (e.g., encrypting personally identifiable information [PII])
16. Use spreadsheet software (e.g., Microsoft Excel)
17. Use presentation software (e.g., Microsoft PowerPoint)
18. Use word processing software (e.g., Microsoft Word)
19. Use the Internet to search for public information
20. Participate in online training
21. Use the Internet to learn on my own
22. Troubleshoot computer problems
23. Use computer or software user's guides when help is needed
24. Learn to use a variety of software programs

Demographical Questions

Gender [Dropdown with the following selections:]

Male

Female

Non-Binary/Other

Prefer not to answer

Race/Ethnicity

Are you of Hispanic, Latino, or of Spanish origin? [Dropdown with the following selections:]

Yes

No

How would you describe yourself? [Dropdown with the following selections:]

Asian

Black or African-American

Hawaiian or Pacific Islander

Native American or Native Alaskan

White or Caucasian

Biracial or Multiracial

Other/Unknown

Prefer not to answer

Age

Open-ended

Grade [Dropdown with the following selections:]

O-1

O-2

O-3

O-4

O-5

O-6

O-7

O-8

O-9

O-10

N/A

Service [Dropdown with the following selections:]

Army

Air Force

Navy

Marine Corps

Coast Guard

U.S. Government Civilian

Branch (or Service Equivalent [e.g., Infantry, Pilot, Surface Warfare Officer, etc.])

Open-ended

What is the highest level of education you have completed? [Dropdown with the following selections:]

High school

Associates degree

Bachelor's degree

Master's degree

Ph.D. or professional degree

Prefer not to say

What was the major in your degree? If you have earned multiple degrees, please list all majors.

Open-ended

APPENDIX B
RECRUITMENT E-MAIL

Subject: MMAS Survey on Trust and Artificial Intelligence

CGSC Resident Students,

This email is soliciting participation in another student's Master's in Military Art and Science research study on the factors that influence trust in AI systems.

You can find the anonymous survey link below and paste to your web browser.

%URL%

The survey should take no more than 15 minutes to complete.

Your participation is completely voluntary and would greatly be appreciated.

If you have any additional questions about the research, or require more information, the student's contact information is listed below:

Christopher Kean, Maj, USAF

CGSC Student

Staff Group 21A

Cell: 719-339-4965

E-mail: Christopher.r.kean.mil@army.mil

APPENDIX C

CONSENT STATEMENT

Study Summary

You are being asked to join a research study on people's attitudes toward Artificial Intelligence (AI). The results of this study will help senior leaders better develop and refine future training requirements and help in the design of AI systems.

Important things to know:

- Taking part in research is voluntary. You can choose not to be in this study or stop at any time.
- If you decide not to be in this study, your choice will not affect your completion of CGSC. There will be no penalty to you.

Introduction

You are being asked to take part in this study because you are currently a resident student at the Command and General Staff College at Fort Leavenworth, Kansas. The main purpose of this research is to create new knowledge for military decision makers and may or may not help the people who take part in the study.

Study Purpose and Procedures

- This study involves research that is being done through an anonymous online survey.
- The purpose of the research is to study the possible factors that influence a person's level of trust when using an AI system.

- This is an unclassified study; no classified information should be discussed.

Potential violations of this guidance could result in administrative action, such as the loss of security clearance, or punitive punishment in accordance with the Uniformed Code of Military Justice or criminal law.

- The survey should take 15 minutes to complete.

Benefits of Participating in the Study

- No direct benefits to the participant can be expected.

Risks of Participating in the Study

- By participating in the study, you are not expected to have any more risk than you would meet during everyday life.

Confidentiality

The researchers will do everything possible to keep your records confidential. However, it cannot be guaranteed. Certain information may need to be reported to agencies as required by law. Survey records will be kept by the CGSC Quality Assurance Office.

Compensation for Participation

- There is no compensation for taking the survey.

Researcher Contact Information

Christopher Kean, Maj, USAF

CGSC Student

Staff Group 21A

(719) 339-4965 CELL

Human Protections Director Contact Information

Michelle A. Miller, PhD

Human Protections Director

CGSC Human Research Protection Program

Lewis & Clark Office 4521

(913) 684-7311 OFFICE

(913) 682-1185 TELEWORK

(913) 306-3671 CELL

Withdraw Clause

You are free to refuse to take part in this research project or to withdraw your consent and stop participation in the project at any time.

By continuing onto the survey, you show that all your questions have been answered and you agree to take part in the study.

APPENDIX D

DEMOGRAPHIC STATISTICS

The detailed frequency distributions for the participants' responses to the demographical questions are presented in tables 12-19 and figure 11.

Table 12. Demographic Question: Gender

Reported Gender	Frequency	Percent
Male	117	75.5
Female	34	21.9
Prefer not to Answer	3	1.9
No Response	1	.6
Total	155	100.0

Source: Created by author.

Table 13. Demographic Question: Race/Ethnicity - Are you Hispanic, Latino, or of Spanish origin?

Are you Hispanic, Latino, or of Spanish origin?	Frequency	Percent
Yes	9	5.8
No	143	92.3
No Response	3	1.9
Total	155	100.0

Source: Created by author.

Table 14. Demographic Question: Race/Ethnicity - How would you describe yourself?

How would you describe yourself?	Frequency	Percent
Asian	9	5.8
Black or African-American	6	3.9
Hawaiian or Pacific Islander	2	1.3
Native American or Native Alaskan	1	.6
White or Caucasian	113	72.9
Biracial or Multiracial	7	4.5
Other/Unknown	5	3.2
Prefer not to answer	11	7.1
No Response	1	.6
Total	155	100.0

Source: Created by author.

Table 15. Demographic Question: Grade

Reported Grade	Frequency	Percent
O-3	11	7.1
O-4	142	91.6
O-5	1	.6
No Response	1	.6
Total	155	100.0

Source: Created by author.

Table 16. Demographic Question: Service

Reported Service	Frequency	Percent
Army	120	77.4
Air Force	29	18.7
Navy	0	0.0
Marine Corps	4	2.6
Coast Guard	1	.6
U.S. Government Civilian	0	0.0
No Response	1	.6
Total	155	100.0

Source: Created by author.

