2022-12

EVALUATING ARTIFICIAL INTELLIGENCE FOR OPERATIONS IN THE INFORMATION ENVIRONMENT

Jhong, Kelley Y.

Monterey, CA; Naval Postgraduate School

https://hdl.handle.net/10945/71483

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

EVALUATING ARTIFICIAL INTELLIGENCE FOR OPERATIONS IN THE INFORMATION ENVIRONMENT

by

Kelley Y. Jhong

December 2022

Thesis Advisor: Timothy C. Warren
Second Reader: Nicholas Dew

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** December 2022	**3. REPORT TYPE AND DATES COVERED** Master's thesis

4. TITLE AND SUBTITLE
EVALUATING ARTIFICIAL INTELLIGENCE FOR OPERATIONS IN THE INFORMATION ENVIRONMENT

5. FUNDING NUMBERS

6. AUTHOR(S) Kelley Y. Jhong

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release. Distribution is unlimited.

12b. DISTRIBUTION CODE
A

13. ABSTRACT (maximum 200 words)

Recent advances in artificial intelligence (AI) portend a future of accelerated information cycles and intensified technology diffusion. As AI applications become increasingly prevalent and complex, Special Operations Forces (SOF) face the challenge of discerning which tools most effectively address operational needs and generate an advantage in the information environment. Yet, SOF currently lack an end user–focused evaluation framework that could assist information practitioners in determining the operational value of an AI tool. This thesis proposes a practitioner's evaluation framework (PEF) to address the question of how SOF should evaluate AI technologies to conduct operations in the information environment (OIE).

The PEF evaluates AI technologies through the perspective of the information practitioner who is familiar with the mission, the operational requirements, and OIE processes but has limited to no technical knowledge of AI. The PEF consists of a four-phased approach—prepare, design, conduct, recommend—that assesses nine evaluation domains: mission/task alignment; data; system/model performance; user experience; sustainability; scalability; affordability; ethical, legal, and policy considerations; and vendor assessment. By evaluating AI through a more structured, methodical approach, the PEF enables SOF to identify, assess, and prioritize AI-enabled tools for OIE.

14. SUBJECT TERMS
artificial intelligence (AI), operations in the information environment (OIE), psychological operations (PSYOP), human-machine teaming (HMT), technology adoption, evaluations

15. NUMBER OF PAGES
219

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

EVALUATING ARTIFICIAL INTELLIGENCE FOR OPERATIONS IN THE INFORMATION ENVIRONMENT

Kelley Y. Jhong
Major, United States Army
BS, United States Military Academy, 2011
MA, Oklahoma University, 2015

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN INFORMATION STRATEGY AND POLITICAL WARFARE

from the

NAVAL POSTGRADUATE SCHOOL
December 2022

Approved by: Timothy C. Warren
Advisor

Nicholas Dew
Second Reader

Carter Malkasian
Chair, Department of Defense Analysis

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Recent advances in artificial intelligence (AI) portend a future of accelerated information cycles and intensified technology diffusion. As AI applications become increasingly prevalent and complex, Special Operations Forces (SOF) face the challenge of discerning which tools most effectively address operational needs and generate an advantage in the information environment. Yet, SOF currently lack an end user–focused evaluation framework that could assist information practitioners in determining the operational value of an AI tool. This thesis proposes a practitioner's evaluation framework (PEF) to address the question of how SOF should evaluate AI technologies to conduct operations in the information environment (OIE).

The PEF evaluates AI technologies through the perspective of the information practitioner who is familiar with the mission, the operational requirements, and OIE processes but has limited to no technical knowledge of AI. The PEF consists of a four-phased approach—prepare, design, conduct, recommend—that assesses nine evaluation domains: mission/task alignment; data; system/model performance; user experience; sustainability; scalability; affordability; ethical, legal, and policy considerations; and vendor assessment. By evaluating AI through a more structured, methodical approach, the PEF enables SOF to identify, assess, and prioritize AI-enabled tools for OIE.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AI	artificial intelligence
AGI	artificial general intelligence
API	application programming interface
AUC	area under the curve
CDAO	Chief Digital and Artificial Intelligence Office
COTS	commercial-off-the-shelf
C-TAM-TPB	Combined Technology Acceptance Model and Theory of Planned Behavior
DARPA	Defense Advanced Research Projects Agency
DIU	Defense Innovation Unit
DOD	Department of Defense
DOI	Diffusion of Innovation Theory
DOTMLPF-P	doctrine, organization, training, material, leadership and education, personnel, facilities, policy
DPRK	Democratic People's Republic of Korea
GAN	generative adversarial network
GOTS	government-off-the-shelf
GUI	graphical user interface
HMT	human-machine teaming
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
INCAS	Influence Campaign Awareness and Sensemaking
ISO	International Organization for Standardization
JAIC	Joint Artificial Intelligence Center
KPI	key performance measure
MISO	military information support operations
ML	machine learning

MM	Motivational Model
MOE	measures of effectiveness
MPCU	Model of PC Utilization
NLP	natural language processing
OIE	operations in the information environment
OK	Odnoklassniki
PAI	publicly available information
PEF	Practitioner's Evaluation Framework
PSYOP	psychological operations
R&D	research and development
RAI	responsible artificial intelligence
RAITE	responsible artificial intelligence test and evaluation
ROC	receiver operator characteristics
SBIR	Small Business Innovation Research Program
SCT	Social Cognitive Theory
SMART	specific, measurable, achievable, relevant, time-bound
T&E	test and evaluation
TA	target audience
TAM	Technology Acceptance Model
TEMP	test and evaluation master plan
TPB	Theory of Planned Behavior
TRA	Theory of Reasoned Actions
TOE	Technology-Organization-Environment (framework)
UFE	Utilization-Focused Evaluation
UTAUT	Unified Theory of Acceptance and Use of Technology
UX	user experience
VK	VKontakte
XAI	explainable artificial intelligence

EXECUTIVE SUMMARY

Recent advances in artificial intelligence (AI) portend a future of accelerated information cycles and intensified technology diffusion. Special Operations Forces (SOF) currently lack an end user–focused evaluation framework that could assist information practitioners in determining the operational value of an AI tool. This thesis proposes a practitioner's evaluation framework (PEF) to address the question of how SOF should evaluate AI technologies to conduct operations in the information environment (OIE).

The PEF evaluates AI technologies through the perspective of the information practitioner who is familiar with the mission, the operational requirements, and OIE processes but has limited to no technical knowledge of AI. The framework consists of a simple four-phased approach—prepare, design, conduct, recommend—that assesses nine evaluation domains, as shown in Figure 1.

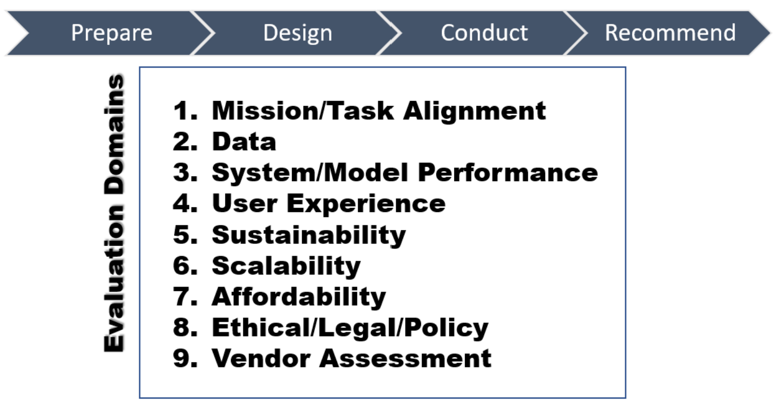


Figure 1. Practitioner's Evaluation Framework

Drawing upon extant literature and interviews with AI and OIE experts within the U.S. government, industry, and academia, the PEF was developed from an analysis of four major research areas:

1. **Potential AI applications for OIE.** The thesis finds that AI can assist practitioners in addressing four traditional challenges in OIE: analyzing the information environment, enabling in-house product development,

enhancing timeliness and scale of information dissemination, and improving the ability to measure effectiveness.

2. **Key principles and considerations for human-machine teaming (HMT)**. Research highlights the importance of "justified confidence" for effective HMT.[1] Information forces face additional complexity in HMT due to the cognitive, human-centric nature of OIE, which necessitates further transparency and explainability of AI systems.

3. **Technology acceptance and adoption theories**. Theoretical analysis reveals the prominent role of relative advantage, compatibility, and complexity in technology adoption.[2] Trust in the technology and time available to users also impact the adoption potential of AI for OIE. Organizational-level factors such as readiness, management support, and government policies are also important considerations.[3]

4. **Ongoing initiatives to increase transparency of AI**. Analysis of existing frameworks—the Defense Innovation Unit's *Responsible AI (RAI) Guidelines*, model cards, datasheets, FactSheets, and System Cards—reveals key elements to consider for transparency and assessing the effectiveness of AI. These elements include having a clear understanding of intended use, data provenance, model performance, limitations of the model, and ethical considerations.

To test the feasibility of the proposed framework, the PEF was used to evaluate Pulse, a data collection and engagement platform currently used by OIE units within U.S. Army Special Operations.[4] The evaluation finds that the PEF enables the practitioner to

[1] DOD Responsible AI Working Council, U.S. Department of Defense Responsible Artificial Intelligence Strategy and Implementation Pathway (Washington, DC: Department of Defense, 2022), 9, https://www.ai.mil/docs/RAI_Strategy_and_Implementation_Pathway_6-21-22.pdf.

[2] Everett M. Rogers, *Diffusion of Innovations*, 4th ed. (New York, NY: The Free Press, 1995), 221.

[3] Louis G. Tornatzky, Mitchell Fleischer, and Alok K. Chakrabarti, The Processes of Technological Innovation (Lexington, MA: Lexington Books, 1990); Ali Al Hadwer et al., "A Systematic Review of Organizational Factors Impacting Cloud-Based Technology Adoption Using Technology-Organization-Environment Framework," Internet of Things 15 (September 2021): 1–10, https://doi.org/10.1016/j.iot.2021.100407.

delineate both the advantages as well as the areas that require extra consideration when using an AI-enabled tool. In particular, evaluations of system/model performance and user experience highlight two key points. First, the level of complexity associated with the system requires a focused understanding of who the intended users should be. Second, to properly evaluate AI, model performance metrics—which have not been traditionally communicated to end users—should be accessible and interpretable to practitioners.

This thesis recommends that U.S. Special Operations leverage the PEF as a guideline for practitioners to conduct an initial evaluation of AI technologies. By evaluating AI through a more structured, methodical approach, this framework enables SOF to identify, assess, and prioritize AI-enabled tools that effectively address operational needs and generate an advantage in the information environment. The PEF also ensures that practitioners consider evaluation criteria that incorporate the DOD's *RAI Strategy* and complement the Defense Innovation Unit's *RAI Guidelines*.[5]

---

[4] Two Six Technologies, "Pulse: Enabling Data Collection and Two-Way Engagement," Two Six Technologies | Advanced Technology Solutions for Critical Missions, 2022, https://twosixtech.com/products/pulse/.

[5] DOD Responsible AI Working Council, *U.S. Department of Defense Responsible Artificial Intelligence Strategy and Implementation Pathway*; Jared Dunnmon et al., *Responsible AI Guidelines in Practice* (Mountain View, CA: Defense Innovation Unit, 2021), https://www.diu.mil/responsible-ai-guidelines.

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

My heart is filled with gratitude for everyone who has helped me along this journey of learning and discovery. I started with minimal awareness of this topic. However, thanks to the many individuals within government, industry, and academia who shared their precious time, expertise, and experience with me, I was able to expand my knowledge far beyond my expectations. While these individuals are not listed by name in this thesis, they are the ones who made this thesis possible.

All my professors at the Naval Postgraduate School played a critical part in helping me gain the necessary skills and knowledge to complete this thesis. Nonetheless, I would like to especially thank my advisors, Dr. Camber Warren and Dr. Nick Dew, for their guidance and wisdom throughout the process. I also want to thank Dr. Ryan Maness and Rebecca Lorentz for their support in allowing me to attend Phoenix Challenge; this series of conferences enabled me to connect with individuals who were key to both my research and professional development. Furthermore, I want to express my immense appreciation for the people at the Defense Innovation Unit, who have graciously given me the opportunity to learn and observe the real-world application of evaluations, rapid prototyping, and adoption of critical technologies.

Finally, I want to thank my family and friends for their love, patience, and motivation. Words cannot describe how important their support is to me.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

> The prospect of adversaries using machine learning, planning, and optimization to create systems to manipulate citizens' beliefs and behavior in undetectable ways is a gathering storm. Most concerning is the prospect that adversaries will use AI to create weapons of mass influence to use as leverage during future wars, in which every citizen and organization becomes a potential target.
>
> —The National Security Commission on Artificial Intelligence[1]

As early as 3,000 years ago when Homer described intelligent machines possessing "phrenes" (thoughts) and "noos" (minds) in the *Iliad,* artificial intelligence (AI) has been the subject of awe as well as anxiety within our society.[2] Recent advancements in machine learning (ML) have engendered effusive optimism in the seemingly boundless potential of AI, while also rousing consternation about the prospects for misuse. Combined with unprecedented access to data and information, AI may offer new opportunities to improve efficiency and effectiveness across a wide range of tasks and operations. Yet, its deep integration within a growing number of applications across both the military and the broader society creates vulnerabilities that can be easily exploited by adversaries.

AI's transformational effects are already being felt within the information environment. Advertisers, streaming services, and social media companies are leveraging AI to tailor content and provide personalized recommendations. Social bots and deepfakes enable adversaries to inflame disinformation. As the technology continues to develop, AI will accelerate the pace of information cycles, amalgamate human-machine interaction, and intensify the diffusion of technology and information capabilities.

[1] National Security Commission on Artificial Intelligence, *National Security Commission on Artificial Intelligence Final Report* (Washington, DC: National Security Commission on Artificial Intelligence, 2021), 46, https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf.

[2] Genevieve Liveley and Sam Thomas, "Homer's Intelligent Machines: AI in Antiquity," in *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*, ed. Stephen Cave, Kanta Dihal, and Sarah Dillon (Oxford, UK: Oxford University Press, 2020), 37.

AI empowers information forces to operate with greater speed, effectiveness, and reach within the information environment. U.S. Special Operations Forces (SOF) recognize that an AI-driven environment poses additional challenges in obtaining an information advantage.[3] At the same time, improvements in AI fields such as computer vision and natural language processing (NLP) present new opportunities to enhance the effectiveness of operations in the information environment (OIE). Recognizing the profound impact that AI can have on operational effectiveness, U.S. Special Operations Command is increasing its effort to obtain commercial-off-the-shelf (COTS) and government-off-the-shelf (GOTS) AI tools.[4]

The adoption of AI, however, introduces new challenges—namely, the need for technical expertise, data management, and new policies and processes—that information forces must contend with. As with other novel technologies, AI tools are often acquired without an adequate understanding of how the technology fits into organizational processes, doctrine, and culture, nor is it clear to what extent the technology actually enhances SOF's ability to conduct OIE. To be effective, the OIE community needs to address these challenges and take a more structured approach in evaluating which AI technologies should be acquired to support OIE.

A. RESEARCH QUESTION

Although test and evaluation (T&E) occur as part of the DOD acquisitions process, T&E criteria are typically developed by program managers or software developers with minimal input from the end user. Currently, there is no end user-focused evaluation framework that could assist the information practitioner in evaluating the

[3] Patrick Tucker, "How AI Will Soon Change Special Operations," *Defense One*, May 18, 2020, https://www.defenseone.com/technology/2020/05/how-ai-will-soon-change-special-operations/165487/.

[4] David Vergun, "Special Operations Strives to Use the Power of Artificial Intelligence," *U.S. Department of Defense News*, December 7, 2020, https://www.defense.gov/News/News-Stories/Article/Article/2438076/special-operations-strives-to-use-the-power-of-artificial-intelligence/.

utility and adoption potential of an AI tool.[5] User involvement throughout the development and acquisition process is crucial for successful adoption and acceptance of the technology.[6] Thus, this thesis seeks to develop a framework to evaluate AI technology through the perspective of the information practitioner, who is familiar with the mission, the operational requirements, and OIE processes, but has limited to no technical knowledge of AI. Hence, this thesis examines the following research question: how should SOF evaluate AI technologies to conduct operations in the information environment?

B. APPROACH

This thesis contains six primary sections (Chapters II through VII). Chapter II establishes the foundation by first explaining the key concepts and definitions of AI and OIE. The chapter also discusses the challenges practitioners face in planning, executing, and assessing OIE and identifies AI technologies that can address some of these challenges. Chapter III expounds upon the critical role of trust in human-machine teaming, which impacts the effective use of AI for military operations. Chapter IV dives into technology acceptance and adoption theories to ascertain important factors that influence the adoptability of AI technology within OIE units. Chapter V assesses existing frameworks that facilitate increased transparency of AI systems and can be leveraged by SOF to evaluate the technology. Qualitative analysis was conducted to identify common themes across 50 model cards found through open-source research. The findings from

[5] The term, "information practitioner" is used to refer to information forces down to the individual level. *JP 3-04 Information in Joint Operations* defines "information forces" as "those Active Component and Reserve Component forces of the Services specifically organized, trained, and equipped to create effects in the IE. These forces provide expertise and specialized capabilities that leverage information and can be aggregated as components of an OIE unit to conduct OIE." Joint Chiefs of Staff, *Information in Joint Operations*, JP 3-04 (Washington, DC: Department of Defense, 2022), xi.

[6] Department of Defense, *DOD Enterprise DevSecOps Fundamentals* (Washington, DC: Department of Defense, 2021), 19, https://dodcio.defense.gov/Portals/0/Documents/Library/DoDEnterpriseDevSecOpsFundamentals.pdf; Amela Karahasanovic et al., "User Involvement in the Design of ML-Infused Systems," in *CHI Greece 2021: 1st International Conference of the ACM Greek SIGCHI Chapter* (Athens, Greece: ACM, 2021), https://doi.org/10.1145/3489410.3489421; DOD Responsible AI Working Council, *U.S. Department of Defense Responsible Artificial Intelligence Strategy and Implementation Pathway* (Washington, DC: Department of Defense, 2022), 24, https://www.ai.mil/docs/RAI_Strategy_and_Implementation_Pathway_6-21-22.pdf; Individual within the Chief Digital and Artificial Intelligence Office, personal communication, June 10, 2022.

Chapters IV and V are used to build the practitioner's evaluation framework (PEF) proposed in Chapter VI. In Chapter VII, the framework is then used to evaluate Pulse—an AI tool currently employed by OIE units. Chapter VIII concludes with overall recommendations and suggestions for future research.

II. APPLICATIONS OF AI FOR OPERATIONS IN THE INFORMATION ENVIRONMENT

> A machine which simulated human behavior in detail would indeed tell us the "Inside Story."
>
> —B.F. Skinner, *Contingencies of Reinforcement* [7]

A. CHAPTER OVERVIEW

Although the term, "artificial intelligence," is frequently used in the vernacular, a more precise understanding of AI is needed to determine what factors practitioners should consider when evaluating or utilizing an AI tool. This chapter first breaks down the field of AI into different categories to identify some of the capabilities and limitations of AI models. It then provides an overview of operations in the information environment (OIE) and narrows the discussion of OIE to influence activities aimed at the cognitive dimension—as opposed to attacks on information systems, protection of friendly information, or activities primarily aimed at informing audiences. Finally, the chapter discusses challenges facing the OIE community, specifically highlighting the difficulties of understanding a dynamic, noisy information environment, the limited in-house product development capabilities, the shortfalls in timeliness and scale of messaging, and the enduring issues of measuring effectiveness. Several applications of AI are proposed as potential solutions for mitigating the identified challenges. The intent is not to provide a comprehensive review of AI or OIE, but to highlight the broad range of potential use cases of AI for OIE.

B. DEFINING ARTIFICIAL INTELLIGENCE

There is no standard definition of "artificial intelligence." John McCarthy—one of the original founders of the AI discipline and the neologist of the term—defined AI as the "science and engineering of making intelligent machines, especially intelligent

---

[7] Burrhus Frederic Skinner, *Contingencies of Reinforcement: A Theoretical Analysis* (Englewood Cliffs, NJ: Prentice-Hall, 1969), 295.

computer programs."[8] This definition, however, did not result in a uniform conception of AI. Debates over fundamental questions such as what constitutes "intelligence" remain unresolved. Nonetheless, AI can be viewed as a "constellation of technologies" or a field of study rather than a specific type of hardware or software.[9] It is a foundational, general-purpose technology that has a wide range of applications. The National Security Commission on AI describes it as the "quintessential 'dual-use' technology," or as Michael Horowitz, the Director of the Emerging Capabilities Policy Office at the DOD, puts it, an "ultimate enabler."[10]

Artificial intelligence is categorized into two typologies: general and narrow. Narrow AI—or "weak AI"—performs a specific, defined task that augments human intelligence.[11] Artificial general intelligence (AGI), on the other hand, seeks to attain, at a minimum, human-level intelligence that can be applied across multiple domains or environments.[12] AGI, which is often referred to as "strong AI," has long been the subject of science fiction and remains largely aspirational. Although there have been noteworthy breakthroughs in AI research over the past decade, most experts agree that the realization of AGI remains in the distant future; others also argue that true AGI is impossible, given that computers are unable to develop human reasoning or

[8] John McCarthy, "What Is Artificial Intelligence?," November 12, 2007, 2, http://jmc.stanford.edu/articles/whatisai/whatisai.pdf.

[9] National Security Commission on Artificial Intelligence, National Security Commission on Artificial Intelligence Final Report, 31.

[10] National Security Commission on Artificial Intelligence, 7; Alex Wilner and Casey Babb, "New Technologies and Deterrence: Artificial Intelligence and Adversarial Behaviour," in *NL ARMS Netherlands Annual Review of Military Studies 2020*, ed. F. Osinga and T. Sweijs (The Hague: T.M.C. Asser Press, 2020), 406, https://doi.org/10.1007/978-94-6265-419-8_21.

[11] Dave Martinez et al., Artificial Intelligence: Short History, Present Developments, and Future Outlook Final Report (Cambridge, MA: Massachusetts Institute of Technology, 2019), 9.

[12] Artificial intelligence that surpasses human intelligence is sometimes referred to as "artificial super intelligence," which is sometimes viewed as a third type of AI.

Pei Wang and Ben Goertzel, "Introduction: What Is the Matter Here?," in *Theoretical Foundations of Artificial General Intelligence*, ed. Pei Wang and Ben Goertzel, vol. 4, Atlantis Thinking Machines (Paris: Atlantis Press, 2012), 2, https://doi.org/10.2991/978-94-91216-62-6_1; Serap Uğur and Gulsun Kurubacak, "Artificial Intelligence to Super Artificial Intelligence, Cyber Culture to Transhumanist Culture: Change of the Age and Human," in *Handbook of Research on Learning in the Age of Transhumanism*, ed. Serap Sisman-Ugur and Gulsun Kurubacak (Hershey, PA: IGI Global, 2019), 3, https://doi.org/10.4018/978-1-5225-8431-5.

articulate tacit human knowledge.[13] Thus, currently existing AI technologies fall into the category of narrow AI.[14]

1. Rules-Based Expert Systems

Expert systems and machine learning are two prominent categories within narrow AI. Expert systems leverage programmed sets of rules representative of human knowledge to process information.[15] These systems were part of the "first wave" of AI; they included programs such as the General Problem Solver, ELIZA natural language processing tool, and the Air Force's tactical expert mission planner system (TEMPLAR).[16] Unfortunately, expert systems are limited by their lack of learning capability, poor handling of uncertainty, and restrictive focus on narrowly defined problems.[17] Expert systems derive their knowledge base from specific domains that are not easily transferable to other fields, often lacking interoperability between systems.[18] They also exhibit rigidity due to their reliance on static facts and specified human input.[19] Nonetheless, advancements in computing have greatly improved the performance of expert systems over the past two decades, leading to widespread adoption

---

[13] Ragnar Fjelland, "Why General Artificial Intelligence Will Not Be Realized," *Palgrave Communications* 7, no. 1 (June 2020): 3, https://doi.org/10.1057/s41599-020-0494-4.

[14] Nanyi Fei et al., "Towards Artificial General Intelligence via a Multimodal Foundation Model," *Nature Communications* 13, no. 1 (December 2022): 1, https://doi.org/10.1038/s41467-022-30761-2.

[15] Greg Allen, *Understanding AI Technology*, AD1099286 (Washington, DC: Joint Artificial Intelligence Center (JAIC), 2020), 6, https://apps.dtic.mil/sti/citations/AD1099286.

[16] Ann Miller, "Expert Systems: The Structure, History, and Future of Successful AI Applications," *IEEE Potentials* 5, no. 3 (October 1986): 12, https://doi.org/10.1109/MP.1986.6500801; Caroline Bassett, "The Computational Therapeutic: Exploring Weizenbaum's ELIZA as a History of the Present," *AI & SOCIETY* 34, no. 4 (December 1, 2019): 803, https://doi.org/10.1007/s00146-018-0825-9; Alton K. Marsh, *Guide to Defense & Aerospace Expert Systems* (Arlington, VA: Pasha Publications Inc., 1986), 23, https://apps.dtic.mil/sti/pdfs/ADA636820.pdf.

[17] John Launchbury, "A DARPA Perspective on Artificial Intelligence" (Presentation, DARPA, Arlington County, VA, February 15, 2017), 5, https://www.darpa.mil/attachments/AIFull.pdf.

[18] Chuleeporn Changchit, "Expert Systems," in *Encyclopedia of Information Technology Curriculum Integration:*, ed. Lawrence A. Tomei (Hershey, PA: IGI Global, 2008), 319, https://doi.org/10.4018/978-1-59904-881-9.

[19] Changchit, 318.

in areas that require a consistent level of consultation or knowledge management.[20] Furthermore, expert systems are frequently integrated into larger intelligent systems—or hybrid systems—that leverage a combination of AI methods to provide more advanced or optimized solutions.[21]

2. Machine Learning

The application of machine learning (ML) is the primary subject of this thesis. Recent excitement over the potential of AI is typically in reference to ML, which has been fueled by an explosive growth in datasets, infrastructure, and graphics processing units.[22] ML systems differ from expert systems in that they can automatically learn from data and generate their own set of rules—unlike expert systems that depend on human inputted rules.[23] ML systems are trained through one of four techniques: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning (Figure 1). Distinguishing between the different methods of learning is important because certain techniques are more appropriate for addressing certain tasks. For more details on the different ML techniques, see Appendix A.

---

[20] Christine Strauss et al., eds., Database and Expert Systems Applications: 32nd International Conference, DEXA 2021, Virtual Event, September 27–30, 2021, Proceedings, Part I, vol. 12923, Lecture Notes in Computer Science (Cham, Denmark: Springer International Publishing, 2021), https://doi.org/10.1007/978-3-030-86472-9; Fatima M. Salman and Samy S. Abu-Naser, "Expert System for COVID-19 Diagnosis," 2020, http://dspace.alazhar.edu.ps/xmlui/handle/123456789/588.

[21] Edward E. Brent, "Expert Systems," in *SAGE Research Methods Foundations* (London, UK: SAGE Publications Ltd, 2020), https://doi.org/10.4135/9781526421036842862; Larry R. Medsker, *Hybrid Neural Network and Expert Systems* (Boston, MA: Springer U.S., 1994), 220–21, https://doi.org/10.1007/978-1-4615-2726-8.

[22] Muhammad Usama et al., "Unsupervised Machine Learning for Networking: Techniques, Applications and Research Challenges," *IEEE Access* 7 (2019): 65579–80, https://doi.org/10.1109/ACCESS.2019.2916648.

[23] Allen, Understanding AI Technology, 7.

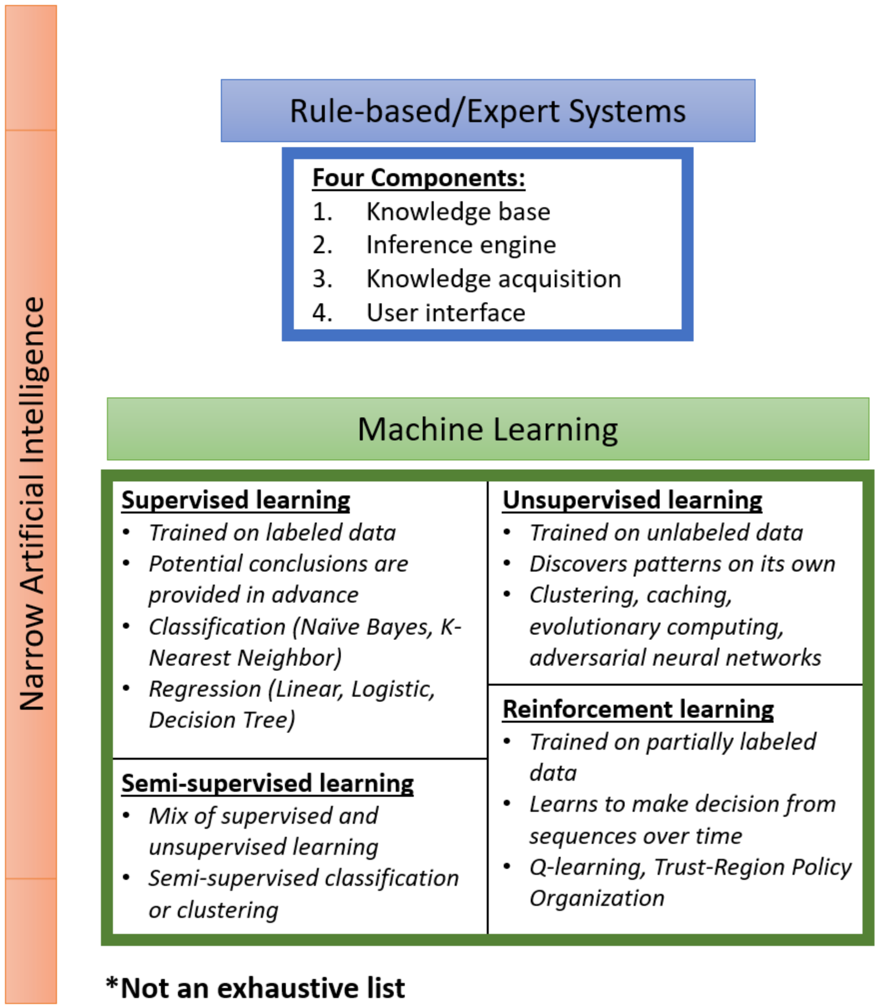


Figure 1. Categories of Narrow Artificial Intelligence

3. Neural Networks and Deep Learning

Another important ML concept is deep learning and the use of artificial neural networks. Modelled after the human brain, neural networks consist of layers of interconnected nodes that have an assigned value and weight.[24] Input data is adjusted based on the respective weights at each layer and then evaluated on whether it meets a

[24] Amit Kumar Tyagi and G. Rekha, "Challenges of Applying Deep Learning in Real-World Applications," in *Challenges and Applications for Implementing Machine Learning in Computer Vision*, ed. Ramgopal Kashyap and A.V. Senthil Kumar (Hershey, PA: IGI Global, 2020), 93, http://www.igi.global.com/chapter/challenges-of-applying-deep-learning-in-real-world-applications/242103.

threshold before moving onto the next layer. The model undergoes an iterative process of optimization and refinement as it moves through the layers, while correcting for errors. The term, “deep,” refers to the multiple “hidden layers”—layers between the input and output layers—within the neural network.[25] Deep learning, which traditionally leverages supervised learning but increasingly uses unsupervised learning, is able to process unstructured data and is less reliant on human intervention.[26] This advantage allows the model to perform more complex tasks in an efficient manner. Yet, despite the excitement surrounding the application of deep learning for tasks such as image, signal, or language processing, there are several drawbacks. Neural networks require significant allocation of energy and time as well as a massive amount of data to train. Its computational complexity also exacerbates the black box problem.[27]

4. Natural Language Processing

Interviews with industry, R&D, and information practitioners indicate that natural language processing (NLP) is currently the most prevalently utilized AI application within the OIE community. NLP is a major area of research that cuts across multiple disciplines, and while the field has existed for decades, it was not until the last five years that significant progress was made.[28] This technology automates the processing, categorization, and analysis of human language and is applied to a wide range of different applications to include document classification, content analysis, language translation, and virtual assistance.[29] One of the most well-known applications of NLP is sentiment analysis. This technology automatically extracts subjective information from texts,

[25] Michael Nielsen, *Neural Networks and Deep Learning* (Determination Press, 2015), 11, http://neuralnetworksanddeeplearning.com/.

[26] Juergen Schmidhuber, “Deep Learning in Neural Networks: An Overview,” *Neural Networks* 61 (January 2015): 7–9, https://doi.org/10.1016/j.neunet.2014.09.003.

[27] Tyagi and Rekha, “Challenges of Applying Deep Learning in Real-World Applications,” 102.

[28] Bonan Min et al., “Recent Advances in Natural Language Processing via Large Pre-Trained Language Models: A Survey,” November 1, 2021, 1, http://arxiv.org/abs/2111.01243.

[29] Peter Schirmer et al., *Natural Language Processing: Security- and Defense-Related Lessons Learned* (Santa Monica, CA: RAND Corporation, 2021), 2, https://www.rand.org/pubs/perspectives/PEA926-1.html.

enabling an analyst to perceive the social sentiment—positive, negative, or neutral—of an online user.[30] Further developments in NLP can remove language and cultural barriers by accurately discerning the context, emotion, and nuance within text.[31]

The training of ML systems involves an iterative, non-linear process of collecting and cleaning data as well as constructing, training, fine-tuning, and evaluating the model. Distinguishing between the various training processes of AI/ML systems illuminates the variability in performance. Each learning method has its advantages and disadvantages, which makes it more apt for certain tasks. Although many of today's AI applications use a combination of different ML algorithms and learning methods, understanding the distinction between the techniques enables users to recognize the limitations of the technology. This understanding is critical if the intent is to keep the human in the loop when utilizing an AI system.

C. UNDERSTANDING OPERATIONS IN THE INFORMATION ENVIRONMENT

Information, a foundational component of human activity, poses a particular challenge for policymakers and practitioners due to its extensive scope and multifaceted application across multiple disciplines.[32] Over the years, the DOD has used a variety of terms to refer to operations that impact the information environment. The evolution of these terms and their usage have been shaped by political, operational, and cultural factors. Despite the DOD's ongoing struggle with terminology, establishing a common conceptual understanding of information's role in military operations is central to the formulation of strategy and policy and provides a baseline for discourse and organized action.[33]

---

[30] Yabing Wang et al., "Refined Global Word Embeddings Based on Sentiment Concept for Sentiment Analysis," *IEEE Access* 9 (2021): 37075, https://doi.org/10.1109/ACCESS.2021.3062654.

[31] Michael Horowitz et al., *Artificial Intelligence and International Security* (Washington, DC: Center for a New American Security, 2018), 6, https://www.cnas.org/publications/reports/artificial-intelligence-and-international-security.

[32] Sandra Braman, "Defining Information," *Telecommunications Policy* 13, no. 3 (September 1989): 233, https://doi.org/10.1016/0308-5961(89)90006-2.

[33] Braman, 234.

Recognizing the omnipresence of information within every warfighting domain and its criticality across a range of operations, the Joint Chiefs of Staff have pivoted away from the term, "information operations," in favor of "operations in the information environment." The recently released *Joint Publication 3-04: Information in Joint Operations* defines OIE as "military actions involving the integrated employment of multiple information forces to affect drivers of behavior by informing audiences; influencing foreign relevant actors; attacking and exploiting relevant actor information, information networks, and information systems; and protecting friendly information, information networks, and information systems."[34]

The doctrinal shift in terminology facilitates a more expansive view of the application of informational power. Rather than narrowly focusing on the coordination, integration, and employment of various information-related capabilities, OIE recognizes "the inherent informational aspects of all activities."[35] Although the intent is to foster greater appreciation of information across the competition continuum, the broadened scope of OIE creates challenges when trying to discuss the concept with greater precision. OIE could range from kinetic operations to cyber operations or public affairs. Though they may all affect the information environment, each type of operation involves unique capabilities, processes, and effects.

Therefore, to discuss the role of AI systems within OIE with more granularity, this thesis focuses on a subset of OIE: information activities aimed at *influencing* foreign relevant actors. While these activities may span across different types of information forces (e.g., psychological operations forces, civil affairs, public affairs, combat camera), the purpose is to maintain, prevent, or change behavior of a foreign target audience.[36]

[34] Joint Chiefs of Staff, *Information in Joint Operations*, VII–1.

[35] Joint Chiefs of Staff, I–9.

[36] To avoid entering the unresolved terminology quagmire, this thesis resorts to using one term—operations in the information environment (OIE)—to discuss all information activities aimed at influencing foreign actors. This thesis acknowledges that many of the activities discussed can fall into more specific categories (i.e., military information support operations) or have been discussed in other terms previously (i.e., psychological warfare, information warfare). Although important, an explication of these terms, which vary in interpretation across the military services, the interagency community, and the general public, is beyond the scope of this thesis.

These activities contrast with others that simply inform audiences (i.e., enable the release of accurate and timely information) or protect, attack, or exploit information systems and networks. Influence activities target the cognitive—rather than the "functional" or technical—component of a military objective.[37]

The information environment has been traditionally viewed through three interrelated dimensions: physical, informational, and cognitive.[38] The cognitive dimension involves "individuals' or groups' information processing, perception, judgement, and decision making," which are shaped by psychographic, environmental, and other factors.[39] Influencing the cognitive dimension is inherently a complex endeavor, requiring a deep understanding of human psychology, communications, and social sciences. Ascertaining the true conditions that lead to behavior change is an intractable—if not impossible—problem. In addition to the lack of visibility and tangibility of mental processes, the numerous layers or "channels" present within the conveyance of messages complicate the sender's ability to obtain and interpret feedback.[40] The challenge of assessing the effectiveness of messaging is exacerbated at the group and societal level, where direct feedback is generally absent, and indicators of success are inferential in nature.[41]

D. CHALLENGES WITHIN OIE AND POTENTIAL AI APPLICATIONS

In addition to the inherent complexity of operating within the cognitive dimension, the OIE community faces a number of longstanding challenges ranging from concerns over information force structure to the lack of a cohesive national-level

[37] U.S. Marine Corps, *MCDP 8 Information*, PCN 142 000018 00 (Washington, DC: Department of the Navy, Headquarters United States Marine Corps, 2022), 2–18, https://www.marines.mil/Portals/1/Publications/MCDP%208.pdf?ver=6gIvEcD0CUuPAgTSmyDNag%3d%3d.

[38] Joint Chiefs of Staff, *Information Operations*, JP 3-13 (Washington, DC: Department of Defense, 2014), I–1, https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_13.pdf.

[39] Joint Chiefs of Staff, I–3.

[40] Wilbur Schramm, "How Communication Works," in *Readings in Psychological Operations*, ST 33–151 (Fort Bragg, NC: U.S. Army Special Warfare School, 1963), III-I–7.

[41] Schramm, III-I–16.

information strategy.[42] Organizational politics and deeply ingrained biases have contributed to risk aversion and an inconsistent application of informational power. The nature of the modern information environment generates further complications. Information forces are faced with a fast-moving, saturated information environment, where competition for attention is at unprecedented levels. Although technology is largely responsible for this increased complexity, it also offers an opportunity for information practitioners to break through some of the traditional obstacles within OIE.

The following section identifies four major challenges facing the OIE community—difficulties in sensing the information environment, limited in-house product development, shortfalls in timeliness and scale of messaging, and persistent challenges in measuring effectiveness—and discusses how AI could potentially address these concerns. This section is intended to provide a general overview of how existing and future applications of AI could augment OIE; it does not provide an exhaustive list of all the potential AI applications and tools that could impact the information environment, which could be quickly outdated due to the pace of technological advancement.

1. Difficulties in Sensing the Information Environment

As with all operations, gaining situational understanding of the environment is critical for the planning and execution of OIE. The speed of information diffusion and the inundation of noise within the information environment creates serious challenges in parsing relevant information to inform decision-making. The vast troves of publicly available information (PAI) can induce information overload, and coupled with high operational tempo, may lead to simplified mental models that generate suboptimal decisions (i.e., bounded rationality).[43] Furthermore, the cognitive dimension poses an

---

[42] Christopher J. Lamb, *Review of Psychological Operations Lessons Learned from Recent Operational Experience* (Washington, DC: National Defense University, 2005), 4, https://apps.dtic.mil/sti/citations/ADA445151; Individuals within the Department of Defense and OIE community, personal communications and panel discussions during the 2022 Phoenix Challenge workshop, April 25, 2022.

[43] Herbert A. Simon, "Bounded Rationality," in *Utility and Probability*, ed. John Eatwell, Murray Milgate, and Peter Newman (London, UK: Palgrave Macmillan, 1990), 15, https://doi.org/10.1007/978-1-349-20568-4_5.

even greater challenge for sensemaking due to the paucity of observable indicators and the reliance on inferential analysis.

To obtain effective situational awareness of the information environment, information practitioners require close integration with the intelligence community.[44] The inextricable linkage between intelligence and OIE during all phases of an influence campaign—from planning to assessment—has been well documented over the years.[45] Unfortunately, the vexations of the OIE community in its "consumer role" persist today. A 2020 MITRE study identified "poor integration with intel" as the "number one concern" among members within the Special Operations OIE community.[46] This concern over insufficient intelligence support was a common theme across several studies.[47] Requests to the intelligence community for additional information regarding a target audience (TA) have been known to get redirected back to the originator of the request.[48] Despite this longstanding issue, the OIE community remains in a continual struggle for dedicated intelligence resources.[49] This competition for intelligence support is unsurprising, given that the demand stretches across multiple warfighting functions. As a result, information forces often resort to relying on internal resources and capabilities to fill intelligence gaps. The problem, however, is that in addition to lacking the appropriate

---

[44] Christopher Paul et al., *Improving C2 and Situational Awareness for Operations in and Through the Information Environment* (Santa Monica, CA: RAND Corporation, 2018), 60–62, https://doi.org/10.7249/RR2489.

[45] Raymond J. Barrett, "PSYOP What Is It," in *The Art and Science of Psychological Operations: Case Studies of Military Application. Volume Two*, ed. Ronald De McLaurin, Carl F. Rosenthal, and Sarah A. Skillings (Washington, DC: American Institutes for Research, 1976), 57, https://apps.dtic.mil/sti/citations/ADA030140.

[46] Anthony L. Hinen et al., Data Guide for Special Operations Forces (SOF) Military Information Support Operations (MISO) (Tampa, FL: MITRE, 2020), 8.

[47] Paul et al., Improving C2 and Situational Awareness for Operations in and Through the Information Environment, xix; Hinen et al., Data Guide for Special Operations Forces (SOF) Military Information Support Operations (MISO), 8.

[48] Individuals within the Department of Defense and OIE community, personal communications and panel discussions during the 2022 Phoenix Challenge workshop.

[49] Michael Schwille et al., Intelligence Support for Operations in the Information Environment: Dividing Roles and Responsibilities Between Intelligence and Information Professionals (Santa Monica, CA: RAND Corporation, 2020), xiv, https://doi.org/10.7249/RR3161.

means or authorities for intelligence collection, analysis, and production, information forces have significantly less personnel than other military or intelligence specialty fields.