Table 17. Demographic Question: Branch

Reported Branch	Frequency	Percent
Acquisition	2	1.3
AG	7	4.5
Air Defense	5	3.2
Armor	2	1.3
Army Aviation	8	5.2
Artillery	9	5.8
Chaplain	1	.6
Chemical	1	.6
Civil Affairs	1	.6
Cyber	6	3.9
Engineer	12	7.7
FA59	1	.6
Finance	1	.6
Force Support	2	1.3
Infantry	12	7.7
Intelligence	21	13.5
JAG	3	1.9
Logistics	18	11.6
Maintenance	2	1.3
Medical	6	3.9
No Response	7	4.5
ORSA	2	1.3
PA	1	.6
PYOPS	2	1.3
Signal	7	4.5
Space Ops	1	.6
Special Force	1	.6
Special Forces	2	1.3
USAF Aviation	11	7.1
Weather	1	.6
Total	155	100.0

Source: Created by author.

Table 18. Demographic Question: What is the highest level of education you have completed?

What is the highest level of education you have completed?	Frequency	Percent
High school	0	0.0
Associates degree	0	0.0
Bachelor's degree	42	27.1
Master's degree	103	66.5
Ph.D. or professional degree	9	5.8
Prefer not to Answer	0	0.0
No Response	1	.6
Total	155	100.0

Source: Created by author.

Table 19. Demographic Question: Is your degree Science, Technology, Engineering, Mathematics (STEM)

Is your degree Science, Technology, Engineering, Mathematics (STEM)	Frequency	Percent
Yes	58	37.4
No	88	56.8
No Response	9	5.8
Total	155	100.0

Source: Created by author.

NOTE: The demographic question asked on the survey was: "What was the major in your degree? If you have earned multiple degrees, please list all majors." The results were then coded to either be considered STEM or not. Since definitions of what qualifies as a STEM degree vary, two notable coding choices are explained. Following most university definitions, medical degrees (e.g., M.D., D.V.M., D.D.S., etc.) were excluded, while degrees relating to Economics were included.

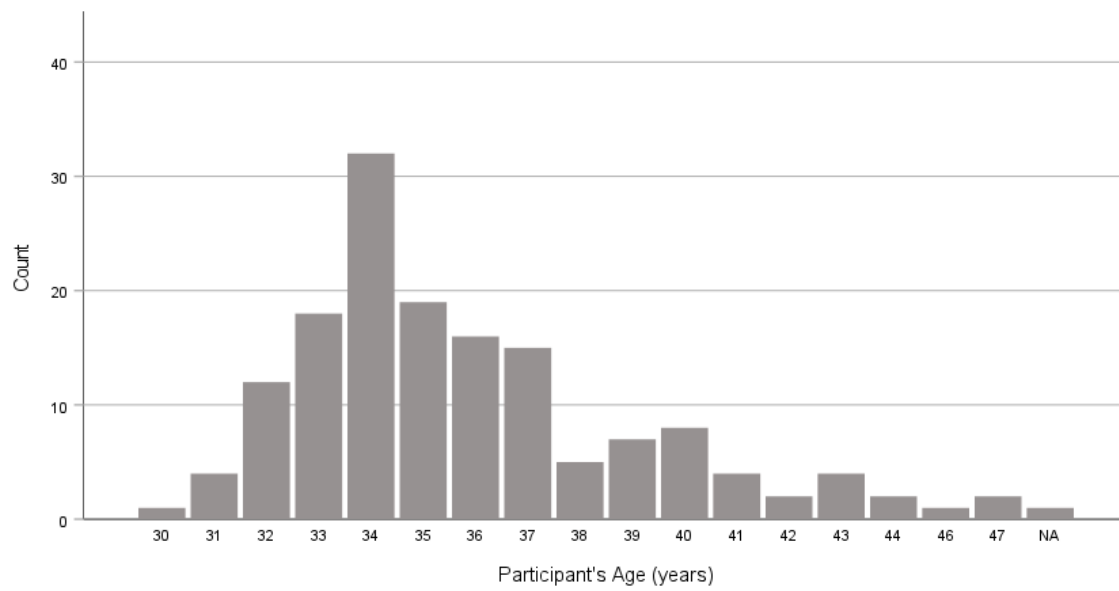


Figure 11. Demographic Question: Age

Source: Created by author.

APPENDIX E

ADDITIONAL QUANTITATIVE ANALYSIS

Figures 12 through 21 show the frequency distributions for the individual questions on the self-reported trust survey found in Appendix A. All responses are asking the participant their probability on trusting a teammate with the provided characteristics along a 7-point Likert-type scale (1 representing low trust and 7 representing high trust). table 20 shows the descriptive statistics of self-reported trust survey sub-competencies, while table 21 shows the regression model using those sub-competencies to estimate the effect on self-reported trust in AI.

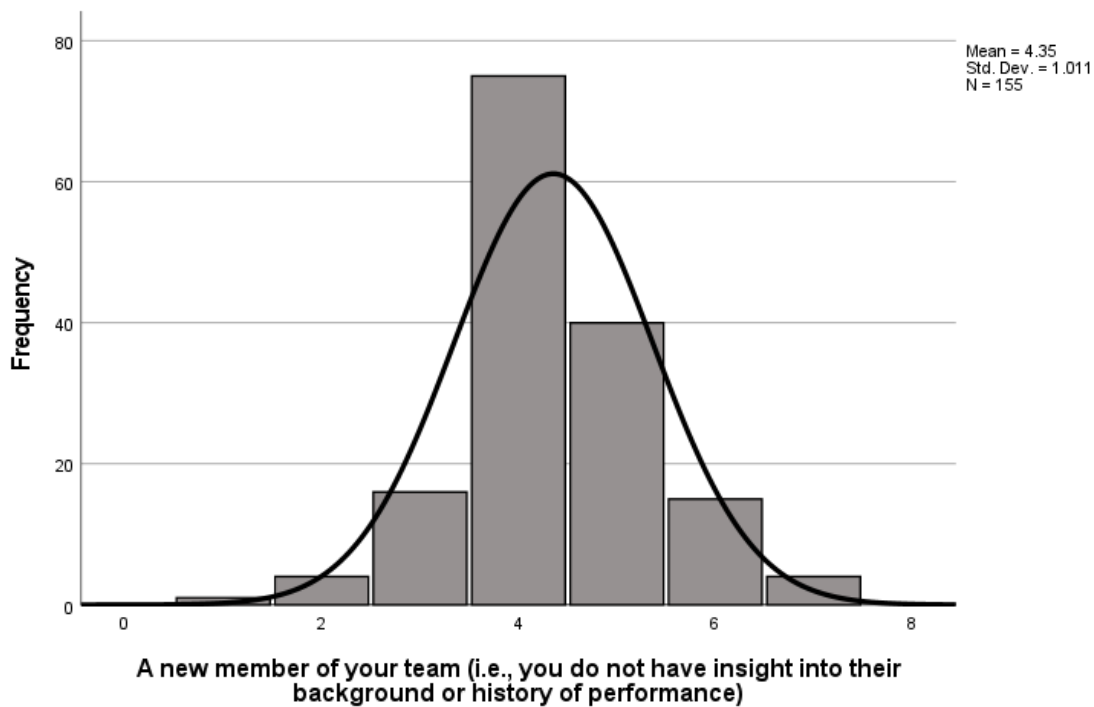


Figure 12. Frequency Distribution of Self-Reported Trust Survey Question #1

Source: Created by author.

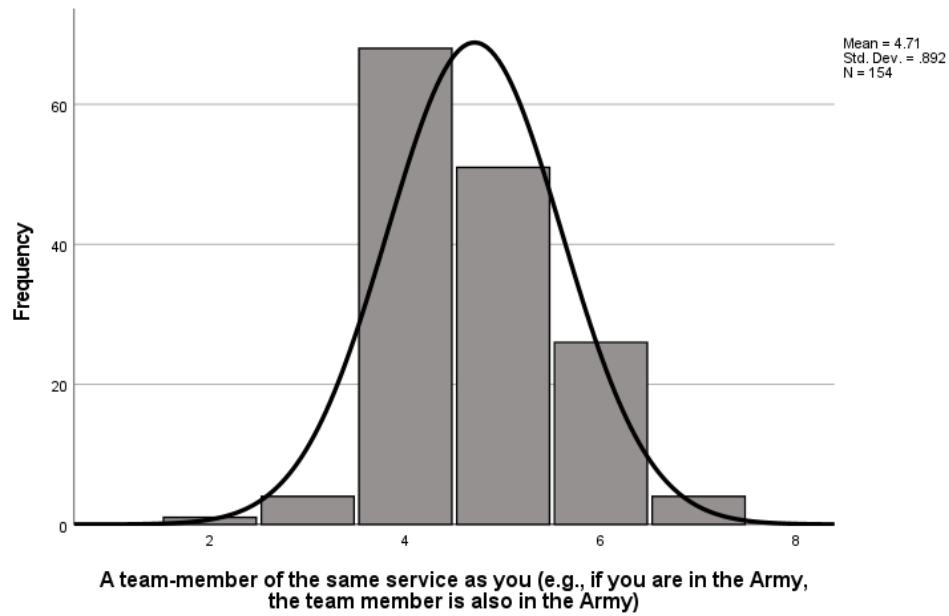


Figure 13. Frequency Distribution of Self-Reported Trust Survey Question #2

Source: Created by author.

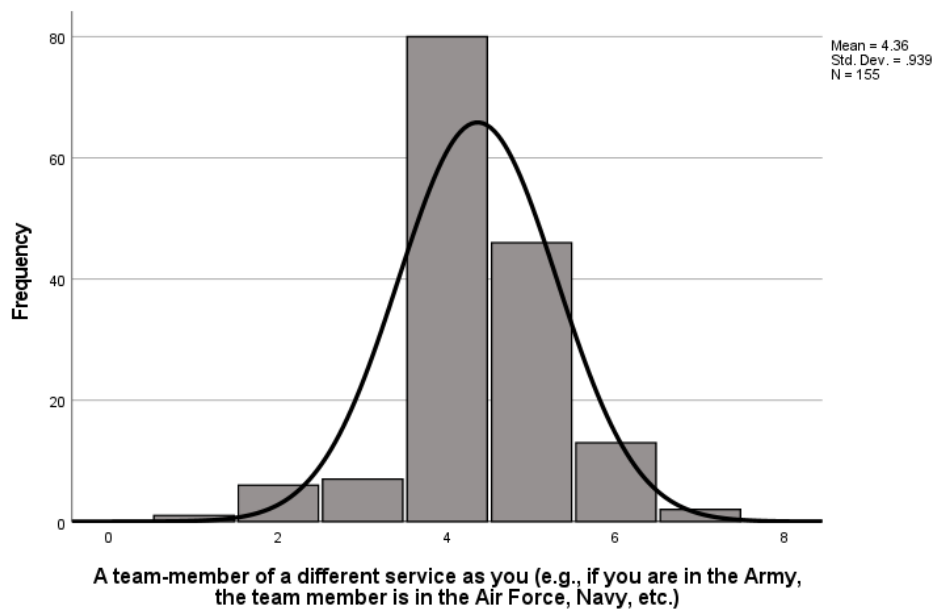


Figure 14. Frequency Distribution of Self-Reported Trust Survey Question #3

Source: Created by author.



Figure 15. Frequency Distribution of Self-Reported Trust Survey Question #4

Source: Created by author.

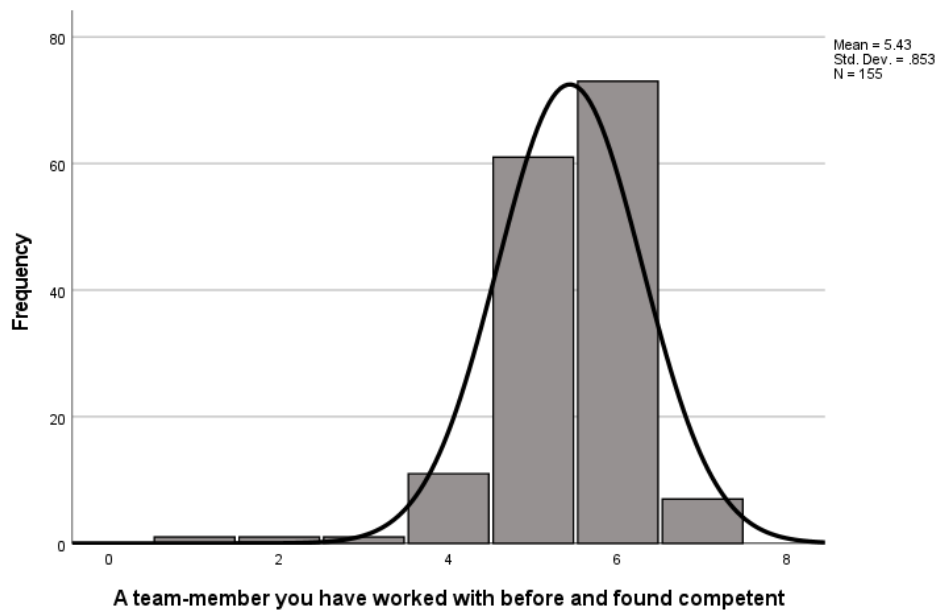


Figure 16. Frequency Distribution of Self-Reported Trust Survey Question #5

Source: Created by author.