Given the growing demand for increased situational understanding of the information environment, the DOD is beginning to explore various tools and technologies to assist in research and analysis of the information environment. One of the ongoing efforts is the "Influence Campaign Awareness and Sensemaking" (INCAS) program run by the Defense Advanced Research Projects Agency (DARPA). This program seeks to leverage emerging AI techniques to "detect, characterize, and track geopolitical influence campaigns with quantified confidence."[50] The Joint Artificial Intelligence Center (JAIC), which has been recently integrated into the Chief Digital and Artificial Intelligence Office (CDAO), has also been working on an AI tool called Entropy, which aims to ingest text and video data streams and provide practitioners with real-time trend analysis.[51]

In addition to acquiring a general understanding of the information environment, the analysis of TAs is a prerequisite for effective influence operations. AI technology could support this critical step within the OIE process—one that is formally ingrained in the planning processes within psychological operations, civil affairs, and public affairs—by enhancing the breadth and depth of understanding of a TA's vulnerabilities and susceptibility to influence.[52] Information practitioners can leverage many of the tools that are currently being explored by the intelligence community such as Palantir Gotham, Better Extraction from Text Toward Enhanced Retrieval (BETTER), Machine

[50] Brian Kettler, "Influence Campaign Awareness and Sensemaking," Defense Advanced Research Projects Agency, accessed August 29, 2022, https://www.darpa.mil/program/influence-campaign-awareness-and-sensemaking.

[51] Mark Pomerleau, "Pentagon's AI Center to Field New Psychological Operations Tool," *C4ISRNet*, September 11, 2020, sec. Artificial Intelligence, https://www.c4isrnet.com/artificial-intelligence/2020/09/11/pentagons-ai-center-to-field-new-psychological-operations-tool/.

[52] Joint Chiefs of Staff, *Civil-Military Operations*, JP 3-57 (Washington, DC: Department of Defense, 2018), A-A-6, https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_57.pdf; Joint Chiefs of Staff, *Military Information Support Operations*, JP 3-13.2 (Washington, DC: Department of Defense, 2014), V–4, https://irp.fas.org/doddir/dod/jp3-13-2.pdf; Joint Chiefs of Staff, *Joint Intelligence Preparation of the Operational Environment* (Washington, DC: Department of Defense, 2014), III–22, https://irp.fas.org/doddir/dod/jp2-01-3.pdf; Joint Chiefs of Staff, *Public Affairs*, JP 3-61 (Washington, DC: Department of Defense, 2016), III–23, https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_61.pdf.

Translation for English Retrieval of Information in Any Language (MATERIAL), Open Source Indicators (OSI), and Hybrid Forecasting Competition (HFC).[53]

2. Limited In-House Capability to Develop High-Quality Products

Some view World War II as the glory days of American information prowess, when the U.S. government was able to enlist great talents such as John Steinbeck and Frank Capra to develop products supporting U.S. propaganda efforts.[54] This period, however, was not without some failures and flops—as described by journalist Gladwin Hill who wrote a scathing assessment of the American film program in Europe.[55] The struggle to develop high quality OIE products continues into more recent times. A 2005 *Review of Psychological Operations Lessons Learned from Recent Operational Experience Studies* states that the "lack of sufficiently high-quality PSYOP products" has been a consistent issue over the past two decades.[56] Although the benchmark for "high quality" may be viewed subjectively, the sophistication of an OIE product would ideally be on par with industry standards to compete for salience in a crowded information environment.

Developing a quality product, however, is not an easy task. In addition to the extensive analysis required to understand the TA, ingenuity and skill are needed to generate high quality media, which often require extensive training and experience. The military, however, faces the challenge of continual personnel turnover, averaging every one to three years. Although each branch of service has career information professionals, it is rare for an individual to remain dedicated to a particular TA or mission for the

[53] Daniel Ish, Jared Ettinger, and Christopher Ferris, *Evaluating the Effectiveness of Artificial Intelligence Systems in Intelligence Analysis* (Santa Monica, CA: RAND Corporation, 2021), 14–16, https://www.rand.org/pubs/research_reports/RRA464-1.html.

[54] John Steinbeck, *The Moon Is Down* (New York, NY: Penguin Books, 1995); Mark Harris, *Five Came Back* (New York, NY: Penguin Books, 2014).

[55] Gladwin Hill, "The American Films' Program in Occupied Germany," in *A Psychological Warfare Casebook*, ed. William E. Daugherty and Morris Janowitz (Baltimore, MD: The Johns Hopkins Press, 1958), 575.

[56] Lamb, Review of Psychological Operations Lessons Learned from Recent Operational Experience, 12.

duration of an influence campaign. Thus, given the limited resources, personnel, and expertise within the military, development and production of the OIE products are often contracted out to external entities.

Although higher production value may result from handing off the task to other organizations with the requisite capital, there are a few downsides. First, the information practitioner loses a certain amount of flexibility and control over the development process. Second, the utilization of an external resource could be a time-consuming activity, since it would require the practitioner to facilitate the contracting process and undergo continual negotiations with the vendor. The third downside is the cost of outsourcing the job. Adequate funding is needed to research, create, disseminate, and assess products. Yet, compared to other entities that conduct influence campaigns (e.g., marketing and advertising firms, commercial companies, political action committees), information forces face minimal funding and resources. In comparison to Amazon's $10 billion ad spending in 2021 or the $1.7 billion ad spending during the 2020 presidential election, the U.S. military had a budget of $228 million in FY21 for theater-level military information support operations and a $185 million budget request in FY22.[57]

Over the years, technology has aided in the improvement of product quality across various forms of media. New printers, high-quality cameras, upgraded loudspeaker systems, advanced video editing and graphic design software offer advanced capabilities for product development. Emerging AI technology has the potential to revolutionize product development through the use of large language models and synthetic media (also referred to as manipulated digital content or deepfakes). Language models such as

[57] Bradley Johnson, "How Marketers Got Back in the Game: Ad Spending Surged Last Year with the Biggest Increase Seen since the 1970s. But Budgets Could Come under Pressure amid Signs of a Looming Recession," *Advertising Age* 93, no. 10 (June 27, 2022): 14–14; Travis N. Ridout, Erika Franklin Fowler, and Michael M. Franz, "Spending Fast and Furious: Political Advertising in 2020," *The Forum* 18, no. 4 (March 1, 2021): 475, https://doi.org/10.1515/for-2020-2109; Office of the Under Secretary of Defense (Comptroller)/Chief Financial Officer, *Operation and Maintenance Overview: Fiscal Year 2022 Budget Request* (Washington, DC: Department of Defense, 2021), 85–86, https://comptroller.defense.gov/Portals/45/Documents/defbudget/FY2022/FY2022_OM_Overview.pdf.

DALLE, AudioLM, and Generative Pre-trained Transformer (GPT) 3 can create original images, audio, or human-like text from relatively short prompts.[58]

Although deepfakes are popularly known for their nefarious use to support disinformation, synthetic media—which leverages generative adversarial networks (GANs)—can be used for visual content creation and editing. Gaming and entertainment industries are known to have used this technology to create synthetic faces that are not just indistinguishable from real ones, but also viewed as more trustworthy.[59] The use of GANs for image translation, enhancement, restoration, and inpainting allows an editor to manipulate the image in a way that is imperceptible to humans.[60] Though GANs are difficult and costly to train, they are being leveraged for a growing number of applications and will likely increase in sophistication and accessibility. The availability of this technology to the military could significantly enhance audiovisual product quality.

3. Shortfalls in Timeliness and Scale of Information Dissemination

While the quality of the product is important, the timing and scale of its release can make the difference between a product going viral or the TA never getting exposed to the message. As the speed of information diffusion continues to grow exponentially, the explosion of content within the modern information environment has exacerbated the competition for audience attention. Technology has reshaped traditional communication

58 "DALL·E 2," OpenAI, 2022, https://openai.com/dall-e-2/; Zalán Borsos et al., "AudioLM," Google Research, September 7, 2022, https://google-research.github.io/seanet/audiolm/examples/; "OpenAI API," OpenAI, accessed November 21, 2022, https://beta.openai.com; Luciano Floridi and Massimo Chiriatti, "GPT-3: Its Nature, Scope, Limits, and Consequences," *Minds and Machines* 30, no. 4 (December 1, 2020): 681–94, https://doi.org/10.1007/s11023-020-09548-1.

59 Sophie J. Nightingale and Hany Farid, "AI-Synthesized Faces Are Indistinguishable from Real Faces and More Trustworthy," *Proceedings of the National Academy of Sciences* 119, no. 8 (February 22, 2022), https://doi.org/10.1073/pnas.2120481119.

60 Ming-Yu Liu et al., "Generative Adversarial Networks for Image and Video Synthesis: Algorithms and Applications," *Proceedings of the IEEE* 109, no. 5 (November 30, 2020): 6–11, https://doi.org/10.1109/JPROC.2021.3049196.

models, blurring the lines between mass and personal communications.[61] Information practitioners face the challenge of penetrating through the noise and gaining the attention of their targeted audience. In an era when information cycles move in quick, simultaneous iterations, the timing, scale, and method of the information dissemination should be carefully considered during the execution of OIE.

Developments in AI technology could augment the ability for information practitioners to increase the speed and scale of product dissemination. While existing literature often refers to the malicious use of microtargeting and social bots, these two potential applications of AI can be leveraged to tailor messaging efforts to selected TAs with greater efficiency.

Marketing and advertising firms as well as political campaigns have been utilizing a variety of techniques to identify, analyze, and direct tailored messaging to particular TAs. One such technique involves the application of microtargeting or content personalization. Microtargeting—the use of personal and demographic data to shape messaging that resonates with a small target group or individual—is not a new phenomenon. Advertising companies and political campaigns have employed this practice for decades.[62] Yet, today's expansion in computing power and the availability of data significantly enhances the ability to target individuals with more personalized messaging and improve product differentiation.[63] With sophisticated data analytics, psychometric assessments, and pattern recognition, AI could develop individual psychological profiles and generate highly personalized content to increase the

[61] Katherine Ognyanova, "Multistep Flow of Communication: Network Effects," in *The International Encyclopedia of Media Effects*, ed. Patrick Rössler, Cynthia A. Hoffner, and Liesbet Zoonen, 1st ed. (Wiley, 2017), 5, https://doi.org/10.1002/9781118783764.wbieme0056; Patrick B O'Sullivan and Caleb T Carr, "Masspersonal Communication: A Model Bridging the Mass-Interpersonal Divide," *New Media & Society* 20, no. 3 (March 1, 2018): 1162, https://doi.org/10.1177/1461444816686104.

[62] Oana Barbu, "Advertising, Microtargeting and Social Media," *Procedia - Social and Behavioral Sciences*, International Conference on Communication and Education in Knowledge Society, 163 (December 19, 2014): 44, https://doi.org/10.1016/j.sbspro.2014.12.284.

[63] Mathieu Lavigne, "Strengthening Ties: The Influence of Microtargeting on Partisan Attitudes and the Vote," *Party Politics* 27, no. 5 (September 1, 2021): 966, https://doi.org/10.1177/1354068820918387.

susceptibility of the TA by playing into human heuristics and biases.[64] Recommender systems, like those utilized by social media platforms and video streaming services, can learn from the user's preferences and filter relevant information to the TA.[65] Despite ethical concerns over the potential abuse of microtargeting, the demand for more efficient social marketing and political advertising will likely remain unabated.[66] Although empirical evidence of the impacts of microtargeted messages has varied across studies, the technique could provide information practitioners with an opportunity to improve message reach, customization, and availability to the TA.[67]

The use of social bots is another controversial but increasingly prolific application of AI. Although most social bots today remain relatively rudimentary in their capabilities, it is likely that they will grow in sophistication and eventually mimic genuine human behavior by learning, adapting, and engaging with dynamic online communities.[68] Social botnets can serve as vectors to conduct information campaigns. A "botmaster" can establish command and control over a network of bot operations by sending message content to disseminate across multiple social media platforms.[69] As technologies in

---

[64] Amelia Arsenault, "Microtargeting, Automation, and Forgery: Disinformation in Age of Artificial Intelligence" (major research, Ottawa, Canada, University of Ottawa, 2020), 43, http://ruor.uottawa.ca/handle/10393/40495.

[65] Sachi Nandan Mohanty et al., Recommender System with Machine Learning and Artificial Intelligence: Practical Tools and Applications in Medical, Agricultural and Other Industries (Newark, NJ: John Wiley & Sons, Incorporated, 2020), 3, 6, http://ebookcentral.proquest.com/lib/ebook-nps/detail.action?docID=6225842.

[66] Filipe N. Ribeiro et al., "On Microtargeting Socially Divisive Ads: A Case Study of Russia-Linked Ad Campaigns on Facebook," in *Proceedings of the Conference on Fairness, Accountability, and Transparency* (FAT* '19: Conference on Fairness, Accountability, and Transparency, Atlanta, GA: ACM, 2019), 140, https://doi.org/10.1145/3287560.3287580; Johanna Schäwel, Regine Frener, and Sabine Trepte, "Political Microtargeting and Online Privacy: A Theoretical Approach to Understanding Users' Privacy Behaviors," *Media and Communication* 9, no. 4 (2021): 160, http://dx.doi.org.libproxy.nps.edu/10.17645/mac.v9i4.4085.

[67] Lavigne, "Strengthening Ties," 967–68; Alexander L. Metcalf et al., "More 'Bank' for the Buck: Microtargeting and Normative Appeals to Increase Social Marketing Efficiency," *Social Marketing Quarterly* 25, no. 1 (March 1, 2019): 34, https://doi.org/10.1177/1524500418818063.

[68] Arsenault, "Microtargeting, Automation, and Forgery: Disinformation in Age of Artificial Intelligence," 44.

[69] Sharma Makino, M. Shrivastava, and B. Agarwal, "Denial-of-Service and Botnet Analysis, Detection, and Mitigation," in *Forensic Investigations and Risk Management in Mobile and Wireless Communications* (Hershey, PA: IGI Global, 114AD).

automation and NLP improve over time, social bots may require few specific instructions, allowing for the human to provide general themes to a network of bots, which could automatically tailor messages to specific TAs by platform.[70]

While the idea of a "botmaster" can lead to instant antipathy—or worse, feed into conspiracy theories—there could be appropriate use cases. Social bots can assist in situations where human security and safety are at risk. They can offer increased flexibility in quickly responding to emerging situations. For instance, bots can simultaneously suppress social media posts that threaten U.S. operational security and hijack a hashtag utilized by an adversary, while also amplifying pro-U.S. or allied messaging. Sleeper bots can also provide surge capability during periods of crisis. Even if some bots were to be exposed or proven to be ineffective, the relative ease with which new bots are created enables a more resilient response.[71] Therefore, social bots armed with emerging AI technologies could enhance the speed and scale of messaging efforts in an increasingly complex, fast-moving information environment.

4. The Persistent Challenge of Measuring Effectiveness

Assessing the effectiveness of OIE has been an age-old challenge for the military. As Daugherty and Janowitz stated in 1958, "no problem…is more basic than the requirement that periodic attempts be made to assess results obtained in past operations. Yet no requirement is more difficult to implement satisfactorily."[72] While the DOD is hardly the only organization to wrestle with the challenges of measuring operational effectiveness, its recent experiences in Afghanistan revealed a broken assessment system marred by overoptimism, junk arithmetic, logic failures, improper metric collection, and

70 William Marcellino et al., *Counter-Radicalization Bot Research: Using Social Bots to Fight Violent Extremism*, RR2705 (Santa Monica, CA: RAND Corporation, 2020), 10, https://doi.org/10.7249/RR2705.

71 Wajeeha Ahmad, *Why Botnets Persist: Designing Effective Technical and Policy Interventions*, IPRI(2019)02 (Boston, MA: Massachusetts Institute of Technology, 2019), 5, https://internetpolicy.mit.edu/publications-ipri-2019-02.

72 William E. Daugherty and Morris Janowitz, *A Psychological Warfare Casebook* (Baltimore, MD: The Johns Hopkins Press, 1958), 681.

simplistic representations of data.[73] Doctrinal deficiencies, inadequate training, poor processes and products, commander disinterest, and lack of advocacy have fed into a failure cycle that perpetuates inaccurate representations of operational effects.[74]

The challenge of assessing OIE harks back to the familiar predicament within causal analysis and applying quantifiable scientific measurements to the complexities of human and social behavior. Lengthy timelines to observe behavior change create a particular challenge for assessing the effects of influence operations. This difficulty can lead to an overreliance on measures of performance to explain success rather than ensuring that valid, useful measurements are collected to inform a theory of change and advance the program's objectives and intended effects.[75]A corpus of social science and market research exists to help inform various methodologies in ascertaining behavior change and campaign success, but the military fails to apply these methods in a cohesive, consistent manner. Furthermore, OIE assessments—as opposed to battle damage assessments—pose a particular challenge in that they rely less on physical evidence and indicators.[76]

The problems that arise from assessing OIE relate directly to the issues raised in the aforementioned sections. The onslaught of information, the speed with which it travels, and the lack of resources complicate the ability for information practitioners to determine whether a particular influence activity or product achieved its intended effect. In joint doctrine, assessments are often discussed separately from the rest of the planning process, and in particular, military information support operations (MISO) doctrine lists "assessments" as a distinct step (step seven) from planning and target audience analysis

[73] Stephen Downes-Martin, "Operations Assessment in Afghanistan Is Broken—What Is to Be Done?," *Naval War College Review*, no. 4 (2011): 107–17.

[74] Jonathan Schroden, "WHY OPERATIONS ASSESSMENTS FAIL: It's Not Just the Metrics," *Naval War College Review* 64, no. 4 (2011): 96, https://www.jstor.org/stable/26397245.

[75] Christopher Paul et al., Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Handbook for Practitioners (Santa Monica, CA: RAND, 2015), 7, 89.

[76] D. H. Dearth, "Implications and Challenges of Applied Information Operations," *Journal of Information Warfare* 1, no. 1 (2001): 13, http://www.jstor.org/stable/26485919.

(steps one and two).[77] This distinction, however, has more to do with the intrinsic role that assessments play throughout all phases of operations. Gaining situational awareness of the information environment, target audience analysis, and assessments are interdependent tasks that often run concomitantly once an influence campaign begins.

Assessments require an iterative process that continually monitors and reevaluates assumptions, current conditions, and relevant indicators. Yet, the unavailability of dedicated intelligence assets inhibits the ability to collect longitudinal data to determine the progress of an influence campaign. Furthermore, there are two important elements within the assessment process that are often neglected or done in a cursory manner.[78] First, outlining a theory of change or a logic model is essential to articulating the inputs and activities required to achieve specific outputs that will lead to desired effects.[79] Second, establishing a baseline is a foundational component of an effective assessment plan. Without an ex-ante understanding of the TA's conditions, attitudes, and behaviors, it would be disingenuous to claim any measures of effectiveness.

Recent advancements in AI open new possibilities to addressing some of these longstanding issues. Although not a panacea for all the challenges associated with assessments, AI offers added capabilities that could automate processes, provide new insights, and facilitate more proactive analyses.[80] Various applications of AI can assist at different stages of the assessment process. AI-enabled modeling and simulation can capture "emergent phenomenon" and describe large-scale complex systems, which would

---

[77] Joint Chiefs of Staff, *Joint Operations*, JP 3-0 (Washington, DC: Department of Defense, 2022), II–8, https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_13.pdf; Joint Chiefs of Staff, *Joint Planning*, JP 5-0 (Washington, DC: Department of Defense, 2020), VI–1, https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_13.pdf; Joint Chiefs of Staff, *Information Operations*, VI–1.

[78] Information practitioner, personal communication, March 8, 2022.

[79] Paul et al., Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade, 31.

[80] David Kiron, "AI Can Change How You Measure - and How You Manage," *MIT Sloan Management Review* 63, no. 3 (Spring 2022): 25, https://www.proquest.com/docview/2655629383/abstract/5D3E21D5094F4D31PQ/1.

provide practitioners with a better picture of potential outcomes as well as help inform logic models.[81]

Improvements in AI's ability to scan for information and identify features within text, images, and video can address the perennial challenges of collecting, analyzing, and synthesizing data into meaningful information.[82] Through automatic ingestion and fusion of data, AI can mitigate the time-consuming and costly nature of data collection both pre and post operations. NLP tasks such as named-entity recognition and topic modeling allow users to identify and categorize key pieces of information as well as extract meaning from a large corpus of texts.[83] Furthermore, continuous monitoring by AI systems enables trend analysis, anomaly detection, and the application of quantifiable metrics.[84] These types of AI applications are already being used in the marketing and advertising industry to conduct research on consumer behavior, campaign execution, and forecasting.[85]

With proper infrastructure, AI can enable the tracking of longitudinal data and account for a larger array of variables. Thus, AI can assist practitioners in identifying pertinent indicators by drawing attention to changes in the information environment. In some cases, AI could be used to remove cognitive bias and facilitate a more data-driven decision-making process.[86] By capturing and processing data from multiple sources, the

---

[81] Wenhui Fan et al., "Multi-Agent Modeling and Simulation in the AI Age," *Tsinghua Science and Technology* 26, no. 5 (October 2021): 609–10, https://doi.org/10.26599/TST.2021.9010005.

[82] Haoran Li et al., "Read, Watch, Listen, and Summarize: Multi-Modal Summarization for Asynchronous Text, Image, Audio and Video," *IEEE Transactions on Knowledge and Data Engineering* 31, no. 5 (May 2019): 996–1009, https://doi.org/10.1109/TKDE.2018.2848260.

[83] Patrick Rafail and Isaac Freitas, "Natural Language Processing," in *SAGE Research Methods Foundations*, ed. Paul Atkinson et al. (London, UK: SAGE Publications Ltd, 2020), https://dx.doi.org/9781529749120.

[84] M. Thottan and Chuanyi Ji, "Anomaly Detection in IP Networks," *IEEE Transactions on Signal Processing* 51, no. 8 (August 2003): 2192, https://doi.org/10.1109/TSP.2003.814797.

[85] Srikrishna Chintalapati and Shivendra Kumar Pandey, "Artificial Intelligence in Marketing: A Systematic Literature Review," *International Journal of Market Research* 64, no. 1 (January 2022): 44, https://doi.org/10.1177/14707853211018428.

[86] Merlin Stone et al., "Artificial Intelligence (AI) in Strategic Marketing Decision-Making: A Research Agenda," *The Bottom Line* 33, no. 2 (2020): 188, https://doi.org/10.1108/BL-03-2020-0022.

practitioner will be able to obtain a more accurate assessment of the information environment.

E. CONCLUSION

This chapter discussed how emerging and existing applications of AI could mitigate some of the challenges within OIE. AI can assist practitioners in analyzing the information environment, enabling in-house product development, enhancing timeliness and scale of information dissemination, and improving the ability to measure effectiveness. Most AI-enabled tools currently used by the DOD for OIE address the first challenge of analyzing the information environment and by extension, address some aspects of improving assessments of OIE. While there are several emerging AI technologies that can enhance product development and dissemination capabilities, adoption and employment of these technologies are incipient.

As technology continues to evolve, new AI approaches and tools will expand capabilities and potentially improve the information force's ability to plan, execute, and assess OIE. Yet, AI continues to face significant limitations in its ability to understand abstract or subjective information. It can also be rather inflexible and encounter problems when applied to different conditions or problem sets. Thus, collaboration between the human user and AI is necessary to achieve greater efficiency and effectiveness. Despite the hype and growing integration of AI technology into everyday life, there continues to be limited understanding of how humans and machines should partner to conduct and evaluate OIE.[87] If information practitioners are seeking to adopt AI tools into their organizations, a deeper understanding of human-machine teaming is necessary, which is the subject of the next chapter.

---

[87] Ipek Ozkaya, "The Behavioral Science of Software Engineering and Human–Machine Teaming," *IEEE Software* 37, no. 6 (November 2020): 4, https://doi.org/10.1109/MS.2020.3019190.

III. TRUST AND HUMAN-MACHINE TEAMING

> Interviewer: HAL, you have an enormous responsibility on this mission, in many ways perhaps the greatest responsibility of any single mission element. You're the brain, and central nervous system of the ship, and your responsibilities include watching over the men in hibernation. Does this ever cause you any lack of confidence?
>
> HAL: Let me put it this way, Mr. Amor. The 9000 series is the most reliable computer ever made. No 9000 computer has ever made a mistake or distorted information. We are all, by any practical definition of the words, foolproof and incapable of error.
>
> -2001: A Space Odyssey[88]

A. CHAPTER OVERVIEW

Stanley Kubrick and Arthur C. Clarke's HAL may be a fictional character, but it has inspired fear and fascination over the potentiality of AI while raising questions about trust and ethics in AI systems. These questions, which are critical to human-machine teaming (HMT), are the subject of this chapter. While research in this area is still at its early stages, trust and HMT frame the broader discourse of successful adoption and integration of AI into OIE. How and where the AI interacts with the human can significantly inhibit or enhance their ability to accomplish a task. These concepts are intricate and highly context and user dependent, but they underlie the effective use of the technology. An appropriate level of trust—also referred to as "calibrated trust" or "justified confidence"—is needed to ensure that the user is leveraging the advantages offered by the AI but at the same time, providing the necessary human input to ensure responsible outputs.[89]

[88] IMDb, "2001: A Space Odyssey (1968)," accessed November 13, 2022, http://www.imdb.com/title/tt0062622/characters/nm0706937.

[89] National Security Commission on Artificial Intelligence, *National Security Commission on Artificial Intelligence Final Report*, 133; Joint Artificial Intelligence Center, "Test and Evaluation Framework for AI-Enabled Systems" (Department of Defense, Washington, DC, April 27, 2022), 19, https://testscience.org/wp-content/uploads/formidable/13/Final-change-1-20220427-DataWORKS-Session-1A-TE-for-AI-Enabled-Systems-Lt-Col-Woolley.pdf.

B. HUMAN-MACHINE TEAMING

Human-machine teaming is not merely the act of an individual using a system as a tool to complete a task. It involves extensive interaction, coordinated action, shared understanding of the goal, and an interdependency between the human and the machine.[90] The challenge is determining the right balance between human involvement and increased automation. This interaction can be viewed as a "symbiotic continuum."[91] At one end, the AI is augmenting the human who is taking a more active role in completing the task. Toward the other end of the spectrum, the human is augmenting the AI and allowing for greater automation. Finding the right balance in this collaboration maximizes the efficiency and effectiveness of the system.

Saenz et al. propose a slightly different framework that configures different HMT capabilities into a quad chart based on the level of openness in the decision-making process and the level of risk (Figure 2).[92] Each category—machine-based, sequential machine-human, cyclic machine-human, and human-based AI systems—requires a different level of engagement and role for the human and the machine.[93] In a structured situation with well-defined, fixed variables and lower risk, humans can afford to take a more passive, supervisory role. An open process with high risk would warrant a human-based AI system that places the full authority of decision-making on the human.[94]

---

[90] Ozkaya, "The Behavioral Science of Software Engineering and Human–Machine Teaming," 4.

[91] Ming Qian and Davis Qian, "Defining a Human-Machine Teaming Model for AI-Powered Human-Centered Machine Translation Agent by Learning from Human-Human Group Discussion: Dialog Categories and Dialog Moves," in *Artificial Intelligence in HCI*, ed. Helmut Degen and Lauren Reinerman-Jones, Lecture Notes in Computer Science (Cham, Denmark: Springer International Publishing, 2020), 71, https://doi.org/10.1007/978-3-030-50334-5_5; *Harnessing Artificial Intelligence 2.0 - Human Machine Interaction AI*, 2021, https://www.youtube.com/watch?v=eMx2_gZqs7U.

[92] Maria Jesus Saenz, Elena Revilla, and Cristina Simón, "Designing AI Systems With Human-Machine Teams," *MIT Sloan Management Review* 61, no. 3 (Spring 2020): 2, https://www.proquest.com/docview/2392464050/abstract/2BE117B7C37847A0PQ/1.

[93] Saenz, Revilla, and Simón, 3.

[94] Saenz, Revilla, and Simón, 5.

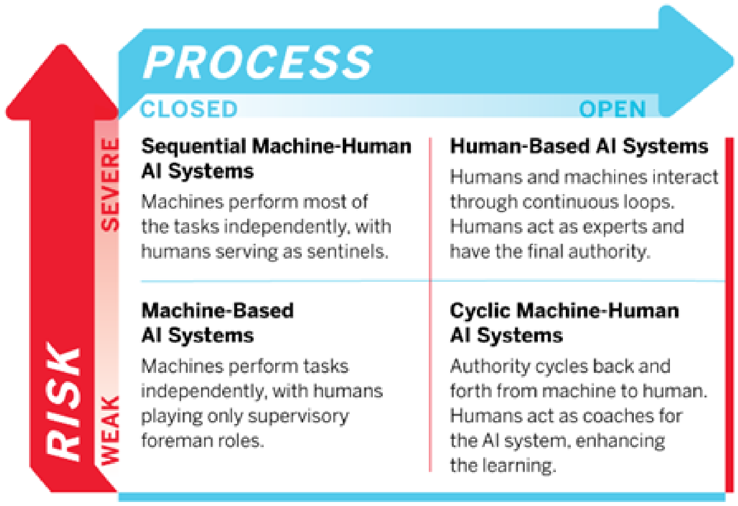


Figure 2. Human-Machine Teaming Capabilities.[95]

Where to insert human oversight within the AI life cycle has been a subject of much discussion. How and at what point the human is involved not only impact the extent to which the human-machine team can achieve the desired end state, but they can also affect the perceived trustworthiness of the system. One way to frame human involvement is by analyzing where in the "loop" human intervention is needed. "Human-before-the-loop" entails human involvement during the planning and design phase of the AI life cycle; this is where requirements, standards, and expectations are specified for the technology.[96] "Human-in-the loop"—arguably the most commonly referenced term—requires human participation in "data collection, model development, testing, and deployment of the AI system."[97] "Human-over-the-loop" consists of oversight mechanisms in which humans—specifically users or domain experts—provide feedback, serve as an arbiter, and make necessary adjustments to the system.[98]

---

[95] Source: Saenz, Revilla, and Simón, "Designing AI Systems With Human-Machine Teams."

[96] Davinder Kaur et al., "Trustworthy Artificial Intelligence: A Review," *ACM Computing Surveys* 55, no. 2 (January 18, 2022): 39:8, https://doi.org/10.1145/3491209.

[97] Kaur et al., 39:8.

[98] Kaur et al., 39:8.

Although these approaches reiterate the importance of human validation throughout the AI life cycle, they hardly address the question of how a researcher should develop an AI system to support effective HMT. Initially, the challenge of HMT was viewed primarily as a design problem focused on user experience (UX) and the way that the AI communicates back to the user.[99] Design, however, is only a subset of the HMT process, which first starts with an analysis of the problem that the HMT is trying to solve. Clearly articulating the problem is an essential prerequisite to apprehending the appropriate level of human-machine integration and applying the right AI capability in an effective manner.

Once the purpose of the system is established, task and functional analyses provide greater fidelity in determining the appropriate balance in workload and interaction. Proper task allocation recognizes the strengths and limitations of the machine versus the human. Certain tasks such as calculations, categorizing large amounts of data, or conducting other tedious tasks are much easier for machines to perform than understanding context or reacting appropriately to unanticipated situations, which often come naturally to humans. Zerilli et al. distinguish between "adaptable" and "adaptive" task allocation.[100] Within an adaptable allocation strategy, users establish the division of labor a priori, allowing the human to select which specific tasks should be given to the machine. An adaptive strategy, in contrast, enables the machine to dynamically modulate the allocation of tasks, which may be preferred in cases involving multiple tasks and requiring continual vigilance (e.g., air traffic control).[101]

To conduct a utilitarian assessment of how the AI should interact with the human, a user should establish a clear understanding of the problem, the tasks required to address the problem, and the competencies and limitations of the AI system. Some key questions to consider include how, when, and in what form should the machine interrupt or notify

[99] Harnessing Artificial Intelligence 2.0 - Human Machine Interaction AI, 2021, https://www.youtube.com/watch?v=eMx2_gZqs7U.

[100] John Zerilli, Umang Bhatt, and Adrian Weller, "How Transparency Modulates Trust in Artificial Intelligence," *Patterns* 3, no. 4 (April 8, 2022): 5, https://doi.org/10.1016/j.patter.2022.100455.

[101] Zerilli, Bhatt, and Weller, 6.

the human user? How should the AI depict uncertainty and error? What level of detail is required for explaining a particular decision? These questions continue to challenge AI developers, T&E personnel, and operational users throughout the development and deployment of the technology.[102] Answers vary based on context, mission, task, type of AI system, and the user. Therefore, UX must be continually reassessed and modified based on user feedback.

C. ROLE OF TRUST IN HUMAN-MACHINE TEAMING

To address these challenges of HMT, greater consideration is needed to account for the human element. As with any other team or partnership, trust is a foundational component of effective HMT. From facial recognition to ride sharing apps and internet searches, AI is already embedded within our daily lives.[103] The convenience and interoperability offered by these AI-enabled tools have led to widespread acceptance and use of these technologies. Despite the growing enthusiasm and hype about the potentiality of AI capabilities, trust in AI remains a key factor in assessing the integration of AI into society.[104] In the military, trust is of particular importance, given the possible strategic consequences—as well as the impact on human lives—that may arise from AI-enabled decision making. Thus, when evaluating the suitability and adoption potential of AI, trustworthiness of the technology must be a critical consideration for the warfighter.

Trust is an involuted, multilayered concept and the subject of a diverse corpus of research. While no single definition exists for trust or trustworthiness, the literature points to several key elements of trust. At a foundational level, it consists of the belief that the

[102] Ozkaya, "The Behavioral Science of Software Engineering and Human–Machine Teaming."

[103] George Hurlburt, "How Much to Trust Artificial Intelligence?," *IT Professional* 19, no. 4 (2017): 7, https://doi.org/10.1109/MITP.2017.3051326.

[104] Francesca Rossi, "Building Trust in Artificial Intelligence," *Journal of International Affairs* 72, no. 1 (2018): 128, http://www.jstor.org/stable/26588348.

other party will not cause harm.[105] It is often associated with playing by the rules, acting reasonably, meeting expectations, and fulfilling commitments.[106] Cook et al. highlight how risk and uncertainty underlie situations in which trust emerges.[107] Similarly, trust involves the inherent acceptance of vulnerability by opening oneself to the expectation of another party's behavior.[108]

Rational choice theory is often applied to the context of trust in AI, given that one side is not human. Konaev et al. characterize trust as the human's confidence in the reliability of the AI to accomplish defined tasks.[109] This logical determination based on the output of a task may seem to be the clearest interpretation of trust in a machine. One can argue that as long as the machine provides the expected outputs, the interests of the human and machine align, and thus follow the logic of Russell Hardin's theory of "trust as encapsulated interest."[110] Yet, viewing trust solely through the lens of rational choice theory neglects a more nuanced understanding of human-machine interaction. Human feelings and attitudes toward the technology (i.e., affective trust) as well as the normative expectations of behavior have a profound impact on the perceived trustworthiness of the AI.[111]

---

105 Mark S. Granovetter, *Society and Economy: Framework and Principles* (Cambridge, Massachusetts: The Belknap Press of Harvard University Press, 2017), 58; Margaret Foddy and Toshio Yamagishi, "Group-Based Trust," in *Whom Can We Trust?: How Groups, Networks, and Institutions Make Trust Possible*, ed. Karen S. Cook, Margaret Levi, and Russell Hardin (New York. NY: Russell Sage Foundation, 2009), 17.

106 Katherine Hawley, *TRUST: A Very Short Introduction*, First edition (Oxford, UK: Oxford University Press, 2012), 6; Onora O'Neill, *A Question of Trust* (Cambridge, UK: Cambridge University Press, 2003), 23.

107 Karen S. Cook et al., "Assessing Trustworthiness in Providers," in *ETrust: Forming Relationships in the Online World* (New York. NY: Russell Sage Foundation, 2009), 189–90, https://www.jstor.org/stable/10.7758/9781610446082.

108 Denise Rousseau et al., "Not So Different After All: A Cross-Discipline View of Trust," *Academy of Management Review* 23 (July 1, 1998): 395, https://doi.org/10.5465/AMR.1998.926617.

109 Margarita Konaev, Tina Huang, and Husanjot Chahal, *Trusted Partners: Human-Machine Teaming and the Future of Military AI* (Washington, DC: Center for Security and Emerging Technology, 2021), 10, https://cset.georgetown.edu/publication/trusted-partners/.

110 Russell Hardin, *Trust and Trustworthiness* (New York: Russell Sage Foundation, 2002), 1.

111 Joel E. Thompson, "Influencing Trust in Human and Artificial Intelligence Teaming through Heuristics" (master's thesis, Monterey, CA, Naval Postgraduate School, 2021), 44, https://calhoun.nps.edu/handle/10945/67823.

A wide range of literature has sought to identify factors that determine the extent of trust between humans and machines. Thiebes et al. identify five principles of trustworthy AI: beneficence, non-maleficence, autonomy, justice, and explicability.[112] Similarly, Smith emphasizes that the AI system must be respectful, secure, honest, usable, and accountable to humans as well as explicit about the risks and limitations of using the system.[113] The European Commission's *Ethics Guidelines for Trustworthy Artificial Intelligence* also highlight seven key requirements for trustworthy AI: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity and non-discrimination; environmental and societal well-being; and accountability.[114]

In addition to the identified principles that contribute toward trustworthiness, Sethumadhavan discusses the importance of considering five specific factors when ascertaining the right level of trust in AI: dispositional factors (i.e., user characteristics such as age, culture, gender, and personality); internal factors such as mood, workload, and working memory capacity; environmental or situational factors; learned factors (e.g., reliability, reputation of the system, error factors, and perceived use of the system based on past experiences); and design factors.[115] These factors could influence human-machine teaming in a multitude of different ways. For example, several studies

---

[112] Scott Thiebes, Sebastian Lins, and Ali Sunyaev, "Trustworthy Artificial Intelligence," *Electronic Markets* 31, no. 2 (June 1, 2021): 447, https://doi.org/10.1007/s12525-020-00441-4.

[113] Carol J. Smith, "Designing Trustworthy AI: A Human-Machine Teaming Framework to Guide Development," in *AAAI FSS-19: Artificial Intelligence in Government and Public Sector* (Arlington, Virginia: arXiv, 2019), 2, https://doi.org/10.48550/arXiv.1910.03515.

[114] Heike Felzmann et al., "Transparency You Can Trust: Transparency Requirements for Artificial Intelligence between Legal Norms and Contextual Concerns," *Big Data & Society* 6, no. 1 (January 1, 2019): 10, https://doi.org/10.1177/2053951719860542.

[115] Arathi Sethumadhavan, "Trust in Artificial Intelligence," *Ergonomics in Design* 27, no. 2 (April 1, 2019): 34, https://doi.org/10.1177/1064804618818592.

demonstrate that applying anthropomorphic features on an AI agent increases trust in the machine.[116]

Along similar lines, Kaplan et al. conduct a meta-analysis to examine whether certain human-related, AI-related, or contextual factors could be considered as significant predictors of trust in AI.[117] Their findings demonstrate that human competency, understanding, and expertise along with AI performance and reliability are positive predictors of trust. Greater length of relationship between the human and machine as well as similar speech patterns are also associated with more trust, while riskier situations decrease the level of trust. Their study also illustrates that several characteristic-based (e.g., culture, gender, personality) and attribute-based (e.g., AI behavior, reputation, transparency) factors are significant. Level of trust correlate with particular personality traits (e.g., "innovative" versus "lonely" individuals), culture, and gender.[118] AI systems that are more anthropomorphic and utilize teamwork and human-centered language are viewed as more trustworthy. Transparency, as well as the perception that the AI is honest and a benevolent rule-follower, also increases trust.[119]

Several common themes emerge from this body of literature. First, trust in AI is a complex mix of diverse elements to include cognitive, emotional, and situational factors. While transparency is often highlighted as a primary factor, it is but one component of developing a trustworthy system. Second, building an effective HMT and trustworthy AI requires continual refinement and assessment of human-machine interaction. The establishment of a constant feedback loop is especially important, given the advancements in modern AI technology. Unlike the narrow expert systems of the past,

---

[116] Lingyun Qiu and Izak Benbasat, "Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems," *Journal of Management Information Systems* 25, no. 4 (2009): 233, https://www.jstor.org/stable/40398956; Qiu and Benbasat, 145; Richard Pak et al., "Decision Support Aids with Anthropomorphic Characteristics Influence Trust and Performance in Younger and Older Adults," *Ergonomics* 55, no. 9 (September 1, 2012): 1059, https://doi.org/10.1080/00140139.2012.691554.

[117] Alexandra D. Kaplan et al., "Trust in Artificial Intelligence: Meta-Analytic Findings," *Human Factors*, May 28, 2021, 1, https://doi.org/10.1177/00187208211013988.

[118] Kaplan et al., 7.

[119] Kaplan et al., 8.

today's machines display greater capacity and capability to conduct more sophisticated tasks such as predictive analysis, anomaly detection, and deep reinforcement learning. Third, determining or measuring the right level of trust needed for a particular system remains a significant challenge due to the abstract, subjective nature of trust.

D. THE DEPARTMENT OF DEFENSE'S VIEW ON TRUST AND AI

Despite the inherent difficulty of measuring trust, the U.S. government and research community have been working to narrow the broad concept of trust to a more specific and measurable application. The National Security Commission on AI states in its final report that "AI systems must be developed and fielded with justified confidence."[120] The term, "justified confidence"—taken from the International Organization for Standardization (ISO), International Electrotechnical Commission (IEC), and the Institute of Electrical and Electronics Engineers (IEEE) international standards—involves the "proper calibration of trust" in which "the amount or level of trust humans place in machines is appropriate, given the machine's capabilities at that particular time and context."[121] Understanding this calibration is key to assessing the relationship between trust and the overall effectiveness of HMT to achieve operational objectives.

In 2020, the Joint Artificial Intelligence Center (JAIC) established an AI ethical principles framework to facilitate the development of responsible AI (RAI) and to assist the DOD in navigating through ethical ambiguities and risks.[122] The five principles—responsible, equitable, traceable, reliable, and governable—echo those proposed by extant literature.[123] The JAIC and now the Chief Digital Artificial Intelligence Office

120 National Security Commission on Artificial Intelligence, National Security Commission on Artificial Intelligence Final Report, 133.

121 Konaev, Huang, and Chahal, Trusted Partners: Human-Machine Teaming and the Future of Military AI, 15.

122 Joint Artificial Intelligence Center, *Responsible Artificial Intelligence (RAI): Transforming the Department of Defense Through AI* (Washington, DC: Department of Defense, 2022), 1, https://www.ai.mil/docs/rai_slick_sheet-dopsr_appvd_220317.pdf.

123 Joint Artificial Intelligence Center, 1.

(CDAO) have aligned efforts toward six tenets of RAI implementation: RAI Governance, Warfighter Trust, AI Product and Acquisition Life cycle, Requirements Validation, Responsible AI Ecosystem, and AI Workforce.[124] These tenets are delineated by lines of effort within the DOD's *Responsible Artificial Intelligence Strategy and Implementation Pathway*, which was released in June 2022.[125] Adding to these initiatives, the Defense Innovation Unit (DIU), along with members of the Software Engineering Institute at Carnegie Mellon University, developed a set of RAI guidelines to operationalize the *DOD Ethical Principles for AI*.[126] These guidelines, which are discussed in Chapter V, provide a practical way forward toward implementing the broad principles outlined by the DOD.

Test and evaluation are integral parts of not only building confidence in the system, but also improving the performance of HMT. The JAIC and CDAO are working to establish a responsible AI test and evaluation (RAITE) framework that allows DOD personnel to "generate defensible drafts of RAI evaluation plans" from which experts can "review and fine-tune…across the broader program portfolios."[127] This framework seeks to integrate "real-time monitoring, algorithm confidence metrics, and user feedback" to ensure verification and validation of the AI system.[128] The JAIC's T&E framework outlines four types of testing: algorithmic test, system integration, operational test, and human system integration (Figure 3).[129]

[124] DOD Responsible AI Working Council, U.S. Department of Defense Responsible Artificial Intelligence Strategy and Implementation Pathway, 9.