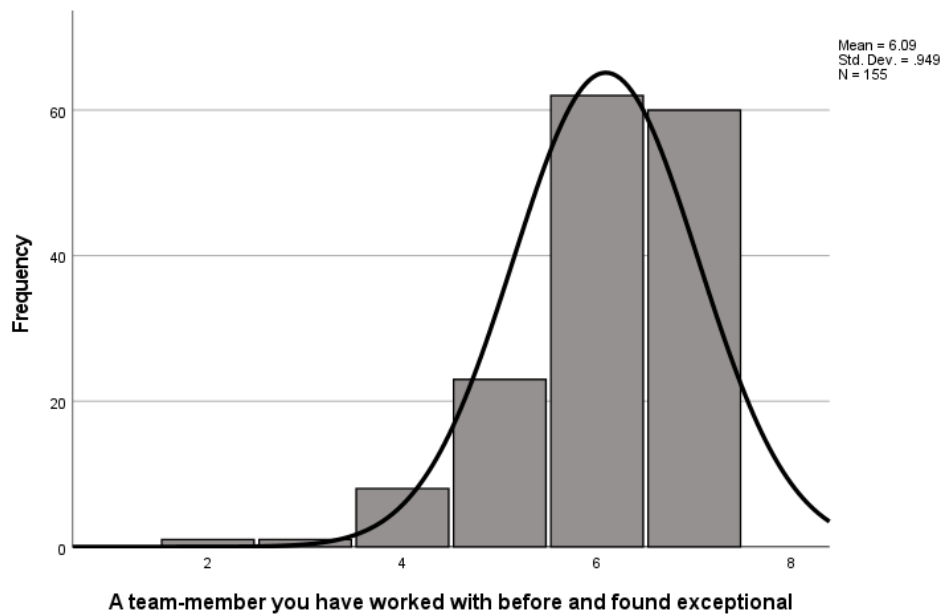


Figure 17. Frequency Distribution of Self-Reported Trust Survey Question #6

Source: Created by author.

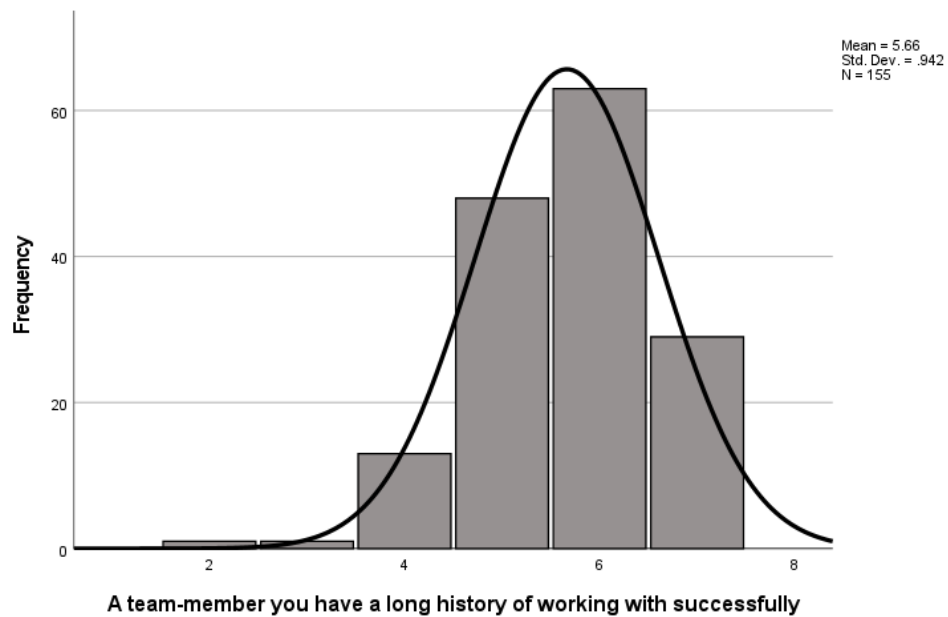


Figure 18. Frequency Distribution of Self-Reported Trust Survey Question #7

Source: Created by author.

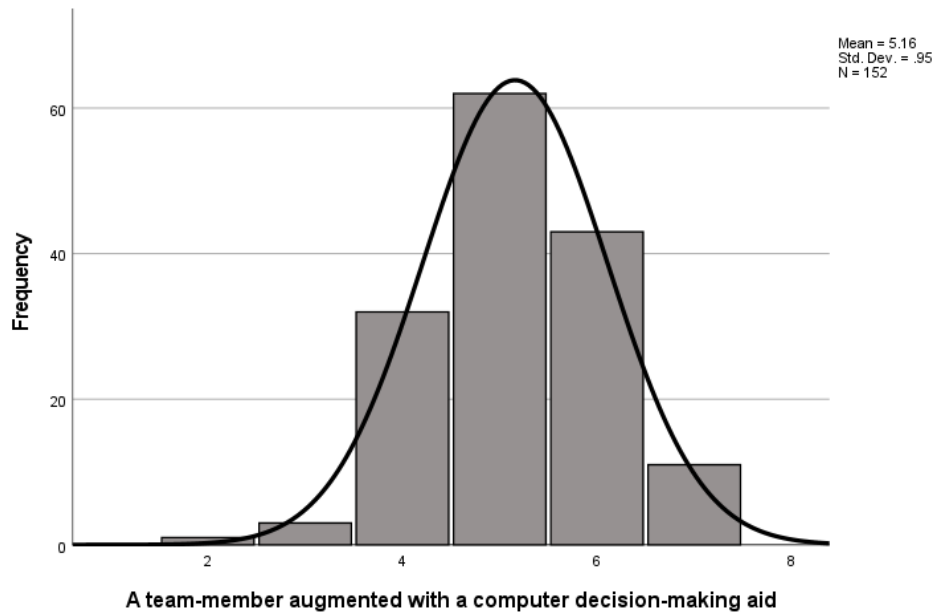


Figure 19. Frequency Distribution of Self-Reported Trust Survey Question #8

Source: Created by author.

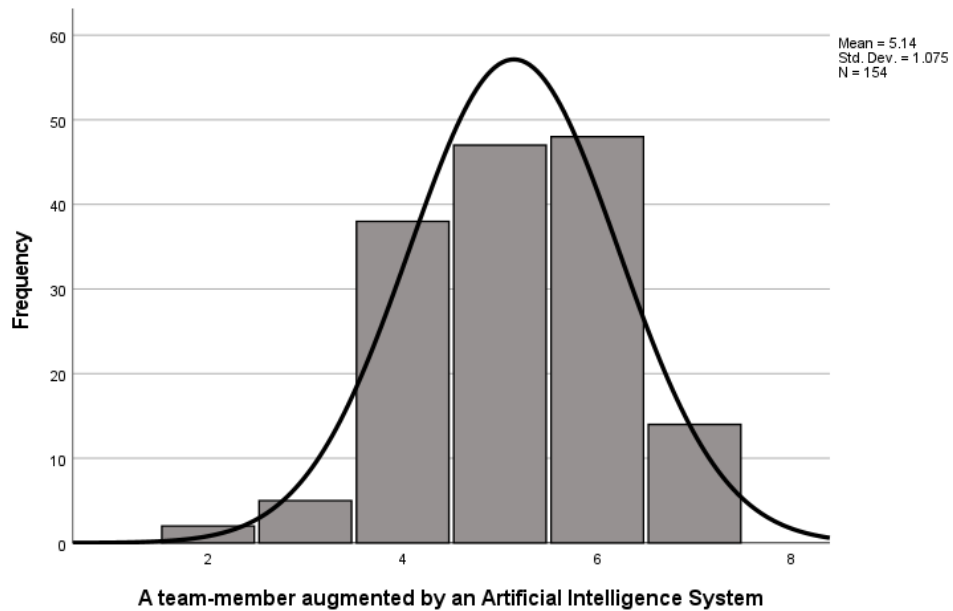


Figure 20. Frequency Distribution of Self-Reported Trust Survey Question #9

Source: Created by author.

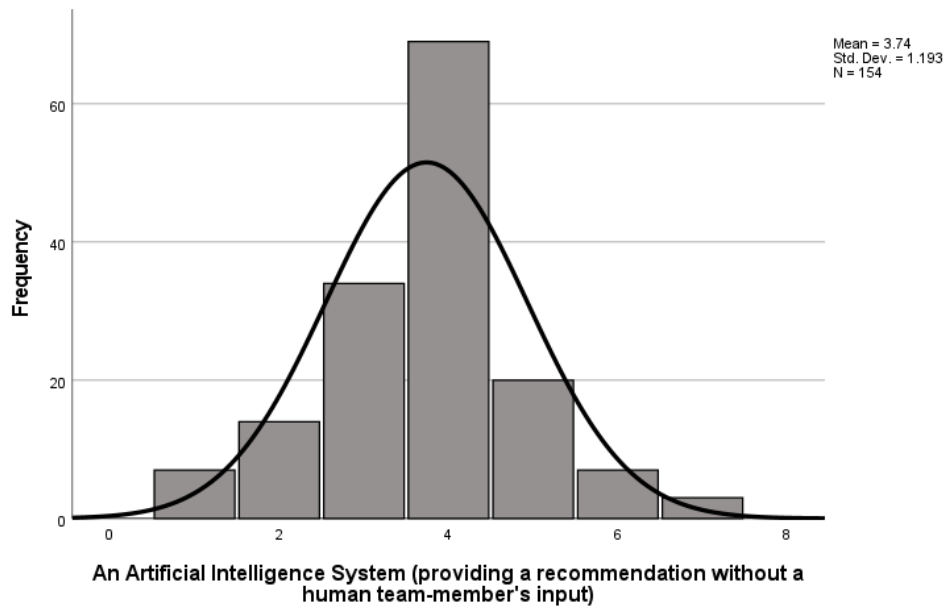


Figure 21. Frequency Distribution of Self-Reported Trust Survey Question #10

Source: Created by author.

Table 20. Descriptive Statistics of Self-Reported Trust Survey Sub-Competencies

Sub-competency	N	Range	Mean	Std. Dev.
Appropriate Social Media Use	152	[2.0, 4.0]	3.76	.37
Information Literacy	155	[2.0, 4.0]	3.44	.50
Communication Medium Choice	153	[2.5, 4.0]	3.42	.49
Technological Exploration	153	[1.0, 4.0]	2.53	.72
Information Protection	155	[1.33, 4.0]	3.18	.60
Software Proficiency	155	[2.67, 4.0]	3.73	.43
Active Internet Use	155	[2.67, 4.0]	3.66	.40
General Computer Skills	155	[1.67, 4.0]	3.15	.62

Source: Created by author.

Table 21. Linear Regression Results of Predictors of Participant's Trust in AI (Mean)

Variable	<i>B</i>	<i>SE</i>	<i>p</i>	95% CI	R ²
Digital Literacy	.349	.209	.095	[-.051, .760]	.015
Appropriate Social Media Use	-.009	.357	.977	[-.535, .532]	.000
Information Literacy	.266	.157	.088	[-.027, .578]	.022
Communication Medium Choice	.045	.145	.759	[-.235, .325]	.001
Technological Exploration	.146	.096	.123	[-.047, .354]	.014
Information Protection	-.119	.125	.348	[-.369, .105]	.007
Software Proficiency	.146	.148	.313	[-.135, .425]	.005
Active Internet Use	.258	.163	.112	[-0.77, .595]	.014
General Computer Skills	.208	.130	.102	[-.043, .496]	.021

95% bias corrected and accelerated confidence intervals and standard errors based on 1000 bootstrap samples

Source: Created by author.

APPENDIX F

QUALITATIVE ANALYSIS

Table 22. Qualitative Code Descriptions and Examples

Code	Description	Example(s)
Evaluation Factors Driven (AI Wargaming Method)	Described choice as being influenced by the evaluation factors displayed by the AI.	"Moderate CI, greatest probably of success, least risk to force"; "Higher probability of success and lower loss to forces"
Evaluation Criteria Driven (Human Wargaming Method)	Described choice as being influenced by the evaluation criteria presented by the human team. Not necessarily related to choosing the human COA (e.g., could have valued the criteria of flexibility and chosen the COA with the highest score)	"It meets 3 of the evaluation criteria"; "COA1's score against the evaluation criteria is best"; "COA 1 has more criteria that are green."
Deliberate Compromise	Described a choice that was a compromise between the two different wargaming results	"COA 2 balances both results"; "happy medium"; "we should go with the middle option which both human and AI thinks is acceptable."
Mismatch Between Wargaming Evaluation Methods	Expressed a dislike relating to the fact that the wargame results were not comparing the same things in during there methods. Apples vs oranges argument related to wargaming results	"the two do not appear to be using similar metrics which poses a challenge"; "A percentage of success does not readily translate into weighting a decision versus 5 different points of consideration."