[125] DOD Responsible AI Working Council, U.S. Department of Defense Responsible Artificial Intelligence Strategy and Implementation Pathway.

[126] Jared Dunnmon et al., *Responsible AI Guidelines in Practice* (Mountain View, CA: Defense Innovation Unit, 2021), 4, https://www.diu.mil/responsible-ai-guidelines; Kathleen H. Hicks, "Implementing Responsible Artificial Intelligence in the Department of Defense" (official memorandum, Washington, DC: Department of Defense, 2021).

[127] Joint Artificial Intelligence Center, "Responsible Artificial Intelligence Test & Evaluation (RAITE)" (Washington, DC: Department of Defense, April 11, 2022), https://www.ai.mil/docs/raite_slick_220411.pdf.

[128] Hicks, "Implementing Responsible Artificial Intelligence in the Department of Defense," 2.

[129] Joint Artificial Intelligence Center, "Test and Evaluation Framework for AI-Enabled Systems," 10.

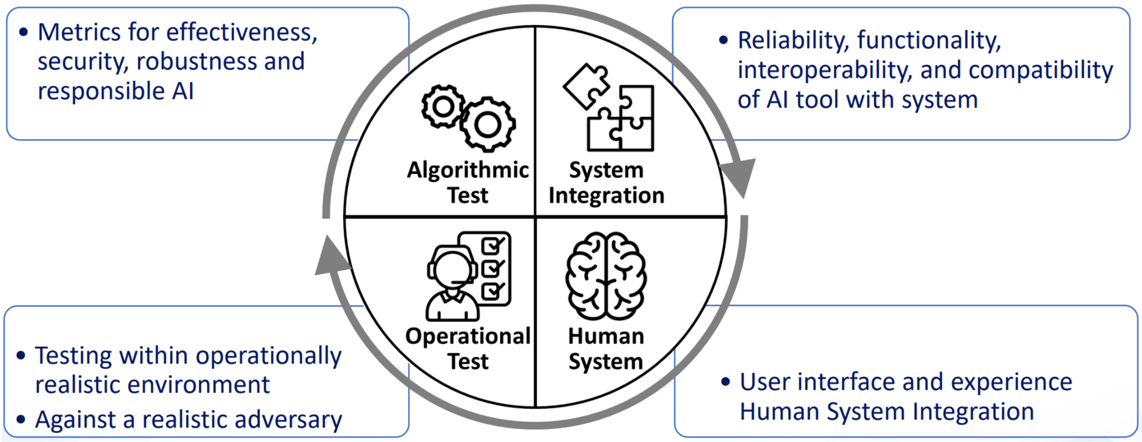


Figure 3. Joint AI Center Test and Evaluation Process.[130]

Each aspect of the process consists of different requirements and T&E approaches, but they all drive toward an overall understanding of the system's capabilities, limitations, and risks. The purpose of the technology, the identified users, and the linking of operational needs to testable requirements inform the development of the T&E strategy. The test and evaluation master plan (TEMP) integrates the framework into the overall test strategy and serves as a living document for the project throughout its life cycle.[131] Details about test requirements, evaluation approaches, performance metrics, and test schedules are laid out in the TEMP.

Domain experts and operational users are involved throughout the entire AI life cycle and T&E process, but they are likely most involved during operational testing to provide feedback on how the system should perform in a realistic scenario as well as during human system integration to evaluate user interface and UX. Operational testing examines the effectiveness, suitability, robustness, and safety of the system within a mission context. It validates the system's ability to produce the desired outputs within a real-world exercise and allows the assessors to verify that the outputs contribute to the overall objective. There are two notable challenges associated with operational testing. The first challenge is creating a realistic threat environment to test the machine. Not only

[130] Source: Joint Artificial Intelligence Center, 10.

[131] Joint Artificial Intelligence Center, 11.

is anticipating adversarial behavior inherently difficult, but the DOD also lacks sufficient data and appropriate infrastructure to mimic realistic conditions.[132] The second challenge is establishing operationally relevant and informative metrics. Given the complexity of the operational environment, creating standard metrics is a formidable task. Furthermore, translating metrics into operational effects has been a long-standing issue for the DOD.[133]

Nonetheless, evaluating human system integration can be even more convoluted than operational testing. As discussed previously, establishing an objective calibration of trust—or justified confidence—is difficult to do, given the wide range of factors and human variability when assessing trust in a system. Relevant metrics need to be identified to generate qualitative and quantitative assessments that inform how the user interacts and understands the AI system, whether HMT creates a net benefit, and how to address the evolution of HMT as both the user and machine adapt over time. Ensuring that the machine provides the right level of explainability at the right time is more of an art than a science, requiring continual operator feedback.[134] Yet, this also raises additional questions such as what is the optimal way for the user to provide feedback and how does this feedback get processed and analyzed to inform the continued development of the system? While the T&E of human system integration may vary widely depending on the specific user and system, these issues are important to consider across all cases of HMT.

E. CONSIDERATIONS FOR HUMAN-MACHINE TEAMING IN OIE

Although most of the concepts discussed thus far apply to a broad spectrum of AI technologies, there are several points that are worth emphasizing regarding the pairing of humans and AI for OIE.

---

[132] Michele A. Flournoy, Avril Haines, and Gabrielle Chefitz, *Building Trust through Testing* (Washington, DC: Center for Security and Emerging Technology, 2020), 9, https://cset.georgetown.edu/wp-content/uploads/Building-Trust-Through-Testing.pdf.

[133] Daniel Egel et al., *Leveraging Machine Learning for Operation Assessment* (Santa Monica, CA: RAND Corporation, 2022), 1, https://www.rand.org/pubs/research_reports/RR4196.html.

[134] Patrick Hall, "On the Art and Science of Machine Learning Explanations," in *2019 KDD XAI Workshop* (Anchorage, AK: arXiv, 2020), http://arxiv.org/abs/1810.02909.

First, OIE is inherently a cognitive, human-centric activity. The ultimate aim is to influence individual and group behavior and decision-making, which are shaped by a multitude of factors including cultural beliefs, emotions, vulnerabilities, and experience. AI is particularly weak in its ability to create cognitive solutions or manage subjectivity and the dynamic nature of the human domain.[135] Therefore, greater reliance is placed on the human to conduct analysis and decision-making. In this case, it is even more critical for the human to understand how the AI is processing data and generating outputs.[136] The explainability of the system ensures that the human can make the most informed and optimal decisions.

Yet, even more fundamental to the effective and responsible employment of HMT is the proper maintenance and safeguard of data that is being fed into the AI system. Simply obtaining voluminous amounts of data does not automatically lead to more advanced AI systems. The quality of the data matters; it must be preprocessed and cleaned to provide utility. Ingrained flaws and biases may be embedded within the dataset, and it is not always easy for the human to discern these potential problems. The dataset could be missing values, include incorrect or duplicate inputs, or consist of incompatible data that varies in units or formatting.

In addition to the basic challenges of processing data, the availability of data is another issue facing the OIE community. The variance in data availability affects the representativeness of the dataset and impacts the utility of the AI's output. Although information forces rely heavily on publicly available information (PAI), cost and data collection restrictions emplaced by companies impact the information practitioners'

---

[135] Dinesh C Verma, Archit Verma, and Utpal Mangla, "Addressing the Limitations of AI/ML in Creating Cognitive Solutions," in *2021 IEEE Third International Conference on Cognitive Machine Intelligence (CogMI)*, 2021, 189, https://doi.org/10.1109/CogMI52975.2021.00033.

[136] Dawn Branley-Bell, Rebecca Whitworth, and Lynne Coventry, "User Trust and Understanding of Explainable AI: Exploring Algorithm Visualisations and User Biases," in *Human-Computer Interaction. Human Values and Quality of Life*, ed. Masaaki Kurosu (Cham, Denmark: Springer International Publishing, 2020), 383, https://doi.org/10.1007/978-3-030-49065-2_27.

ability to leverage AI-enabled tools.[137] Furthermore, data from individual platforms may not be representative of the population that the information practitioner is trying to analyze—as PAI does not include data within private pages nor does it include the views of individuals who are not engaging in the online space. Online activity and behavior do not necessarily translate into the physical dimension. While the number of Internet users is growing, there is still a significant majority (such as those within the Global South) who do not use social media as the primary means of communication.[138]

The potential ease with which an AI system can provide an output or product—such as providing summaries of online narratives—can offer convenience and efficiency for the information practitioner. It can, however, also amplify human heuristics and lead to automatic reliance or over-trust. Therefore, conscious considerations about data and the general logic behind AI outputs are needed for effective HMT during OIE.

Another challenge facing information practitioners is determining how to modulate between greater human oversight versus reliance on AI, or more fundamentally, whether the application of HMT is benefiting operations. For information forces, this challenge is particularly troublesome because assessing MOE is very convoluted, given the variety of indicators as well as noise that exist within the information environment.[139] The lack of clear metrics to assess OIE generates a compounding problem for AI T&E. Users are confronted with the problem of trying to evaluate the success of HMT despite the inherent ambiguity of determining the effects of the overall inform and influence activity. Even with the incredible progress of AI technology over recent years, the information practitioner will continue to face the complex task of extracting relevant, valid outputs from the machine while evaluating other sources of

[137] Katharina E. Kinder-Kurlanda and Katrin Weller, "Perspective: Acknowledging Data Work in the Social Media Research Life cycle," *Frontiers in Big Data* 3 (December 2020): 2, https://doi.org/10.3389/fdata.2020.509954.

[138] Simon Kemp, "Digital 2022: Global Overview Report," DataReportal – Global Digital Insights, January 26, 2022, https://datareportal.com/reports/digital-2022-global-overview-report.

[139] Paul et al., Improving C2 and Situational Awareness for Operations in and Through the Information Environment, 28.

information to build a comprehensive understanding of the information environment. How the machine conveys information to the human can significantly affect the validity of analysis and the effectiveness of the operation.

Finally, perceived risks associated with using AI for OIE can vary widely across different users and stakeholders. Some may view the consequences of applying AI to OIE as having greater strategic risks compared to other AI applications such as weapon systems, drone swarms, or intelligence process optimization.[140] The effects of OIE stretch across the continuum of competition and conflict as well as beyond the scope of military operations. As demonstrated by adversaries who are already using AI and automation to spread propaganda, amplify disinformation campaigns, and conduct microtargeting, these types of AI usage could reshape society over the long-term.[141] More aggressive applications of AI—such as leveraging social bots and synthetic media—incur significant risk, given that its exposure could cause reputational damage to the United States—a country that not only prides itself in transparency but also denounces adversarial troll and bot farms. It could also generate new conspiracy theories and negatively affect the norms of authentic online engagement.

Nonetheless, others may view the risk of using AI for OIE as relatively low, given that human lives are not directly involved (unlike lethal autonomous weapon systems). In fact, many of the tools being offered to information practitioners today are focused on assisting users with a narrow set of tasks (e.g., analysis of text), which are less likely to engender immediate or severe controversy.

---

[140] Christopher Telley, "The Influence Machine: Automated Information Operations as a Strategic Defeat Mechanism," *The Institute of Land Warfare*, no. 121 (October 2018): v, 1.

[141] Ronan Ó Fathaigh et al., "Microtargeted Propaganda by Foreign Actors: An Interdisciplinary Exploration," *Maastricht Journal of European and Comparative Law* 28, no. 6 (December 1, 2021): 856–77, https://doi.org/10.1177/1023263X211042471; Arsenault, "Microtargeting, Automation, and Forgery: Disinformation in Age of Artificial Intelligence"; Cristian Vaccari and Andrew Chadwick, "Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News," *Social Media + Society* 6, no. 1 (January 1, 2020), https://doi.org/10.1177/2056305120903408; Nightingale and Farid, "AI-Synthesized Faces Are Indistinguishable from Real Faces and More Trustworthy."

Given these additional considerations within OIE, one can assume that most HMT for OIE will fall into either one of two categories: “Human-Based AI Systems” or “Cyclic Machine-Human AI Systems” (as laid out in Figure 3).[142] While the risk level may vary depending on what specific task the human-machine team is working on, decision-making will typically remain an open process, since problems within the information environment are often ill-defined and entail a profusion of varying factors. In either case, transparency and explainability are paramount in facilitating effective HMT. As risk level increases, greater vigilance and situational awareness is required by the human to mitigate potential negative effects.[143]

F. CONCLUSION

This chapter touches only the surface of research required to understand the dynamics of HMT within the DOD context. Designing an effective human-machine team is a highly complex and multifaceted challenge, requiring an intricate understanding of the mission, the AI system, the human user, and their interactions. Trust is an underlying component of HMT and will impact the adoption of AI. Establishing objective measurements of trust remains an elusive task, particularly within OIE. Nonetheless, the principles and guidelines provided by DOD entities such as the CDAO and the DIU establish a foundation which the OIE community can build upon to implement responsible and effective AI.

---

142 Saenz, Revilla, and Simón, “Designing AI Systems With Human-Machine Teams,” 3.

143 Saenz, Revilla, and Simón, 7.

IV. ASSESSING AI ADOPTION FOR OIE

A. CHAPTER OVERVIEW

Predicting whether a certain technology will be successfully adopted within an organization can be an exceedingly difficult task. Nonetheless, understanding the factors that affect adoption potential is critical when acquiring new technologies. This chapter first discusses three overarching challenges found in the AI adoption literature. To conduct a deeper analysis of adoption factors, three theoretical frameworks will be highlighted: the Diffusion of Innovation (DOI) theory, the Unified Theory of Acceptance and Use of Technology (UTAUT), and the Technology-Organization-Environment (TOE) framework. The chapter will then utilize these frameworks to assess the potential for AI adoption within OIE at the individual and organizational levels.

B. CHALLENGES IN AI ADOPTION

Developments in AI technology over the last several years have led to heightened enthusiasm and greater AI adoption across various industries.[144] Yet, despite the widespread desire to leverage the technology, companies continue to struggle with the integration of AI systems into their core practices.[145] Many organizations fail to realize the expected benefits of the technology and often terminate AI initiatives before completion.[146] A diverse set of literature explores various factors impacting the adoption potential of AI. Recent studies focus on adoption challenges within specific sectors such

---

[144] Joe McKendrick, "AI Adoption Skyrocketed Over the Last 18 Months," *Harvard Business Review*, September 27, 2021, https://hbr.org/2021/09/ai-adoption-skyrocketed-over-the-last-18-months; IBM Watson, *IBM Global AI Adoption Index 2022* (Armonk, NY: IBM Corporation, 2022), https://www.ibm.com/watson/resources/ai-adoption.

[145] Tim Fountaine, Brian McCarthy, and Tamim Saleh, "Building the AI-Powered Organization," *Harvard Business Review*, July 1, 2019, https://hbr.org/2019/07/building-the-ai-powered-organization.

[146] "Businesses Are Finding AI Hard to Adopt," *The Economist*, June 11, 2020, 70, https://www.economist.com/technology-quarterly/2020/06/11/businesses-are-finding-ai-hard-to-adopt; Thomas H. Davenport, Jeff Loucks, and David Schatsky, *The 2017 Deloitte State of Cognitive Survey* (Deloitte, 2017), 12, https://www2.deloitte.com/content/dam/Deloitte/us/Documents/deloitte-analytics/us-da-2017-deloitte-state-of-cognitive-survey.pdf.

as healthcare, construction, human resources, education, and supply chain.[147] Other studies tailor their research based on the size of companies or the type of AI application (e.g., chatbots, recommendation systems).[148] Surveys conducted by IBM, McKinsey, O'Reilly Media, the National Science Foundation, and others assess AI adoption levels across different industries and use cases.[149]

Although exact figures and analyses may vary, these studies highlight several common themes regarding the challenges in AI adoption. First, the lack of understanding of AI capabilities can be a serious impediment to the proper integration of the technology into organizational processes. Misperceptions about the capabilities and limitations of AI can lead to improper task alignment or create unwarranted expectations that result in disappointment and abandonment of the initiative due to perceptions of failure. As a

---

147 Ahmad Khanijahani et al., "Organizational, Professional, and Patient Characteristics Associated with Artificial Intelligence Adoption in Healthcare: A Systematic Review," *Health Policy and Technology* 11, no. 1 (March 1, 2022), https://doi.org/10.1016/j.hlpt.2022.100602; Massimo Regona et al., "Opportunities and Adoption Challenges of AI in the Construction Industry: A PRISMA Review," *Journal of Open Innovation: Technology, Market, and Complexity* 8, no. 1 (March 2022), https://doi.org/10.3390/joitmc8010045; Alpana Agarwal, "AI Adoption by Human Resource Management: A Study of Its Antecedents and Impact on HR System Effectiveness," *Foresight*, January 1, 2022, https://doi.org/10.1108/FS-10-2021-0199; Chatterjee Sheshadri and Kalyan Kumar Bhattacharjee, "Adoption of Artificial Intelligence in Higher Education: A Quantitative Analysis Using Structural Equation Modelling," *Education and Information Technologies* 25, no. 5 (September 2020): 3443–63, https://doi.org/10.1007/s10639-020-10159-7; Johannes Hangl, Viktoria Joy Behrens, and Simon Krause, "Barriers, Drivers, and Social Considerations for AI Adoption in Supply Chain Management: A Tertiary Study," *Logistics* 6, no. 3 (2022), https://doi.org/10.3390/logistics6030063.

148 Anuj Kumar and Anjali Kalse, "Usage and Adoption of Artificial Intelligence in SMEs," *Materials Today: Proceedings*, February 26, 2021, https://doi.org/10.1016/j.matpr.2021.01.595; Markus Bauer, Clemens van Dinther, and Daniel Kiefer, "Machine Learning in SME: An Empirical Study on Enablers and Success Factors," in *Americas Conference on Information Systems (AMCIS 2020)* (Salt Lake City, UT: ResearchGate, 2020), https://www.researchgate.net/publication/344651203_Machine_Learning_in_SME_An_Empirical_Study_on_Enablers_and_Success_Factors; Rajasshrie Pillai and Brijesh Sivathanu, "Adoption of AI-Based Chatbots for Hospitality and Tourism," *International Journal of Contemporary Hospitality Management* 32, no. 10 (2020): 3199–3226, https://doi.org/10.1108/IJCHM-04-2020-0259; Si Shi, Yuhuang Gong, and Dogan Gursoy, "Antecedents of Trust and Adoption Intention toward Artificially Intelligent Recommendation Systems in Travel Planning: A Heuristic–Systematic Model," *Journal of Travel Research* 60, no. 8 (November 1, 2021): 1714–34, https://doi.org/10.1177/0047287520966395.

149 IBM Watson, *IBM Global AI Adoption Index 2022*; Michael Chui et al., "The State of AI in 2021," December 8, 2021, http://ceros.mckinsey.com/global-ai-survey-2020-a-desktop-3-1; Mike Loukides, "AI Adoption in the Enterprise 2022," O'Reilly Media, March 31, 2022, https://www.oreilly.com/radar/ai-adoption-in-the-enterprise-2022/; Hodan Omaar, "NSF Data Shows AI Adoption in the United States Remains Low But Big Companies Are Leading the Way," *Center for Data Innovation* (blog), March 17, 2022, https://datainnovation.org/2022/03/nsf-data-shows-ai-adoption-in-the-united-states-remains-low-but-big-companies-are-leading-the-way/.

general-purpose technology, AI has the potential to create transformational effects across a multitude of applications, but its broad utility requires focused specificity. Kiron and Schrage stress that AI should support clearly defined objectives and key performance measures (KPIs).[150] Tying the technology to specific KPIs is critical, given the complexity of the development, deployment, and maintenance of AI systems.[151]

The second challenge involves resource constraints. Although AI could assist with cost-cutting measures in the long term, successful adoption requires significant investment in human capital, data, and other necessary supporting systems and infrastructure. A wide range of expertise and skill is needed throughout the AI life cycle. The demand for AI developers, engineers, and data scientists has led to a fierce competition for talent. Access to quality data is typically a differentiator between effective and ineffective AI implementation, but data is costly and often requires supporting architecture that enables data sharing, processing, and quality assurance.[152] Given these resource requirements, it is unsurprising that the gap in AI adoption between large and small companies has grown significantly.[153]

The third challenge is the limited transparency and perceived trustworthiness of AI. The "black box" nature of AI and the difficulty in clearly quantifying tangible benefits of the technology generate cost-benefit uncertainty. Although efforts are underway to improve explainability and transparency of AI systems, the inherent complexity of AI models makes this a particular challenge. "Model instability" and

---

150 David Kiron and Michael Schrage, "Strategy For and With AI," *MIT Sloan Management Review* 60, no. 4 (Summer 2019): 30–35, http://www.proquest.com/docview/2273705001/abstract/9489BC4062544CF0PQ/1.

151 Lucas Baier, Fabian Jöhren, and Stefan Seebacher, "Challenges in the Deployment and Operation of Machine Learning in Practice," in *27th European Conference on Information Systems* (Stockholm: ResearchGate, 2019), 10–11, https://www.researchgate.net/publication/332996647_CHALLENGES_IN_THE_DEPLOYMENT_AND_OPERATION_OF_MACHINE_LEARNING_IN_PRACTICE.

152 Jan Jöhnk, Malte Weißert, and Katrin Wyrtki, "Ready or Not, AI Comes— An Interview Study of Organizational AI Readiness Factors," *Business & Information Systems Engineering* 63, no. 1 (February 2021): 13–14, https://doi.org/10.1007/s12599-020-00676-7.

153 Omaar, "NSF Data Shows AI Adoption in the United States Remains Low But Big Companies Are Leading the Way"; IBM Watson, *IBM Global AI Adoption Index 2022*, 4.

potential adversarial manipulation are other areas of concern that could lead to uncertainty about the AI system's technical performance.[154] This uncertainty could affect the expected benefits of the technology and thus lead to greater reluctance in adopting AI-enabled systems.[155]

While most studies focus on the adoption of AI within the commercial sector, the challenges identified within these studies are highly pertinent to the government and the DOD. Although the DOD is placing greater priority on AI investments and integration, the DOD's "AI ecosystem—the complex network of technology, people, computing infrastructure, data, and policy—is underdeveloped."[156] Misconceptions of what AI could provide, which often manifest as initial overhype, continue to exist. Tactical commanders may recognize the revolutionary impact that AI could have on the battlefield, but there continues to be little understanding of how the technology should be employed and maintained.

The 2021 *National Security Commission on Artificial Intelligence Final Report* states that the DOD and the intelligence community "face an alarming talent deficit," which is the "greatest impediment to the U.S. being AI-ready by 2025."[157] Obtaining the required expertise and talent is especially challenging for the government, not only because of its inability to match private sector salaries but also due to the perceived (and real) bureaucratic inertia within government—as opposed to a fast-paced innovative culture in the commercial sector. Furthermore, only recently did the DOD officially recognize data as a "strategic asset" that requires developed architecture, standards, and

---

154 Jens Westenberger, Kajetan Schuler, and Dennis Schlegel, "Failure of AI Projects: Understanding the Critical Factors," *Procedia Computer Science* 196 (January 1, 2022): 74, https://doi.org/10.1016/j.procs.2021.11.074.

155 Peter Coughlan, Nicholas Dew, and William Gates, *Crossing the Technology Adoption Chasm: Implications for DOD* (Monterey, CA: Naval Postgraduate School, 2008), 22, https://calhoun.nps.edu/handle/10945/434.

156 Lindsey Sheppard, *Accelerating the Defense Department's AI Adoption* (New York, NY: Council on Foreign Relations, 2020), http://www.jstor.org/stable/resrep29909.

157 National Security Commission on Artificial Intelligence, *National Security Commission on Artificial Intelligence Final Report*, chap. 6.

governance.[158] While the DOD owns and generates massive amounts of data, significant work is still needed to ensure the accessibility, interoperability, and linking of data for operational use.[159]

The introduction of AI systems into the DOD can be further complicated by the existence of numerous legacy systems that are difficult to upgrade due to issues of security, contracts, and permissions.[160] Budget constraints and inflexible acquisition processes have also impeded the DOD's ability to acquire and develop AI capabilities. The DOD also faces the issue of fostering transparency and trust in AI—not only among its warfighters but also the general public. While significant strides have been made in generating policies and strategies to address concerns of trust and transparency, actual implementation remains in the nascent stages.

The unique characteristics of AI create additional challenges for organizational and individual adoption of the technology. Nonetheless, there are notable similarities between the factors that affect AI adoption and those that affect other technologies. Thus, the substantial body of research on technology adoption offers additional insight into the issues of adoption for AI. Grounding this analysis in theory enables a more granular examination of the factors that affect the adoption process.

C. TECHNOLOGY ADOPTION THEORIES

Theoretical models provide a framework for understanding complex systems by linking concepts and assumptions within a construct.[161] They offer a systematic way of analyzing key variables and their relationships. Decades of research have led to the introduction of numerous models that attempt to explain technology acceptance and

---

[158] Department of Defense, *DOD Data Strategy* (Washington, DC: Department of Defense, 2020), 3, 5, https://media.defense.gov/2020/Oct/08/2002514180/-1/-1/0/DOD-DATA-STRATEGY.PDF.

[159] Department of Defense, 6–7.

[160] Sheppard, Accelerating the Defense Department's AI Adoption.

[161] Gloria Barczak, "From the Editor," *Journal of Product Innovation Management* 31, no. 5 (September 2014): 878–878, https://doi.org/10.1111/jpim.12219.

utilization.[162] This thesis focuses on three prevalent models to identify critical factors that affect the acceptance and adoption of technology: the Diffusion of Innovation (DOI) theory, the Unified Theory of Acceptance and Use of Technology (UTAUT), and the Technology-Organization-Environment (TOE) framework. These theories enable a more detailed understanding of the factors that affect AI adoption and the challenges mentioned in the previous section.

1. Diffusion of Innovation Theory

Introduced by Everett Rogers in 1962, the Diffusion of Innovation (DOI) theory provides a foundational framework for understanding the adoption potential of a technology or innovation.[163] The theory seeks to explain the process by which new ideas, practices, or technologies are communicated among potential users within a social system over time. According to the DOI theory, there are five stages of the adoption process:

1. Knowledge (awareness of the innovation)
2. Persuasion (seeking additional information)
3. Decision (consideration of whether to try the innovation)
4. Implementation (use of the innovation)
5. Confirmation (the adoption of the innovation)

Ram and Sheth highlight the importance of persuasion; educating potential adopters can overcome functional and psychological barriers.[164] MacVaugh and

[162] Davit Marikyan and Savvas Papagiannidis, "Unified Theory of Acceptance and Use of Technology: A Review," in *TheoryHub Book*, ed. Savvas Papagiannidis, 2021, 16, https://open.ncl.ac.uk/theories/2/unified-theory-of-acceptance-and-use-of-technology/.

[163] Everett M. Rogers, *Diffusion of Innovations*, 4th ed. (New York, NY: The Free Press, 1995), 5.

[164] Sundaresan Ram and Jagdish N. Sheth, "Consumer Resistance To Innovations: The Marketing Problem A," *The Journal of Consumer Marketing* 6, no. 2 (Spring 1989), https://www.proquest.com/docview/220132547/abstract/725F9F68A45C4184PQ/1.

Schiavone analyze how technological, social, or learning conditions could inhibit or facilitate the adoption process.[165]

Rogers outlines five factors that influence adoption—relative advantage, compatibility, complexity, observability, and trialability—which account for 49–87% of the variance in the rate of innovation adoption.[166] These factors are shaped by the perceptions of the potential adopter and may shift over time. Evaluating AI adoption potential based on these five factors alone will likely be difficult. The hype surrounding the technology complicates the evaluation of relative advantage. AI could help reduce costs, increase efficiency, and provide a competitive advantage, but to attain these benefits, changes must be made to help integrate AI into existing systems and processes. This integration will test the level of compatibility. As discussed previously, the complexity of AI will continue to be a significant challenge for adoption. The degree of observability will vary depending on the type of AI application. For instance, autonomous systems or the production of synthetic media will be more observable than AI-enabled analytic tools. Trialability will also differ by AI system, depending on the level of risk associated with the technology.

Nonetheless, the DOI theory provides a foundational framework to examine innovation adoption in a variety of contexts (e.g., sociology, public health, economics, technology) and levels of analysis (e.g., individual, organizational, and societal).[167] Although the theory is sometimes viewed as too simplistic, the application of the DOI theory does provide a predictive assessment and highlights the likely difficulties in adopting AI. Furthermore, it underscores key concepts that are used to extend both the UTAUT and the TOE framework.

---

[165] Jason MacVaugh and Francesco Schiavone, "Limits to the Diffusion of Innovation: A Literature Review and Integrative Model," *European Journal of Innovation Management* 13, no. 2 (2010): 207, https://doi.org/10.1108/14601061011040258.

[166] Rogers, Diffusion of Innovations, 221.

[167] Joe Tidd, *Gaining Momentum: Managing The Diffusion Of Innovations* (Singapore: World Scientific Publishing Company, 2010), 6–7, http://ebookcentral.proquest.com/lib/ebook-nps/detail.action?docID=731340.

2. Unified Theory of Acceptance and Use of Technology

The Unified Theory of Acceptance and Use of Technology (UTAUT) is derived from eight theories of technology acceptance: the Theory of Reasoned Actions (TRA), the Theory of Planned Behavior (TPB), the Technology Acceptance Model (TAM), the combined TAM and TPB (C-TAM-TPB), the Model of PC Utilization (MPCU), the Motivational Model (MM), the Social Cognitive Theory (SCT) and the DOI theory .[168] As proposed by the TRA, TPB, and TAM, the UTAUT ties "behavioral intention" with the actual manifestation of a behavioral action—the use of the technology.[169] In other words, certain motivational factors (e.g., attitudes, norms) influence the likelihood of an individual using the technology.

The original UTAUT model postulates four main factors—"performance expectancy, effort expectancy, social influence, and facilitating conditions"—as well as four moderators—"age, gender, experience, and voluntariness."[170] Performance expectancy is "the degree to which an individual believes that using the system will help him or her to attain gains in job performance."[171] This concept is rooted in earlier theories discussing perceived usefulness (TAM and C-TAM-TPB), extrinsic motivation (MM), job-fit (MPCU), relative advantage (DOI), and outcome expectancy (SCT). Performance expectancy is a key consideration especially within the military context. When provided with new technology or equipment, service members are concerned with

[168] Viswanath Venkatesh et al., "User Acceptance of Information Technology: Toward a Unified View," *MIS Quarterly* 27, no. 3 (2003): 425–78, https://doi.org/10.2307/30036540.

[169] Viswanath Venkatesh and Fred D. Davis, "A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies," *Management Science* 46, no. 2 (2000): 187, http://www.jstor.org/stable/2634758.

[170] Venkatesh et al., "User Acceptance of Information Technology."

[171] Venkatesh et al., 447.

whether it will enhance their ability to accomplish a mission or a task. Several studies have identified performance expectancy as the main predictor of behavioral intention.[172]

Effort expectancy is also viewed as a strong predictor of technology acceptance—although this factor becomes less significant over extended usage and periods of time.[173] Effort expectancy is often equated to ease of use, or the perceived degree of effort required to utilize the technology. The complexity of the system influences effort expectancy by affecting the user's perception of the relative difficulty in understanding and using the technology.

Social influence relates closely with the concepts of subjective norms, social factors, and image, which are discussed in the TRA, TPB, C-TAM-TPB, MPCU, and DOI. It involves an individual's perception of whether others believe he or she should use the technology. Social influence can be broken down into three distinct cognitive processes: compliance, identification, and internalization.[174] Compliance is when an individual seeks to gain approval (or reward) and avoid disapproval (or punishment) by conforming to perceived expectations and behaviors, regardless of his or her personal beliefs. Identification occurs when an individual adopts a behavior to establish or maintain a desired relationship. Internalization occurs when an individual aligns his or her belief structure with others and views the behavior as "intrinsically rewarding."[175] Studies show that social influence can have a significant effect in determining the

---

[172] Lisa Cornelissen et al., "The Drivers of Acceptance of Artificial Intelligence–Powered Care Pathways Among Medical Professionals: Web-Based Survey Study," *JMIR Formative Research* 6, no. 6 (June 21, 2022): 7, https://doi.org/10.2196/33368; Alaa Momani, "The Unified Theory of Acceptance and Use of Technology: A New Approach in Technology Acceptance," *International Journal of Sociotechnology and Knowledge Development* 12 (July 1, 2020): 87, https://doi.org/10.4018/IJSKD.2020070105.

[173] Venkatesh et al., "User Acceptance of Information Technology," 450.

[174] Lorenz Graf-Vlachy, Katharina Buhtz, and Andreas König, "Social Influence in Technology Adoption: Taking Stock and Moving Forward," *Management Review Quarterly* 68, no. 1 (February 2018): 40–41, https://doi.org/10.1007/s11301-017-0133-3.

[175] Herbert C. Kelman, "Compliance, Identification, and Internalization Three Processes of Attitude Change," *Journal of Conflict Resolution* 2, no. 1 (March 1958): 53, https://doi.org/10.1177/002200275800200106.

perceived usefulness and ease of use.[176] Analyses by Hartwick and Barki as well as Venkatesh and Davis differentiate between voluntary and mandatory use of technology, finding that social influence has a stronger effect in mandatory use settings due to compliance.[177] Posterior literature, however, indicates that the linkages between social influence measures and varying settings are more convoluted.[178]

Facilitating conditions affect an individual's impression of whether existing organizational and technical infrastructure exist to support the usage of the technology.[179] Related constructs include compatibility and perceived behavioral control, derived from the TPB, C-TAM-TPB, MPCU, and DOI. The degree to which resources (i.e., infrastructure and training), guidance, knowledge, and opportunity are available to a potential user can impact self-efficacy and whether a user thinks adequate support exists to adopt the technology. Venkatesh et al. posit that unlike the other three factors, facilitating conditions directly affect use behavior rather than just influencing behavioral intention.[180]

---

[176] Li Qin et al., "The Effects of Social Influence on User Acceptance of Online Social Networks," *International Journal of Human-Computer Interaction* 27, no. 9 (September 2011): 885–99, https://doi.org/10.1080/10447318.2011.555311.

[177] John Hartwick and Henri Barki, "Explaining the Role of User Participation in Information System Use," *Management Science* 40, no. 4 (1994): 440–65, http://www.jstor.org/stable/2632752; Venkatesh and Davis, "A Theoretical Extension of the Technology Acceptance Model."

[178] Graf-Vlachy, Buhtz, and König, "Social Influence in Technology Adoption," 48.

[179] Venkatesh et al., "User Acceptance of Information Technology," 453.

[180] Venkatesh et al., 454.

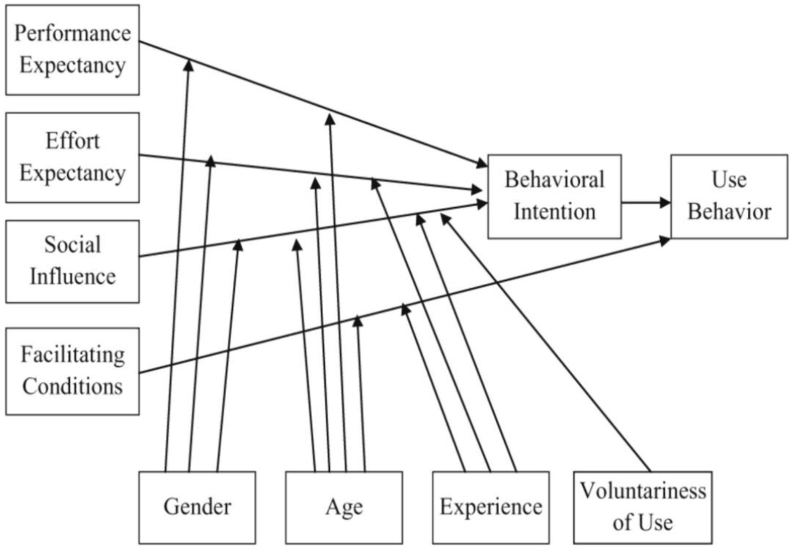


Figure 4. Unified Theory of Acceptance and Use of Technology.[181]

A growing number of studies are utilizing the UTAUT model to investigate factors that underlie behavioral intention and usage of AI technologies. Empirical research upholds major facets of the model, confirming that performance expectancy and effort expectancy are critical antecedents to AI adoption.[182] Individuals are more likely to use AI tools if they feel that the tool enhances their performance, while requiring minimal effort to use and integrate into their workflow. Although several studies omit social influence and facilitating conditions from their analyses, Jain et al.'s empirical study indicates a positive effect between AI adoption and social influence as well as

181 Source: Venkatesh et al., 447.

182 Ruchika Jain, Naval Garg, and Shikha N. Khera, "Adoption of AI-Enabled Tools in Social Development Organizations in India: An Extension of UTAUT Model," *Frontiers in Psychology* 13 (June 20, 2022): 9–10, https://doi.org/10.3389/fpsyg.2022.893691; Oliver Alexander Gansser and Christina Stefanie Reich, "A New Acceptance Model for Artificial Intelligence with Extensions to UTAUT2: An Empirical Study in Three Segments of Application," *Technology in Society* 65 (May 1, 2021): 9–10, https://doi.org/10.1016/j.techsoc.2021.101535; Sheshadri Chatterjee et al., "Assessing Organizational Users' Intentions and Behavior to AI Integrated CRM Systems: A Meta-UTAUT Approach," *Information Systems Frontiers*, August 13, 2021, https://doi.org/10.1007/s10796-021-10181-1.

facilitating conditions.[183] Similarly, Chatterjee et al. confirms that facilitating conditions impact attitudes, intentions, and use-behavior for AI-integrated systems.[184]

Many of the studies have also extended the UTAUT to include additional variables along with the four main factors from the original UTAUT model. In their study of AI applications within the mobility, household, and health sectors, Gansser and Reich test five additional variables to the base model, concluding that health, convenience comfort, and sustainability have "valid predictive power for performance expectancy" with convenience and comfort having the overall greatest influence.[185] Jain et al. consider the role of algorithmic aversion, finding that increased AI aversion influences performance and effort expectancies and thus, decreases the use of the technology. They also extend the model by including perceived collaboration as an outcome variable, demonstrating that AI-enabled tools facilitate increased collaboration and employee engagement.[186] Hasija and Esper examine the unique aspects of AI technologies within supply chain management and expand the model by including AI trustworthiness as a significant consideration in social influence.[187] Bedué and Fritzche opt for a more substantial modification of the UTAUT in order to conduct a detailed examination of trust and the influence of perceived benefits and risks.[188]

---

[183] Jain, Garg, and Khera, "Adoption of AI-Enabled Tools in Social Development Organizations in India," 10.

[184] Chatterjee et al., "Assessing Organizational Users' Intentions and Behavior to AI Integrated CRM Systems."

[185] Gansser and Reich, "A New Acceptance Model for Artificial Intelligence with Extensions to UTAUT2," 9.

[186] Jain, Garg, and Khera, "Adoption of AI-Enabled Tools in Social Development Organizations in India," 10–11.

[187] Abhinav Hasija and Terry L. Esper, "In Artificial Intelligence (AI) We Trust: A Qualitative Investigation of AI Technology Acceptance," *Journal of Business Logistics* 43, no. 3 (February 1, 2022): 402, https://doi.org/10.1111/jbl.12301.

[188] Patrick Bedué and Albrecht Fritzsche, "Can We Trust AI? An Empirical Investigation of Trust Requirements and Guide to Successful AI Adoption," *Journal of Enterprise Information Management* 35, no. 2 (January 1, 2021): 544, https://doi.org/10.1108/JEIM-06-2020-0233.

The prominence of the UTAUT within extant literature indicates the generalizability of the model in explaining the main factors of technology adoption.[189] Yet, as subsequent variations of the model demonstrate, the original UTAUT requires modifications to account for individual factors that may be important in explaining technology acceptance within specific contexts. Furthermore, the UTAUT is concerned with *individual* intention and behavior toward technology usage rather than the organizational adoption of a technology such as AI.[190]

3. Technology-Organization-Environment Framework

The Technology-Organization-Environment (TOE) framework takes an organizational-level perspective in analyzing innovation adoption.[191] Originally developed by Tornatzky and Fleischer in 1990, the framework consists of three elements that influence how organizations adopt innovations.[192] The first is the technological dimension, which considers the availability and characteristics of all technologies relevant to the organization, including ones that are already in use internally and those that exist in the external marketplace. In this context, technology can be equipment, processes, or practices. The second element is the organizational dimension that includes the size, structure, amount of slack, culture, and other characteristics of the organization. Studies indicate that managerial leadership and communication are key factors as well as organizational scale and resource availability.[193] The third element—the environmental dimension— considers the broader conditions of the market, regulatory environment, and

---

[189] Viswanath Venkatesh, James Y. L. Thong, and Xin Xu, "Unified Theory of Acceptance and Use of Technology: A Synthesis and the Road Ahead," *Journal of the Association for Information Systems* 17, no. 5 (May 2016): 332, https://www.proquest.com/docview/1794948207/abstract/112891861CE74109PQ/1.

[190] Ali Al Hadwer et al., "A Systematic Review of Organizational Factors Impacting Cloud-Based Technology Adoption Using Technology-Organization-Environment Framework," *Internet of Things* 15 (September 2021): 2–3, https://doi.org/10.1016/j.iot.2021.100407.

[191] Al Hadwer et al., 2–3.

[192] Louis G. Tornatzky and Mitchell Fleischer, *The Processes of Technological Innovation* (Lexington, MA: Lexington Books, 1990), 152.

[193] Al Hadwer et al., "A Systematic Review of Organizational Factors Impacting Cloud-Based Technology Adoption Using Technology-Organization-Environment Framework," 7–8.

competition as well as external infrastructure such as the availability of skilled labor and supporting industries.

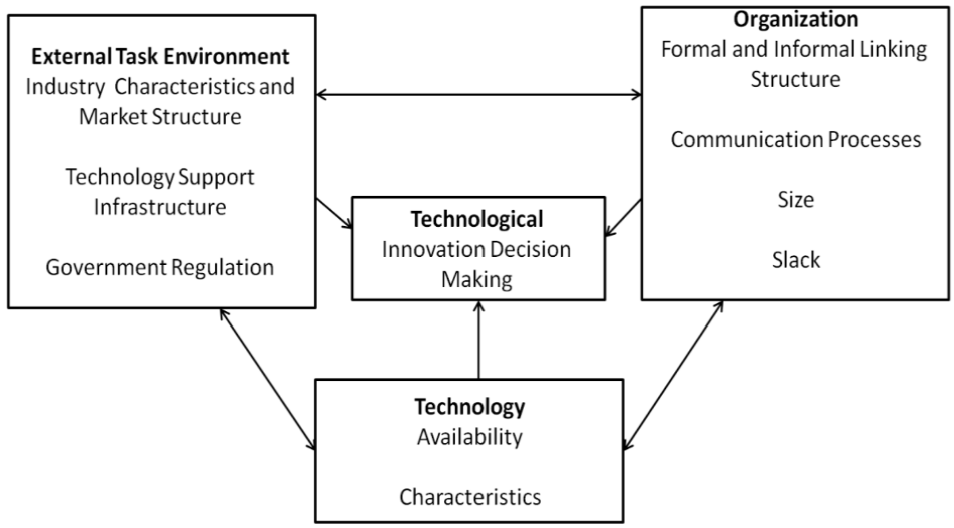


Figure 5. Technology-Organization-Environment Framework.[194]

As with the other theories of innovation adoption, the TOE framework has been widely applied and adapted to a variety of contexts and technologies, ranging from open systems to e-business and AI.[195] AlSheibani et al. and Pumplun et al. combine elements of DOT into the TOE framework to explore AI readiness factors at the firm level. They assess technological readiness based on relative advantage and compatibility of the technology.[196] Being clear about what problem the AI is supposed to solve is an important prerequisite to accurately assessing technological readiness. Pumplun et al. emphasize that AI is not a "panacea" and should be compared to other conventional

[194] Source: Tornatzky and Fleischer, *The Processes of Technological Innovation*, 153.