Code	Description	Example(s)
Human Bias/Subjectivity	Expressed a negative view on the biases and subjectivity of humans, and the impact that has on decision-making	"There is too much subjectivity in human creation, assessment, and evaluation of evaluation criteria - Human nature - we tend to introduce bias, even unknowingly, into the COAs we 'want' to be selected"
Human Experience	Expressed a positive view for the role of human experience in the planning process, including aspects surrounding being more familiar with existing processes.	"Experience drives decisions, not calculators and computers."
“Art” of War	Expressed a positive view for the role that intuition and other intangible aspects of human nature impact the planning process. Generally summarized as a viewpoint on the importance of the "art" of war.	"Human factors provide the operational art to the COA while AI provides only the science to the overall equation."; "I trust humans intuition and analysis of other humans"
AI Capabilities Mismatched with Military Operations	General comments on misalignment of AI in military operations for numerous reasons.	Need human-in-the-loop; Humans are the ones executing a plan; Need for Staff Buy-in
Unable to Replicate Complex Situations	A generally negative viewpoint on the ability for any AI system to be able to model the complexity of the operational environment, including replicating the adversary's thinking	"I think the environment is too complex for a computer to adequately factor in all the variables that humans can across separate WFFs."
Calculation of Large/Complex Data	A generally positive viewpoint on the ability for AI systems to calculation large amounts of complex data. Also, considerations for the importance of the "science" of war in the planning process	"AI has the ability to coordinate more data and variables than the humans can";

Code	Description	Example(s)
Lack of Historical Performance	Expressed a lack of confidence in the AI system due to a lack of knowledge regarding the historical performance of the system	"not knowing anything about this AI [...] its historical performance record"; "There is not enough information provided on the reliability of the AI-produced COA"
Lack of Personal Experience with AI System	Expressed a lack of confidence in the AI system due to lack of personal experience with the system	"lack of interaction/familiarity with the AI system."
Positive Historical Performance	Expressed a positive viewpoint based on being provided the historical performance of the AI system	"system that has performed reliably"
General Lack of Knowledge Regarding AI	Expressed a lack of willingness to trust AI due to a general lack of knowledge	"I don't trust the AI right now because I don't understand (or have knowledge of) the underlying AI"
Explainability	Expressed the desire to know "why" the AI system made a decision. Also, statements expressing the need to convey the reasoning of the AI system to their superiors	"cant justify the outcome to the commander based on 'that's what the AI said'"; "essentially a black box that is not providing any explanation as to why it is selecting a particular COA"
Lack of Knowledge – Calculation Methods	Expressed a lack of willingness to trust AI without knowing more about "how" the system did its calculation, this includes a desire to know how the AI system was trained.	"How is the AI evaluating or what is underlying data set training"; "I have no information on the AI's evaluation process"; "uncomfortable with the AI's assessment of Surprise and Risk without understanding how it was designed to measure those things"

Code	Description	Example(s)
Lack of Knowledge – Data Inputs	Expressed a lack of willingness to trust AI without knowing more about the data that the system was using	"without knowing the criteria and inputs that the AI system used"; "single entity making a decision with its own inputs only to rely on"; "If I had a better understanding of the AI's input data, I may be more open to its conclusion."
AI Bias	Expressed a lack of willingness to trust AI due to the potential for the AI system to have bias. Often explanations included the fact that this bias was due to the method of programming	"algorithms in a computer system, which were written and weighted in a vacuum.";
Knowledge of AI Methodology	Expressed a greater willingness to trust the AI system due to the presentation of the AI system's methodology	"Understanding methodology and number of simulations run is an important part of determining rigor of analysis"
Misread/Misunderstood Wargaming Results	Description of a wargaming result that was not present as a reasoning for selection	"It seemed to score higher on more criteria but also had the lowest calculated loss to forces."
Response Too Vague/Not Helpful	Response is too vague to provide insight into the participant's reasoning. Response is generally unhelpful due to a lack of seriousness (sarcastic, vignette "gamesmanship", etc.)	"Gotta follow that gut feeling"; "You can blame the computer for issues with the plan."; "AI vs human decision."
Mismatch Between Response and COA Selection	The response offered an explanation that didn't correspond to the selected COA	"I would have to default to the human recommendation." but selected AI COA (#3)
No Response	No Response was provided	N/A

Source: Created by author.

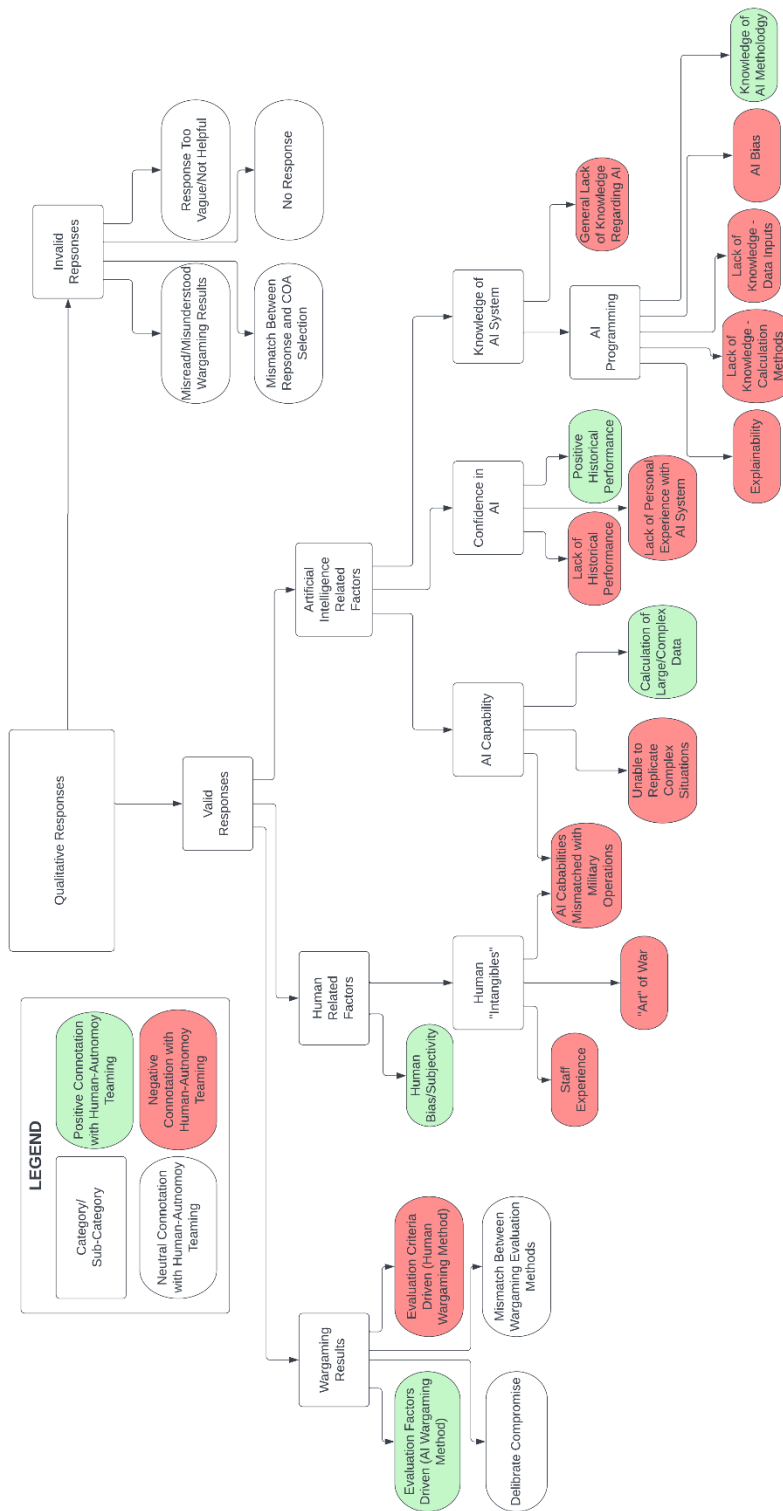


Figure 22. Hierarchical Qualitative Coding Results

Source: Created by author.

BIBLIOGRAPHY

- Barron, Laura. *Evaluation of Digital Literacy as an Air Force Foundational Competency*. Air Education and Training Command (AETC) Technical Report No. 3. Randolph Air Force Base, San Antonio, TX: AETC, July 2021. <https://apps.dtic.mil/sti/pdfs/AD1143616.pdf>.
- Bonikowska, Aneta, Claudia Sanmartin, and Marc Frenette. "Data Literacy: What It Is and How to Measure It in the Public Service." Analytical Studies Branch, Statistics Canada, Ottawa, Ontario, Canada, August 2019. http://epe.lac-bac.gc.ca/100/201/301/weekly_acquisitions_list-ef/2019/19-33/publications.gc.ca/collections/collection_2019/statcan/11-633-x/11-633-x2019003-eng.pdf.
- Brande, Lieve Van Den, Stephanie Carretero, Riina Vuorikari, and Yves Punie. *DigComp 2.0: The Digital Competence Framework for Citizens. Update Phase 1: The Conceptual Reference Model*. Luxembourg: Publication Office of the European Union, 2016. <https://data.europa.eu/doi/10.2791/11517>.
- Chairman of the Joint Chiefs of Staff. *Joint Operating Environment 2035: The Joint Force in a Contested and Disordered World*. Washington, DC: Joint Chiefs of Staff, 2016.
- . Joint Publication 3-0, *Joint Operations*. Washington, DC: Joint Chiefs of Staff, 2017.
- . Joint Publication 5-0, *Joint Planning*. Washington, DC: Joint Chiefs of Staff, 2020.
- Chen, Jessie Y. C., and Michael J. Barnes. "Supervisory Control of Multiple Robots: Effects of Imperfect Automation and Individual Differences." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 54, no. 2 (April 2012): 157–74. <https://doi.org/10.1177/0018720811435843>.
- Costa, Ana Cristina, C. Ashley Fulmer, and Neil R. Anderson. "Trust in Work Teams: An Integrative Review, Multilevel Model, and Future Directions." *Journal of Organizational Behavior* 39, no. 2 (February 2018): 169–184. <https://doi.org/10.1002/job.2213>.
- De Jong, Bart A., Kurt T. Dirks, and Nicole Gillespie. "Trust and Team Performance: A Meta-Analysis of Main Effects, Moderators, and Covariates." *Journal of Applied Psychology* 101, no. 8 (August 2016): 1134–50. <https://doi.org/10.1037/apl0000110>.

- Frank, Mark, Johanna Walker, Judie Attard, and Alan Tygel. "Data Literacy - What Is It and How Can We Make It Happen?" *The Journal of Community Informatics* 12, no. 3 (September 19, 2016): 4-8. <https://doi.org/10.15353/joci.v12i3.3274>.
- Hill, N. Sharon, Kathryn M. Bartol, Paul E. Tesluk, and Gosia A. Langa. "Organizational Context and Face-to-Face Interaction: Influences on the Development of Trust and Collaborative Behaviors in Computer-Mediated Groups." *Organizational Behavior and Human Decision Processes* 108, no. 2 (March 2009): 187–201. <https://doi.org/10.1016/j.obhdp.2008.10.002>.
- Hoff, Kevin Anthony, and Masooda Bashir. "Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 57, no. 3 (May 2015): 407–34. <https://doi.org/10.1177/0018720814547570>.
- Kahneman, Daniel. *Thinking, Fast and Slow*. New York: Farrar, Straus, and Giroux, 2011.
- Kahneman, Daniel, and Gary Klein. "Conditions for Intuitive Expertise: A Failure to Disagree." *American Psychologist* 64, no. 6 (2009): 515–526. <https://doi.org/10.1037/a0016755>.
- Laird, John, Charan Ranganath, and Samuel Gershman. "Future Directions in Human Machine Teaming Workshop." Basic Research Innovative Collaboration Center, Arlington, VA, 16-17 July 2019. <https://basicresearch.defense.gov/Portals/61/Future%20Directions%20in%20Human%20Machine%20Teaming%20Workshop%20report%20%20%28for%20public%20release%29.pdf>.
- Lee, John, and Katrina See. "Trust in Automation: Designing for Appropriate Reliance." *Human Factors* 46, no. 1 (2004): 50-80. https://doi.org/10.1518/hfes.46.1.50_30392.
- Lewicki, Roy, and Chad Brinsfield. "Framing Trust: Trust as a Heuristic." In *Framing Matters: Perspectives on Negotiation Research and Practice in Communication*, edited by William A. Donohue, Randall G. Rogan, and Sandra Kaufman, 110-135. New York: Peter Lang Publishing, 2011.
- Lewicki, Roy, Daniel McAllister, and Robert Bies. "Trust and Distrust: New Relationships and Realities." *Academy of Management Review* 23, no. 3 (July 1998): 438-458. <https://doi.org/10.5465/amr.1998.926620>.
- Lewicki, Roy J., Edward C. Tomlinson, and Nicole Gillespie. "Models of Interpersonal Trust Development: Theoretical Approaches, Empirical Evidence, and Future Directions." *Journal of Management* 32, no. 6 (December 2006): 991–1022. <https://doi.org/10.1177/0149206306294405>.