[195] Tiago Oliveira and Maria Fraga Martins, "Literature Review of Information Technology Adoption Models at Firm Level," *Electronic Journal of Information Systems Evaluation* 14, no. 1 (January 1, 2011): 113–16, https://academic-publishing.org/index.php/ejise/article/view/389.

[196] Sulaiman AlSheibani, Yen Cheung, and Chris Messom, "Artificial Intelligence Adoption: AI-Readiness at Firm-Level," in *Pacific Asia Conference on Information Systems*, vol. 37 (Japan, 2018), https://aisel.aisnet.org/pacis2018/37.

systems according to specific use cases.[197] Furthermore, compatibility will depend on whether the organization is able to adjust their systems and processes to enable integration of the technology.

Technological considerations relate to organizational factors that impact whether an organization is able to successfully adopt AI. Several studies find top management support as one of the most significant factors affecting AI adoption.[198] The level of commitment by the leadership to introduce AI into an organization will directly affect resourcing and the ability to make organizational changes that may be necessary for successful implementation. Yet, effective top management support also requires adequate understanding of the technology, which can be challenging, given the inherent complexity of AI systems.[199] Neumann et al. also consider AI strategy, collaboration, and origin of project initiative as potential factors for AI adoption within the organizational context.[200] Although Jöhnk et al. assert that the existence of an AI strategy influence AI adoption, Neumann et al. found that organizations did not always have strategic documents promoting or regulating AI use.[201] In their study, AI projects were initiated based on technological rather than strategic considerations. Collaboration

[197] Luisa Pumplun, Christoph Tauchert, and Margareta Heidt, "A New Organizational Chassis for Artificial Intelligence-Exploring Organizational Readiness Factors," in *27th European Conference on Information Systems (ECIS)* (Stockholm, Sweden: ResearchGate, 2019), 7, https://aisel.aisnet.org/ecis2019_rp/106.

[198] Al Hadwer et al., "A Systematic Review of Organizational Factors Impacting Cloud-Based Technology Adoption Using Technology-Organization-Environment Framework," 7; Shavneet Sharma et al., "Why Do SMEs Adopt Artificial Intelligence-Based Chatbots?," *IEEE Transactions on Engineering Management*, 2022, 7, https://doi.org/10.1109/TEM.2022.3203469; Tatjana Vasiljeva, Ilmars Kreituss, and Ilze Lulle, "Artificial Intelligence: The Attitude of the Public and Representatives of Various Industries," *Journal of Risk and Financial Management* 14, no. 8 (July 21, 2021): 14, https://doi.org/10.3390/jrfm14080339.

[199] Pumplun, Tauchert, and Heidt, "A New Organizational Chassis for Artificial Intelligence-Exploring Organizational Readiness Factors," 7.

[200] Oliver Neumann, Katharina Guirguis, and Reto Steiner, "Exploring Artificial Intelligence Adoption in Public Organizations: A Comparative Case Study," *Public Management Review*, March 20, 2022, 7, https://doi.org/10.1080/14719037.2022.2048685.

[201] Jöhnk, Weißert, and Wyrtki, "Ready or Not, AI Comes— An Interview Study of Organizational AI Readiness Factors," 11–12; Neumann, Guirguis, and Steiner, "Exploring Artificial Intelligence Adoption in Public Organizations," 15.

was determined as an important factor—particularly if it involved financing from external partners—but not as decisive as the strategic management of AI adoption.[202]

External pressures from industry, government, or society can also affect environmental conditions surrounding AI adoption.[203] Government regulation and guidelines can support or inhibit AI adoption. The 2018 General Data Protection Regulation, which controls the collection and processing of personal data, may make the use of AI "excessively laborious," but it can alleviate public pressure for ethical use of the technology and potentially increase overall trustworthiness.[204] Competition is also an important factor driving organizations to launch AI initiatives. The DOD—unlike the public organizations within Neumann et al.'s study, which did not see themselves in a competitive situation—is facing a highly competitive environment.[205] In addition to the ongoing competition with adversaries, the DOD is in a continual struggle for funding and human capital.

D. APPLYING THE ADOPTION MODELS TO AI FOR OIE

The technology adoption models provide a baseline for understanding crucial factors influencing the use of AI for OIE. The UTAUT elucidates important factors that should be considered when introducing new technology to individual users. Facilitating the adoption of AI at the user level is an important prerequisite to organizational adoption.[206] The TOE provides a framework to understand the conditions that could influence an organization's adoption of AI tools, while the DOI theory underlies both the

---

[202] Neumann, Guirguis, and Steiner, "Exploring Artificial Intelligence Adoption in Public Organizations," 18.

[203] Cindy Schaefer et al., "Truth or Dare? – How Can We Influence the Adoption of Artificial Intelligence in Municipalities?," in *Hawaii International Conference on System Sciences* (Honolulu, HI: University of Hawaiʻi at Mānoa, 2021), 2353–54, https://doi.org/10.24251/HICSS.2021.286.

[204] Neumann, Guirguis, and Steiner, "Exploring Artificial Intelligence Adoption in Public Organizations," 10; Schaefer et al., "Truth or Dare?," 2354.

[205] Neumann, Guirguis, and Steiner, "Exploring Artificial Intelligence Adoption in Public Organizations," 16.

[206] Viswanath Venkatesh, "Adoption and Use of AI Tools: A Research Agenda Grounded in UTAUT," *Annals of Operations Research* 308, no. 1/2 (January 2022): 642, https://doi.org/10.1007/s10479-020-03918-9.

UTAUT and TOE framework. In this section, we examine how these frameworks apply to AI technologies that can impact OIE.

1. Individual-Level Analysis (Applying UTAUT)

Perceptions and attitudes about the characteristics of the technology will affect performance expectancy and effort expectancy. Model performance, reliability, and error rates will impact performance expectancy, but for the information practitioner, the AI's contribution to the overall operation carries more weight. The ability to clearly define KPIs and understand how the technology fits into existing workflows will impact perceived usefulness. Performance expectancy, however, is not only contingent upon whether the AI tool is potentially useful but also whether it provides a relative advantage over other technological solutions and even more importantly, over human labor. Thus, assessing task alignment is a critical component of performance expectancy. If the AI-enabled tool does not seem to provide an articulable benefit, then it is less likely that the user would adopt the tool. User experience (UX) and the level of explainability (i.e., interpretability, understandability) of the AI system are important considerations for effort expectancy. For a practitioner that has limited technical knowledge, UX will likely play an important part in shaping the attitudes of the user.

Although the model separates facilitating conditions as a separate construct from the other three factors, there is much overlap between effort expectancy and facilitating conditions when applying the model in a practical context. Accessibility to the AI tool will impact perceived ease of use. Due to security concerns, certain websites, software, or tools have only been accessible through government networks or domains; or inversely, some tools are not accessible on government systems, given their commercial or open-source features. Restrictions in accessibility, although at times warranted, could inhibit the use of the technology.

Whether users view the AI tool as interoperable with other existing or impending systems will also affect effort expectancy. The ability of a practitioner to seamlessly integrate the tool into his or her workflow depends not only on interpretation and interactions with the AI tool, but also on technical interoperability—the ability to integrate and transfer data or information across systems. New tools will need to be able to not only communicate with legacy systems but also prospective tools and data sources. Companies often specialize in specific AI/ML approaches or obtain data from particular sources. For example, Peakmetrics offers the ability to conduct sentiment analysis of online media data, while Predata focuses more on network traffic metadata (primarily from Wikipedia) to identify emerging trends within the information environment.[207] Each tool provides a piece of the puzzle. It would be quixotic to expect that one company will be able to provide an all-encompassing solution that can last over time. Thus, current and future interoperability considerations should be taken into account.

As previous studies indicate, social influence is a complex factor that has varying effects on technology adoption.[208] Given the culture of the military, compliance could play a greater role in influencing individual adoption of AI. Yet, commanders tend to be more concerned with the ability to achieve an operational objective than directing specific means to conduct operations. There have been cases in which technological tools—albeit not AI—were acquired by OIE units but failed to attain widespread adoption among users despite support from leadership. For example, the Susceptibility and Vulnerability Analysis Network Tool (SAVANT), which was intended to assist with the mapping of population behaviors and various lines of persuasion, was acquired for the PSYOP community but was ultimately not used by individual practitioners.[209]

One possible explanation is that social influence emanating from a network of peers could have a stronger effect than compliance with the leadership's expectations. Another possible explanation is that social influence may be a weaker factor in explaining technology acceptance and adoption within the OIE community. In these

[207] Peakmetrics representatives, personal communication, August 2, 2022; Predata representatives, personal communication, August 18, 2022.

[208] Graf-Vlachy, Buhtz, and König, "Social Influence in Technology Adoption," 66.

[209] John Boiney and David Foster, *Progress and Promise: Research and Engineering for Human Sociocultural Behavior Capability in the U.S. Department of Defense*, 13–2042 (McLean, VA: MITRE Corporation, 2013), 29–30, https://apps.dtic.mil/sti/pdfs/ADA587451.pdf; Information practitioner, personal communication, November 22, 2022.

cases, concerns affecting performance expectancy and effort expectancy along with facilitating conditions could override constructs within social influence.

Although the original UTAUT provides a solid foundation for analysis, several adjustments are needed to tailor the model to the specific context of AI adoption among information practitioners. As advocated by other researchers (e.g., Cornelissen et al., Hasija and Esper), trust is a crucial factor in AI technology acceptance and should be an extension of the UTAUT construct.[210] The original model also proposes gender, age, experience, and voluntary use as its four moderators, but in the context of the military—and more specifically, the OIE community—experience, rank, and time may have a greater effect. Time, in particular, could have a defining role in how the user perceives the utility and ease of use of the AI tool. The OIE community faces high operational demand but limited supporting personnel. A decrease in time available could affect the degree to which the user thinks he or she requires additional organizational or technical support (i.e., facilitating conditions).

2. Organizational-Level Analysis (Applying the TOE Framework)

Along with individual factors of technological acceptance, organizational factors should be considered when assessing the adoption potential of a technology. Even if a group of individuals become strong proponents of an AI tool, crossing the "chasm" from early adopters to more mainstream adoption within a community can be a challenge if factors such those discussed within the TOE framework (e.g., top management support, regulatory restrictions) are not assessed.[211]

a. Technological Dimension

For OIE units seeking to integrate AI technology, relative advantage, compatibility, and complexity will remain key factors in assessing the viability of AI

[210] Cornelissen et al., "The Drivers of Acceptance of Artificial Intelligence–Powered Care Pathways Among Medical Professionals"; Hasija and Esper, "In Artificial Intelligence (AI) We Trust."

[211] Geoffrey A. Moore, Crossing the Chasm: Marketing and Selling Technology Project (HarperCollins e-Books, 2014).

tools. Yet, measuring these factors is not a straightforward process; they each consist of multiple sub-components and rely on a significant degree of subjectivity. They also involve dynamic assessments, given the constant evolution of technologies. While there is broad recognition of the aspirational benefits of AI, there is still significant uncertainty regarding the value and direct impact of currently existing AI technologies for OIE. This uncertainty is perpetuated by the limited knowledge of the technology, which often makes determinations of relative advantage less credible. For compatibility and complexity, the main concern is whether the AI tool can fit into existing organizational processes and culture, raising the question of how adaptable the organization is. Given the complexity of AI and the indurate nature of military organizations, making the necessary adjustments to integrate AI technology is far from guaranteed.

The cost of the technology is another important consideration for OIE units who face limited (and inflexible) budgets. Although AI is often touted as a cost-reducing solution, that is not necessarily the case within every context. Although leveraging AI tools to detect narratives or sentiment of key audiences may enhance the ability of OIE units to assess the information environment, it is not a necessarily a cost-saving measure. The alternative would have been anecdotal analysis of the information environment conducted by existing personnel. Management of personnel within the military differs from the commercial sector where organizations have greater flexibility in hiring, dismissing, or diverting personnel. The continual monitoring and maintenance of AI systems will require dedicated personnel (most likely contracted) and maintaining access to required data sources will likely incur substantial costs.

b. Organizational Dimension

As discussed in other TOE studies, leadership support will be an important component in enacting necessary changes to integrate new technology. Commanders will play a key role in ensuring that objectives and KPIs are clearly defined, that supporting (internal and external) infrastructure is established and maintained, and that enough slack is given to enable service members to receive necessary training as well as experiment with the new capability. Information practitioners, however, face an additional complication. The upper echelon of their chain of command (i.e., top management) consists of commanders from different communities. This distinction does not necessarily mean that top leadership will be unsupportive of integrating the new technology for OIE. In fact, it is possible that they may be more committed to the initiative. The complication arises from differences in culture, which generates disparate expectations. Therefore, while top leadership remains critical to driving innovation adoption, "middle management" (i.e., field grades) will also have significant sway in whether (and how) AI is successfully integrated into the organization.

Apart from leadership, a variety of other organizational factors could influence AI adoption within an OIE unit. Alignment in doctrine and the ability to take advantage of training and education opportunities should be viewed as essential conditions for AI integration. Doctrine offers guiding principles that enable "coordinated and integrated action toward a common objective."[212] This provides a starting point for practitioners to develop their workflow—whether through the joint planning process, the targeting cycle, or the seven step PSYOP process. Therefore, AI usage should be compatible with doctrine. Suitable infrastructure, which includes physical facilitates and appropriate architecture to enable data flow, is needed to ensure that the AI tool is accessible to information practitioners. Supporting personnel—experts provided by the vendor or contracted in-house—are critical to ensuring the sustainability and reliability of the AI system. These experts also play an important role in assisting information practitioners with interpreting the outputs of the AI tool.

c. Environmental Dimension

Sensitives surrounding influence operations, data protection, and privacy have led to greater scrutiny and oversight of OIE—more specifically, MISO. Authorities and permissions have been age-old issues for the OIE community. Traditional risk-aversion has generated laborious processes that inhibit the ability of information practitioners to react flexibly and quickly to emerging situations. Despite growing recognition of the

[212] Joint Chiefs of Staff, "Joint Doctrine Publications," accessed October 10, 2022, https://www.jcs.mil/Doctrine/Joint-Doctine-Pubs/.

criticality and the dynamic nature of the information environment, public attitudes remain apprehensive toward influence operations. The use of AI technology for OIE could cause further complications and increase public pressure. Potential fallout could occur from aggressively utilizing technology for OIE. As an example, in August 2022, Graphika and the Stanford Cyber Policy Center reported on pro-Western covert influence operations involving the use of fake social media accounts attributed to the U.S. military.[213] The report led to public outcry and an audit of PSYOP online activities.[214] Given these sensitivities, many of the policy restrictions enacted by Congress and the DOD leadership will likely remain, and while these policies may be necessary, in their current form, they are inhibitors to AI adoption—particularly for applications such as social bots, synthetic media, and microtargeting.

E. CONCLUSION

Decades of research illustrate the challenges of technology adoption and clearly suggest that adopting AI into OIE units is far from easy. Understanding the factors that influence technology acceptance and adoption is critical when seeking to integrate AI into OIE. The UTAUT highlights four individual-level factors—performance expectancy, effort expectancy, facilitating conditions, and social influence—that affect technology acceptance and should be considered when evaluating AI. Additionally, trust and time available to the user can significantly influence a user's perception of the technology, which can affect the level of adoption within OIE units. The TOE framework identifies key organizational-level factors such as organizational readiness (e.g., leadership support, training, doctrine) and the impact of policy and public perception. Both the UTAUT and TOE notably overlap with the DOI's five factors (relative advantage, compatibility, complexity, trialability, and observability), but it is relative advantage, compatibility, and

---

213 Graphika, Stanford Internet Observatory, *Unheard Voice: Evaluating Five Years of pro-Western Covert Influence Operations* (Stanford, CA: Stanford Cyber Policy Center, 2022).

214 Ellen Nakashima, "Pentagon Opens Sweeping Review of Clandestine Psychological Operations," *The Washington Post*, September 19, 2022, https://www.washingtonpost.com/national-security/2022/09/19/pentagon-psychological-operations-facebook-twitter/.

complexity that play a prominent role and are especially pertinent to the evaluation of AI technology.

The theories covered in this chapter underscore key concepts that organizations need to account for when assessing whether the technology under evaluation is one that aligns with existing structure, processes, and culture as well as whether it could be easily adopted by individual users. Furthermore, these theories assist with an organizational assessment of its "AI readiness" and determine what changes are required to successfully integrate the technology.

THIS PAGE INTENTIONALLY LEFT BLANK

V. EXISTING FRAMEWORKS TO EVALUATE AI

A. CHAPTER OVERVIEW

To address the high complexity and the growing ubiquity of AI systems, AI researchers and developers have offered a number of different proposals to enhance the transparency of AI models, systems, and data. Despite the immense diversity in AI models and applications, there is general recognition that some level of standardization is required to explain, document, and evaluate the technology. This chapter highlights five existing frameworks used to facilitate greater transparency of AI: the Defense Innovation Unit's (DIU) *Responsible AI Guidelines*, model cards, datasheets, FactSheets, and System Cards. Analyses of these frameworks reveal key areas to consider for evaluations and enable the adaptation of existing concepts to the needs of the information practitioner.

B. THE DEFENSE INNOVATION UNIT'S *RESPONSIBLE AI GUIDELINES*

Among the various HMT and responsible AI frameworks being developed, the DIU's *Responsible AI Guidelines* have been viewed as the "de facto standard" for establishing accountability in the development and adoption of AI systems into the DOD.[215] The guidelines provide practical steps to operationalize the *DOD Ethical Principles for Artificial Intelligence* throughout the planning, development, and deployment phases of the AI life cycle. Workflows, supplemented with more detailed worksheets, outline key lines of inquiry that should be considered within each phase (Figure 6).[216] The provided worksheets help drive conversations to ensure the deployment of an effective AI system that adheres to the RAI principles. The Phase I:

[215] Andy Ilachinski and David Broyles, "Face/Off," November 19, 2021, in *AI with AI*, produced by CNA, podcast, MP3 audio, 40:13, https://www.cna.org/our-media/podcasts/ai-with-ai/season-5/5-3.

[216] Dunnmon et al., Responsible AI Guidelines in Practice, 7.

Planning Worksheet involves critical input from the DOD stakeholder and is especially pertinent to practitioners.[217]

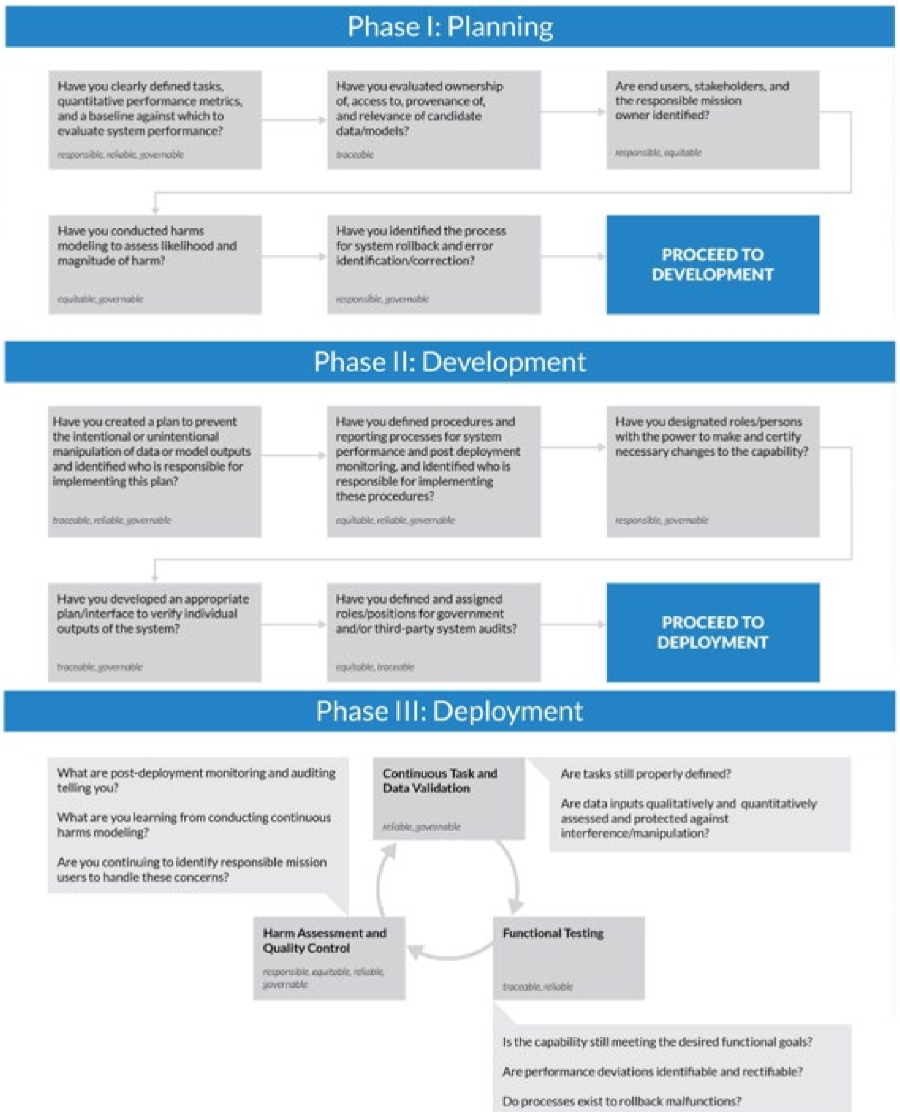


Figure 6. The Defense Innovation Unit *Responsible AI Guidelines*.[218]

[217] Jared Dunnmon et al., "Phase I: Planning Worksheet for DIU AI Guidelines" (Defense Innovation Unit, November 14, 2021), https://assets.ctfassets.net/3nanhbfkr0pc/1vJvimVkijueLbJzqcaOMr/39691b01cc00ca98a295804e269d6f51/Planning_Worksheet_DIU_AIGuidelines.pdf.

[218] Source: Dunnmon et al., Responsible AI Guidelines in Practice.

While the DIU's guidelines provide a useful tool for increasing collaboration and understanding among various stakeholders, the process and listed questions are constructed in a manner that assumes technical expertise. Yet, although end users are continually considered throughout the entire process, the primary implementers of these guidelines are AI companies, program managers, and upper-level DOD stakeholders who can comprehend the technical elements and have access to requisite expertise. The guidelines aim to elicit enough detail to concretely determine whether the AI system meets rigorous vetting standards.[219] Given the technical nature of many of the questions—such as detailed inquiries about root access or error modes—general users without technical knowledge will likely dismiss or fail to comprehend the questions in the worksheets. Thus, end users will likely require a modification of these guidelines to adapt to their needs.

C. MODEL CARDS

The increased emphasis on transparency has led to the recognition that standardized documentation is needed to explain highly complex AI models. Google's 2019 seminal whitepaper, "Model Cards for Model Reporting," serves as a template for the adoption of model cards to communicate important information such as intended use and model performance to a wide audience.[220] Model cards "aim to standardize ethical practice and reporting—allowing stakeholders to compare candidate models for deployment across not only traditional evaluation metrics but also along the axes of ethical, inclusive, and fair considerations."[221] Model cards enable users to discern the capabilities and limitations of the AI system through a digestible format.

As depicted in Figure 7, the original proposal consists of nine categories within the model card: model details, intended use, factors, metrics, evaluation data, training

219 Dunnmon et al., 7.

220 Margaret Mitchell et al., "Model Cards for Model Reporting," in *Proceedings of the Conference on Fairness, Accountability, and Transparency* (FAT* '19: Conference on Fairness, Accountability, and Transparency, Atlanta GA: ACM, 2019), 1, https://doi.org/10.1145/3287560.3287596.

221 Mitchell et al., 2.

data, quantitative analysis, ethical considerations, and caveats and recommendations. Open-source access to Google's Tensorflow Model Card Toolkit has allowed other developers to autogenerate model cards by leveraging the JSON schema to integrate into the ML pipeline and automatically populate relevant information into the model card fields.[222]

While Google's template provides a good baseline for the creation of model cards, actual content within the model cards can vary based on the type of model and use case.[223] This thesis conducted an analysis of 50 publicly available model cards (listed in Appendix B). Although most of the categories within Mitchell et al.'s template are reflected in the majority of the analyzed model cards, there are notable variations in the level of detail, size, and presentation of information among the different model cards—even among the ones that were developed by the same company.

In conducting a qualitative analysis of the 50 model cards, this thesis took an inductive approach to coding, initially using an in vivo coding system to assign labels to major sections, themes, and key points within the model cards. The initial batch of codes were then compared to the categories outlined in Mitchell et al.'s template. Some of the codes were organized into sub-categories, while others required their own separate category. Although Mitchell et al. designate "factors," "metrics," "evaluation data," "training data," and "quantitative analyses" as separate sections of the model card, the distinction was not always clear within majority of the analyzed model cards. Furthermore, none of the model cards had a standalone section for "quantitative analyses"; instead, related information would fall under "metrics" or "performance."

---

[222] "Model Card Toolkit | Responsible AI Toolkit," TensorFlow, February 9, 2021, https://www.tensorflow.org/responsible_ai/model_card_toolkit/guide.

[223] Iain Barclay et al., "A Framework for Fostering Transparency in Shared Artificial Intelligence Models by Increasing Visibility of Contributions," *Concurrency and Computation: Practice and Experience* 33, no. 19 (October 10, 2021): 4, https://doi.org/10.1002/cpe.6129.

Model Card

- **Model Details.** Basic information about the model.
  - Person or organization developing model
  - Model date
  - Model version
  - Model type
  - Information about training algorithms, parameters, fairness constraints or other applied approaches, and features
  - Paper or other resource for more information
  - Citation details
  - License
  - Where to send questions or comments about the model
- **Intended Use.** Use cases that were envisioned during development.
  - Primary intended uses
  - Primary intended users
  - Out-of-scope use cases
- **Factors.** Factors could include demographic or phenotypic groups, environmental conditions, technical attributes, or others listed in Section 4.3.
  - Relevant factors
  - Evaluation factors
- **Metrics.** Metrics should be chosen to reflect potential real-world impacts of the model.
  - Model performance measures
  - Decision thresholds
  - Variation approaches
- **Evaluation Data.** Details on the dataset(s) used for the quantitative analyses in the card.
  - Datasets
  - Motivation
  - Preprocessing
- **Training Data.** May not be possible to provide in practice. When possible, this section should mirror Evaluation Data. If such detail is not possible, minimal allowable information should be provided here, such as details of the distribution over various factors in the training datasets.
- **Quantitative Analyses**
  - Unitary results
  - Intersectional results
- **Ethical Considerations**
- **Caveats and Recommendations**

Figure 7. Model Card Template.[224]

Figure 8 depicts codes that were used within the model code analysis. To facilitate visual comprehension, not all sub-codes are displayed within Figure 8. The full list of codes is provided in Appendix B. As the graph depicts, most model cards include a description of intended use, limitations, training data, performance metrics, and ethical considerations.

[224] Source: Mitchell et al., "Model Cards for Model Reporting."

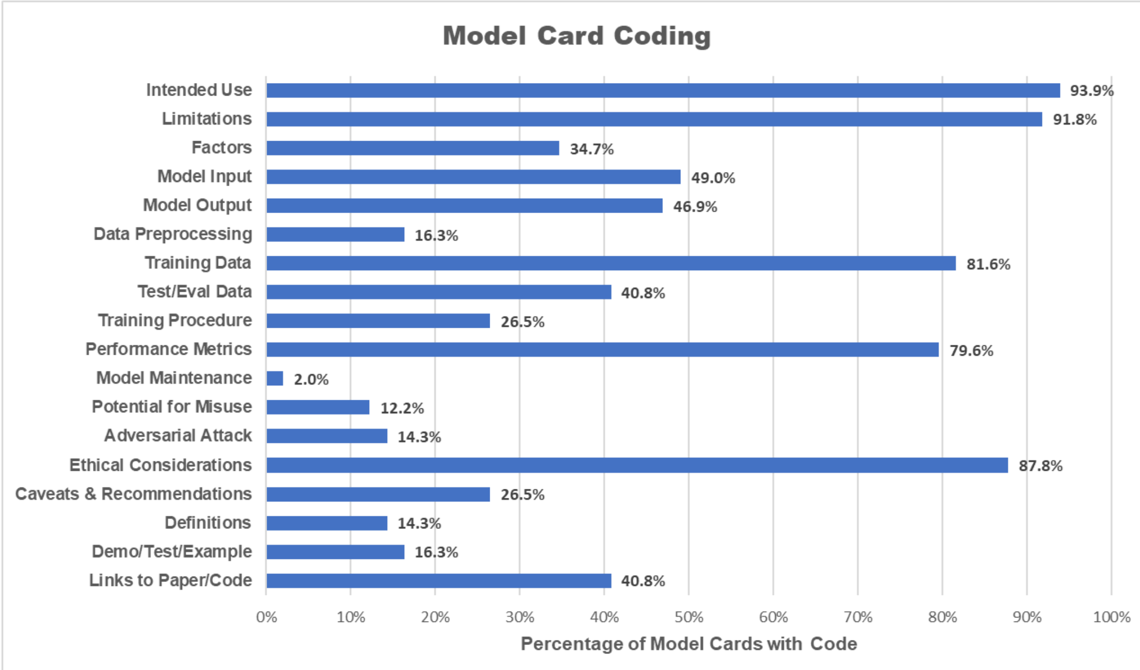


Figure 8. Model Card Themes

A comparative analysis was conducted to assess how similar the model cards are across the 15 different model card creators in this study. Given that multiple model cards originated from the same producers (e.g., Google, Salesforce, OpenAI, etc.), this thesis opted to select one model card from each producer to generate the comparative analysis. One could assume that companies would produce relatively identical model cards, but this was found not necessarily to be the case. Variations did occur based on type of model or the specific author of the model card. Therefore, a similarity analysis was conducted within each group of producers to determine which model card would be most representative of the group. Table 1 provides an example of a similarity analysis conducted on model cards produced by Salesforce. Appendix B provides further details on how calculations were made to determine the degree of similarity.

Table 1. Similarity Analysis of Salesforce Model Cards

Salesforce Model Cards	Total	Detect Sent.	Detector for CTRL	Einstein Engage. Freq.	Einstein Engage. Scor. for Mobile	Einstein Messag. Insights	Einstein OCR	Sim. Card: Found.	CTRL
Einstein Messaging Insights	5.05	0.56	0.52	0.70	0.90	1.00	0.48	0.39	0.50
Einstein Engagement Scoring for Mobile	4.99	0.56	0.52	0.70	1.00	0.90	0.48	0.33	0.50
CTRL	4.77	0.67	0.63	0.50	0.50	0.50	0.57	0.40	1.00
Detect Sentiment	4.76	1.00	0.42	0.50	0.56	0.56	0.64	0.41	0.67
Einstein OCR	4.55	0.64	0.52	0.41	0.48	0.48	1.00	0.45	0.57
Einstein Engagement Frequency	4.50	0.50	0.45	1.00	0.70	0.70	0.41	0.24	0.50
Detector for CTRL	4.41	0.42	1.00	0.45	0.52	0.52	0.52	0.35	0.63
Simulation Card: Foundation	3.57	0.41	0.35	0.24	0.33	0.39	0.45	1.00	0.40

This example indicates that the Einstein Messaging Insight model card had the closest comparative alignment to the seven other model cards produced by Salesforce. This analysis to determine which model card would be the most representative of each group was repeated across the different model card producers. Similarity analysis was then conducted among the selected model cards and positioned on a document map utilizing a distance matrix (Figure 9). The documents were then assigned to a cluster group based off their calculated distances with respect to the coding system.

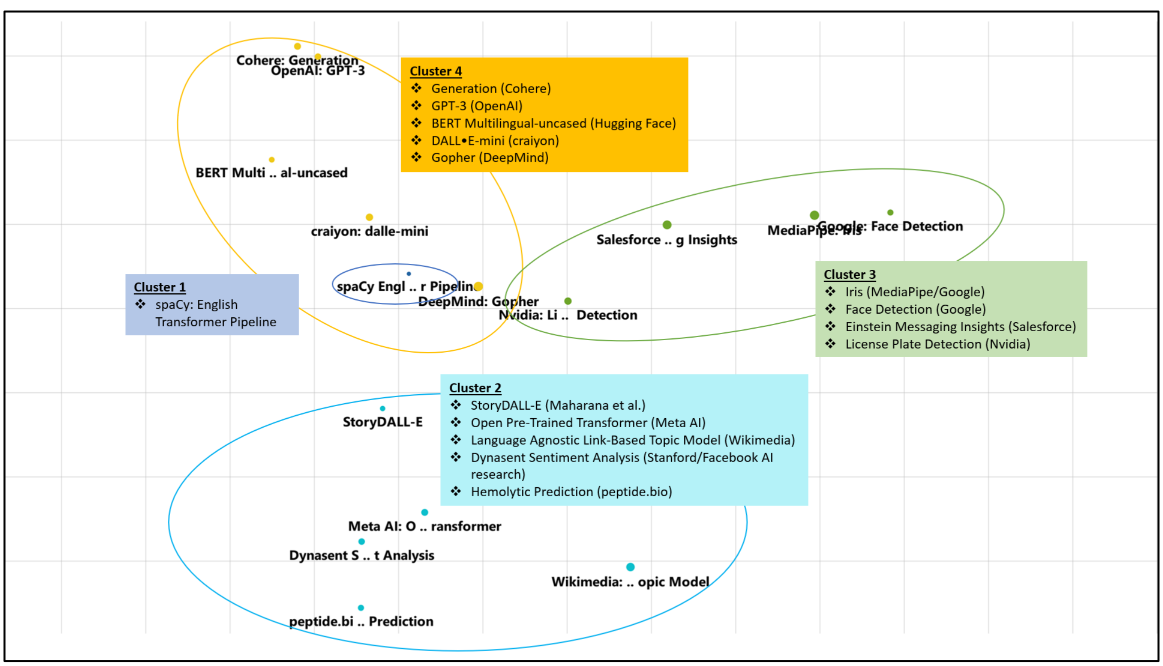


Figure 9. Similarity Analysis Across Various Model Card Producers

Although a small sample, the clusters of documents highlight similarities and differences in key topics covered within their respective model cards. One immediately notices that spaCy's English Transformer Pipeline model card stands alone from the other documents. This distinction is due to its condensed format, which includes only training data and performance metrics. In contrast, model cards within the other cluster groups cover a larger number of topics. All documents within Clusters 2 and 4 provide a description of the model's intended use and information on its datasets, but Cluster 2 distinguishes between training and test (or evaluation) data—unlike most documents within the other clusters. Cluster 2 also provides a distinct section covering caveats and recommendations as well as access to a demo or test example. While the model cards within Cluster 2 reference ethical considerations, details within this section are noticeably absent—with no specific reference to bias, fairness, or risk. On the other hand, fairness is prominently featured within Cluster 3 model cards. Cluster 3 also provides information on model input and output as well as more explicit inclusion of relevant factors such as group or environmental factors that may affect model performance. Limitations are also described in greater detail with most model cards describing out-of-use cases and tradeoffs. In Cluster 4, all model cards contain a more robust discussion about bias. Furthermore, unlike the other clusters, Cluster 4 includes documents that reference safety, potential for misuse, and adversarial attack.

The takeaway from this comparative analysis is that even though all 50 model cards reference Mitchell et al.'s template as a source document, they vary in content, presentation, and level of detail. Yet, the model cards do adopt several overarching elements. Descriptions of intended use, limitations, training data, and performance metrics exist across the different model cards. Furthermore, while discussions of risk, bias, toxicity, and safety varied in specificity, the vast majority of the documents referenced ethical considerations in some capacity.

Another interesting observation is the limited discussion of potential adversarial attacks or the robustness of the model. Of the 50 analyzed model cards, only four accounted for manipulation by an adversary. Another topic that was largely absent was the necessity of maintaining the model over time. There are several possible explanations that could account for this general absence. First, the model cards are provided in an open forum and most seem to be tailored toward other AI/ML developers or data scientists; thus, there could be an assumption that the adopter of the model will determine the details of how the model will be maintained. Second, some of the models are still undergoing development and discussions of model maintenance over time may be premature. Finally, adversarial manipulation and sustainability of the model may be viewed as outside the scope of a model card, which is primarily seeking to explain how the model was developed and its suitability for a particular task. Nonetheless, discussions about potential adversarial attacks and model maintenance are critical, especially within the defense or military context.

Although this thesis would have benefitted from obtaining a larger corpus of model cards to conducts its analysis, not all model cards are publicly available; some are only provided upon formal request or viewed as proprietary. Furthermore, the model card is a relatively new framework (proposed in 2019) and thus, its adoption is nascent. Nevertheless, model cards offer one potential framework to increase transparency and explainability of AI tools. The model card could serve as a baseline for discussion during the evaluation process of an AI tool. Questions should be raised regarding limitations, performance metrics, ethical considerations, and other key points identified by model cards—especially if a model card is not provided by the AI developer.

Nonetheless, despite their utility in assessing AI models, model cards may contain information that is too narrow in scope or too technically focused for a general user to understand. Additionally, practitioners are often presented with AI-enabled tools that consist of a combination of different models. While model cards may facilitate the comparison of different AI models, they provide only one reference point for determining the utility of the AI system in achieving operational requirements. Given the limitations of model cards, others have proposed alternative documentation techniques to explain AI/ML models. The next sections briefly discuss two other proposals: FactSheets created by IBM and System Cards created by Meta AI.

D. FACTSHEETS

Rather than focusing on individual AI models for documentation, members of the IBM Research team propose the creation of FactSheets to convey information about the overall AI system, which may include several different AI models. In their proposal, Arnold et al. take the view of AI as a "service," emphasizing the "functional perspective" of the output or application that is accessed by the consumer of the AI service.[225] FactSheets facilitate greater transparency by enabling consumers to obtain information about the class of algorithms used as well as an explanation of what algorithmic decisions led to a particular output.[226]

The IBM Research team argues that their proposal is distinct from model cards in that it takes a more general approach by addressing several additional concerns. First, developers do not always interface with all components of the AI system, given that many AI applications retrieved from an API consist of multiple models and datasets. Second, the level of expertise varies throughout the AI life cycle and thus, documentation is needed to capture the facts along its various points of development. Third, the use of "trusted models" does not necessarily translate to trustworthiness and transparency of the AI service as a whole. Therefore, assessments in overall performance and safety are required.[227]

IBM Research maintains a repository of resources and example FactSheets along with an explanation on the methodology of creating one.[228] Recognizing that the format and information presented on a FactSheet will depend on context, the IBM site offers three example formats—slide, tabular, and full report—to adjust content based on the intended audience. Closer examination of these formats reveals that there are

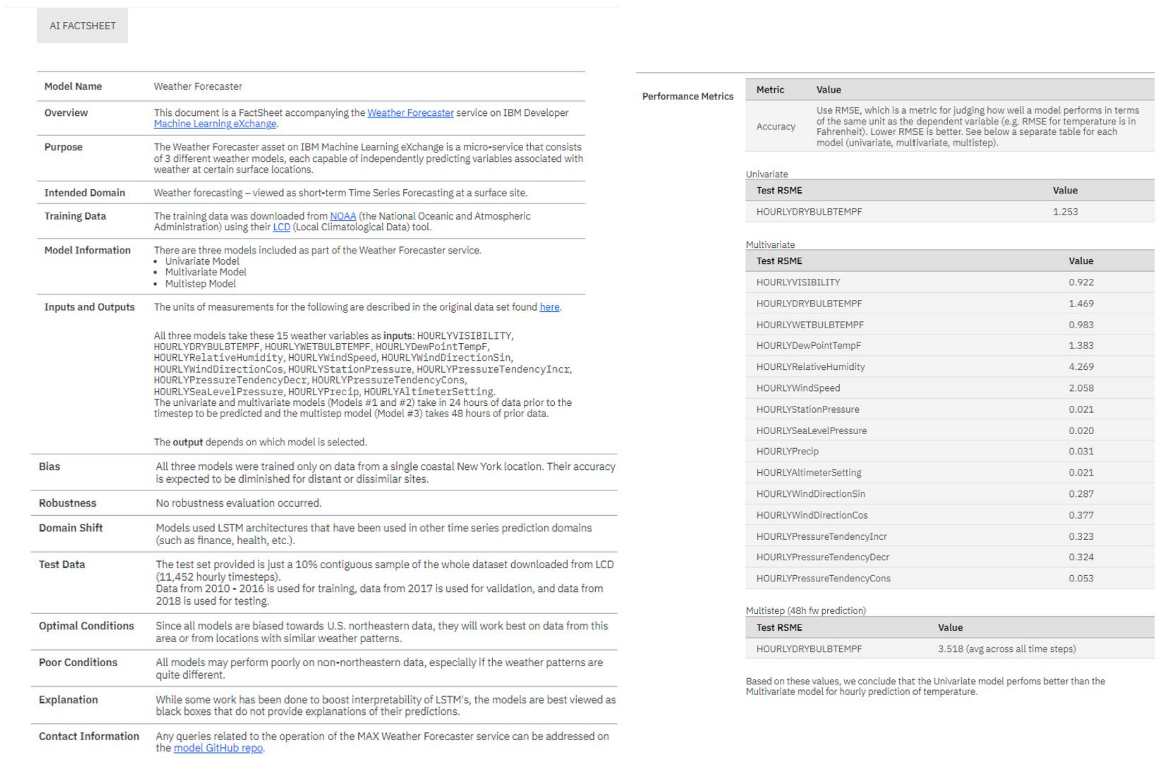
[225] M. Arnold et al., "FactSheets: Increasing Trust in AI Services through Supplier's Declarations of Conformity," *IBM Journal of Research and Development* 63, no. 4/5 (July 2019): 1–2, https://doi.org/10.1147/JRD.2019.2942288.

[226] Arnold et al., 7.

[227] John Richards et al., "A Human-Centered Methodology for Creating AI FactSheets," *Bulletin of the Technical Committee on Data Engineering*, December 2021, 47–48.

[228] IBM Research, "AI FactSheets 360," IBM Research AI FactSheets 360, accessed October 2, 2022, https://aifs360.mybluemix.net/.

considerable similarities between FactSheets and model cards. As with model cards, FactSheets provide information on model details, training data, performance metrics, relevant factors (i.e., conditions), and ethical considerations such as risk and bias. The difference is that FactSheets are organized by their intended function (e.g., text sentiment classifier, object detector, image caption generator). If the AI system includes multiple models—like the one in Figure 11, model information and performance metrics would be provided for all models.