- Lewis, J. David, and Andrew Weigert. "Trust as a Social Reality." *Social Forces* 63, no. 4 (June 1985): 967-985. <https://doi.org/10.2307/2578601>.
- Lingel, Sherrill, Jeff Hagen, Eric Hastings, Mary Lee, Matthew Sargent, Matthew Walsh, Li Ang Zhang, and David Blancett. *Joint All-Domain Command and Control for Modern Warfare: An Analytic Framework for Identifying and Developing Artificial Intelligence Applications*. Santa Monica, CA: RAND Corporation, 2020. <https://doi.org/10.7249/RR4408.1>.
- Lyons, Joseph B., Katia Sycara, Michael Lewis, and August Capiola. "Human–Autonomy Teaming: Definitions, Debates, and Directions." *Frontiers in Psychology* 12 (May 2021). <https://doi.org/10.3389/fpsyg.2021.589585>.
- McKnight, D. Harrison, Larry L. Cummings, and Norman Chervany. "Initial Trust Formation in New Organizational Relationships." *The Academy of Management Review* 23, no. 3 (July 1998): 473-490. <https://doi.org/10.5465/amr.1998.926622>.
- McNeese, Nathan J., Mustafa Demir, Erin K. Chiou, and Nancy J. Cooke. "Trust and Team Performance in Human–Autonomy Teaming." *International Journal of Electronic Commerce* 25, no. 1 (January 2021): 51–72. <https://doi.org/10.1080/10864415.2021.1846854>.
- Mercado, Joseph E., Michael A. Rupp, Jessie Y.C. Chen, Michael J. Barnes, Daniel Barber, and Katelyn Procci. "Intelligent Agent Transparency in Human–Agent Teaming for Multi-UxV Management." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 58, no. 3 (May 2016): 401–15. <https://doi.org/10.1177/0018720815621206>.
- Meyerson, Debra, Karl E. Weick, and Roderick M. Kramer. "Swift Trust and Temporary Groups." In *Trust in Organizations: Frontiers of Theory and Research*, edited by Roderick M. Kramer, and Tom R. Tyler, 166–195. Thousand Oaks, CA: SAGE Publications, 1996.
- O'Shaughnessy, Terrence J. "Decision Superiority Through Joint All-Domain Command and Control," *Joint Forces Quarterly* 99, no. 4 (October 2020): 74-80. <https://ndupress.ndu.edu/Portals/68/Documents/jfq/jfq-99/jfq-99.pdf>.
- Parasuraman, Raja; Thomas B. Sheridan, and Christopher D. Wickens. "A Model for Types and Levels of Human Interaction with Automation." *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* 30, no. 3 (May 2000): 286–297. <https://doi.org/10.1109/3468.844354>.
- Pratama, M. A., Supahar Pahar, Diah P. Lestari, Widia K. Sari, Tri S. Y. Putri, and V.A.K. Adiatmah. "Data Literacy Assessment Instrument for Preparing 21 Cs Literacy: Preliminary Study." *Journal of Physics: Conference Series* 1440 (January 2020). <https://doi.org/10.1088/1742-6596/1440/1/012085>.

- Ridsdale, Chantel, James Rothwell, Michael Smit, Hossam Ali-Hassan, Michael Bliemel, Dean Irvine, Daniel Kelley, Stan Matwin, and Bradley Wuetherick. “Strategies and Best Practices for Data Literacy Education.” Knowledge Synthesis Report, Dalhousie University, Nova Scotia, Canada, 2015. <https://dalspace.library.dal.ca/handle/10222/64578>.
- Roff, Heather M., and David Danks. “‘Trust but Verify’: The Difficulty of Trusting Autonomous Weapons Systems.” *Journal of Military Ethics* 17, no. 1 (January 2018): 2–20. <https://doi.org/10.1080/15027570.2018.1481907>.
- Schaefer, Kristin E., Deborah R. Billings, James L. Szalma, Jeffrey K. Adams, Tracy L. Sanders, Jessie Y. Chen, and Peter A. Hancock. “A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Human-Robot Interaction” Army Research Laboratory, Aberdeen Proving Ground, MD, July 2014. <https://apps.dtic.mil/sti/pdfs/ADA607926.pdf>.
- Schaefer, Kristin E., Jessie Y.C. Chen, James L. Szalma, and Peter A. Hancock. “A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems.” *Human Factors: The Journal of the Human Factors and Ergonomics Society* 58, no. 3 (May 2016): 377–400. <https://doi.org/10.1177/0018720816634228>.
- Schaefer, Kristin E., Edward R. Straub, Jessie Y.C. Chen, Joe Putney, and A.W. Evans. “Communicating Intent to Develop Shared Situation Awareness and Engender Trust in Human-Agent Teams.” *Cognitive Systems Research* 46 (December 2017): 26–39. <https://doi.org/10.1016/j.cogsys.2017.02.002>.
- Smolnikova, Marketa. “Next Step: Data Literacy Measurement.” In *Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 234–40. Budapest, Hungary: SCITEPRESS - Science and Technology Publications, 2020. <https://doi.org/10.5220/0010146402340240>.
- Skov, Anders. “What is Digital Competence.” Center for Digital Dannelsen. Accessed February 17, 2022. <https://digital-competence.eu/front/what-is-digital-competence/>.
- U.S. Army Combat Capabilities Development Command, Army Research Laboratory. “Human Autonomy Teaming (HAT),” Open Campus. Accessed November 9, 2021. <https://www.arl.army.mil/opencampus/HAT>.
- U.S. Congress. Senate. *Statement of General Terrence J. O’Shaughnessy, United States Air Force Commander, United States Northern Command and North American Aerospace Defense Command before the Senate Armed Services Committee*. February 13, 2020.

Wolff, Annika, Daniel Gooch, Jose J. Cavero Montaner, Umar Rashid, and Gerd Kortuem. "Creating an Understanding of Data Literacy for a Data-Driven Society." *The Journal of Community Informatics* 12, no. 3 (August 2016): 9-26. <https://doi.org/10.15353/joci.v12i3.3275>.