AI FACTSHEET

Model Name	Weather Forecaster
Overview	This document is a FactSheet accompanying the Weather Forecaster service on IBM Developer Machine Learning eXchange.
Purpose	The Weather Forecaster asset on IBM Machine Learning eXchange is a micro-service that consists of 3 different weather models, each capable of independently predicting variables associated with weather at certain surface locations.
Intended Domain	Weather forecasting – viewed as short-term Time Series Forecasting at a surface site.
Training Data	The training data was downloaded from NOAA (the National Oceanic and Atmospheric Administration) using their LCD (Local Climatological Data) tool.
Model Information	There are three models included as part of the Weather Forecaster service. • Univariate Model • Multivariate Model • Multistep Model
Inputs and Outputs	The units of measurements for the following are described in the original data set found here. All three models take these 15 weather variables as **inputs**: `HOURLYVISIBILITY`, `HOURLYDRYBULBTEMPF`, `HOURLYWETBULBTEMPF`, `HOURLYDewPointTempF`, `HOURLYRelativeHumidity`, `HOURLYWindSpeed`, `HOURLYWindDirectionSin`, `HOURLYWindDirectionCos`, `HOURLYStationPressure`, `HOURLYPressureTendencyIncr`, `HOURLYPressureTendencyDecr`, `HOURLYPressureTendencyCons`, `HOURLYSeaLevelPressure`, `HOURLYPrecip`, `HOURLYAltimeterSetting`. The univariate and multivariate models (Models #1 and #2) take in 24 hours of data prior to the timestep to be predicted and the multistep model (Model #3) takes 48 hours of prior data. The **output** depends on which model is selected.
Bias	All three models were trained only on data from a single coastal New York location. Their accuracy is expected to be diminished for distant or dissimilar sites.
Robustness	No robustness evaluation occurred.
Domain Shift	Models used LSTM architectures that have been used in other time series prediction domains (such as finance, health, etc.).
Test Data	The test set provided is just a 10% contiguous sample of the whole dataset downloaded from LCD (11,452 hourly timesteps). Data from 2010 - 2016 is used for training, data from 2017 is used for validation, and data from 2018 is used for testing.
Optimal Conditions	Since all models are biased towards U.S. northeastern data, they will work best on data from this area or from locations with similar weather patterns.
Poor Conditions	All models may perform poorly on non-northeastern data, especially if the weather patterns are quite different.
Explanation	While some work has been done to boost interpretability of LSTM's, the models are best viewed as black boxes that do not provide explanations of their predictions.
Contact Information	Any queries related to the operation of the MAX Weather Forecaster service can be addressed on the model GitHub repo.

Performance Metrics

Metric	Value
Accuracy	Use RMSE, which is a metric for judging how well a model performs in terms of the same unit as the dependent variable (e.g. RMSE for temperature is in Fahrenheit). Lower RMSE is better. See below a separate table for each model (univariate, multivariate, multistep).

Univariate

Test RSME	Value
HOURLYDRYBULBTEMPF	1.253

Multivariate

Test RSME	Value
HOURLYVISIBILITY	0.922
HOURLYDRYBULBTEMPF	1.469
HOURLYWETBULBTEMPF	0.983
HOURLYDewPointTempF	1.383
HOURLYRelativeHumidity	4.269
HOURLYWindSpeed	2.058
HOURLYStationPressure	0.021
HOURLYSeaLevelPressure	0.020
HOURLYPrecip	0.031
HOURLYAltimeterSetting	0.021
HOURLYWindDirectionSin	0.287
HOURLYWindDirectionCos	0.377
HOURLYPressureTendencyIncr	0.323
HOURLYPressureTendencyDecr	0.324
HOURLYPressureTendencyCons	0.053

Multistep (48h fw prediction)

Test RSME	Value
HOURLYDRYBULBTEMPF	3.518 (avg across all time steps)

Based on these values, we conclude that the Univariate model perfoms better than the Multivariate model for hourly prediction of temperature.

Figure 10. Example FactSheet by IBM Research.[229]

Nonetheless, the differences between model cards and FactSheets seem relatively marginal. The distinction rests primarily on perspective—whether one is trying to convey information about a single model or trying to take a more holistic perspective in including

229 Source: IBM Research, "Weather Forecaster FactSheet," IBM Research AI FactSheets 360, accessed September 29, 2022, https://aifs360.mybluemix.net/examples/max_weather_forecaster.

multiple models for a particular function. The broader perspective is undoubtedly useful, especially for an end user who is more interested in the AI's impact on a mission or task. FactSheets would require further refinement based on the consumer of the information. Studies by Hind et al. and Piorkowski et al. involving various stakeholder groups of an AI service confirm the significant diversity in how and what information is needed to meet the requirements of various consumers.[230]

E. SYSTEM CARDS

In February 2022, Meta AI introduced another form of documentation called "System Cards."[231] Building upon the work done on model cards and FactSheets, Meta AI aims to deliver an "integrated solution" that provides "both an overview and detailed information" about an AI/ML-based system so that expert and non-expert stakeholders can gain a more in-depth understanding of how the system operates.[232] The rationale behind the creation of System Cards is essentially the same as IBM's FactSheets. The intent is to provide a more comprehensive explanation of the entire AI system to account for the complexity of multiple models as well as the inclusion of non-AI components within the system.[233]

Meta attempts to distinguish System Cards from traditional AI documentation through the prominence of its interactive component and the ability to "walkthrough with an example input."[234] While Google's Face Detection and Object Detection model cards offer details on performance metrics alongside the ability for consumers to upload their

[230] Michael Hind et al., "Experiences with Improving the Transparency of AI Models and Services," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20 (New York, NY: Association for Computing Machinery, 2020), 1–8, https://doi.org/10.1145/3334480.3383051; David Piorkowski, John Richards, and Michael Hind, "Evaluating a Methodology for Increasing AI Transparency: A Case Study," *ArXiv*, January 24, 2022, http://arxiv.org/abs/2201.13224.

[231] Nekesha Green et al., "System Cards, a New Resource for Understanding How AI Systems Work," Meta AI, February 23, 2022, https://ai.facebook.com/blog/system-cards-a-new-resource-for-understanding-how-ai-systems-work/.

[232] Chavez Procope et al., *System-Level Transparency of Machine Learning* (Menlo Park, CA: Meta AI, 2022), 2, https://ai.facebook.com/research/publications/system-level-transparency-of-machine-learning.

[233] Green et al., "System Cards, a New Resource for Understanding How AI Systems Work."

[234] Procope et al., System-Level Transparency of Machine Learning, 3.

own images to test the model, Meta avoids technical jargon and instead provides a step-by-step demonstration of the system's process so it could be understood by a layperson. Although Meta's System Cards project remains in the pilot phase, its Instagram Feed Ranking System Card is available in the public domain.[235] The card utilizes animated and unified modeling language diagrams to explain the components and process of its ranking system. Given the exclusion of model details or performance metrics, the Instagram System Card appears to be tailored toward non-technical users of the service who are curious about how the feed ranking system works.

The current version of the System Card provides an easy-to-consume visual representation of the system architecture, but the lack of detail makes it difficult to discern the model's performance, limitations, and other key considerations such as security. In other words, the System Card offers a high-level explanation that may be useful for a user to understand the basic process of how the system works, but it provides less utility in evaluating and comparing the system and its underlying models. Nonetheless, the value of the System Card's interactive component should not be overlooked. As demonstrated in Chiang and Yin's experimental study on machine learning literacy among laypeople, interactivity could improve user understanding of model performance and limit overreliance on the model.[236]

F. DATASHEETS

Datasheets are another important form of documentation that closely complement the frameworks mentioned previously. Given the criticality of data for ML, particular attention is needed to understand the source, structure, and appropriateness of the data

[235] Meta AI, "Instagram Feed Ranking System Card," February 23, 2022, https://ai.facebook.com/tools/system-cards/instagram-feed-ranking.

[236] Chun-Wei Chiang and Ming Yin, "Exploring the Effects of Machine Learning Literacy Interventions on Laypeople's Reliance on Machine Learning Models," in *27th International Conference on Intelligent User Interfaces* (IUI '22: 27th International Conference on Intelligent User Interfaces, Helsinki Finland: ACM, 2022), 148–61, https://doi.org/10.1145/3490099.3511121.

being fed into the model. Datasheets dive deeper into questions of data provenance and allow creators and consumers of the dataset to identify limitations and impacts to the broader system. Gebru et al. propose seven categories of questions that should be considered in the development of a datasheet: motivation for dataset creation; dataset composition; collection process; preprocessing; distribution; maintenance; and legal and ethical considerations.[237] The open-ended format of Gebru et al.'s datasheet closely aligns with Mitchell et al.'s model card template; and like model cards, several variations of datasheets have emerged.

Bender and Friedman propose "data statements" for NLP, differentiating between long and short form data statements—similar to IBM's Factsheets "full report" and "tabular" formats.[238] The Dataset Nutrition Label, created by the Data Nutrition Project in 2018, is another method of evaluating and presenting information about the underlying datasets. Inspired by the nutrition label scheme, the Dataset Nutrition Label provides a modular format through a web-based GUI, consisting of three distinct categories of information: Overview, Use Cases & Alerts, and Dataset Info.[239] While data scientists are the primary intended audience for this label, any stakeholder should be able to ascertain key characteristics of the dataset from the Dataset Nutrition Label.

[237] Timnit Gebru et al., "Datasheets for Datasets," *Communications of the ACM* 64, no. 12 (December 1, 2021): 86–92, https://www.microsoft.com/en-us/research/publication/datasheets-for-datasets/.

[238] Emily M. Bender and Batya Friedman, "Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science," *Transactions of the Association for Computational Linguistics* 6 (December 2018): sec. 5, https://doi.org/10.1162/tacl_a_00041; IBM Research, "AI FactSheets 360."

[239] Kasia S. Chmielinski et al., "The Dataset Nutrition Label (2nd Gen): Leveraging Context to Mitigate Harms in Artificial Intelligence" (arXiv, March 10, 2022), 2–3, https://doi.org/10.48550/arXiv.2201.03954.

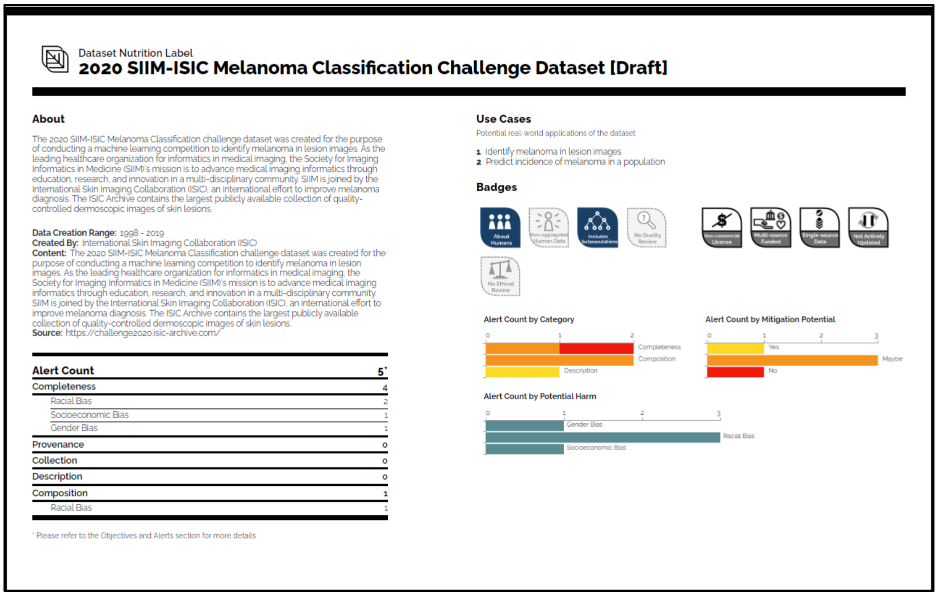


Figure 11. Example Overview Section of the Dataset Nutrition Label.[240]

G. CONCLUSION

The frameworks outlined in this section demonstrate an ongoing effort to improve the transparency of AI systems and models. They are, by no means, all encompassing. Other variations of AI documentation exist including Gilbert et al.'s "Reward Reports" (addressing reinforcement learning algorithms), Shen et al.'s "Value Cards" (using a combination of model, persona, and checklist cards), Tagliabue et al.'s "Directed Acyclic Graph (DAG) Cards" (focused on ML pipelines rather than just models), and Naja et al.'s "Knowledge Graphs" (a semantic approach to collecting accountability information).[241]

240 International Skin Imaging Collaboration (ISIC), "Dataset Nutrition Label: 2020 SIIM-ISIC Melanoma Classification Challenge Dataset [Draft]," Data Nutrition Project, 2020, https://datanutrition.org/labels/isic-2020/.

241 Thomas Krendl Gilbert et al., *Choices, Risks, and Reward Reports: Charting Public Policy for Reinforcement Learning Systems* (Berkeley, CA: UC Berkeley Center for Long-Term Cybersecurity, 2022), https://cltc.berkeley.edu/wp-content/uploads/2022/02/Choices_Risks_Reward_Reports.pdf; Hong Shen et al., "Value Cards: An Educational Toolkit for Teaching Social Impacts of Machine Learning through Deliberation," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event Canada: ACM, 2021), 850–61, https://doi.org/10.1145/3442188.3445971; Jacopo Tagliabue et al., "DAG Card Is the New Model Card," in *Thirty-Fifth Conference on Neural Information Processing Systems* (arXiv, 2021), https://doi.org/10.48550/arXiv.2110.13601; Iman Naja et al., "Using Knowledge Graphs to Unlock Practical Collection, Integration, and Audit of AI Accountability Information," *IEEE Access* 10 (2022): 74383–411, https://doi.org/10.1109/ACCESS.2022.3188967.

Other proposals have advocated for checklists or toolkits to facilitate accountability and to ensure that ethical responsibility and trustworthiness are considered.[242] Furthermore, initiatives aimed at evaluating AI technologies are not necessarily new. Disinfo Cloud, which was sponsored by the Global Engagement Center, developed a list of criteria to assess "counter propaganda and disinformation" tools.[243]

The relatively recent inception of these varying frameworks makes it difficult to discern which ones will be widely adopted. There are, however, strong indications that a growing number of organizations in diverse fields are advocating for the generation of datasheets and model cards.[244] Additionally, DIU's *RAI Guidelines* are likely to serve as a foundational document for assessing AI technology within government, given the organization's central role in adopting commercial technologies for the DOD, its close relationship with the CDAO, and the overall positive reception of its *RAI Guidelines*.

242 Office of the Director of National, "Artificial Intelligence Ethics Framework for the Intelligence Community," INTEL.gov, June 2020, https://www.intelligence.gov/artificial-intelligence-ethics-framework-for-the-intelligence-community; Carol Smith, "Designing Trustworthy AI: A User Experience (UX) Framework," in *RSA Conference 2020* (Pittsburgh, PA: Carnegie Mellon University, 2020), https://doi.org/10.1184/R1/12198321.v1; DJ Patil, Hilary Mason, and Mike Loukides, "Of Oaths and Checklists," O'Reilly Media, July 17, 2018, https://www.oreilly.com/radar/of-oaths-and-checklists/; DrivenData, "An Ethics Checklist for Data Scientists," Deon, accessed October 3, 2022, https://deon.drivendata.org/; Henriette Cramer et al., "Translation, Tracks & Data: An Algorithmic Bias Effort in Practice," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19: CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland: ACM, 2019), 1–8, https://doi.org/10.1145/3290607.3299057; European Commission, "Ethics Guidelines for Trustworthy AI | Shaping Europe's Digital Future," Shaping Europe's digital future, April 8, 2019, 26–31, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai; Digital Catapult, "Ethics Framework," Machine Intelligence Garage, 2022, https://migarage.digicatapult.org.uk/ethics/ethics-framework/; David Anderson et al., "Ethics & Algorithms Toolkit (Beta)," accessed October 3, 2022, http://ethicstoolkit.ai/; Natasha Duarte, "Digital Decisions Tool," *Center for Democracy and Technology* (blog), August 8, 2017, https://cdt.org/insights/digital-decisions-tool/; Shannon Vallor, *Ethics in Tech Practices: A Toolkit* (Santa Clara, CA: Markkula Center for Applied Ethics at Santa Clara University, 2018), https://www.scu.edu/ethics-in-technology-practice/ethical-toolkit/.

243 Global Engagement Center Technology Engagement Team, "Disinfo Cloud" (U.S. Department of State, 2022), https://www.state.gov/wp-content/uploads/2021/04/Disinfo-Cloud-flyer-April-042921.pdf.

244 Spiros Baxevanakis et al., "The MeVer DeepFake Detection Service: Lessons Learnt from Developing and Deploying in the Wild," in *Proceedings of the 1st International Workshop on Multimedia AI against Disinformation* (ICMR '22: International Conference on Multimedia Retrieval, Newark, NJ: ACM, 2022), 59–68, https://doi.org/10.1145/3512732.3533587; Daniel Hershcovich et al., "Towards Climate Awareness in NLP Research," in *2022 Conference on Empirical Methods in Natural Language Processing* (Abu Dhabi: arXiv, 2022), http://arxiv.org/abs/2205.05071; Vlad Stirbu, Tuomas Granlund, and Tommi Mikkonen, "Continuous Design Control for Machine Learning in Certified Medical Systems," *Software Qual J*, September 13, 2022, https://doi.org/10.1007/s11219-022-09601-5.

These varying frameworks, however, will inevitably evolve over time to fit new applications of AI technologies.

These different approaches to communicating the details of an AI system serve as a useful reference for developing evaluation criteria for practitioners. The questions and topics covered within the varying frameworks provide a starting point for thinking through important considerations in assessing AI systems for OIE. These questions are particularly important if the AI developer or provider is unable to furnish documentation that is equivalent to a model card or a FactSheet. Having a single, standardized document for evaluations facilitates quicker, more effective on-boarding when people who assess the technology are no longer available—a common issue within the military.

Yet, there are inherent challenges to creating standardized AI documentation such as concerns over time, cost, maintenance, and oversharing of proprietary information. Most existing frameworks are not designed for end users who will be utilizing the technology to make key decisions. Either the components of the framework are too technical and tailored to the AI developer, or they are too high-level or abstract, making it difficult for practitioners to operationalize or conduct an effective evaluation of the AI system. Therefore, even if a model card (or equivalent) was provided by the AI developer, practitioners should have their bespoke evaluation framework to ensure their concerns and questions are adequately addressed. Of course, this does not call for a reinvention; many of the key elements identified in other existing frameworks should be incorporated in user evaluations.

THIS PAGE INTENTIONALLY LEFT BLANK

VI. DEVELOPING AN AI EVALUATION SYSTEM FOR INFORMATION PRACTITIONERS

A. CHAPTER OVERVIEW

The purpose of this chapter is to outline a framework that can be used by information practitioners to evaluate AI tools. Although there are a growing number of initiatives within government, industry, and academia aimed at evaluating AI technology, an evaluation framework tailored for the end user—or more specifically, the warfighter—has not yet been developed. Nonetheless, as discussed in Chapter V, these ongoing initiatives lay the groundwork to identify key considerations for increased transparency of AI throughout the evaluation process. This thesis draws upon these initiatives along with extant literature discussed in the previous chapters to develop an evaluation framework for information practitioners.

B. UTILIZATION-FOCUSED EVALUATION

The utility of evaluation frameworks derives from their systematic approach in assessing performance, identifying key issues, and determining what improvements can be made in the tool, system, or program. A wide variety of evaluation methods exist within government, industry, and academia. These methods range from formative and process evaluations to performance monitoring and meta-evaluations.[245] Patton's Utilization-Focused Evaluation (UFE) provides a useful framework for practitioners, given its focus on "specific intended primary users for specific intended uses."[246] Rather than prescribing a particular method or theory, the UFE offers a guideline through a series of non-linear steps, seen in Figure 12. The framework allows intended users to

[245] Joseph S. Wholey, Harry P. Hatry, and Kathryn E. Newcomer, eds., *Handbook of Practical Program Evaluation* (Hoboken, NJ: Wiley, 2004).

[246] Michael Quinn Patton, *Essentials of Utilization-Focused Evaluation* (Los Angeles, CA: SAGE, 2012), 39.

"feel ownership of the evaluation" and to modify the content and method of evaluation based on a specific situation.[247]

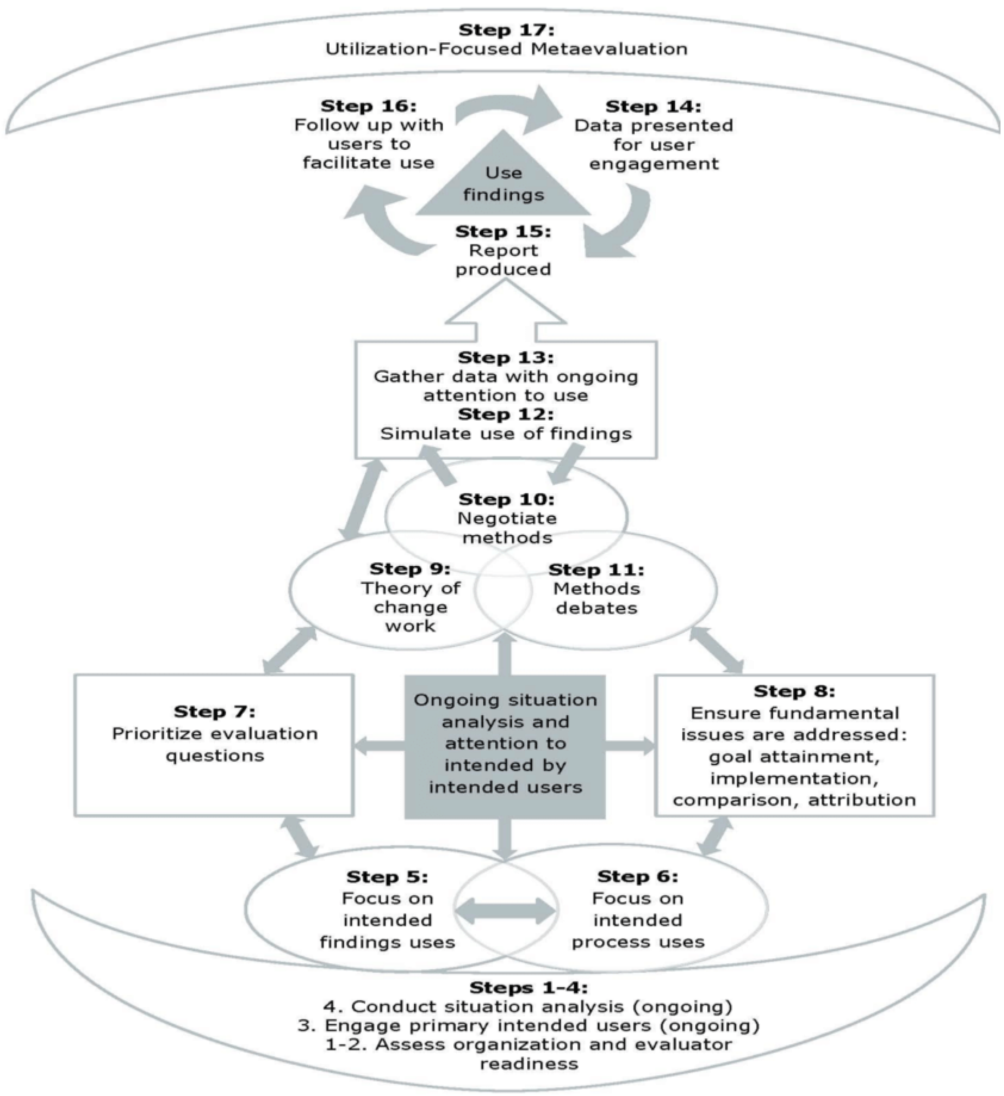


Figure 12. Utilization-Focused Evaluation.[248]

[247] Ricardo Ramírez and Dal Brodhead, *Utilization Focused Evaluation: A Primer for Evaluators.* (Penang, Malaysia: Southbound, 2013), 89–90.

[248] Source: Patton, Essentials of Utilization-Focused Evaluation.

Yet, despite the UFE's practical approach, the implementation of a methodical evaluation system becomes increasingly dubious when placed in the hands of a practitioner whose primary role is to plan and execute OIE—not to evaluate a piece of technology. Typically, professional evaluators have the requisite knowledge, skills, and experience to design, manage, and implement an evaluation program. In an ideal world, OIE units would have resident technical and domain experts who could methodically collect and analyze the data, simulate the findings, and conduct a detailed meta evaluation. Unfortunately, the reality reflects a much more constrained environment. Given the high rate of personnel turnover in the military, expertise is difficult to retain. Practitioners are often unable to conduct systematic evaluations due to operational demand, lack of resources, and limited time.

An evaluation framework designed for practitioners should account for these "on the ground" realities. To increase the likelihood of adoption, the framework should be consistent with existing organizational processes (i.e., compatible) and minimally complex. Ramírez and Brodhead distill the original 12-step UFE model into five categories: 1) preparing for evaluation; 2) analyzing the situation; 3) designing evaluation; 4) undertaking evaluation; and 5) reflecting on evaluation done.[249] These categories provide a starting point for further modification of the UFE to ensure that the framework is tailored to the needs of a practitioner who is evaluating a piece of technology—rather than other programs or processes. This simplification permits the removal of inapplicable steps and allows for the addition of AI specific evaluation considerations.

C. PRACTITIONER'S EVALUATION FRAMEWORK

Figure 13 outlines a four-phase process for evaluations. These phases combine Patton's 17-step UFE model and closely mirror Ramírez and Brodhead's five categories. While this framework can be applied to almost any type of evaluation, the intent is to provide broad guidelines that can be easily understood and implemented by a

[249] Ramírez and Brodhead, *Utilization Focused Evaluation*, 1–2.

practitioner, while at the same time allow for sufficient flexibility to incorporate greater rigor in the evaluation process. Each phase produces specific outputs that feed into the next phase of the process. The following sections discuss each phase in the context of AI evaluations for information practitioners.

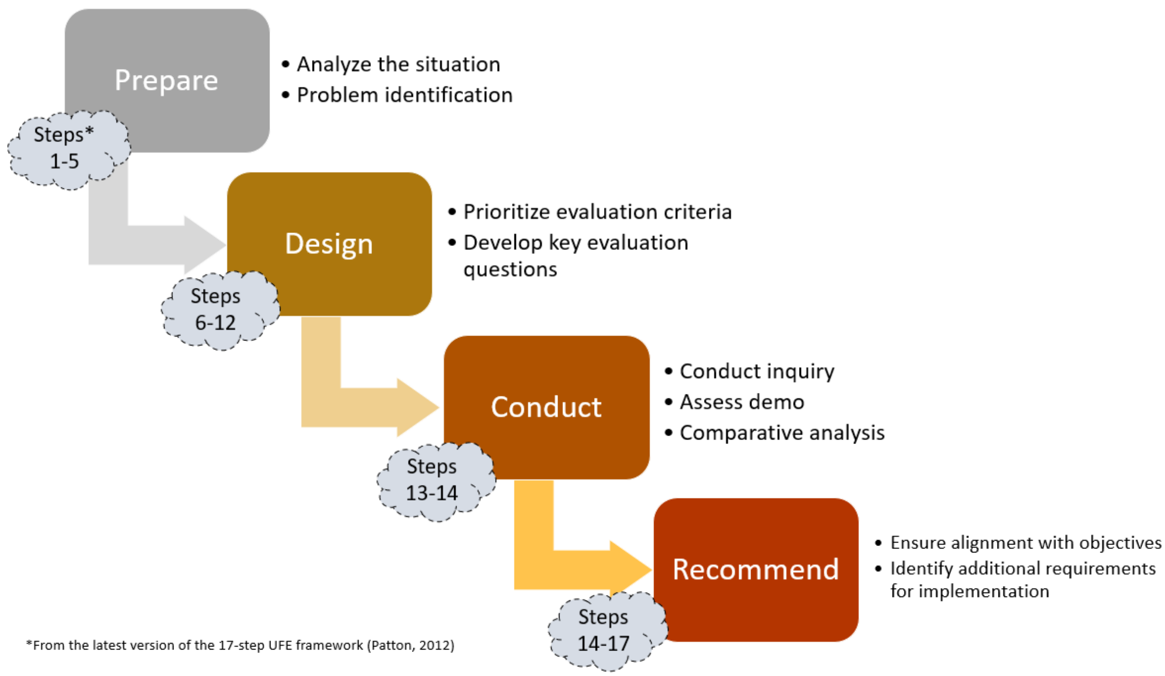


Figure 13. Evaluation Process for Practitioners

1. Phase 1: Prepare

As underscored within systems engineering, project management, and military planning, clearly defining the problem is the most important step in any decision-making process. The evaluation of AI tools is no different. As a general-purpose technology, AI offers a vast number of potential applications, which continue to expand exponentially with the advancement of the technology. New models are developed every day, promising improved performance, novel solutions, and greater insight. Yet, the technology's wide applicability does not necessarily translate to effective utilization or adoption—as discussed in Chapter IV. Failure to precisely identify what problem or capability gap the AI is trying to address leads to wasted investment. For complex, open-ended problems, AI may not be the most appropriate solution and can be counterproductive, requiring significant computing power and data.[250]

To avoid the temptation of tailoring the solution toward a specific technology, platform, or tool rather than addressing the problem, evaluators should first define the problem—agnostic to the technological solution. In other words, before thinking about the specific technology, *what problem or gap are we trying to address and why does it require a technical solution?* Exploration of this question requires analysis of the situation—more specifically, an understanding of current capabilities and challenges as well as an assessment of the organization's readiness to adopt a new solution (e.g., support from leadership, availability of time and resources). Establishing a problem statement grounds the evaluation within the larger strategic context and ensures that the developed solutions are evaluated against the objectives that the organization cares about.

Once there is a clear understanding of the problem, information practitioners can then examine what aspect of the problem they are expecting the AI to solve. In the current state of technology, it is highly unlikely that a single piece of AI would be able to solve all facets of the identified problem. The solution may require a constellation of different technologies. In some cases, AI may not be the most appropriate or cost-effective solution for a given problem. Therefore, the following questions must be asked: *what benefit does the AI technology provide and how does it address the problem or gap?* A careful examination of this question will enable users to identify specific tasks that they are expecting the AI tool to do. This examination then raises two additional questions: *is the proposed AI technology capable of being implemented and can it address the problem or gap within the given organizational and environmental context?* Proper consideration of these questions requires a certain level of knowledge about the technology, which may not exist organically within the OIE unit conducting the evaluation. The criticality of this question, however, necessitates additional time invested in researching the technology or seeking expert advice.

[250] U.S. Government Accountability Office, *Artificial Intelligence: Status of Developing and Acquiring Capabilities for Weapon Systems*, GAO-22-104765 (Washington, DC: U.S. Government Accountability Office, 2022), 13, https://www.gao.gov/assets/gao-22-104765.pdf.

The DIU's Planning Worksheet provides a useful guideline in defining appropriate AI tasks, noting that AI may not be the optimal solution for tasks that require complete predictability, transparency, interpretability, explainability, or assurance.[251] Determining what tasks within OIE could or should be conducted by AI is not always straightforward. The planning and execution of OIE involves significant subjectivity, due to its direct connection with human reasoning, perception, and emotion. It would be inadvisable to rely solely on AI to conduct target audience analysis or assess human reaction in a face-to-face engagement. Even AI's ability to identify disinformation can be limited, despite the recent progress and investment made in this area.

Assessing the problem as well as the needs and feasibility of the organization may seem like an obvious step—one that is reflected in existing acquisition pathways and other planning processes.[252] Yet, AI raises additional challenges because of its complexity and breadth of application. As discussed previously, the failure to adopt AI is often attributed to misalignment between organizational needs and the actual capability offered by the technology.[253] Thus, greater emphasis is needed on Phase 1 than the other three phases. Given a clear understanding of the linkages between the problem and the technology, organizations can develop an initial set of key performance indicators—generated during Phase 2—that can drive the rest of the evaluation process.

2. Phase 2: Design

Program or evaluation design can be an elaborate process that includes theory testing, experiments, or analyses of causal relationships between variables. The design of an evaluation will vary by context and organization. The DOD employs formal T&E programs that provide specific guidelines in designing and conducting developmental,

[251] Dunnmon et al., "Phase I: Planning Worksheet for DIU AI Guidelines."

[252] Office of the Under Secretary of Defense for Acquisition and Sustainment, *DOD Instruction 5000.02 Operation of the Adaptive Acquisition Framework* (Washington, DC: Department of Defense, 2022), https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500002p.pdf.

[253] "Businesses Are Finding AI Hard to Adopt"; Kiron and Schrage, "Strategy For and With AI."

operational, and live fire T&E.[254] Similarly, the CDAO is working to develop a tailored T&E framework for AI. A/B testing, often used by marketers and content creators to assess online user engagement, is also an attractive option to test technology within the information environment. The assumption underlying all these evaluative designs is that adequate resources, time, and expertise are available to support the evaluation process. Yet, these assumptions do not always reflect reality. Although maintaining a dedicated T&E team is preferable, organizations—particularly at lower echelons—struggle to establish and maintain dedicated T&E teams to assess new technological applications.

The growing accessibility to emerging technologies requires end users to be able to conduct practical evaluations to inform program managers and higher echelon leadership on whether to acquire or invest in a particular tool. This bottom-up approach is especially pertinent within the Special Operations community, which seeks to be on the leading edge of innovation. Given the potential limitations of the practitioner, the design of the evaluation should emphasize simplicity and necessity. Thus, step 7 of the UFE—prioritize evaluation questions—should be the primary focus within this phase, as opposed to steps 8–12, which are less applicable or too time-consuming to be adopted by practitioners.[255] Section D below—Evaluation Domains—provides an initial set of criteria and evaluation questions from which the adopting organization can tailor to fit the specific use case.

3. Phase 3: Conduct

The third phase can be viewed as the execution phase, which involves direct interaction with the vendor or developer. Vendors will typically provide a demo of their tool and may even allow users to obtain a trial account to independently test the tool's

[254] Defense Acquisition University, "CH 8 Test & Evaluation," in *Defense Acquisition Guidebook* (Fort Belvoir, VA: The Defense Acquisition University Press, 2020), https://www.dau.edu:443/pdfviewer?Guidebooks/DAG/DAG-CH-8-Test-and-Evaluation.pdf.

[255] Steps 8–12 are as follows: 8) Check that fundamental areas for evaluation inquiry are being adequately addressed; 9) Determine what intervention model or theory of change is being evaluated; 10) Negotiate appropriate methods to generate credible findings that support intended use by intended users; 11) Make sure intended users understand potential methods controversies and their implications; 12) Simulate use of findings.

features. During this phase, the quality of effort placed in the prior two phases will become readily apparent. For a given problem or task, a myriad of different tools exist that can potentially address the needs of the organization and user. By identifying what capabilities and features are most important to the user, the evaluator will have an easier time differentiating between the different tools. Ideally, units would evaluate multiple vendors within a fixed timeframe to generate a comparison across the different tools, using the same evaluation criteria. However, time constraints or emerging opportunities may generate situations in which units are obliged to assess a tool separately. In these cases, documenting the evaluation will be key to maintaining objectivity and providing the means for comparison when similar technologies are presented in the future.

4. Phase 4: Recommendation

The recommendation phase consolidates the analysis from Phase 3 and presents the findings to key stakeholders. The level of detail and format of the recommendation will vary depending on the intended audience as well as the time and resources available to conduct the evaluation. Nonetheless, the recommendation should explain how the selected solution fits into the broader context of the organization's mission, existing processes, and culture. It should also highlight any additional requirements needed to facilitate the adoption of the AI tool such as training, support personnel, or infrastructure.

D. EVALUATION DOMAINS

The generation of quality questions requires a methodical approach and should go beyond an extemporaneous inquiry. Determining the "right" evaluation questions can be a significant challenge, given the inherent complexity of AI. The synthesis of prominent themes within the AI adoption literature, related AI frameworks, and expert interviews offers a way to navigate through the complexity and highlights critical areas—or evaluation *domains*—that should be considered during the evaluation process.[256]

[256] Rebecca M. Teasdale, "Evaluative Criteria: An Integrated Model of Domains and Sources," *American Journal of Evaluation* 42, no. 3 (September 1, 2021): 356, https://doi.org/10.1177/1098214020955226.

Evaluation domains assist in grouping and ordering questions; they serve as a "prompt to check that different areas are investigated and receive the appropriate degree of emphasis."[257]

This thesis finds that nine evaluation domains emerge as leading factors in the effective employment and adoption of AI. Figure 14 depicts how these domains fit within the TOE framework and relate to prominent factors within the DOI and UTAUT models. Performance expectancy and effort expectancy—along with their related constructs of relative advantage, complexity, and compatibility—are highlighted in the figure due to their particular relevance to AI adoption. The *DOD Ethical Principles for AI*, listed at the top of Figure 14, serve as guidelines throughout the evaluation process. Similarly, mission/task alignment should be a driving consideration. Several of the evaluation domains fall within multiple categories of the technology adoption model. For example, sustainability sits within all three dimensions of the TOE. The ability to maintain the system impacts relative advantage and compatibility of the technology, while organizational and environmental factors such as available infrastructure, budget, public attitudes toward the technology will affect the willingness and ability to leverage the AI tool over time. Although not considered an evaluation domain in this framework, organizational readiness is an important reference when considering domains such as sustainability, scalability, and affordability.

---

[257] Anne Markiewicz and Ian Patrick, *Developing Monitoring and Evaluation Frameworks* (Thousand Oaks, CA: SAGE Publications, Inc, 2016), chap. 5, https://doi.org/10.4135/9781071878774.

Evaluation domain is often used interchangeably with *evaluation criteria*, but Teasdale specifies that a domain is only one aspect of criteria; the other is the source, which refers to the organization, interest group, or document that the criteria is drawn from. The source is a critical consideration during evaluations—hence the discussion about other AI evaluation frameworks and guidelines as well as the emphasis on mission alignment. The focus of this section, however, is identifying the domain or substance of the evaluation criteria.

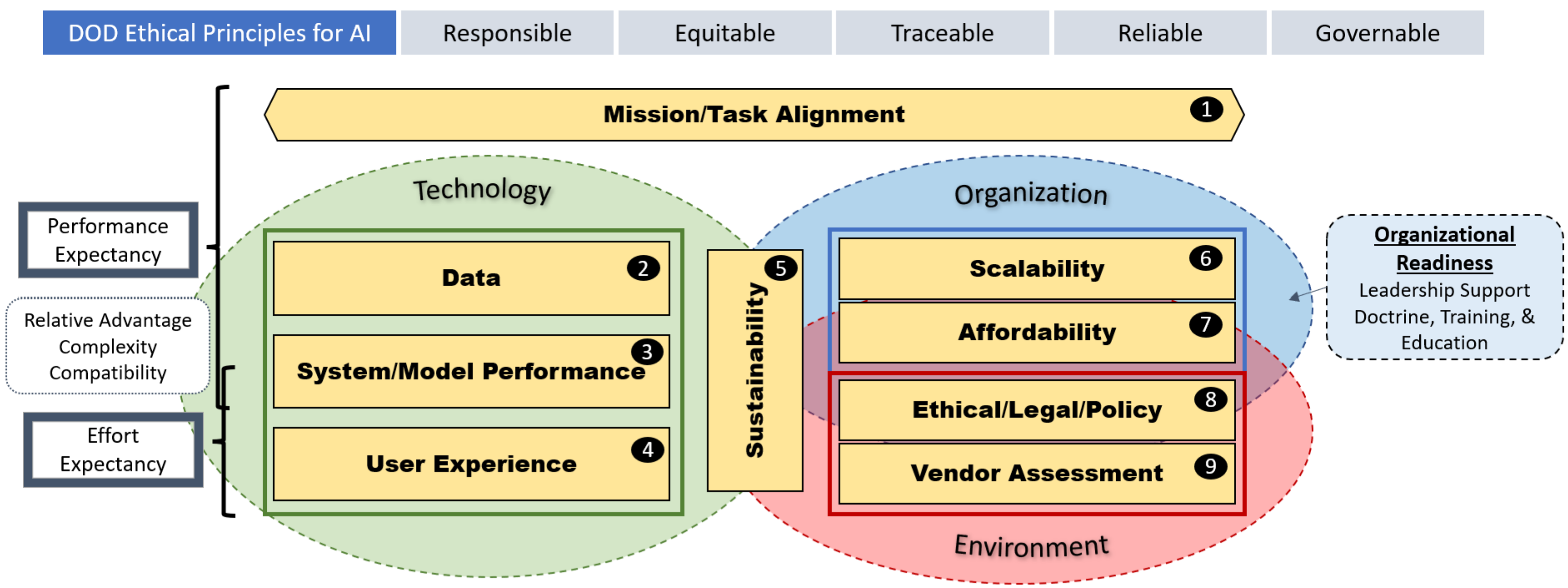


Figure 14. Evaluation Domains for Practitioners

1. Mission/Task Alignment

While the preparation phase (Phase 1) examines the overall appropriateness of AI as a solution to a defined problem, evaluation questions (as listed in Figure 15) should pinpoint whether the intended use of the specific AI tool aligns with the mission or task that the organization is seeking to fulfill. While, in many cases, this may be a simple question, there are situations—particularly when dealing with COTS—where the connection between the tool's original use and the organization's problem statement may not be readily apparent. Therefore, evaluators should avoid outright acceptance or complete disregard of the tool's applicability to the operational requirement, and instead, opt for a discussion about *why* the tool may or may not align with the unit's objectives.

Relative advantage was identified as a common factor across all technology adoption theories and should be a paramount consideration during the evaluation. The evaluator should distinguish the unique characteristics or capabilities of the tool to determine whether the tool provides greater value over current capabilities or other existing solutions. Relative advantage extends beyond the attributes of the tool; it includes assessing the tool's impact on productivity. Therefore, an initial conceptualization of how the tool will fit into a user's current workflow is needed to ascertain its actual utility. A prerequisite to this conceptualization, however, is the identification of the specific intended users, which directly impacts intended use of the tool. The context in which the other evaluation domains are assessed will differ based on user groups, even those from the same organization. For example, a practitioner based in the United States, conducting analysis in preparation for a deployment will perceive the utility of a tool differently than someone who is deployed with limited time, infrastructure, or support. Therefore, intended use should be specified to the greatest level of detail possible.

- What is the AI's intended use/purpose?
  - Does this align with the mission/task?
  - How does it fit into the existing workflow?
- What relative advantage does this tool provide?
- What are the risks of adopting this tool?

Figure 15. Evaluation Questions: Mission/Task Alignment

2. Data

Data fundamentally drives the quality of the AI model and influences its outcomes. Several characteristics should be considered when assessing the data. First is the size of the dataset. AI models need large amounts of data to train on to improve their performance, given that greater amounts of data generally increase the degree of statistical reliability of the model. As an example, Figure 16 shows how the accuracy of a random forest classifier model increases with the size of the dataset.

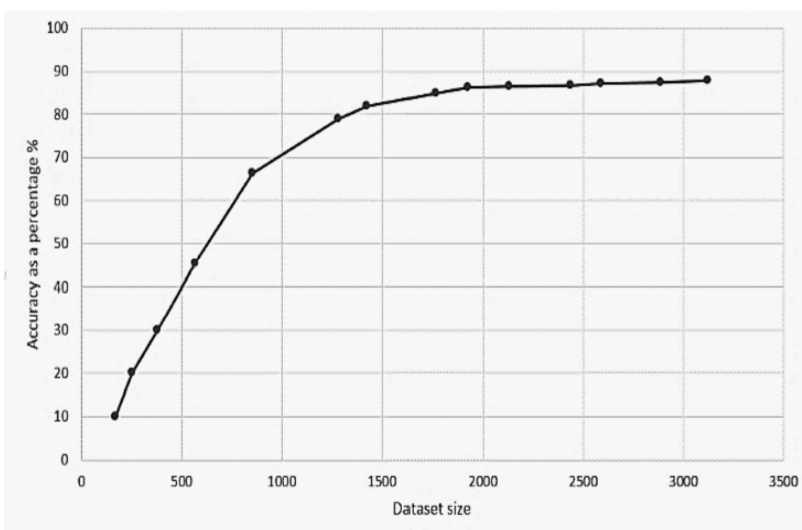


Figure 16. Learning Curve of Machine Learning Model with the Size of Dataset Used for Testing and Training.[258]

[258] Source: Lakmal Meegahapola et al., "Random Forest Classifier Based Scheduler Optimization for Search Engine Web Crawlers," 2018, https://doi.org/10.1145/3185089.3185103.

There is no hard minimum number for how much data is required for AI, since it will vary depending on the complexity of the model. Nonetheless, the general rule of thumb is that the amount of input data should number at least an order of magnitude more than the number of trainable parameters.[259] Having a large enough dataset allows for models to discern meaningful relationships.

Large datasets, however, do not always equate to better models. Massive or not, datasets could be affected by a range of issues including missing, incorrect, or duplicate values.[260] Incompatibility in units or format is also a classic problem. Depending on the ML technique, nonnumeric data may have to be converted to numeric data or vice versa. These issues highlight the importance of cleaning and preprocessing the data. For supervised learning models, attention should be placed on the reliability of data labeling. Approaches to data labeling can range from the use of existing labels to crowdsourcing or "weak supervision."[261] Precision and accuracy in data labeling are essential, since the labeled dataset guides the model toward the ground truth.

Therefore, proper evaluation requires awareness of data provenance—its origin, lineage, and preprocessing methods.[262] Ethical and accountability concerns such as bias, fairness, and privacy often evolve from the characteristics of the data source. How and what data are collected impacts whether it is representative of the conditions that the AI will operate in. Werder et al. discuss how biases from data sources (e.g., sampling or measurement issues, inappropriate data repurposing or augmentation) and data processing

---

[259] Google Developers, "The Size and Quality of a Data Set | Machine Learning," Machine Learning, July 18, 2022, https://developers.google.com/machine-learning/data-prep/construct/collect/data-size-quality.

[260] Ethem Alpaydin, *Machine Learning* (Cambridge, MA: The MIT Press, 2016), 154.

[261] Yuji Roh, Geon Heo, and Steven Euijong Whang, "A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective," *IEEE Transactions on Knowledge and Data Engineering* 33, no. 4 (April 2021): sec. 3, https://doi.org/10.1109/TKDE.2019.2946162.

[262] James Cheney, Laura Chiticariu, and Wang-chiew Tan, "Provenance in Databases: Why, How, and Where," *Foundations and Trends in Databases* 1 (January 1, 2009): 382, https://doi.org/10.1561/1900000006; Dunnmon et al., *Responsible AI Guidelines in Practice*, 21.

errors (e.g., dataset shifts, opaque preprocessing) can undermine RAI.[263] The DOD's *Data Strategy* outlines guiding principles to facilitate "high quality, accurate, complete, timely, protected, and trustworthy" data.[264]

Caution is also needed to recognize overfitted models, which rely too strongly on the details of the training dataset and thus run into problems when applied in a general context.[265] In other words, the model may be finely attuned to the provided training data, but it may be overvaluing the significance of irrelevant data or noise. One example of overfitting is when an AI model trains on millions of pictures to classify weapon systems such as javelins, but unbeknownst to the AI developer, the model learns to associate the soldier shouldering the weapon or other features such as the sky or vegetation with the classification of the javelin. As a result, the model is unable to recognize the weapon system if it is depicted in non-deployed circumstances (e.g., a production or display setting). Another example is when a model is trained to identify drones, but the training data only included military unmanned aerial vehicles that are known to be in existence; thus, the model would have trouble identifying commercially made, modified, or newly designed drones, since these images were not part of its training dataset.

Overfitting can be mitigated through several means. The most obvious is ensuring that the training data is truly representative of real-world conditions. Other ways include data augmentation to artificially increase the sample size or regularization to reduce the impact of noise.[266] A good practice to detect overfitting is partitioning the data to create a separate test set that can validate the trained model.

[263] Karl Werder, Balasubramaniam Ramesh, and Rongen (Sophia) Zhang, "Establishing Data Provenance for Responsible Artificial Intelligence Systems," *ACM Transactions on Management Information Systems* 13, no. 2 (March 10, 2022): 22:4-8, https://doi.org/10.1145/3503488.

[264] Department of Defense, *DOD Data Strategy*, 3–6.

[265] Ian H. Witten et al., eds., *Data Mining: Practical Machine Learning Tools and Techniques*, Fourth Edition (Amsterdam, Netherlands: Elsevier, 2017), 31.

[266] Simukayi Mutasa, Shawn Sun, and Richard Ha, "Understanding Artificial Intelligence Based Radiology Studies: What Is Overfitting?," *Clinical Imaging* 65 (September 2020): 96–99, https://doi.org/10.1016/j.clinimag.2020.04.025.

Beyond the characteristics of the dataset itself, practitioners should consider the management of the data to include ownership, access, maintenance, and security. Ownership not only affects access to the data but also informs how and by whom the data will be maintained, as well as what measures need to be in place to ensure that sensitive data is secured and protected. The government, however, does not always have to be the owner of the data, which would be an expensive and likely impossible feat. Data can be procured through commercial means and may include non-public or proprietary information, involving more restrictive data rights.[267] In other cases (such as those involving PAI), ownership of the data is not as much of an issue as continual access and automatic updating of ingested data. Determining what aspects of data management are important to the organization will enable the practitioner to screen out less desirable options.

Although practitioners may not understand every detail, questions about data provenance, processing, and maintenance (Figure 17) remain critical to the evaluation of the AI tool.

- What data is the AI using?
- Is the data relevant, current, and representative of the actual operational environment?
- Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?
- What restrictions have been placed on the data?
- *Data Security/Integrity*
    - What steps have/will be taken to ensure the data is appropriately secure during and after the project?
    - What steps will be taken to check for potential errors, noise, bias, and/or redundancies?
- *Data Maintenance*
    - Who is maintaining the dataset?
    - How will it be updated? (e.g., automatic ingestion via API)

Figure 17. Evaluation Questions: Data

---

[267] Emanuel Trunzer et al., "A Flexible Architecture for Data Mining from Heterogeneous Data Sources in Automated Production Systems," in *2017 IEEE International Conference on Industrial Technology (ICIT)* (Toronto, Canada: IEEE, 2017), 1109, https://doi.org/10.1109/ICIT.2017.7915517.

3. System/Model Performance

While data may serve as the fuel for AI, AI algorithms are what drive the model toward a solution. One of the basic steps of evaluating an AI tool is assessing its performance. Methods for evaluating performance can vary based on perspective and type of AI. Evaluators can examine the performance metrics of individual models (as provided by model cards) or at the system level, which are more concerned with whether the final output meets the expectations and needs of the user (i.e., performance expectancy). System—or "end-to-end"—evaluations may be viewed as more meaningful for practitioners in quantifying the tool's effectiveness, especially since most AI tools consist of multiple models and components.[268] Yet, individual model performance should not be discounted. Errors can have a cascading effect, with each model impacting a subsequent component of the system.

Within NLP, multiple models are needed to perform intermediate functions, such as tokenization, word embeddings, or paraphrasing to conduct various NLP tasks (e.g., semantic analysis, machine translation, summarization), since most learning algorithms require the conversion of raw texts and sentences before ingestion.[269] Thus, researchers propose the use of intrinsic and extrinsic evaluations to account for the different levels. Intrinsic evaluations compare the outputs of a particular NLP component against a pre-determined criteria separately from other system components, while extrinsic evaluations assess performance based on its impact on the larger system or other NLP functions.[270] For example, questions within an intrinsic evaluation of a paraphrase generation system

[268] Alexander Clark, Chris Fox, and Shalom Lappin, eds., *The Handbook of Computational Linguistics and Natural Language Processing*, Blackwell Handbooks in Linguistics (Malden, MA: Wiley-Blackwell, 2010), chap. 11.

[269] Yong Shi et al., "Intrinsic or Extrinsic Evaluation: An Overview of Word Embedding Evaluation," in *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2018, 1255, https://doi.org/10.1109/ICDMW.2018.00179.

[270] Clark, Fox, and Lappin, The Handbook of Computational Linguistics and Natural Language Processing; Hanna Suominen, "Performance Evaluation Measures for Text Mining," in Handbook of Research on Text and Web Mining Technologies, ed. Min Song and Yi-Fang Brook Wu (Hershey, PA: IGI Publishing, 2009), 726, 10.4018/978-1-59904-990-8.ch041.

would include, “Does the generated paraphrase convey meaning of the original text?”[271] On the other hand, extrinsic evaluations may entail questions such as, “Do the incorporated paraphrases significantly improve performance of a question-answering model?”[272]

Once again, extrinsic evaluations may provide more value to practitioners since they assess the performance of the overall NLP task. Nonetheless, researchers stress the importance of incorporating both intrinsic and extrinsic evaluation strategies.[273] Some suggest that intrinsic evaluations may offer a more practical option, given that it is more straightforward and easier to incorporate with automatic evaluation techniques.[274] They can facilitate more useful comparisons among foundation models, which are becoming increasingly prevalent in a wide range of applications.[275] Intrinsic evaluations also enable individuals within a technical team to diagnose issues within the subsystem. Thus, even if limited time or technical knowledge inhibit the practitioners’ ability to assess the AI tool at the subtask level, practitioners should be aware of both types of evaluations.

The metrics used to evaluate AI performance can also vary, based on model type and operational requirements. Given its relative simplicity, accuracy is a commonly used metric, but it can also be a misleading measure of performance. For example, a binary classification model can claim that it is 99% accurate in identifying hostile text messages; this would mean that out of 1000 text messages, the model was able to classify 990

[271] Tulu Tilahun Hailu, Junqing Yu, and Tessfu Geteye Fantaye, “Intrinsic and Extrinsic Automatic Evaluation Strategies for Paraphrase Generation Systems,” *Journal of Computer and Communications* 8, no. 2 (2020): 2, https://doi.org/10.4236/jcc.2020.82001.

[272] Hailu, Yu, and Fantaye, 2.

[273] Shi et al., “Intrinsic or Extrinsic Evaluation”; Paula M. Franceschini, Henrique D. P. dos Santos, and Renata Vieira, “Intrinsic and Extrinsic Evaluation of the Quality of Biomedical Embeddings in Different Languages,” in *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, 2020, 271–76, https://doi.org/10.1109/CBMS49503.2020.00058.

[274] Hailu, Yu, and Fantaye, “Intrinsic and Extrinsic Automatic Evaluation Strategies for Paraphrase Generation Systems,” 12; Clark, Fox, and Lappin, *The Handbook of Computational Linguistics and Natural Language Processing*, chap. 11.

[275] Rishi Bommasani et al., *On the Opportunities and Risks of Foundation Models* (Stanford, CA: Stanford University Human-Centered Artificial Intelligence, 2022), 91–95, https://crfm.stanford.edu/report.html.

messages correctly as either hostile or non-hostile. Yet, this metric can be problematic when there is an imbalance of classes. If the corpus of text messages included 990 non-hostile messages and only 10 hostile messages, a model can simply conclude that all messages are non-hostile, which would amount to a 99% accuracy level but would misclassify all the hostile messages.

To address this shortcoming, recall and precision are often used to evaluate the performance of classification models. Recall (also known as sensitivity) is the proportion of correctly identified positive cases to the total number of actual positive cases.[276] To apply the aforementioned example, recall would be calculated by dividing the number of correctly identified hostile messages by the total number of hostile messages existing within the dataset. Precision is the proportion of correctly identified positive cases to the total number of positive cases identified by the model.[277] Thus, precision is calculated by dividing the number of correctly identified hostile messages over the total number of cases that the model has identified as hostile. These two metrics can be combined into an F1 score, which calculates the weighted average of precision and recall across all predicted classes.[278] The F1 score is particularly useful, since it provides a single metric to facilitate univocal comparisons across systems.

Another way of visualizing these metrics is through a confusion matrix, shown in Figure 18. A confusion matrix provides a breakdown of the errors and the types of errors found in the model and assists in the calculation of precision and recall as well as specificity, which measures the proportion of identified negative cases to the total number of actual negative cases.[279]

---

[276] Bhuvan Unhelkar and Tad Gonsalves, *Artificial Intelligence for Business Optimization: Research and Applications* (Boca Raton, FL: CRC Press, 2021), 95, https://doi.org/10.1201/9781003120926.

[277] Unhelkar and Gonsalves, 95.

[278] Neeraj Mohan et al., eds., Artificial Intelligence, Machine Learning, and Data Science Technologies: Future Impact and Well-Being for Society 5.0 (Boca Raton, FL: CRC Press, 2021), 57, https://doi.org/10.1201/9781003153405.

[279] Ajay Kulkarni, Deri Chong, and Feras A. Batarseh, “Foundations of Data Imbalance and Solutions for a Data Democracy,” in *Data Democracy* (Amsterdam, Netherlands: Elsevier, 2020), sec. 3.1, https://doi.org/10.1016/B978-0-12-818366-3.00005-8.

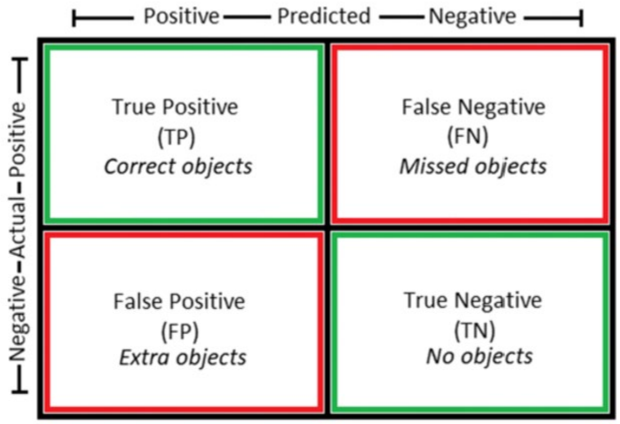


Figure 18. Confusion Matrix.[280]

The confusion matrix can also be used to plot the receiver operator characteristics (ROC) curve, which depicts the tradeoff between recall (sensitivity) and specificity.[281] The utility of the ROC curve is that it illustrates the classification (or decision) threshold, which affects the number of correctly classified cases.[282] Determining the appropriate threshold will depend on whether one cares more about including all the positive cases or whether it is more important to get the negative cases correct. The area under the curve (AUC), which "provides an aggregate measure of performance across all classification thresholds," is calculated to enable comparative analysis among different algorithms.[283] A higher AUC value signifies a better performing classification model.

Accuracy, recall, precision, F1 score, confusion matrix, and ROC AUC are by no means the only ways to describe the performance of a system or model. AI regression

280 Source: Narinder Punn and Sonali Agarwal, "Automated Diagnosis of COVID-19 with Limited Posteroanterior Chest X-Ray Images Using Fine-Tuned Deep Neural Networks," *Applied Intelligence* 51 (May 1, 2021): 2697, https://doi.org/10.1007/s10489-020-01900-3.

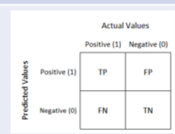
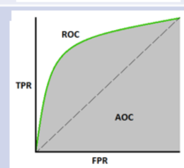
281 Helen R. Sofaer, Jennifer A. Hoeting, and Catherine S. Jarnevich, "The Area under the Precision-Recall Curve as a Performance Metric for Rare Binary Events," *Methods in Ecology and Evolution* 10, no. 4 (2019): 565–77, https://doi.org/10.1111/2041-210X.13140.

282 Zhe Hui Hoo, Jane Candlish, and Dawn Teare, "What Is an ROC Curve?," *Emergency Medicine Journal* 34, no. 6 (June 2017): 357–58, https://doi.org/10.1136/emermed-2017-206735.

283 Google Developers, "Classification: ROC Curve and AUC | Machine Learning," Machine Learning, July 18, 2022, https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc.

models, which aim to predict continuous values rather than binary classes, utilize different evaluation metrics such as mean absolute error or root mean squared error. Tables 2 and 3 provides a list of common performance metrics.

Table 2. Algorithm Metrics for Classification Models.[284]

Metrics	Definition	Formula
Accuracy	Accuracy equals the number of correct predictions over the total number of predictions.	$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)}$
Mean Average Precision (mAP)	Quantifies the mean accuracy and precision of model queries.	$\text{mAP} = \frac{1}{N}\sum_{i=1}^{N} \text{AP}_i$
Precision	It is the number of accurate positives a model claims compared to the number of positives it actually claims.	$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$
Recall	It is the number of positives a model claims compared to the actual number of positives there are throughout the data.	$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$
F1 Score	The F1 score is the harmonic mean of the precision and recall.	$F_1 = 2 * \frac{precision * recall}{precision + recall}$
Confusion Matrix	It is a table that reports the number of false positives (FP), false negatives (FN), true positives (TP), and true negatives (TN).	
Receiver Operator Characteristic Area Under the Curve (ROC AUC)	Quantifies the model's ability to separate classes by capturing the count of accurate positive predictions against the count of incorrect positive predictions at different thresholds.	

284 Source: Joint Artificial Intelligence Center, "JAIC T&E Workshop" (presentation, Department of Defense, October 28, 2021).

Table 3. Algorithm Metrics for Regression Models.[285]

Metrics	Definition	Formula
Mean Absolute Error (MAE)	Measure of the average distance of errors in a set of predictions.	$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} \lvert \boldsymbol{x}_i - \boldsymbol{x}_i \rvert$
Mean Squared Error (MSE) or Mean Squared Deviation (MSD)	The average squared difference between the estimated values and actual values.	$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \tilde{y}_i)^2$
Root Mean Squared Error (RMSE)	Represents how close model predictions fit a data set. Provides a high weight to large errors relative to MAE.	$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$
Coefficient of Determination (R^2)	Provides a measure of how well observed outcomes are replicated by the model.	$r^2 = \frac{\mathrm{RSS}}{\mathrm{TSS}} = \frac{\sum(\hat{Y} - \bar{Y})^2}{\sum(Y - \bar{Y})^2}$
Mean Absolute Deviation (MAD)	The average distance between each data point x and the mean value of the data $\bar{x}$ indicating the variability in the dataset.	$\mathrm{MAD} = \frac{\sum \lvert x_i - \bar{x} \rvert}{n}$
Mean Absolute Percentage Error (MAPE)	Also known as Mean Average Percentage Deviation (MAPD). This metric is typically used in forecasting applications to measure accuracy over a time period.	$\mathrm{M} = \frac{1}{n}\sum_{t=1}^{n} \left\lvert \frac{A_t - F_t}{A_t} \right\rvert$

Although these metrics play an important role in validating a model or comparing performance across different AI systems, practitioners may have trouble understanding the significance of the measures, especially for ones that are more complex such as perplexity or BLEU (bilingual evaluation understudy).[286] One can argue that without understanding the full context of what the metric is assessing, these metrics provide little practical value for the practitioner. For example, being told without additional context that a model has an F1 score of 85% will likely mean little to a practitioner. Furthermore,

---

[285] Source: Joint Artificial Intelligence Center.

[286] Ben Hutchinson et al., “Evaluation Gaps in Machine Learning Practice,” in *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT ‘22)* (Seoul, Republic of Korea: ACM, 2022), 6, https://doi.org/10.1145/3531146.3533233.

variance in performance between different algorithms can be rather minimal.[287] Even if a practitioner was able to compare scores between two models (e.g., BERT-Based-Cased model with an accuracy of 94.147% and F1 score of 85.687% versus the XLNet-Based-Cased model with an accuracy of 94.214% and an F1 of 85.292%), the scores may not differ enough to be viewed as significant by the practitioner—even though the difference may be of interest to a researcher.[288]

Nonetheless, these performance metrics remain an important consideration during the evaluation process even for practitioners because they provide some level of standardization and objective measure to assess the value of the AI tool. While high accuracy or F1 scores do not necessarily equate to a more desirable tool, low performance scores do hint at diminished value. Therefore, practitioners should request that vendors share their system's performance metrics as well as additional context such as interpretations of what would be an acceptable minimum performance threshold and what types of errors would cause serious problems. These questions (shown in Figure 19) are important for increasing user understanding and facilitating greater transparency of the AI system.

- ❖ What metrics are used to measure system performance? Why are those the correct metrics?
- ❖ What is an acceptable minimum performance threshold?
- ❖ What is the process for detecting and correcting errors? What types of errors would cause serious problems?

Figure 19. Evaluation Questions: System/Model Performance

[287] Neil C. Rowe, "Algorithms for Artificial Intelligence," *Computer* 55, no. 7 (July 2022): 102, https://doi.org/10.1109/MC.2022.3169360.

[288] Muhammad Zohaib Khan, "Comparing the Performance of NLP Toolkits and Evaluation Measures in Legal Tech" (master's thesis, Passau, Germany, Universität Passau, 2021), 55–60, 69–71, http://arxiv.org/abs/2103.11792.

4. User Experience

User experience (UX)—perceptions and responses resulting from the interaction with a product or system—is influenced by an array of system factors such as presentation and functionality as well as individual user attributes such as experience, beliefs, and emotions.[289] Given its "dynamic, context-dependent, and subjective" nature, UX can be difficult to assess in a standardized way, but it is not something that is "overly subjectivistic, where prediction of and design for experience would become futile."[290] Usability tests offer a formal method to evaluate the effectiveness, efficiency, and satisfaction of human-AI interaction.[291] Albert et al. discuss various metrics used to assess UX (shown in Table 4).

Evaluating UX can be an extensive process—as indicated by the number of suggested UX metrics in Table 4 and the ample existing literature on usability testing. Usability tests involving a controlled environment with human subjects are ideal for collecting data and isolating variables, but they are unrealistic in an operational context. Rather than relying on a perfectly constructed experiment or test, the more practical approach would be to make an educated initial assessment and then allow for continual prototyping and iteration. Thus, a top consideration should be what kind of feedback mechanism exists between the user, the tool, and the AI developer. Feedback is not only essential for improving the user interface, but it is also important for maintaining and improving the performance of the AI, given that it is a learning system. The way that feedback occurs will depend on the particular system and the capacity of the vendor or AI developer. Nonetheless, the ease in which a user can share his or her

[289] International Organization for Standardization, "ISO 9241–11 Ergonomics of Human-System Interaction — Part 11: Usability: Definitions and Concepts," ISO Online Browsing Platform (OBP), 2018, sec. 3.2.3, https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en.

[290] Effie Lai-Chong Law et al., "Understanding, Scoping and Defining User Experience: A Survey Approach," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY: Association for Computing Machinery, 2009), 722, https://doi.org/10.1145/1518701.1518813.

[291] Morten Hertzum, "Usability Testing: A Practitioner's Guide to Evaluating the User Experience," *Synthesis Lectures on Human-Centered Informatics* 1, no. 1 (March 9, 2020): 10, https://doi.org/10.2200/S00987ED1V01Y202001HCI045.

concerns and recommendations as well as the responsiveness of the developer will impact the overall UX.

Table 4. Usability Metrics.[292]

Type of Metric	Method/Measure
Performance	Task success
	Time on task
	Errors
	Efficiency
	Learnability
Issue-Based	In-person studies (e.g., think aloud protocol)
	Automated studies
	Severity ratings
	Frequency of issues
	Issues by category or task
Self-Reported	Rating scales
	Post-task ratings
	Postsession ratings
	System Usability Scale (SUS)
	Online opinion surveys
	Assessing specific attributes/elements
	Open-ended questions
	Checking awareness, comprehension, usefulness gaps
Behavioral & physiological	Eye tracking
	Measuring emotion
	Stress (e.g., heart rate, skin conductance)
Combined & Comparative	Combining metrics based on target goals, percentages, Z scores
	Single usability scores
	Usability scorecards
	Comparison to goals and expert performance

The ability to test the tool directly (i.e., trialability) is a key factor for successful adoption and it enables practitioners to assess UX more effectively. Therefore, the OIE unit should strive to obtain trial accounts during the evaluation process. Getting access to

292 Source: Bill Albert, Tom Tullis, and William Albert, *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics* (San Francisco, UNITED STATES: Elsevier Science & Technology, 2013), http://ebookcentral.proquest.com/lib/ebook-nps/detail.action?docID=1204543.

these accounts, however, may not be possible at the early stages of an evaluation, and practitioners may have to settle for a demo instead. Even if practitioners are unable to test the tool directly, they should consider several key characteristics during their initial appraisal of the tool's expected UX. These characteristics can be grouped into three overlapping categories: ease of use, compatibility, and explainability.

Ease of use—a term taken from technology acceptance literature—assesses the amount of effort required to utilize the system.[293] A demo alone will not be able to satisfy the level of fidelity needed to determine ease of use, but several questions can probe into the features that affect ease of use. For instance, what kind and how long is the required training for the tool? Is the tool accessible in different environments (e.g., external and DOD networks)? What collaboration features are available? Formulating questions aimed at ease of use will prompt evaluators to think through which UX features should be valued higher within a specific use case.

Closely related to ease of use is compatibility. An influential component of adoption and the intensity of technology use, compatibility is defined as "the degree to which a product is consistent with existing values and experiences."[294] This interpretation of compatibility can be applied broadly and could imply significant subjectivity. Karahanna et al. disaggregate compatibility into four constructs: compatibility with preferred work style, existing work practices, prior experience, and values.[295] For the evaluation of UX, how the tool integrates into existing workflows

[293] Fred D. Davis, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly* 13, no. 3 (1989): 320, https://doi.org/10.2307/249008.

[294] John T. Gourville, "Eager Sellers and Stony Buyers: Understanding the Psychology of New-Product Adoption," *Harvard Business Review*, June 1, 2006, 2, https://hbr.org/2006/06/eager-sellers-and-stony-buyers-understanding-the-psychology-of-new-product-adoption; Mohammad Daradkeh, "Determinants of Self-Service Analytics Adoption Intention: The Effect of Task-Technology Fit, Compatibility, and User Empowerment," *Journal of Organizational and End User Computing (JOEUC)* 31, no. 4 (2019): 35, https://doi.org/10.4018/JOEUC.2019100102.

[295] Elena Karahanna, Ritu Agarwal, and Corey M. Angst, "Reconceptualizing Compatibility Beliefs in Technology Acceptance Research," *MIS Quarterly* 30, no. 4 (2006): 781–804, https://doi.org/10.2307/25148754.

(i.e., compatibility with existing work practices) is of particular interest, since it affects to what extent a user has to change processes and apply effort to use the system. Assessing this criterion requires an understanding of where and how the tool can be applied in operations. The practitioner should examine to what extent the tool enhances efficiency while also keeping in mind potential tradeoffs with effectiveness. Compatibility can also apply in a software or structural context, where the concern is interoperability with existing data, systems, and infrastructure.

Explainability is another key component in evaluating UX of an AI tool. As with ease of use and compatibility, explainability is inherently subjective and will differ based on different types of users. The topic encompasses a broad range of approaches, techniques, and assessments. Although explainability is ultimately concerned with allowing users to comprehend and trust the AI system, much of the literature focuses on incorporating technical improvements to models to produce explainable outputs. DARPA's explainable AI (XAI) program sought to develop new or modified ML and explanation techniques by exploring approaches such as tractable probabilistic models, causal models, visual saliency maps, and GAN dissection.[296]

Nonetheless, from the practitioner's perspective, the specific technique selected by the developer matters less than the user's actual comprehension of the reasons for the system's outputs. To determine whether the system is providing the right level of explainability, there needs to be an understanding of how the output is supposed to connect back to the larger task or mission requirement. An AI tool that is providing predictive analysis of a TA's behavior will require greater detailed explanation of the factors that the model is considering than an AI generated image used for PSYOP product development.

Given these key considerations for UX, practitioners should ensure that questions listed in Figure 20 are assessed during the evaluation process.

---

[296] David Gunning et al., "DARPA's Explainable AI (XAI) Program: A Retrospective," *Applied AI Letters* 2, no. 4 (2021): 3, https://doi.org/10.1002/ail2.61.

- *Ease of Use*
  - Does the tool require any special knowledge or expertise?
  - What training is required to use the tool effectively?
  - In what environments is the tool accessible (e.g., external or DOD networks)?
- *Compatibility*
  - Is the AI tool able to exchange and share data/information with other existing systems?
  - Does the usage of the AI tool interfere with any existing systems?
  - What collaboration features are incorporated into the tool?
- *Explainability*
  - Does the tool provide explanations or context to its outputs?
  - Does the explanation improve the user's decision/task performance?
- What feedback mechanisms exist?

Figure 20. Evaluation Questions: User Experience

5. Sustainability

In addition to assessing the existing capabilities of the AI tool, practitioners should think about the sustainability of the system—or the ability to maintain its performance and provide value to the user and the organization. AI systems are especially sensitive to changes in data or the environment.[297] Thus, continual monitoring is necessary to ensure that the system maintains its expected performance levels over time and within degraded environments. Sustainability can be assessed in three ways: reliability, robustness, and resilience. Reliability is the ability to complete a task in a satisfactory manner under specified operating conditions.[298] Robustness refers to the system's ability to continue functioning despite adversarial attacks or faults within its components.[299] A system that can maintain its robustness over time can be considered

297 Paul-Lou Benedick, Robert Jérémy, and Yves Le Traon, "A Systematic Approach for Evaluating Artificial Intelligence Models in Industrial Settings," *Sensors* 21, no. 18 (2021): 9, https://doi-org.libproxy.nps.edu/10.3390/s21186195.

298 Georges Zissis, "The R3 Concept: Reliability, Robustness, and Resilience [President's Message]," *IEEE Industry Applications Magazine* 25, no. 4 (July 2019): 5, https://doi.org/10.1109/MIAS.2019.2909374.

299 J. Lee, M. Ghaffari, and S. Elmeligy, "Self-Maintenance and Engineering Immune Systems: Towards Smarter Machines and Manufacturing Systems," *Annual Reviews in Control* 35, no. 1 (April 2011): sec. 4, https://doi.org/10.1016/j.arcontrol.2011.03.007.

reliable. Resilience is similar to robustness in its ability to withstand disruptions, but it includes the ability to "return to a new stable situation" after being exposed to unforeseen events—unlike robustness which deals with anticipated challenges and known risks.[300]

Various studies examine methods to evaluate the reliability, robustness, and resilience of AI models. These studies often focus on the technical, algorithmic, and mathematical underpinnings of specific AI techniques, such as introducing perturbations, adversarial AI, or mitigation (e.g., soft error, permanent fault, timing error) techniques.[301] These technical evaluation methods are likely too complex for an average practitioner to implement. Therefore, similar to the evaluation of performance metrics, the expectation is not to fully investigate the details of the testing procedure but rather to verify that these validations and tests are being done throughout the development phase and into the post-deployment phase. The DIU's Phase III Deployment Worksheet provides a useful template for assessing continuous task and data validation, functional testing, harms assessment, and quality control.[302] Vendors should be able to provide information on how they will test the system for model drift and what actions they will take to rectify performance deviations. In addition, there should be an explicit discussion of the system's limitations so that users can understand what conditions could lead to fallibility. As with UX, the user and AI developer should ensure the existence of an effective feedback mechanism so that both sides can communicate any indications of performance deviations or adversarial attacks. Figure 21 provides a list of questions to evaluate sustainability.

[300] Lee, Ghaffari, and Elmeligy, sec. 4; Harry Jones, "Going Beyond Reliability to Robustness and Resilience in Space Life Support Systems," in *50th International Conference on Environmental Systems* (Lisbon, Portugal: Texas Tech University, 2021), 2, https://ttu-ir.tdl.org/handle/2346/87122.

[301] Hollen Barmer et al., *Robust and Secure AI* (Pittsburgh, PA: Carnegie Mellon University, 2021), https://doi.org/10.1184/R1/16560252.v1; Igor Buzhinsky, Arseny Nerinovsky, and Stavros Tripakis, "Metrics and Methods for Robustness Evaluation of Neural Networks with Generative Models," *Machine Learning*, July 6, 2021, https://doi.org/10.1007/s10994-021-05994-9; Muhammad Abdullah Hanif and Muhammad Shafique, "Dependable Deep Learning: Towards Cost-Efficient Resilience of Deep Neural Network Accelerators against Soft Errors and Permanent Faults," in *2020 IEEE 26th International Symposium on On-Line Testing and Robust System Design (IOLTS)*, 2020, 1–4, https://doi.org/10.1109/IOLTS50870.2020.9159734; Benedick, Jérémy, and Traon, "A Systematic Approach for Evaluating Artificial Intelligence Models in Industrial Settings."

[302] Dunnmon et al., Responsible AI Guidelines in Practice, 30–32.

- ❖ Who is responsible for the maintenance of the system?
- ❖ How will monitoring and auditing occur? How will the system be tested for model drift?
- ❖ Has adversarial disruption been considered?
- ❖ Is there a process for system rollback?

Figure 21. Evaluation Questions: Sustainability

6. Scalability

Along with conducting an individual-level assessment of the AI tool, information practitioners should evaluate the scalability of the tool. The Carnegie Mellow University Software Engineering Institute (SEI) defines "scalable AI" as the "ability of algorithms, data, models, and infrastructure to operate at the size, speed, and complexity required for the mission."[303] The SEI identifies three areas of focus regarding scalable AI: scalable management of data and models; enterprise scalability of AI development and deployment; and scalable algorithms and infrastructure.[304] Although these focus areas are centered mainly on AI engineering enhancements and solutions, they remain pertinent to the broader discussion of scalability. To harness AI technologies across the services, the DOD needs improvements in production pipelines, architecture, and infrastructure, as well as the capacity to recombine and reuse data and models. How the DOD shapes enterprise-level infrastructure and policies will impact operational and tactical level practitioners as they work to integrate AI-enabled tools. Therefore, while addressing the questions in Figure 22, OIE units and evaluators should consider the organization's readiness (e.g., leadership support, doctrine, training) to adopt the AI tool.

---

[303] Hollen Barmer et al., *Scalable AI* (Pittsburgh, PA: Carnegie Mellon University, 2021), https://doi.org/10.1184/R1/16560273.v1.

[304] Barmer et al.

- ❖ How many users should be considered for this tool? Is large-scale usage feasible?
- ❖ What are the infrastructure requirements to deploy this tool at the organizational-level?
- ❖ What additional resources or personnel are required to scale this tool? Can these resources be sustained over the long-term?

Figure 22. Evaluation Questions: Scalability

7. Affordability

As with any other procurement or acquisition, affordability should be a factor during the decision-making process. The cost of acquiring and maintaining the tool will directly affect other evaluation domains, such as sustainability and scalability. Although affordability may seem like a glaringly obvious and simple factor to consider, the characteristics of AI necessitate a more careful review of the costs associated with adopting the technology. Initial discussions about pricing will likely center around the cost of a prototype or a set of licenses. However, OIE units should account for long terms costs, given that AI technologies, like software products, require continual maintenance and updates.

AI raises additional issues, such as the need to access large amounts of data, which can amount to significant costs. For example, training GPT-3, a recent large language model, is estimated to cost almost $5 million, given the massive data needed to train the model.[305] Additional infrastructure requirements to include security monitoring and centralized compute and storage resources can add to costs. These cost-incurring items emphasize the importance of considering the tool's interoperability with existing infrastructure and systems. Less compatibility will lead to the need to create new or substantially modified architecture or configurations that further increase the cost. Data cleaning and labeling are crucial steps in the training and deployment of AI, but they can

[305] Will Knight, "AI's Smarts Now Come With a Big Price Tag," *Wired*, October 14, 2021, https://www.wired.com/story/ai-smarts-big-price-tag/.

often be laborious processes that may require additional resources or external partnerships. Due to these additional considerations, information practitioners should recognize the questions in Figure 23 as a crucial part of the evaluation process.

- ❖ What is the cost for a prototype? What is the maintenance cost?
- ❖ How does the proposed cost compare to other offerings?
- ❖ Are there existing capabilities or contracts that can augment the solution or address certain components of the problem?

Figure 23. Evaluation Criteria: Affordability

8. Ethical/Legal/Policy Considerations

This evaluation domain lies at the cross-section between organizational and environmental dimensions of the TOE framework since ethical, legal, and policy considerations can have a driving effect both internal and external to the organization. As discussed in Chapter III, effective human-machine teaming and adoption of AI require trust. While the concept of trust involves a significant degree of complexity, adherence to ethical principles serves as a bedrock for fostering trust in the technology. Figure 24 illustrates how the DOD's existing ethical, legal, and policy frameworks underlie its *RAI Strategy*, which views trust as its end state.[306]

[306] DOD Responsible AI Working Council, U.S. Department of Defense Responsible Artificial Intelligence Strategy and Implementation Pathway, 7.

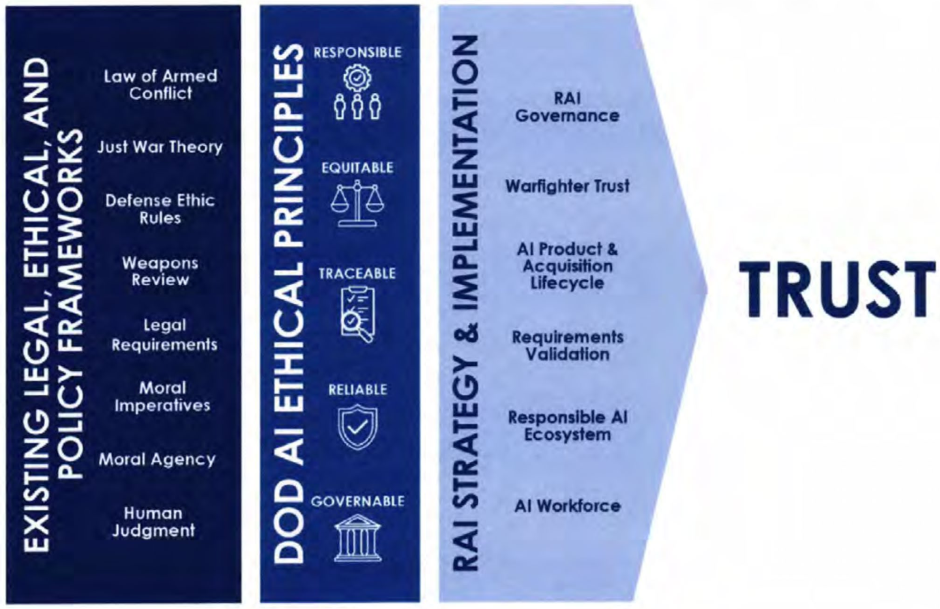


Figure 24. DOD RAI Journey to Trust.[307]

These frameworks and principles—which should guide the development of an AI tool since its inception—serve as cornerstones within the evaluation process. Ensuring that operations stay within the boundaries of law and policy is not a new practice for those in the OIE community. Nonetheless, the rapid advancement of AI technology can introduce new complications not addressed in current legal or policy frameworks. Ethics are paramount in cases where these legal and policy gaps exist. Hence, practitioners should explicitly consider the questions in Figure 25 to identify any ethical, legal, or policy concerns that may inhibit the use of the AI tool or damage trust with the American people, within the DOD, or between humans and the machine.

307 Source: DOD Responsible AI Working Council, 7.

- ❖ Are there any legal or policy restrictions that prevent us from using this tool?
- ❖ What ethical concerns may arise from using this tool? How does the use of this tool impact public or user trust?

Figure 25. Evaluation Criteria: Ethical/Legal/Policy Considerations

9. Vendor Assessment

Along with the technology itself, the quality of service provided by the vendor is an essential factor to consider when evaluating an AI tool. The development of AI is an iterative process requiring continual feedback between users and AI developers. Trust is not only required with the technology but also with the entity developing the technology. This trust with the AI developer is particularly critical because the nature of the technology makes transparency and explainability exceptionally difficult. In some cases, AI could exhibit emergent behavior in which the developers themselves may not fully realize or understand why the AI produced a particular output or decision.[308] Therefore, the user may be left to rely on the developer's trustworthiness and diligence in adhering to the RAI principles and creating the appropriate models to meet operational needs.

Ultimately, the relationship between the vendor and the adopting organization is a partnership. A vendor's willingness to dedicate substantial initial investment and work directly with prospective users can generate advocates within the OIE community, which can increase the likelihood of the AI tool's adoption.[309] The vendor's track record and reputation are key pieces of information to determine whether the partnership would be worth pursuing. Although AI models are becoming increasingly accessible, OIE units will likely continue to rely on significant levels of support, given the lack of available personnel and technical expertise. Some vendors with large human analyst capacity may

[308] Shelley P. Gallup, "Future War at Sea: The U.S. Navy, Autonomy in War at Sea and Emergent Behaviors," in *Emergent Behavior in System of Systems Engineering*, ed. Larry B. Rainey and O. Thomas Holland (Boca Raton, FL: CRC Press, 2022), 68.

[309] Information practitioner, personal communication, October 28, 2022.

focus more on providing finished analytical products rather than expecting practitioners to utilize the tool directly. The decision to select a more service-oriented solution or obtain an easy-to-use tool that practitioners can leverage directly will depend on organizational preference and operational requirements. Nonetheless, adopting organizations should formally assess vendors using the questions in Figure 26, especially since vendor reputations are often shared via word-of-mouth.

- ❖ Is the company committed to adhering to DOD's Ethical Principles for AI?
- ❖ What is the company's reputation for working with the DOD?
- ❖ What support is the vendor willing to provide to effectively integrate the tool into the organization?

Figure 26. Evaluation Criteria: Vendor Assessment

E. CONCLUSION

The evaluation guidelines presented in this chapter are not meant to be all-encompassing, nor do they suggest that one set of evaluation criteria is unequivocally superior in all instances. The primary purpose is to spur discussion, assist with decision-making, and raise additional points of concern that could impact the effective use and adoption of AI technologies. Practitioners are not expected to comprehend all the technical intricacies of AI, but baseline recognition of key considerations is needed for informed decision-making.

These guidelines should be nested within the DOD's *RAI Strategy*, acquisition pathways, and T&E framework. OIE units should strive to create multidisciplinary teams to conduct these evaluations. However, scarcity in time, resources, and personnel—specifically, technical expertise—may render this notion difficult. Instead, practitioners may be left to evaluate tools in the absence of full teams of technical experts. Thus, this evaluation framework serves as a guideline for practitioners to make an initial assessment, which can then be expanded by referring to technical experts (who may be only available at higher echelons) for further examination.

Evaluations are iterative processes requiring continual feedback loops to enhance the tool's utility and foster understanding between the developer and the user. The problem statement and key tasks identified during the preparation phase should drive the rest of the evaluation process, as well as the prioritization of the evaluation criteria. Conducting comparative analyses across different tools that address a particular problem enables the evaluator to ascertain what capabilities currently exist within the market. Without this comparative analysis, attempts to determine relative advantage would be less than fully credible. Hence, this evaluation framework provides practitioners with a method to standardize processes and identify key characteristics of the technology that meet operational and user requirements.

THIS PAGE INTENTIONALLY LEFT BLANK

VII. CASE STUDY

A. CHAPTER OVERVIEW

The purpose of this chapter is to test the application of the information practitioner's evaluation framework (PEF) on an AI-enabled OIE tool. Pulse, a tool developed by IST Research (now known as Two Six Technologies), was selected as the subject of this case study. The tool is currently used operationally by OIE units within U.S. Army Special Operations. Although the intent is to use the PEF as an initial assessment prior to the acquisition of a tool, the framework can be modified to support other types of evaluations (e.g., performance monitoring, implementation analysis) that occur later in the post-acquisition or deployment phase.

B. BACKGROUND

Pulse is a platform that enables "data collection and two-way engagement" using "cloud analytics and advanced data processing tools…to scrape open data across the surface, deep, and dark web."[310] The platform was originally created in 2008 by IST Research to enable two-way information sharing, polling of local populations, and data analysis for teams working in developing areas of the world.[311] The tool has been used by a range of government agencies including the U.S. Agency for International Development (USAID), Department of State, Department of Homeland Security, DOD, and the intelligence community.[312]

Over the years, the platform became increasingly designed for information practitioners specifically. In 2012, the Combined Joint Psychological Operations Task

---

[310] Two Six Technologies, "Pulse: Enabling Data Collection and Two-Way Engagement," Two Six Technologies | Advanced Technology Solutions for Critical Missions, 2022, https://twosixtech.com/products/pulse/.

[311] "Pulse," IST Research, 2014, https://web.archive.org/web/20151127082326/http://istresearch.com/pulse/.

[312] Two Six Technologies, "IST Research and Two Six Labs Combine to Form High-Growth Government Technology Platform," GlobeNewswire News Room, February 1, 2021, https://www.globenewswire.com/news-release/2021/02/01/2167580/0/en/IST-Research-and-Two-Six-Labs-Combine-to-Form-High-Growth-Government-Technology-Platform.html.

Force awarded a contract to IST Research to develop an "SMS Text Messaging Program" to deliver tailored SMS messages to target audiences (TAs). Between 2013 and 2020, IST Research was funded by DARPA through several iterations of the Small Business Innovative Research Program (SBIR) Phase 2 to further develop Pulse.[313] The investments were aimed at scaling the platform, building a tool that infers intent from social media, and creating a data collection environment where operators can create an "information collection campaign from any device" and visualize the collected information.[314] Special Operations Command Africa awarded the company a five year Phase 3 SBIR contract in 2020 to engage with the local population, understand sentiment, and measure effectiveness in support of counter disinformation efforts.[315] Currently, OIE units across various commands have access to Pulse.

To facilitate further growth, IST Research and Two Six Labs merged in 2021 to form Two Six Technologies under the Carlyle Group.[316] The expansion has allowed the company to increase their product offerings to include IKE (cyberwarfare platform), M3 (media manipulation monitor), TrustedKeep (security platform), and SIGMA (chemical, biological, radiological, nuclear detection platform) as well as maintain a global presence in 40 countries.[317] In particular, Two Six Technologies is exploring ways to combine the capabilities of M3 with Pulse to monitor information manipulation, gain insight into censored content, and deliver content to global audiences on their own devices.[318]

---

[313] Ryan Paterson, "DARPA Funds Development of Pulse Platform — IST Research," IST Research, September 13, 2013, https://web.archive.org/web/20160125090234/http://istresearch.com/news/2013/9/6/darpa-funds-development-of-pulse-platform; SIBR-STTR, "IST Research Corp.," SIBR-STTR: America's Seed Fund, accessed November 6, 2022, https://www.sbir.gov/sbc/ist-research-llc.

[314] SIBR-STTR, "IST Research Corp."

[315] Dave Nyczepir, "SOCOM Looks to Combat Disinformation in Africa on New Governmentwide Contract," *FedScoop*, July 27, 2020, https://www.fedscoop.com/socafrica-disinformation-ist-research/.

[316] Two Six Technologies, "IST Research and Two Six Labs Combine to Form High-Growth Government Technology Platform."

[317] Two Six Technologies, "Two Six Technologies Announces New Office in San Antonio, TX," Two Six Technologies | Advanced Technology Solutions for Critical Missions, August 18, 2022, https://twosixtech.com/news/two-six-technologies-announces-new-office-in-san-antonio-tx/.

[318] Two Six Technologies representatives, personal communications and email attachment to author, September 1 and November 14, 2022.

Pulse consists of both hardware and software components that conduct population engagement, social listening, and content discovery.[319] Practitioners use the platform to conduct remote information campaigns by pushing messages to devices or communication platforms that are used by the TA. During the early years of adoption within the U.S. Special Operations community, Pulse was primarily used for its SMS message dissemination capability.[320] Over the years, Two Six Technologies expanded its social listening platform, allowing practitioners to collect publicly available information (PAI) from the web, news sources, RSS feeds, and social media platforms such as Twitter, Facebook, Instagram, Telegram, Vkontakte (VK), Odnoklassniki (OK), YouTube, and Reddit. In content discovery, the practitioner utilizes a variety of dashboards and visualizations to analyze the ingested data.

C. APPROACH/DESIGN

Pulse was selected for this evaluation due to three reasons. First, the platform is developed primarily for OIE. Second, it utilizes an AI-enabled data pipeline and custom analytics.[321] Third, I was able to obtain an account to access the platform. Six other AI-enabled platforms were explored as possible options, but sensitivities associated with using the tools as subjects of open academic research inhibited their usability as part of this thesis.

While Pulse offers both passive monitoring and active engagement capabilities, this thesis focuses the evaluation on the social listening and content discovery functions of the platform—as these are the areas where AI technology is currently integrated. Furthermore, due to time, resource, and operational constraints, I was not able to test the hardware components of the Pulse kit nor the population engagement feature of the tool.

---

[319] Nyczepir, "SOCOM Looks to Combat Disinformation in Africa on New Governmentwide Contract."

[320] Three information practitioners, personal communications, August 19, September 14, October 28, 2022. Individuals had experience with Pulse for SMS dissemination but not for PAI analysis.

[321] Two Six Technologies, "Pulse: Enabling Data Collection and Two-Way Engagement."

Prior to this thesis, I was aware that Pulse was a tool used within the PSYOP community, but I did not have direct interaction with the platform during previous operational experience. Therefore, Two Six Technologies recommended that I receive training on the tool before using the platform. Table 5 provides a breakdown of the engagements conducted with Two Six Technologies.

Table 5. Engagements with Two Six Technologies

Engagement	Duration
Initial Introduction/Onboarding	1 hour
Training	6 hours (2 days of 3 hours)
M3 Deep Dive	45 minutes
Follow-up Interview	1 hour
Narrative Comparison Deep Dive	1 hour

To evaluate Pulse, this thesis generally follows the framework proposed in Chapter VI. However, since Pulse is already acquired by multiple organizations within the DOD, an adapted approach was needed in which aspects of the analysis were conducted retrospectively. While the "prepare" phase of the evaluation process requires an examination of the problem or gap before selecting a particular solution, the capability gap within this case study has already been outlined by units in various information papers, concepts of operations, and required operational capability documents. This thesis assumes that adequate planning and discussion have occurred to identify the problem that Pulse is aiming to address. Additionally, this thesis deliberately omits any information that may be considered sensitive or controlled. Typically, robust conversations would occur surrounding specific task alignment, prioritization, use cases, and costs, but these topics are discussed here with limited details to protect proprietary information and operational security.

D. EVALUATION RESULTS

Multiple documents reflecting operational requirements and shortfalls within the PSYOP Regiment show that the main capability gaps include the limited ability to:

1. Conduct digital message dissemination to include surveys and polls (Gap #1)
2. Aggregate, process, query and analyze large volumes of PAI (Gap #2)
3. Determine measures of effectiveness (MOE) in the cyber and information domains.[322] (Gap #3)

The requirements documents also identified several additional characteristics that were important for consideration. First, the tool should provide efficiency in receiving feedback from the TA and analyzing data from multiple sources. Second, the tool should be able to provide the ability to work anonymously within the online environment to ensure security. Third, the tool needs to be accessible in deployed conditions as well as usable by teams in the continental United States. Furthermore, it should be available for teams at the lowest tactical level.

The level of specificity within the requirements documents appears to increase over time. Later dated documents (circa 2021) specify features such as network mapping (e.g., social network analysis), image classification, and narrative amplification. None of the documents, however, indicated a prioritization of requirements, criteria, or desired features of the tool. Several of the documents did address feasibility of adoption by outlining DOTMLPF-P (doctrine, organization, training, material, leadership and education, personnel, facilities, and policy) implications.

1. Mission/Task Alignment

Pulse addresses the major capability gaps identified by the OIE unit. Although not evaluated in detail, interviews and vignettes from the operational force indicate that

[322] Information derived from a collection of requirements documents provided by U.S. Army Special Operations Command. Due to sensitivities, information has been summarized rather than individually cited and quoted.

PSYOP teams have been able to leverage its messaging capability to support various missions and specifically address Gap #1.[323] Pulse's social listening application enables users to define PAI collection criteria utilizing a rules management system. Once the rules are created and deployed, the platform collects, processes, enriches, and stores the data for users to query. The user then leverages the discover application to visualize and analyze the data. Thus, Pulse addresses Gap #2 by creating a way for information practitioners to harness open-source data to answer specific mission requirements. The ability to analyze this data, however, will depend on user proficiency in tailoring collection requirements and interpreting results. Gap #3 is more complicated and requires further examination of how the tool fits into existing workflows and how it supports a practitioner's assessment of MOE. Pulse's data ingestion and analytical applications support the collection of measures of performance (MOP), but the determination of effectiveness will require deeper human analysis that connects relevant online indicators with variables in the physical dimension. The extent to which Pulse is deemed useful for this effort depends on how the tool is used by the practitioner.

Relative advantage is difficult to assess, given that this tool was evaluated independently from others. Nonetheless, several features of Pulse stand out from many of the tools researched for this thesis. The most notable advantage is its ability to offer a combination of active engagement and passive monitoring capabilities. Pulse's core function of directly messaging remote populations in developing countries is a notable feature (one that was not previously leveraged by OIE units). Combined with the platform's social listening and analysis capabilities, it provides a potent tool for fusing information from multiple communication channels. Therefore, even though other analytical platforms also provide the ability to ingest, process, query, and analyze PAI, Pulse's "two-in-one" offering may make the tool more desirable.

[323] Interviews, vignettes, and concepts of operations from U.S. Army Psychological Operations obtained by author, October 28, 2022.

2. Data

Pulse's social listening tool extracts and enriches online data using either application programming interface (API) or web scraping. Two Six Technologies maintains the unique fields collected from the various data sources for any client that is interested. The data undergoes normalization and a combination of data enrichments depending on the type of data and analysis (e.g., sentiment analysis). The enrichment process involves a series of sub-steps beginning with extraction. For each enrichment model, datasets are divided into training and test sets.

The tool does not collect data from the entire internet, but rather focuses its data collection on rules established by the user or curated by Two Six Technologies. The 22 different rule types range from data collection of specific websites to public postings in Telegram, Facebook Twitter, and other social media platforms. While some data is streaming (e.g., Twitter, Telegram), users can specify how often to scrape other data sources; these scrapes can occur every 12 hours, every day, or as a one-time pull. Collection from RSS feeds allows for a refresh rate of every 15 minutes. Geodata can also be collected if the account provides the requisite information in the profile description or if the account enables the sharing of location services while using the social media platform. In an effort to minimize concerns over attribution, Pulse automatically converts usernames into user IDs when the request for collection is submitted.

There are several limitations that users should account for. First, all the data collected is public, which means that private pages or accounts will not be included in the data lake. Second, Pulse only collects on existing and active accounts, pages, or URLs. Therefore, the results do not include pages or accounts that have been taken down by the social media platform. Third, geodata is severely limited (only 1.5% of tweets contain geodata).[324] Therefore, users should expect to obtain limited location data from social media content. Fourth, each social media platform has its own specific collection

[324] Two Six Technologies representatives, personal communication, August 26, 2022.

restriction. For instance, Twitter only allows the historical collection of 1,000 tweets for each rule. These restrictions may vary over time depending on the social media company's policies. Finally, there is a notable absence of platforms such as Weibo, WeChat, TikTok (Douyin), and LINE on its social listening application—although Two Six Technologies is leveraging M3 to conduct analysis in hard-to-reach information domains and there are plans to incorporate Weibo and WeChat in the near future.[325]

3. System/Model Performance

Two Six Technologies utilizes a variety of NLP models to enrich and analyze the data. The company conducts multiple internal tests to identify and address errors or model drift. Unit tests are used to check whether processors are running correctly. Regular tests are also performed to assess if and how the collection process begins to decline. Analysts examine potential reasons for the decline, which can stem from lower post volume or disruptions due to a system update. For evaluations of data enrichments, the company compares the enrichment of a sample group of documents using in-house models with the outputs of the enrichment that is performed with other publicly available tools. Analysts conduct a blind test to determine which models work the best.[326]

During an interview, representatives from Two Six Technologies explained the general process of addressing concerns over errors and performance metrics. The company stated that they utilize F1, recall, precision, and other standardized scores, which are based on "internal tests" and "industry standards."[327] These metrics are communicated internally before the deployment of a model or when conducting comparative analysis with other external models. The company does not provide model cards, largely because none of their current customers have requested them. Although the release of training and evaluation data may be limited (due to proprietary reasons), they stated that information on performance metrics is available and they are willing to

[325] Two Six Technologies representative, email attachment to author, November 14, 2022.

[326] Two Six Technologies representatives, personal communication.

[327] Two Six Technologies representatives.

provide it to customers.[328] Nonetheless, this information was not readily available, and it required the primary customer-facing representative to make a separate request to the Two Six Technologies technical team.

Apart from model performance metrics, a simple user test can provide information about the performance of the tool overall. To test the outputs of the system, I created a project within the social listening application to collect information on the Democratic People's Republic of Korea's (DPRK) recent provocations and missile launches. For the purposes of this test project, I adopted the role of a novice user with minimal training (i.e., six hours) and limited knowledge on data and statistics. In this case, I wanted to assess the extent to which Pulse can be used as a general situational awareness tool for someone who is not familiar with the topic of interest. Therefore, I relied on keyword searches within the social listening application, building out rules based on the keywords in Table 6. The keywords were translated through the combination of Google Translate, Naver's Papago Web Translator, and a review of several Korean news articles. The keywords were then inputted into five sets of seven rules that derived data from a web search, a news search, Twitter, VK, OK, YouTube, and Reddit. Recurrence of the scrape was set for every 12 hours, and the results were to be proxied from the country of the language being searched (e.g., a search of Korean keywords would return results proxied from the Republic of Korea).

Within a seven-day period, 408,566 documents were collected in which 188 of them contained geodata. Twitter accounted for 93.43% of the documents collected. Documents were also derived from web crawl (4.55%), YouTube (1.95%), VK (0.05%), and Reddit (less than 0.01%). Approximately 80% of the documents were in Japanese, while 13.24% were in Chinese.

[328] Two Six Technologies representatives.

Table 6. List of Keywords Used in Pulse

English	Korean	Simplified Chinese	Traditional Chinese	Japanese
North Korea	북한	北朝鲜	北朝鮮	北朝鮮
ballistic missile	탄도 미사일	弹道导弹	彈道導彈	弾道ミサイル
nuclear weapon	핵무기	核武器	核武器	核兵器
artillery	대포	大炮	大炮	大砲
DPRK	공식 명칭	正式名称	正式名稱	公式名称
Pyongyang	평양	平壤	平壤	平壌
Kim Jong-un	김정은	金正恩	金正恩	金正恩
military drill	군사 훈련	军训	軍訓	軍事訓練
US-Korea	한미	韩美	韓美	韓米
missile	미사일	导弹	導彈	ミサイル
short-range ballistic missiles	단거리탄도미사일	短程弹道导弹	短程彈道導彈	短距離弾道ミサイル
East Sea	동해	东海	東海	東海
West Sea	서해	西海	西海	西海
Yellow Sea	황해	黄海	黃海	黄海
Japan	일본	日本	日本	日本
Republic of Korea	대한민국	大韩民国	大韓民國	大韓民国
Vigilant Storm	비질런트 스톰	维吉伦斯风暴	維吉倫斯風暴	ビジレントストーム
Worker's Party	노동당	劳动党	勞動黨	労働党
Pak Jong-chon	박정천	朴正天	樸正天	パク·ジョンチョン
buffer zone	완충지대	缓冲地带	緩衝地帶	緩衝地帯
South Korean military	한국군	韩国军队	韓國軍隊	韓国軍
provocation	도발	挑衅	挑釁	挑発
Joint Chief of Staff	합참	联合参谋本部	聯合參謀本部	合参
military demarcation line	군사분계선	军事分界线	軍事分界線	軍事境界線
combined flight exercise	연합비행훈련	联合飞行训练	聯合飛行訓練	合同飛行訓練
Central Military Commission	중앙군사위원회	中央军事委员会	中央軍事委員會	中央軍事委員会
North Korea's Ministry of Foreign Affairs	북한 외무성	朝鲜外务省	朝鮮外務省	北朝鮮外務省
President Yoon Seok-yeol	윤석열 대통령	尹锡悦 总统	尹錫悅 總統	尹錫悦大統領
National Security Council	국가안전보장회의	国家安全保障会议	國家安全保障會議	国家安全保障会議
DMZ	비무장지대	非军事区	非軍事區	非武装地帯
USS Ronald Reagan				
USS Key West				

At first glance, the content within the documents did not seem to address the topic of DPRK provocations or missiles. 原神 (translated to Genshin, which is a videogame) was clearly the top hashtag by over eight thousand mentions. 〆のチキンラーメン (translated to chicken ramen) ranked second. The images collected for this project reflected these hashtags as well as other Japanese advertisements. However, utilizing the filter features on the dashboard, I was able to find documents that were of greater interest. For example, fengyunshe, GTV26543476, and weizhenshe were three of the top ten mentioned user accounts. Further analysis reveals that these three accounts are staunch opponents of the Chinese Communist Party and post heavily about world news events.

This user test underscores the criticality of clearly specifying the parameters of the data collection. Merely relying on keyword searches (a technique that Two Six Technologies does not recommend) will lead to the inclusion of large amounts of irrelevant data and may not result in data that aligns with the user's original intent. Although the platform's visualizations and filter system enable users to focus on tailored portions of the data as well as conduct analysis down to the individual level, the ability to leverage these features can require advanced skills and may be time consuming; these factors should be considered when assessing user experience.

4. User Experience

Due to the tool's sophisticated functionality, general users would require considerable training to be able to use the tool effectively. According to Two Six Technologies, the length of initial training can vary depending on the desires of the customer and how much of a "deep dive" they are looking to do. Basic user training is typically four days in order to incorporate practical exercises to make sure that the trainees can understand and use the system.[329]

For this thesis, I received a condensed version of the training—six hours over two days, which enabled me to have a solid understanding of how the social listening application works as well as an overview of commonly utilized dashboards and visualizations. During the training, the Two Six Technologies representative provided a step-by-step demo of how one might create a campaign, build and manage collection rules, and set up a dashboard to visualize the collected data. Yet, this training only touches the surface of how a practitioner might use the platform.

The tool does offer several features to increase usability and ease-of-use, including a user guide that gives instructions on the basic functions of the platform. Dashboard and visualization templates are also posted in the discover application. Users can clone the provided templates and modify them based on specific requirements. Each template provides an overview description and instructions on how to start tailoring the

[329] Two Six Technologies representative, personal communication, August 30 and November 2, 2022.

dashboard. Once a dashboard is customized, little effort is needed to maintain it, since the data is automatically updated and fed into the dashboard until the specified stop date. Users, however, should take note of what campaigns or projects are being fed into the dashboard. The ingestion of unnecessary data from irrelevant campaigns or projects could increase the noise of the dataset or skew the analysis.

Using the tool independently—particularly the discover application—post-training proved to be more challenging than expected even though the training provided a good foundation of how to navigate through the system. Users will need to understand Boolean expressions to query or filter the data properly, since the platform utilizes Lucene, Regex, and Kibana Query Language, which "does not support regular expressions or searching with fuzzy terms."[330] Two Six Technologies representatives recognize that the features within the discover application will likely require more advanced knowledge and skills.[331] Therefore, the company offers analytical services that can produce finished products from specific requirements dictated by the user as well as services that allow for a user to sit side-by-side with a data scientist to create tailored dashboards, queries, and visualizations. Two Six Technologies stated that while there is no standardized feedback process embedded within the platform, users should be able to obtain direct access to a Two Six Technologies team that would be able to respond to requests. In the past, users have provided verbal feedback.[332]

Pulse's web-based platform makes it accessible on the DOD and external networks. There are no indications that the tool interferes with any existing systems. Efforts are underway to increase the integration of Pulse with other information systems in the special operations community. There are ongoing discussions about how to leverage the data and analysis from Pulse for other systems as well incorporate other data into the platform. The platform does support the ingestion of third-party data through file

[330] "Kibana Query Language," elastic, 2022, https://www.elastic.co/guide/en/kibana/7.17/kuery-query.html; Two Six Technologies representative, email attachment to author, November 14, 2022.

[331] Two Six Technologies representatives, personal communication, November 4, 2022.

[332] Two Six Technologies representatives.

uploads using delimited text files, JSON, and log files, as well as programmatically if APIs are available.[333]

One area that remains a challenge within Pulse is explainability. The platform allows the user to dive into the granular details of the data and analysis, especially in the discover tab. Users can view all the data fields, enrichment results, scores, and probabilities down to each individual document. Yet, the context or explanation of how results or scores were calculated is not apparent. For example, a user can create several visualizations from the sentiment scores that might work well as a briefing product but there is little understanding of what would make a certain text more positive or negative unless a deeper examination is conducted to look at trends from individual documents. Therefore, new users may have a hard time ascertaining meaning behind the outputs, or they could jump to tenuous conclusions in their analysis.

5. Sustainability

Some elements of sustainability such as the internal monitoring and testing of the dataset and models were mentioned in Section 3 (System/Model Performance). Two Six Technologies will remain responsible for the maintenance of the system.[334] The platform undergoes regular updates and utilizes standard security measures, although there was no explicit discussion of potential adversarial disruption or a process for a system rollback. Although regular testing occurs on processors, data ingestion, and enrichment, there could still be a risk of data breach, misuse, or reverse engineering.

6. Scalability

Pulse's broad customer base suggests that the platform is scalable. The number of users that could use the tool did not seem restricted. While users may be given a warning if they surpass the data ingestion limit of 100 million documents per month, historical usage indicates that users tend to not exceed the limit.[335]

[333] Two Six Technologies representative, email attachment, November 14, 2022.

[334] Two Six Technologies representatives, personal communication.

[335] Two Six Technologies representative, personal communication, August 26, 2022.

As units seek to scale this tool, particular attention should be placed on the management of users and projects. Pulse provides a hierarchical structure in its social listening application in which users can organize rules and projects under a "campaign." The platform also allows for rule tagging and recommends that the tags follow a structured formula utilizing abbreviations and categories. These standardizations allow for collaboration among a wider group of users. Yet, if the number of users were to grow exponentially within an organization, managing a large number of campaigns and dashboards may become increasingly difficult. Although a user can search and filter for different campaigns, visualizations, and dashboards, keeping track of which ones are active, updated, and pertinent to a particular user can become unwieldy, since every user's campaign or dashboard is listed together if they are under the same command. Furthermore, all users within the same organization are able to edit other user's projects and dashboards, which can create problems if someone overwrites an existing project.

7. Affordability

Affordability was not explicitly assessed during this evaluation due to the sensitive nature of the information. Nonetheless, detailed discussions about appropriate costs are part of the normal acquisition process. Historical requirements documents examined in this thesis included a cost analysis, which accounted for licenses and training.

8. Ethical/Legal/Policy Considerations

Legal and policy restrictions appear to be well-established based on what is outlined in the existing contract. Restrictions would include collection or engagement of U.S. persons' data. Guidelines pertaining to authorities, permissions, and intended use are determined and approved prior to any agreement. Although no unique legal or policy restrictions were identified during conversations with the company, interviews with information practitioners reveal that obtaining permission to use the tool operationally could be challenging.[336] One practitioner cited a case in which a PSYOP team had

[336] Two information practitioners, personal communications, October 28 & 31, 2022.

trouble gaining approval from the Department of State to use the tool in country. By the time approval was given, the tool was not needed anymore.[337] These types of challenges are not necessarily new, but they should be taken under advisement.

Two Six Technologies recognizes ethical use as an important principle and there seems to be a conscious effort to maintain transparency of the tool.[338] Responsible and ethical use of the platform, however, is not deliberately discussed during user training. The assumption is that users should have a clear understanding of what the ethical boundaries are when conducting research. Users should be aware that bias could occur on both the collection and analysis side.

9. Vendor Assessment

Two Six Technologies is a well-established company that has been working with the DOD for over a decade under its previous name—IST Research. While the company is aware of the DOD's RAI efforts, it is not specifically familiar with the DIU's *RAI Guidelines*. Customer service appears to be promising and potentially a notable strength. Two Six Technologies representatives have been able to embed themselves within operational units to provide agile support. Recognizing that the platform is relatively complex, the company seems to be moving toward providing increased human analytical support. Analysts from Two Six Technologies are able to build finished products using a combination of Pulse and other analytical tools so that practitioners do not have to conduct the research themselves.

E. DISCUSSION

By considering each evaluation domain, the PEF assists with the delineation of the tool's advantages and areas that require extra consideration or caution when using the AI-enabled tool. Pulse's main advantages include its unique offering of both population engagement (that includes SMS) and online data aggregation and analysis, the level of

---

[337] Information practitioner, personal communication, October 31, 2022.

[338] Two Six Technologies representatives, personal communication, November 4, 2022.

access and visibility the user has with the ingested data, and the customer support provided by Two Six Technologies. These advantages are important because it provides a consolidated tool which deployed teams can use to engage with remote populations while also having U.S.-based personnel fusing the information together with analysis of the online environment.

Nonetheless, one key area which was difficult to ascertain in this evaluation was model performance. Although Two Six Technologies was open to providing these metrics, they were not easily accessible. The absence of performance metrics poses a challenge for the evaluator when trying to assess objective value of the tool. Although it is likely that users will be able to obtain Pulse's performance metrics at some point (especially since the company is already under contract within U.S. Army Special Operations Command), the lack of readily available metrics for the user indicates a broader issue confirmed by interviews with other vendors: communicating performance metrics to users has not been a standard practice.[339]

The evaluation reveals that there are several aspects of the tool that may hinder widespread adoption by individual users. While the tool is being used across multiple DOD organizations, active users of the tool who leverage Pulse's social listening and discover applications appear to remain within a small group of information practitioners. The challenge of gaining permissions to use the tool operationally is likely an influential factor in why wider usage of the tool does not currently exist. The shift in strategic focus from combat operations in Afghanistan and Iraq to global competition exacerbates this challenge. One PSYOP practitioner commented, "Everyone talks about Pulse, but it doesn't feel like anyone actually gets permissions to use it. It was easy to use in a combat environment, not so much in the current environment."[340]

Another important factor involves the complexity of the tool, which can hamper user experience. Without substantial training, an average user would have difficulty using

[339] Four different vendors developing AI-enabled tools, personal communication, August-November 2022.

[340] Information practitioner, personal communication, October 28, 2022.

the platform effectively. A considerable amount of research on the topic of interest is needed before using the tool if users want to tailor their social listening campaigns, reduce noise in the data, and obtain useful results. Time—a precious commodity especially for information practitioners—is also needed to build out campaigns, customize dashboards, and analyze the data results, which often requires in-depth and granular examination of the data to gain useful insights. Therefore, some practitioners may have negative views about the tool's ease of use, affecting effort expectancy. The amount of effort required to properly build social listening campaigns and ensure ingested data is relevant to the task can also affect performance expectancy. A cursory attempt to build a campaign will produce impractical results and could feed into a perception that the tool is not performing to the standard expected by the user.

One way to address this issue is limiting the usage of the tool to "super users" or those who have the skills, time, and interest to use the advanced features of the tool. Customer support provided by Two Six Technologies can also mitigate these challenges and help streamline the research. Nonetheless, organizations should assess whether they prefer a more service-oriented solution, a tool for specialized usage (by a group of advance skilled practitioners), or a tool adopted widely by users. While a service-oriented solution does engender additional concerns such as overdependency on external analysts or diminished visibility over the analytical process, it can greatly assist operational teams who do not have the time to conduct thorough analysis using the tool.

F. CONCLUSION

This case study demonstrated the feasibility of using the PEF to assess Pulse. The PEF's nine evaluation domains were used as a guide for conducting the evaluation and tailoring interview questions. The study did find that there is some overlap among the different evaluation domains. For example, discussions about error detection and remediation within models and datasets fall under three evaluation domains—data, system/model performance, and sustainability. This thesis concludes that this overlap is inevitable, given that the domains are inherently interconnected. The utility in distinguishing between the different evaluation domains is that it allows practitioners to focus their evaluation on specific areas of concern and potentially develop standardized measurements or indicators.

A key takeaway from this study pertains to the challenge of evaluating model performance. The absence of readily available performance metrics for practitioners can hinder the ability to satisfy this evaluation domain. There has been little expectation that vendors would need to provide this information to practitioners, given the end users' traditionally limited understanding and lack of demand for performance metrics.[341] Metrics, however, are a critical part of evaluating AI-enabled systems, since they offer a level of objectivity and comparative reference among different tools. Therefore, they should be part of a practitioner's request to the vendor. The intent is to build the expectation that performance metrics, along with the necessary context to make them interpretable to a practitioner, are a standard reporting criterion.

Certain aspects of the evaluation may seem non-AI specific—such as UX in navigating through the tool. For an average user, the AI component of Pulse may not be readily apparent, given that the technology is primarily used in the data enrichment phase where users have limited visibility. Yet, a holistic evaluation that includes the consideration of all nine evaluation domains is needed as organizations begin to adopt more advanced AI-enabled systems. Practitioners might not have the time or knowledge to evaluate each domain in depth, but they should at least be able to make an informed initial assessment before referring the tool for further evaluation by technical experts.

This case study did not include a rating or grading scale within the evaluation, given that it would produce little additional meaning. Nonetheless, organizations may want to explore the use of a numeric rating system to differentiate between multiple tools under evaluation. Furthermore, although this chapter discussed the results of the evaluation in a narrative format, the results can also be conveyed as a bullet point report or a survey.

[341] Information practitioner, personal communication, August 29, 2022.

VIII. CONCLUSION

Artificial intelligence is a technology that could transform the face of warfare, and it is already being integrated into a growing number of military technologies and operations.[342] Yet, the OIE community is at the early stages of AI adoption. Most AI-related technologies used for OIE involve relatively standard data aggregation or analysis tools. Given the potential advantages that AI can provide in analyzing the information environment, increasing the speed and scale of information dissemination, developing new content, and assessing the effectiveness of OIE, information practitioners will begin to see an increasing number of capabilities fused with AI technology.

Despite the promising capabilities of AI, the crux of OIE is the cognitive dimension. Although there are ongoing efforts to improve transparency and explainability of AI, information practitioners should be wary of over-trusting emerging AI systems. No technology in the near term can replace human analysis and critical thinking. Even if the AI can provide summarization, recommendations, and analysis of large amounts of data, it is up to the human practitioner to not only understand the potential shortfalls and caveats of the AI system but also to make the necessary connections between the outputs of the AI and the desired operational conditions.

A. RECOMMENDATIONS

The rapid development and explosion of AI-enabled tools introduces new capabilities at a pace that makes it difficult to keep abreast of emerging tools. As AI applications become increasingly prevalent and complex, practitioners are faced with the challenge of discerning which tools address operational needs and generate an advantage in the information environment. This thesis proposes a practitioner's evaluation framework (PEF) to provide a more structured, methodical approach to assessing AI-

[342] Michael C. Horowitz, Lauren Kahn, and Laura Resnick Samotin, "A Force for the Future," *Foreign Affairs*, June 2022, https://www.foreignaffairs.com/articles/united-states/2022-04-19/force-future.

enabled tools for OIE. The PEF ensures that practitioners are incorporating evaluation criteria that consider technology adoption factors and the *DOD Ethical Principles for AI*.

An evaluation is an iterative process that should be tailored to the specific needs of a user and operational requirement. Establishing a single standard for evaluation is futile, given the myriad of use cases and varying technology readiness levels of different AI tools. Nonetheless, the PEF offers a flexible but structured guideline for practitioners to conduct an initial evaluation of an AI tool. The expectation is that the framework would undergo continual refinement to ensure that it aligns with the priorities of the adopting organization. The level of detail with which practitioners examine each evaluation domain of the PEF will vary as they become more familiar with the technology.

Some vendors may be hesitant to share certain information with practitioners—particularly prior to an agreed upon contract. Nonetheless, it is important that these discussions and inquiries occur since they provide transparency. Performance metrics, in particular, should be considered a standard request, and vendors should be able to provide context and explain the significance of the metric so that it is understandable to the practitioner. Practitioners, however, should also remain cognizant of the potential proprietary or sensitive nature of such information. Given that the DOD's *RAI Strategy* and *Guidelines* have been published only within the last year and are continuing to undergo further refinement, it is possible that the vendor may not be aware of the standard practice of sharing certain information with the DOD customer.

OIE units should identify a team of individuals that could implement the PEF. The team would ideally consist of technical experts (e.g., data scientists), super users, and basic users. The diversity in experience and knowledge would facilitate a more robust evaluation. Understanding that operational demand and perpetual turnover of personnel could hinder the creation of a designated evaluation team, the number of individuals identified to conduct the evaluation may vary. Nonetheless, having the same individuals conduct the evaluations will assist in maintaining consistency. Their continued exposure to different AI tools can lead to a more effective discernment of emerging trends, anomalies, or points of concern associated with the technology. The identified evaluators should also explore the use of synthetic environments to test and compare multiple AI tools at the same time.

In addition to the proposed evaluation framework, this thesis offers three supplementary recommendations. First, the OIE community should examine what level of external support is appropriate to augment operations. In other words, to what extent should OIE units invest in in-house capabilities? If the intent is to make greater strides in obtaining support from the intelligence community, OIE units may not need to prioritize or invest as heavily in AI analytical capabilities. Furthermore, units should critically assess the pros and cons of relying on a solution that is more service-oriented versus a tool that is expected to be used directly by a practitioner. While a service-oriented solution saves time by providing finished products to the practitioner, there could be increased cost and less visibility associated with this solution. On the other hand, a tool that is intended for direct use by a practitioner may face greater barriers to adoption.

Second, OIE units, along with the rest of the Special Operations community, should explore additional AI training and educational opportunities for the individual warfighter. While there is a DOD-wide push for increased AI training and education, this initiative has yet to be implemented fully down to the tactical level. Individuals with interest in AI may take the initiative to educate themselves on the technology, but there is little organizational support for gaining broader access to AI educational opportunities. Training on specific tools remains important, but if practitioners are to utilize AI technologies effectively and responsibly, they need to understand the basics of the technology. Furthermore, having a foundation in statistics is a prerequisite for not only using AI but for understanding the increasingly data-driven world.

Finally, the OIE community should foster relations with DOD entities who are leading the effort in AI—namely the CDAO and DIU. These organizations provide additional expertise and understanding of current AI developments. If OIE units are seeking to be on the leading edge of leveraging AI technology, it is important that they are aware of what advancements are being made especially in the commercial sector as well as what capabilities are still considered aspirational.

B. RESEARCH LIMITATIONS

The primary limitation for this research is the lack of testing to validate the PEF with other information practitioners. In this thesis, an evaluation was conducted on only one OIE tool by one evaluator—the author of this thesis. To determine whether the PEF should be adopted by the OIE community, the study would need to test the framework in an operational setting, since the culture and working environment of an operational unit can impact attitudes about the utility of the framework. It is also necessary to evaluate other types of AI tools from different vendors to verify whether the framework can be applied to different AI applications such as synthetic media or predictive modeling and simulation.

C. AREAS FOR FUTURE RESEARCH

Many potential areas for future research stem from the subject of AI and OIE. First, additional research is needed on the application of AI technologies for OIE. While this thesis touched on this topic, the technologies mentioned in this thesis only provide a snapshot of potential applications and are far from all-encompassing. More in-depth research on emerging AI technologies may help identify what could be a "killer application" for OIE.[343]

Second, future research should examine specific AI adoption factors related to OIE. A study like Venkatesh's analysis of AI adoption in operations management would be helpful in further examining the unique issues with AI tools for OIE.[344] Similarly, there is an absence of studies related to trust and human-machine teaming in the context of AI and OIE. Further experimentation and research (tailored by mission, AI system, and

343 P. A. Geroski, *The Evolution of New Markets* (New York, NY: Oxford University Press, 2003), 4.

344 Venkatesh, "Adoption and Use of AI Tools."

user) are required to obtain a more accurate assessment of how to properly calibrate human-machine interaction.

Third, training and education remains a core component of ensuring that AI tools are properly evaluated, adopted, and utilized. Future research should examine what kinds of AI training and education are needed for information practitioners to include specifics on how detailed the instruction should be and where it should be incorporated within a practitioner's training pipeline or military education.

Fourth, a study on the legal and policy implications of the increased use of AI for OIE is needed. OIE is already a subject of heightened scrutiny and sensitivity. Novel capabilities enabled by AI raise questions of whether current laws and policies are adequate to address risks that could arise from the increased use of AI within the information environment.

Finally, there are opportunities to conduct further research on each of the evaluation domains to further assess the "right" level of detail needed to make an effective assessment. For example, a closer examination of how AI tools should fit into information practitioner workflows would facilitate a more informed evaluation of user experience. Additional testing of the framework could also identify factors for evaluation that were not considered in this thesis.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A. MACHINE LEARNING TECHNIQUES

A. SUPERVISED LEARNING

Supervised learning is primarily used for classification or regression tasks.[345] The intent of the model is to extrapolate responses that it has learned from its training and correctly apply them to real-world situations. It relies on labeled datasets, which have been tagged with the relevant features to map the system toward a desired output.[346] The dataset is split into a training set and a test set, which should be representative of data that the model may encounter in the future. Once the model is built from the training set, the model utilizes the test set—holdout data—to assess the performance of the algorithm and ensure that the model is not overfitting the training data.[347] As with all ML techniques, supervised learning is an iterative process of evaluation and refinement of the model. While there have been notable advancements in supervised learning (i.e., neural networks), obtaining labelled data remains a major challenge, often incurring significant manpower, cost, and time. Furthermore, it is unlikely that one can obtain all the necessary example responses that are required to represent all situations that the system may encounter.

B. UNSUPERVISED LEARNING

Unsupervised learning—unlike supervised learning—is trained on unlabeled data. Relying only on input data, unsupervised learning discovers patterns on its own without

---

345 Alpaydin, *Machine Learning*, 38.

346 Pádraig Cunningham, Matthieu Cord, and Sarah Jane Delany, "Supervised Learning," in *Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval*, ed. Matthieu Cord and Pádraig Cunningham, Cognitive Technologies (Berlin, Germany: Springer, 2008), 21, https://doi.org/10.1007/978-3-540-75171-7_2.

347 Joshua Menke and Tony Martinez, "Improving Supervised Learning By Adapting the Problem to the Learner," *International Journal of Neural Systems* 19, no. 1 (2009): 3, https://doi.org/10.1142/S0129065709001793.

explicit human instructions.[348] Clustering, the process of taking unlabeled data and placing them into groups based on their similarities, is one of the most common methods of unsupervised learning and is applied to problems such as market segmentation and anomaly detection.[349] Unsupervised learning offers a more flexible, automated process of ML by avoiding the costs, the potential unavailability, and the manual annotations typically required with labeled data.[350] Furthermore, labeled data tends to get stale or outdated due to the dynamic nature of the information environment. However, unsupervised learning algorithms are generally more complex. Although it is easier to obtain unlabeled data, the accuracy of this technique—when compared to supervised learning—may be more in question, given the absence of ground truth to evaluate the results of its training.[351]

C. SEMI-SUPERVISED LEARNING

Semi-supervised learning attempts to mediate the downsides of supervised and unsupervised learning by using a small set of labeled training data along with a larger unlabeled dataset.[352] Thus, semi-supervised learning avoids the need to obtain large labeled datasets, which can be labor intensive and expensive, and furthermore, exposes the system to a greater amount of test data to facilitate a more accurate model.[353]

---

[348] Matthew J. Denny and Arthur Spirling, "Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It," *Political Analysis* 26, no. 2 (2018): 172, https://www.jstor.org/stable/26563824.

[349] Vishnuvarthanan Govindaraj et al., "Automated Unsupervised Learning-Based Clustering Approach for Effective Anomaly Detection in Brain Magnetic Resonance Imaging (MRI)," *IET Image Processing* 14, no. 14 (2020): 3516–26, https://doi.org/10.1049/iet-ipr.2020.0597; Rik van Leeuwen and Ger Koole, "Data-Driven Market Segmentation in Hospitality Using Unsupervised Machine Learning," *Machine Learning with Applications* 10 (December 15, 2022), https://doi.org/10.1016/j.mlwa.2022.100414.

[350] Usama et al., "Unsupervised Machine Learning for Networking," 65580.

[351] Namrata Dhanda, Stuti Shukla Datta, and Mudrika Dhanda, "Machine Learning Algorithms," in *Computational Intelligence in the Internet of Things*, ed. Hindriyanto Dwi Purnomo (Hershey, PA: IGI Global, 2019), 223, https://doi-org.libproxy.nps.edu/10.4018/978-1-5225-7955-7.ch009.

[352] Y Reddy, Viswanath Pulabaigari, and Eswara B, "Semi-Supervised Learning: A Brief Review," *International Journal of Engineering & Technology* 7 (February 9, 2018): 83, https://doi.org/10.14419/ijet.v7i1.8.9977.

[353] Xiangli Yang et al., "A Survey on Deep Semi-Supervised Learning" (arXiv, August 22, 2021), 1, http://arxiv.org/abs/2103.00550.

D. REINFORCEMENT LEARNING

Although reinforcement learning does not rely on labeled data, it differs from unsupervised learning in that its goal is to maximize rewards rather than recognize hidden structure in the data.[354] Reinforcement learning also takes a more holistic approach in which an "agent" interacts with an uncertain environment and learns through trial-and-error over time. The agent will not know what it is supposed to do until it "stumbles" upon a reward, and any action that the agent takes may affect the conditions in any future decision.[355] Thus, reinforcement learning consists of sequential processes very similar to gaming.

[354] Richard S. Sutton and Andrew G. Barto, *Reinforcement Learning: An Introduction*, Second edition (Cambridge, Massachusetts: The MIT Press, 2018), 2.

[355] Yumeng Zhang, "An Overview of the Theory and Application of Reinforcement Learning," in *ICMLCA 2021; 2nd International Conference on Machine Learning and Computer Application*, 2021, 586.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. MODEL CARD ANALYSIS

A. LIST OF ANALYZED MODEL CARDS

1. spaCy English Transformer Pipeline[356]
2. BERT Multilingual-uncased[357]
3. Dalle-mini[358]
4. FaceDetect[359]
5. STT En Conformer-CTC Large[360]
6. License Plate Detection – LPDNet[361]
7. NeMo – Text to Speech[362]
8. PeopleNet[363]
9. PeopleSegNet[364]

---

356 "English · SpaCy Models Documentation," spaCy, October 19, 2022, https://spacy.io/models/en#en_core_web_trf.

357 "Bert-Base-Multilingual-Uncased · Hugging Face," HuggingFace, accessed September 23, 2022, https://huggingface.co/bert-base-multilingual-uncased.

358 "Dalle-Mini/Dalle-Mini · Hugging Face," HuggingFace, 2021, https://huggingface.co/dalle-mini/dalle-mini.

359 "FaceDetect," Nvidia NGC Catalog, January 13, 2022, https://catalog.ngc.nvidia.com/orgs/nvidia/teams/tao/models/facenet.

360 "STT En Conformer-CTC Large," Nvidia NGC Catalog, June 23, 2022, https://catalog.ngc.nvidia.com/orgs/nvidia/teams/nemo/models/stt_en_conformer_ctc_large.

361 "LPDNet," Nvidia NGC Catalog, May 25, 2022, https://catalog.ngc.nvidia.com/orgs/nvidia/teams/tao/models/lpdnet.

362 "NeMo - Text to Speech," Nvidia NGC Catalog, September 22, 2022, https://catalog.ngc.nvidia.com/orgs/nvidia/collections/nemo_tts.

363 "PeopleNet," Nvidia NGC Catalog, May 26, 2022, https://catalog.ngc.nvidia.com/orgs/nvidia/teams/tao/models/peoplenet.

364 "PeopleSegNet," Nvidia NGC Catalog, August 22, 2022, https://catalog.ngc.nvidia.com/orgs/nvidia/teams/tao/models/peoplesegnet.

10. TrafficCamNet[365]

11. VehicleTypeNet[366]

12. Detect Sentiment[367]

13. Detector for CTRL[368]

14. Einstein Engagement Frequency[369]

15. Einstein Engagement Scoring for Mobile[370]

16. Einstein Messaging Insights[371]

17. Einstein Optical Character Recognition[372]

18. Economic Simulation Framework[373]

---

365 "TrafficCamNet," Nvidia NGC Catalog, May 25, 2022, https://catalog.ngc.nvidia.com/orgs/nvidia/teams/tao/models/trafficcamnet.

366 "VehicleTypeNet," Nvidia NGC Catalog, November 7, 2021, https://catalog.ngc.nvidia.com/orgs/nvidia/teams/tao/models/vehicletypenet.

367 CRM Analytics, "Detect Sentiment Model Card," Salesforce Help, August 12, 2020, https://help.salesforce.com/s/articleView?id=sf.bi_integrate_transformation_detectSentimentModelCard.htm&type=5.

368 "Model Card: Detector for CTRL from Salesforce Research," GitHub, October 13, 2020, https://github.com/salesforce/ctrl-detector/blob/f1ec83bdb5fc176f5c42b353b12df94fffce2c87/ModelCard.pdf.

369 Salesforce Marketing Cloud Einstein, "Einstein Engagement Frequency Model Details," Salesforce Help, May 2020, https://help.salesforce.com/s/articleView?id=sf.mc_anb_einstein_engagement_frequency_model_details.htm&type=5.

370 Salesforce Marketing Cloud Einstein, "Einstein Engagement Scoring for Mobile Model Card," Salesforce Help, August 2020, https://help.salesforce.com/s/articleView?id=sf.mc_anb_einstein_engagement_scoring_for_mobile_model_card.htm&type=5.

371 Salesforce Marketing Cloud Einstein, "Einstein Messaging Insights Model Card," Salesforce Help, May 2020, https://help.salesforce.com/s/articleView?id=sf.mc_anb_einstein_messaging_insights_model_card.htm&type=5.

372 Salesforce AI Research, "Einstein OCR Model Card," Einstein Vision and Language, February 2022, https://developer.salesforce.com/docs/analytics/einstein-vision-language/guide/einstein-ocr-model-card.html.

373 "Foundation: An Economic Simulation Framework," GitHub, October 14, 2020, https://github.com/salesforce/ai-economist/blob/a84d5f3fdcabb207d9fde7754d34906903b3e184/Simulation_Card_Foundation_Economic_Simulation_Framework.pdf.

19. Conditional Transformer Language Model for Controllable Generation[374]

20. DALL·E discrete VAE[375]

21. Diffusion Models[376]

22. GLIDE[377]

23. InstructGPT[378]

24. Summarization Model[379]

25. GPT-3[380]

26. CLIP[381]

27. GPT-2[382]

28. Generation (Cohere)[383]

29. Representation (Cohere)[384]

---

[374] "Model Card for CTRL - A Conditional Transformer Language Model for Controllable Generation," GitHub, April 20, 2020, https://github.com/salesforce/ctrl/blob/a089a0c5956a84f229e5ece9d5da0bc4689dded0/ModelCard.pdf.

[375] "Model Card: DALL·E DVAE," GitHub, November 23, 2022, https://github.com/openai/DALL-E/blob/5be4b236bc3ade6943662354117a0e83752cc322/model_card.md.

[376] "Guided-Diffusion," GitHub, July 19, 2021, https://github.com/openai/guided-diffusion/blob/22e0df8183507e13a7813f8d38d51b072ca1e67c/model-card.md.

[377] "GLIDE," GitHub, December 20, 2021, https://github.com/openai/glide-text2im/blob/69b530740eb6cef69442d6180579ef5ba9ef063e/model-card.md.

[378] "InstructGPT Model Card," GitHub, January 2022, https://github.com/openai/following-instructions-human-feedback/blob/5c0534c78136809ab0118c5c1b74d159befa2467/model-card.md.

[379] "Model Card: Summarization Model Trained with Human Feedback," September 2020, https://github.com/openai/summarize-from-feedback/blob/56b6bb613a1b58a8aa7a5e29266f65c7b980ee48/model_card.md.

[380] "GPT-3 Model Card," GitHub, September 2020, https://github.com/openai/gpt-3/blob/d7a9bb505df6f630f9bab3b30c889e52f22eb9ea/model-card.md.

[381] "Model Card: CLIP," GitHub, April 2021, https://github.com/openai/CLIP/blob/d50d76daa670286dd6cacf3bcd80b5e4823fc8e1/model-card.md.

[382] "GPT-2 Model Card," GitHub, November 2019, https://github.com/openai/gpt-2/blob/a74da5d99abaaba920de8131d64da2862a8f213b/model_card.md.

[383] "Generation," Cohere AI, November 4, 2022, https://docs.cohere.ai/docs/generation-card.

[384] "Representation," Cohere AI, November 4, 2022, https://docs.cohere.ai/docs/generation-card.

30. Face Detection[385]

31. Object Detection[386]

32. Pathways Language Model (PaLM)[387]

33. MLKit Selfie Segmentation[388]

34. Conversation Toxicity Model[389]

35. Open Pre-trained Transformer Language Model (OPT)[390]

36. XGLM Multilinguage Model[391]

37. BlenderBot 2.0[392]

38. BlazeFace[393]

---

385 "Face Detection," Google Cloud Model Cards, accessed August 12, 2022, https://modelcards.withgoogle.com/face-detection.

386 "Object Detection," Google Cloud Model Cards, accessed August 12, 2022, https://modelcards.withgoogle.com/object-detection.

387 Aakanksha Chowdhery et al., "PaLM: Scaling Language Modeling with Pathways" (arXiv, October 5, 2022), 73–74, https://doi.org/10.48550/arXiv.2204.02311.

388 "MLKit Selfie Segmentation," ML Kit, September 23, 2022, https://developers.google.com/ml-kit/images/vision/selfie-segmentation/selfie-model-card.pdf.

389 "Toxicity," GitHub, March 9, 2021, https://github.com/conversationai/perspectiveapi/blob/a51953fab94c34be2901c2ecfe4320c4d3b7039a/model-cards/English/toxicity.md.

390 Susan Zhang et al., "OPT: Open Pre-Trained Transformer Language Models" (arXiv, June 21, 2022), https://doi.org/10.48550/arXiv.2205.01068.

391 "XGLM Multilingual Model," GitHub, February 1, 2022, https://github.com/facebookresearch/fairseq/blob/main/examples/xglm/model_card.md.

392 "BlenderBot 2.0 2.7B," GitHub, September 8, 2021, https://github.com/facebookresearch/ParlAI/blob/f427b4cf97e13448a91d02212914591b0a19877f/parlai/zoo/blenderbot2/model_card.md.

393 Yury Kartynnik, "Model Card MediaPipe BlazeFace Full Range" (Google MediaPipe, January 19, 2021), https://drive.google.com/file/d/1jpQt8TB1nMFQ49VSSBKdNEdQOygNRvCP/preview?usp=embed_facebook.

39. Iris[394]

40. Face Mesh[395]

41. Hands[396]

42. Hair Segmentation[397]

43. BlazePose[398]

44. KNIFT[399]

45. Objectron[400]

46. StoryDALL-E[401]

47. Gopher[402]

---

394 Andrey Vakunov et al., "Model Card MediaPipe Iris" (Google MediaPipe, May 13, 2020), https://drive.google.com/file/d/1bsWbokp9AklH2ANjCfmjqEzzxO1CNbMu/preview?usp=embed_facebook.

395 Ivan Grishchenko, Artsiom Ablavatski, and Yury Kartynnik, "Model Card MediaPipe Face Mesh" (Google MediaPipe, October 10, 2018), https://drive.google.com/file/d/1QvwWNfFoweGVjsXF3DXzcrCnz-mx-Lha/preview?usp=embed_facebook.

396 "Model Card Hand Tracking (Lite_Full)" (Google MediaPipe, October 2021), https://drive.google.com/file/d/1-rmIgTfuCbBPW_IFHkh3f0-U_lnGrWpg/preview?usp=embed_facebook.

397 Andrei Tkachenka et al., "Model Card Content: Hair Segmentation" (Google MediaPipe, 2019), https://drive.google.com/file/d/1lPwJ8BD_-3UUor4LayQ0xpa_RIC_hoRh/preview?usp=embed_facebook.

398 Valentin Bazarevsky, Ivan Grishchenko, and Eduard Gabriel Bazavan, "Model Card Media Pipe BlazePose GHUM 3D" (Google MediaPipe, April 16, 2021), https://drive.google.com/file/d/10WlcTvrQnR_R2TdTmKw0nkyRLqrwNkWU/preview?usp=embed_facebook.

399 Zhicheng Wang, "Model Card Content: MediaPipe KNIFT," February 5, 2020, https://drive.google.com/file/d/1RCdA83a8JDV3ZPS-mtpsV8mTj3r0F9s-/view?usp=sharing&usp=embed_facebook.

400 Liangkai Zhang et al., "Model Card MediaPipe Objectron" (Google MediaPipe, October 27, 2020), https://drive.google.com/file/d/1CMhN7Npdq0Dt2j0_z69mai2-m7oUTRKF/preview?usp=embed_facebook.

401 Adyasha Maharana, "Storydalle," GitHub, September 15, 2022, https://github.com/adymaharana/storydalle/blob/615881ace2dd0a5f5757cfeb4bdd1fb2c08f3337/MODEL_CARD.MD.

402 Jack W. Rae et al., "Scaling Language Models: Methods, Analysis & Insights from Training Gopher" (arXiv, January 21, 2022), 49–52, https://doi.org/10.48550/arXiv.2112.11446.

48. DynaSent Model 0 and Model 1[403]

49. Language Agnostic Link-Based Article Topic Model[404]

50. Hemolytic Prediction[405]

B. CODING SYSTEM

Table 7 lists the codes that were used to conduct a qualitative analysis of 50 publicly available model cards. The numbers on the right column count the number of times the category was found within the group of documents.

Table 7. List of Codes and their Frequencies

1 Model Details	
1.1 Context/Background	4
1.2 Description	48
1.3 Architecture	23
1.4 Model Size	2
1.5 Type	30
1.6 Date	27
1.7 Authors	28
1.8 Version/License	37
1.9 Contact Info/Feedback	17
2 Intended Use	45
2.1 Current Uses	1
2.2 Intended Users	14
2.3 Instructions for Use	10
3 Limitations	40

[403] Christopher Potts, "Model Card for DynaSent Model 0 and Model 1," GitHub, January 4, 2021, https://github.com/cgpotts/dynasent/blob/a771f0d344ffa6d21716644ef8766c0ee743fc1b/dynasent_modelcard.md.

[404] "Machine Learning Models/Proposed/Language Agnostic Link-Based Article Topic Model Card," Wikimedia, September 29, 2022, https://meta.wikimedia.org/wiki/Machine_learning_models/Proposed/Language_agnostic_link-based_article_topic_model_card.

[405] Andrew White and Mehrad Ansari, "Hemolytic Prediction," peptide.bio v0.19.0, 2022, https://peptide.bio/.

3.1 Tradeoffs	8
3.2 Out-of-Scope	28
4 Factors	8
4.1 Attributes	4
4.2 Evaluation Factors	2
4.3 Relevant Factors	2
4.4 Instrumentation	5
4.5 Groups	9
4.6 Environment	12
5 Model Input	29
6 Model Output	27
7 Data	
7.1 Data Preprocessing	10
7.2 Training Data	50
7.3 Test/Eval Data	25
8 Training Procedure	17
9 Performance Metrics	49
10 Explainability	4
11 Model Maintenance	3
12 Potential for Misuse	6
13 Domain Shift	3
14 Adversarial Attack	4
14.1 Robustness	8
15 Ethical Considerations	33
15.1 Risk	14
15.2 Toxicity	8
15.3 Bias	38
15.4 Fairness	18
15.5 Safety	13
15.6 Privacy	9
16 Caveats and Recommendations	14
17 Definitions	7
18 Demo/Test/Example	11
19 Links to Paper/Code	24

C. SIMILARITY ANALYSIS CALCULATIONS

Calculations for the similarity analysis are based on code frequency using the Jaccard similarity index, which does not count against the absence—or non-existence—of a code in a document.[406] The Jaccard was used to prevent the overemphasis of non-existing codes, since some codes did not appear in multiple documents. Similarity measures are determined by using Table 8 and calculating $\frac{a}{(a+b+c)}$ across the pairs of documents; these calculations are repeated across the combination of analyzed model cards.

Table 8. Calculating Similarity between Pairs of Documents.[407]

		Document A	
		Code/Variable value *exists*	Code/Variable value *does not exist*
Document B	Code/Variable value *exists*	a	b
	Code/Variable value *does not exist*	c	d

a = Number of codes or variable values that are identical in both documents.

d = Number of codes or variable values that do not exist in both documents.

b and c = Number of codes or variable values that exist in only one document.

The distances between documents within the map shown in Figure 9 were determined utilizing squared Euclidean distance—or "the sum of squared deviations"—based on the similarity analysis.[408] The calculated distances between the different documents (Figure 27) were used to determine group affiliation or clusters. Given the

[406] MAXQDA, "Similarity Analysis for Documents," MAXQDA, 2022, https://www.maxqda.com/help-mx20/mixed-methods-functions/similarity-analysis-for-documents.

[407] Source: MAXQDA.

[408] MAXQDA.

two-dimensional nature of the map's surface, some documents appear closer together even though their actual distance is further apart in a multidimensional space.[409]

Model Card	Transformer Pipeline	BERT Multilingual-uncased	dalle-mini	LPDNet	Einstein Messaging Insights	GPT-3	Generation	Face Detection	Open Pre-trained Transformer	Iris	StoryDALL-E	Gopher	Dynasent Sentiment Analysis	Lang. Agnostic Topic Model	Hemolytic Prediction	X	Y
spaCy English Transformer Pipeline	0.00	0.86	0.80	0.80	0.86	0.75	0.78	0.87	0.90	0.87	0.83	0.87	0.75	0.85	0.71	-0.09	-0.04
BERT Multilingual-uncased	0.86	0.00	0.55	0.67	0.75	0.60	0.64	0.92	0.64	0.83	0.62	0.76	0.73	0.81	0.82	-0.26	-0.18
dalle-mini	0.80	0.55	0.00	0.67	0.59	0.62	0.54	0.87	0.54	0.68	0.75	0.75	0.62	0.72	0.69	-0.14	-0.11
LPDNet	0.80	0.67	0.67	0.00	0.59	0.71	0.73	0.69	0.73	0.75	0.75	0.75	0.62	0.65	0.69	0.10	-0.01
Einstein Messaging Insights	0.86	0.75	0.59	0.59	0.00	0.62	0.72	0.69	0.79	0.47	0.73	0.62	0.62	0.65	0.69	0.22	-0.10
GPT-3	0.75	0.60	0.62	0.71	0.62	0.00	0.45	0.85	0.79	0.79	0.70	0.65	0.67	0.83	0.75	-0.20	-0.31
Generation (Cohere)	0.78	0.64	0.54	0.73	0.72	0.45	0.00	0.86	0.80	0.80	0.73	0.67	0.69	0.84	0.77	-0.22	-0.32
Face Detection (Google)	0.87	0.92	0.87	0.69	0.69	0.85	0.86	0.00	0.86	0.62	0.80	0.90	0.85	0.75	0.92	0.49	-0.12
Open Pre-trained Transformer (OPT)	0.90	0.64	0.54	0.73	0.79	0.79	0.80	0.86	0.00	0.74	0.60	0.74	0.58	0.53	0.67	-0.07	0.25
Iris	0.87	0.83	0.68	0.75	0.47	0.79	0.80	0.62	0.74	0.00	0.82	0.75	0.79	0.67	0.84	0.40	-0.11
StoryDALL-E	0.83	0.62	0.75	0.75	0.73	0.70	0.73	0.80	0.60	0.82	0.00	0.82	0.56	0.71	0.67	-0.12	0.12
Gopher	0.87	0.76	0.75	0.75	0.62	0.65	0.67	0.90	0.74	0.75	0.82	0.00	0.72	0.60	0.71	-0.01	-0.03
Dynasent Sentiment Analysis	0.75	0.73	0.62	0.62	0.62	0.67	0.69	0.85	0.58	0.79	0.56	0.72	0.00	0.60	0.12	-0.15	0.28
Language Agnostic Link-Based Article Topic Model	0.85	0.81	0.72	0.65	0.65	0.83	0.84	0.75	0.53	0.67	0.71	0.60	0.60	0.00	0.57	0.18	0.31
Hemolytic Prediction	0.71	0.82	0.69	0.69	0.69	0.75	0.77	0.92	0.67	0.84	0.67	0.71	0.12	0.57	0.00	-0.15	0.36

Figure 27. Calculated Distances Based on Similarity Analysis

409 MAXQDA, "Document Map: Arranging Documents According to Similarity," MAXQDA, 2022, https://www.maxqda.com/help-mx20/visual-tools/document-map-arranging-documents-according-to-similarity.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. CONSOLIDATED EVALUATION QUESTIONS

Figure 28 depicts a consolidated list of the evaluation questions within the Practitioner's Evaluation Framework. The figure can be used as a reference handout for information practitioners and unit commanders.

Purpose: Serve as a guideline for practitioners to conduct an initial evaluation of AI-enabled tools.

Prior to conducting the evaluation, units should ask the following questions:

- *What problem or gap are we trying to address and why does it require a technical solution?*
- *What benefit does the AI technology provide and how does it address the problem or gap?*
- *Is the proposed AI technology capable of being implemented and can it address the problem or gaps within the given organizational and environmental context?*

Mission/Task Alignment

- What is the AI's intended use/purpose?
    - Does this align with the mission/task?
    - How does it fit into the existing workflow?
- What relative advantage does this tool provide?
- What are the risks of adopting this tool?

Data

- What data is the AI using?
- Is the data relevant, current, and representative of the actual operational environment?
- Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?
- What restrictions have been placed on the data?
- *Data Security/Integrity*
    - What steps have/will be taken to ensure the data is appropriately secure during and after the project?
    - What steps will be taken to check for potential errors, noise, bias, and/or redundancies?
- *Data Maintenance*
    - Who is maintaining the dataset?
    - How will it be updated? (e.g., automatic ingestion via API)

System/Model Performance

- What metrics are used to measure system performance? Why are those the correct metrics?
- What is an acceptable minimum performance threshold?
- What is the process for detecting and correcting errors? What types of errors would cause serious problems?

User Experience

- *Ease of Use*
    - Does the tool require any special knowledge or expertise?
    - What training is required to use the tool effectively?
    - In what environments is the tool accessible (e.g., external or DOD networks)?
- *Compatibility*
    - Is the AI tool able to exchange and share data/information with other existing systems?
    - Does the usage of the AI tool interfere with any existing systems?
    - What collaboration features are incorporated into the tool?
- *Explainability*
    - Does the tool provide explanations or context to its outputs?
    - Does the explanation improve the user's decision/task performance?
- What feedback mechanisms exist?

Sustainability

- Who is responsible for the maintenance of the system?
- How will monitoring and auditing occur? How will the system be tested for model drift?
- Has adversarial disruption been considered?
- Is there a process for system rollback?

Scalability

- How many users should be considered for this tool? Is large-scale usage feasible?
- What are the infrastructure requirements to deploy this tool at the organizational-level?
- What additional resources or personnel are required to scale this tool? Can these resources be sustained over the long-term?

Affordability

- What is the cost for a prototype? What is the maintenance cost?
- How does the proposed cost compare to other offerings?
- Are there existing capabilities or contracts that can augment the solution or address certain components of the problem?

Ethical/Legal/Policy Considerations

- Are there any ethical, legal, or policy restrictions that prevent us from using this tool?
- What ethical concerns may arise from using this tool? How does the use of this tool impact public or user trust?

Vendor Assessment

- Is the company committed to adhering to DOD's Ethical Principles for AI?
- What is the company's reputation for working with the DOD?
- What support is the vendor willing to provide to effectively integrate the tool into the organization?

Figure 28. List of Consolidated Evaluation Questions

LIST OF REFERENCES

Agarwal, Alpana. "AI Adoption by Human Resource Management: A Study of Its Antecedents and Impact on HR System Effectiveness." *Foresight*, January 1, 2022. https://doi.org/10.1108/FS-10-2021-0199.

Ahmad, Wajeeha. *Why Botnets Persist: Designing Effective Technical and Policy Interventions*. IPRI(2019)02. Boston, MA: Massachusetts Institute of Technology, 2019. https://internetpolicy.mit.edu/publications-ipri-2019-02.

Al Hadwer, Ali, Madjid Tavana, Dan Gillis, and Davar Rezania. "A Systematic Review of Organizational Factors Impacting Cloud-Based Technology Adoption Using Technology-Organization-Environment Framework." *Internet of Things* 15 (September 2021): 1–10. https://doi.org/10.1016/j.iot.2021.100407.

Albert, Bill, Tom Tullis, and William Albert. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. San Francisco, UNITED STATES: Elsevier Science & Technology, 2013. http://ebookcentral.proquest.com/lib/ebook-nps/detail.action?docID=1204543.

Allen, Greg. *Understanding AI Technology*. AD1099286. Washington, DC: Joint Artificial Intelligence Center (JAIC), 2020. https://apps.dtic.mil/sti/citations/AD1099286.

Alpaydin, Ethem. *Machine Learning*. Cambridge, MA: The MIT Press, 2016.

AlSheibani, Sulaiman, Yen Cheung, and Chris Messom. "Artificial Intelligence Adoption: AI-Readiness at Firm-Level." In *Pacific Asia Conference on Information Systems*, Vol. 37. Japan, 2018. https://aisel.aisnet.org/pacis2018/37.

Anderson, David, Joy Bonaguro, Miriam McKinney, Andrew Nicklin, and Wiseman. "Ethics & Algorithms Toolkit (Beta)." Accessed October 3, 2022. http://ethicstoolkit.ai/.

Arnold, M., R. K. E. Bellamy, M. Hind, S. Houde, S. Mehta, A. Mojsilović, R. Nair et al. "FactSheets: Increasing Trust in AI Services through Supplier's Declarations of Conformity." *IBM Journal of Research and Development* 63, no. 4/5 (July 2019): 6:1-6:13. https://doi.org/10.1147/JRD.2019.2942288.

Arsenault, Amelia. "Microtargeting, Automation, and Forgery: Disinformation in Age of Artificial Intelligence." Major research, University of Ottawa, 2020. http://ruor.uottawa.ca/handle/10393/40495.

Baier, Lucas, Fabian Jöhren, and Stefan Seebacher. "Challenges in the Deployment and Operation of Machine Learning in Practice." In *27th European Conference on Information Systems*. Stockholm: ResearchGate, 2019. https://www.researchgate.net/publication/332996647_CHALLENGES_IN_THE_DEPLOYMENT_AND_OPERATION_OF_MACHINE_LEARNING_IN_PRACTICE.

Barbu, Oana. "Advertising, Microtargeting and Social Media." *Procedia - Social and Behavioral Sciences*, International Conference on Communication and Education in Knowledge Society, 163 (December 19, 2014): 44–49. https://doi.org/10.1016/j.sbspro.2014.12.284.

Barclay, Iain, Harrison Taylor, Alun Preece, Ian Taylor, Dinesh Verma, and Geeth Mel. "A Framework for Fostering Transparency in Shared Artificial Intelligence Models by Increasing Visibility of Contributions." *Concurrency and Computation: Practice and Experience* 33, no. 19 (October 10, 2021). https://doi.org/10.1002/cpe.6129.

Barczak, Gloria. "From the Editor." *Journal of Product Innovation Management* 31, no. 5 (September 2014): 878–878. https://doi.org/10.1111/jpim.12219.

Barmer, Hollen, Rachel Dzombak, Matthew Gaston, Eric Heim, Vijaykumar Palat, Frank Redner, Tanisha Smith, and Nathan Vanhoudnos. *Robust and Secure AI*. Pittsburgh, PA: Carnegie Mellon University, 2021. https://doi.org/10.1184/R1/16560252.v1.

Barmer, Hollen, Rachel Dzombak, Matthew Gaston, Vijaykumar Palat, Frank Redner, Tanisha Smith, and John Wohlbier. *Scalable AI*. Pittsburgh, PA: Carnegie Mellon University, 2021. https://doi.org/10.1184/R1/16560273.v1.

Barrett, Raymond J. "PSYOP What Is It." In *The Art and Science of Psychological Operations: Case Studies of Military Application. Volume Two*, edited by Ronald De McLaurin, Carl F. Rosenthal, and Sarah A. Skillings. Washington, DC: American Institutes for Research, 1976. https://apps.dtic.mil/sti/citations/ADA030140.

Bassett, Caroline. "The Computational Therapeutic: Exploring Weizenbaum's ELIZA as a History of the Present." *AI & SOCIETY* 34, no. 4 (December 1, 2019): 803–12. https://doi.org/10.1007/s00146-018-0825-9.

Bauer, Markus, Clemens van Dinther, and Daniel Kiefer. "Machine Learning in SME: An Empirical Study on Enablers and Success Factors." In *Americas Conference on Information Systems (AMCIS 2020)*. Salt Lake City, UT: ResearchGate, 2020. https://www.researchgate.net/publication/344651203_Machine_Learning_in_SME_An_Empirical_Study_on_Enablers_and_Success_Factors.

Baxevanakis, Spiros, Giorgos Kordopatis-Zilos, Panagiotis Galopoulos, Lazaros Apostolidis, Killian Levacher, Ipek Baris Schlicht, Denis Teyssou, Ioannis Kompatsiaris, and Symeon Papadopoulos. "The MeVer DeepFake Detection Service: Lessons Learnt from Developing and Deploying in the Wild." In *Proceedings of the 1st International Workshop on Multimedia AI against Disinformation*, 59–68. Newark, NJ: ACM, 2022. https://doi.org/10.1145/3512732.3533587.

Bazarevsky, Valentin, Ivan Grishchenko, and Eduard Gabriel Bazavan. "Model Card Media Pipe BlazePose GHUM 3D." Google MediaPipe, April 16, 2021. https://drive.google.com/file/d/10WlcTvrQnR_R2TdTmKw0nkyRLqrwNkWU/preview?usp=embed_facebook.

Bedué, Patrick, and Albrecht Fritzsche. "Can We Trust AI? An Empirical Investigation of Trust Requirements and Guide to Successful AI Adoption." *Journal of Enterprise Information Management* 35, no. 2 (January 1, 2021): 530–49. https://doi.org/10.1108/JEIM-06-2020-0233.

Bender, Emily M., and Batya Friedman. "Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science." *Transactions of the Association for Computational Linguistics* 6 (December 2018): 587–604. https://doi.org/10.1162/tacl_a_00041.

Benedick, Paul-Lou, Robert Jérémy, and Yves Le Traon. "A Systematic Approach for Evaluating Artificial Intelligence Models in Industrial Settings." *Sensors* 21, no. 18 (2021). https://doi-org.libproxy.nps.edu/10.3390/s21186195.

Boiney, John, and David Foster. *Progress and Promise: Research and Engineering for Human Sociocultural Behavior Capability in the U.S. Department of Defense*. 13–2042. McLean, VA: MITRE Corporation, 2013. https://apps.dtic.mil/sti/pdfs/ADA587451.pdf.

Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein et al. *On the Opportunities and Risks of Foundation Models*. Stanford, CA: Stanford University Human-Centered Artificial Intelligence, 2022. https://crfm.stanford.edu/report.html.

Borsos, Zalán, Raphaël Marinier, Damien Vincent, Eugene Kharitonov, Olivier Pietquin, Matt Sharifi, Olivier Teboul, David Grangier, Marco Tagliasacchi, and Neil Zeghidour. "AudioLM." Google Research, September 7, 2022. https://google-research.github.io/seanet/audiolm/examples/.

Braman, Sandra. "Defining Information." *Telecommunications Policy* 13, no. 3 (September 1989): 233–42. https://doi.org/10.1016/0308-5961(89)90006-2.

Branley-Bell, Dawn, Rebecca Whitworth, and Lynne Coventry. “User Trust and Understanding of Explainable AI: Exploring Algorithm Visualisations and User Biases.” In *Human-Computer Interaction. Human Values and Quality of Life*, edited by Masaaki Kurosu, 382–99. Cham, Denmark: Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-49065-2_27.

Brent, Edward E. “Expert Systems.” In *SAGE Research Methods Foundations*. London, UK: SAGE Publications Ltd, 2020. https://doi.org/10.4135/9781526421036842862.

Buzhinsky, Igor, Arseny Nerinovsky, and Stavros Tripakis. “Metrics and Methods for Robustness Evaluation of Neural Networks with Generative Models.” *Machine Learning*, July 6, 2021. https://doi.org/10.1007/s10994-021-05994-9.

Changchit, Chuleeporn. “Expert Systems.” In *Encyclopedia of Information Technology Curriculum Integration:*, edited by Lawrence A. Tomei, 316–22. Hershey, PA: IGI Global, 2008. https://doi.org/10.4018/978-1-59904-881-9.

Chatterjee, Sheshadri, Nripendra P. Rana, Sangeeta Khorana, Patrick Mikalef, and Anuj Sharma. “Assessing Organizational Users’ Intentions and Behavior to AI Integrated CRM Systems: A Meta-UTAUT Approach.” *Information Systems Frontiers*, August 13, 2021. https://doi.org/10.1007/s10796-021-10181-1.

Cheney, James, Laura Chiticariu, and Wang-chiew Tan. “Provenance in Databases: Why, How, and Where.” *Foundations and Trends in Databases* 1 (January 1, 2009): 379–474. https://doi.org/10.1561/1900000006.

Chiang, Chun-Wei, and Ming Yin. “Exploring the Effects of Machine Learning Literacy Interventions on Laypeople’s Reliance on Machine Learning Models.” In *27th International Conference on Intelligent User Interfaces*, 148–61. Helsinki Finland: ACM, 2022. https://doi.org/10.1145/3490099.3511121.

Chintalapati, Srikrishna, and Shivendra Kumar Pandey. “Artificial Intelligence in Marketing: A Systematic Literature Review.” *International Journal of Market Research* 64, no. 1 (January 2022): 38–68. https://doi.org/10.1177/14707853211018428.

Chmielinski, Kasia S., Sarah Newman, Matt Taylor, Josh Joseph, Kemi Thomas, Jessica Yurkofsky, and Yue Chelsea Qiu. “The Dataset Nutrition Label (2nd Gen): Leveraging Context to Mitigate Harms in Artificial Intelligence.” arXiv, March 10, 2022. https://doi.org/10.48550/arXiv.2201.03954.

Chowdhery, Aakanksha, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham et al. “PaLM: Scaling Language Modeling with Pathways.” arXiv, October 5, 2022. https://doi.org/10.48550/arXiv.2204.02311.

Chui, Michael, Bryce Hall, Alex Singla, and Alex Sukharevsky. "The State of AI in 2021," December 8, 2021. http://ceros.mckinsey.com/global-ai-survey-2020-a-desktop-3-1.

Clark, Alexander, Chris Fox, and Shalom Lappin, eds. *The Handbook of Computational Linguistics and Natural Language Processing*. Blackwell Handbooks in Linguistics. Malden, MA: Wiley-Blackwell, 2010.

Cohere AI. "Generation," November 4, 2022. https://docs.cohere.ai/docs/generation-card.

Cohere AI. "Representation," November 4, 2022. https://docs.cohere.ai/docs/generation-card.

Cook, Karen S., Coye Cheshire, Alexandra Gerbasi, and Brandy Aven. "Assessing Trustworthiness in Providers." In *ETrust: Forming Relationships in the Online World*. New York. NY: Russell Sage Foundation, 2009. https://www.jstor.org/stable/10.7758/9781610446082.

Cornelissen, Lisa, Claudia Egher, Vincent van Beek, Latoya Williamson, and Daniel Hommes. "The Drivers of Acceptance of Artificial Intelligence–Powered Care Pathways Among Medical Professionals: Web-Based Survey Study." *JMIR Formative Research* 6, no. 6 (June 21, 2022). https://doi.org/10.2196/33368.

Coughlan, Peter, Nicholas Dew, and William Gates. *Crossing the Technology Adoption Chasm: Implications for DOD*. Monterey, CA: Naval Postgraduate School, 2008. https://calhoun.nps.edu/handle/10945/434.

Cramer, Henriette, Jean Garcia-Gathright, Sravana Reddy, Aaron Springer, and Romain Takeo Bouyer. "Translation, Tracks & Data: An Algorithmic Bias Effort in Practice." In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–8. Glasgow, Scotland: ACM, 2019. https://doi.org/10.1145/3290607.3299057.

CRM Analytics. "Detect Sentiment Model Card." Salesforce Help, August 12, 2020. https://help.salesforce.com/s/articleView?id=sf.bi_integrate_transformation_detectSentimentModelCard.htm&type=5.

Cunningham, Pádraig, Matthieu Cord, and Sarah Jane Delany. "Supervised Learning." In *Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval*, edited by Matthieu Cord and Pádraig Cunningham, 21–49. Cognitive Technologies. Berlin, Germany: Springer, 2008. https://doi.org/10.1007/978-3-540-75171-7_2.

Daradkeh, Mohammad. "Determinants of Self-Service Analytics Adoption Intention: The Effect of Task-Technology Fit, Compatibility, and User Empowerment." *Journal of Organizational and End User Computing (JOEUC)* 31, no. 4 (2019): 19–45. https://doi.org/10.4018/JOEUC.2019100102.

Daugherty, William E., and Morris Janowitz. *A Psychological Warfare Casebook*. Baltimore, MD: The Johns Hopkins Press, 1958.

Davenport, Thomas H., Jeff Loucks, and David Schatsky. *The 2017 Deloitte State of Cognitive Survey*. Deloitte, 2017. https://www2.deloitte.com/content/dam/Deloitte/us/Documents/deloitte-analytics/us-da-2017-deloitte-state-of-cognitive-survey.pdf.

Davis, Fred D. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology." *MIS Quarterly* 13, no. 3 (1989): 319–40. https://doi.org/10.2307/249008.

Dearth, D. H. "Implications and Challenges of Applied Information Operations." *Journal of Information Warfare* 1, no. 1 (2001): 7–15. http://www.jstor.org/stable/26485919.

Defense Acquisition University. "CH 8 Test & Evaluation." In *Defense Acquisition Guidebook*. Fort Belvoir, VA: The Defense Acquisition University Press, 2020. https://www.dau.edu:443/pdfviewer?Guidebooks/DAG/DAG-CH-8-Test-and-Evaluation.pdf.

Denny, Matthew J., and Arthur Spirling. "Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It." *Political Analysis* 26, no. 2 (2018): 168–89. https://www.jstor.org/stable/26563824.

Department of Defense. "DOD Responsible AI Working Council." *U.S. Department of Defense Responsible Artificial Intelligence Strategy and Implementation Pathway*. Washington, DC: Department of Defense, 2022. https://www.ai.mil/docs/RAI_Strategy_and_Implementation_Pathway_6-21-22.pdf.

Department of Defense. *DOD Data Strategy*. Washington, DC: Department of Defense, 2020. https://media.defense.gov/2020/Oct/08/2002514180/-1/-1/0/DOD-DATA-STRATEGY.PDF.

Department of Defense. *DOD Enterprise DevSecOps Fundamentals*. Washington, DC: Department of Defense, 2021. https://dodcio.defense.gov/Portals/0/Documents/Library/DoDEnterpriseDevSecOpsFundamentals.pdf.

Dhanda, Namrata, Stuti Shukla Datta, and Mudrika Dhanda. "Machine Learning Algorithms." In *Computational Intelligence in the Internet of Things*, edited by Hindriyanto Dwi Purnomo, 210–33. Hershey, PA: IGI Global, 2019. https://doi-org.libproxy.nps.edu/10.4018/978-1-5225-7955-7.ch009.

Digital Catapult. "Ethics Framework." Machine Intelligence Garage, 2022. https://migarage.digicatapult.org.uk/ethics/ethics-framework/.

Downes-Martin, Stephen. "Operations Assessment in Afghanistan Is Broken—What Is to Be Done?" *Naval War College Review*, no. 4 (2011): 103–25.

DrivenData. "An Ethics Checklist for Data Scientists." Deon. Accessed October 3, 2022. https://deon.drivendata.org/.

Duarte, Natasha. "Digital Decisions Tool." *Center for Democracy and Technology* (blog), August 8, 2017. https://cdt.org/insights/digital-decisions-tool/.

Dunnmon, Jared, Bryce Goodman, Peter Kirechu, Carol J. Smith, and Alexandrea Van Deusen. "Phase I: Planning Worksheet for DIU AI Guidelines." Defense Innovation Unit, November 14, 2021. https://assets.ctfassets.net/3nanhbfkr0pc/1vJvimVkijueLbJzqcaOMr/39691b01cc00ca98a295804e269d6f51/Planning_Worksheet_DIU_AIGuidelines.pdf.

Dunnmon, Jared, Bryce Goodman, Peter Kirechu, Carol J. Smith, and Alexandrea Van Deusen. *Responsible AI Guidelines in Practice*. Mountain View, CA: Defense Innovation Unit, 2021. https://www.diu.mil/responsible-ai-guidelines.

Egel, Daniel, Ryan Andrew Brown, Linda Robinson, Mary Kate Adgie, Jasmin Léveillé, and Luke J. Matthews. *Leveraging Machine Learning for Operation Assessment*. Santa Monica, CA: RAND Corporation, 2022. https://www.rand.org/pubs/research_reports/RR4196.html.

Economist, The. "Businesses Are Finding AI Hard to Adopt," June 11, 2020. https://www.economist.com/technology-quarterly/2020/06/11/businesses-are-finding-ai-hard-to-adopt.

elastic. "Kibana Query Language," 2022. https://www.elastic.co/guide/en/kibana/7.17/kuery-query.html.

European Commission. "Ethics Guidelines for Trustworthy AI | Shaping Europe's Digital Future." Shaping Europe's digital future, April 8, 2019. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

Fan, Wenhui, Peiyu Chen, Daiming Shi, Xudong Guo, and Li Kou. "Multi-Agent Modeling and Simulation in the AI Age." *Tsinghua Science and Technology* 26, no. 5 (October 2021): 608–24. https://doi.org/10.26599/TST.2021.9010005.

Fei, Nanyi, Zhiwu Lu, Yizhao Gao, Guoxing Yang, Yuqi Huo, Jingyuan Wen, Haoyu Lu et al. "Towards Artificial General Intelligence via a Multimodal Foundation Model." *Nature Communications* 13, no. 1 (December 2022). https://doi.org/10.1038/s41467-022-30761-2.

Felzmann, Heike, Eduard Fosch Villaronga, Christoph Lutz, and Aurelia Tamò-Larrieux. "Transparency You Can Trust: Transparency Requirements for Artificial Intelligence between Legal Norms and Contextual Concerns." *Big Data & Society* 6, no. 1 (January 1, 2019). https://doi.org/10.1177/2053951719860542.

Fjelland, Ragnar. "Why General Artificial Intelligence Will Not Be Realized." *Palgrave Communications* 7, no. 1 (June 2020). https://doi.org/10.1057/s41599-020-0494-4.

Floridi, Luciano, and Massimo Chiriatti. "GPT-3: Its Nature, Scope, Limits, and Consequences." *Minds and Machines* 30, no. 4 (December 1, 2020): 681–94. https://doi.org/10.1007/s11023-020-09548-1.

Flournoy, Michele A., Avril Haines, and Gabrielle Chefitz. *Building Trust through Testing*. Washington, DC: Center for Security and Emerging Technology, 2020. https://cset.georgetown.edu/wp-content/uploads/Building-Trust-Through-Testing.pdf.

Foddy, Margaret, and Toshio Yamagishi. "Group-Based Trust." In *Whom Can We Trust?: How Groups, Networks, and Institutions Make Trust Possible*, edited by Karen S. Cook, Margaret Levi, and Russell Hardin, 17–41. New York. NY: Russell Sage Foundation, 2009.

Fountaine, Tim, Brian McCarthy, and Tamim Saleh. "Building the AI-Powered Organization." *Harvard Business Review*, July 1, 2019. https://hbr.org/2019/07/building-the-ai-powered-organization.

Franceschini, Paula M., Henrique D. P. dos Santos, and Renata Vieira. "Intrinsic and Extrinsic Evaluation of the Quality of Biomedical Embeddings in Different Languages." In *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, 271–76, 2020. https://doi.org/10.1109/CBMS49503.2020.00058.

Gallup, Shelley P. "Future War at Sea: The U.S. Navy, Autonomy in War at Sea and Emergent Behaviors." In *Emergent Behavior in System of Systems Engineering*, edited by Larry B. Rainey and O. Thomas Holland, 64–92. Boca Raton, FL: CRC Press, 2022.

Gansser, Oliver Alexander, and Christina Stefanie Reich. "A New Acceptance Model for Artificial Intelligence with Extensions to UTAUT2: An Empirical Study in Three Segments of Application." *Technology in Society* 65 (May 1, 2021). https://doi.org/10.1016/j.techsoc.2021.101535.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. "Datasheets for Datasets." *Communications of the ACM* 64, no. 12 (December 1, 2021): 86–92. https://www.microsoft.com/en-us/research/publication/datasheets-for-datasets/.

Geroski, P. A. *The Evolution of New Markets*. New York, NY: Oxford University Press, 2003.

Gilbert, Thomas Krendl, Sarah Dean, Tom Zick, and Nathan Lambert. *Choices, Risks, and Reward Reports: Charting Public Policy for Reinforcement Learning Systems*. Berkeley, CA: UC Berkeley Center for Long-Term Cybersecurity, 2022. https://cltc.berkeley.edu/wp-content/uploads/2022/02/Choices_Risks_Reward_Reports.pdf.

GitHub. "BlenderBot 2.0 2.7B," September 8, 2021. https://github.com/facebookresearch/ParlAI/blob/f427b4cf97e13448a91d02212914591b0a19877f/parlai/zoo/blenderbot2/model_card.md.

GitHub. "Foundation: An Economic Simulation Framework," October 14, 2020. https://github.com/salesforce/ai-economist/blob/a84d5f3fdcabb207d9fde7754d34906903b3e184/Simulation_Card_Foundation_Economic_Simulation_Framework.pdf.

GitHub. "GLIDE." December 20, 2021. https://github.com/openai/glide-text2im/blob/69b530740eb6cef69442d6180579ef5ba9ef063e/model-card.md.

GitHub. "GPT-2 Model Card," November 2019. https://github.com/openai/gpt-2/blob/a74da5d99abaaba920de8131d64da2862a8f213b/model_card.md.

GitHub. "GPT-3 Model Card," September 2020. https://github.com/openai/gpt-3/blob/d7a9bb505df6f630f9bab3b30c889e52f22eb9ea/model-card.md.

GitHub. "Guided-Diffusion," July 19, 2021. https://github.com/openai/guided-diffusion/blob/22e0df8183507e13a7813f8d38d51b072ca1e67c/model-card.md.

GitHub. "InstructGPT Model Card," January 2022. https://github.com/openai/following-instructions-human-feedback/blob/5c0534c78136809ab0118c5c1b74d159befa2467/model-card.md.

GitHub. "Model Card for CTRL - A Conditional Transformer Language Model for Controllable Generation," April 20, 2020. https://github.com/salesforce/ctrl/blob/a089a0c5956a84f229e5ece9d5da0bc4689dded0/ModelCard.pdf.

GitHub. "Model Card: CLIP," April 2021. https://github.com/openai/CLIP/blob/d50d76daa670286dd6cacf3bcd80b5e4823fc8e1/model-card.md.

GitHub. "Model Card: DALL·E DVAE," November 23, 2022. https://github.com/openai/DALL-E/blob/5be4b236bc3ade6943662354117a0e83752cc322/model_card.md.

GitHub. “Model Card: Detector for CTRL from Salesforce Research,” October 13, 2020. https://github.com/salesforce/ctrl-detector/blob/f1ec83bdb5fc176f5c42b353b12df94fffce2c87/ModelCard.pdf.

Github. “Model Card: Summarization Model Trained with Human Feedback,” September 2020. https://github.com/openai/summarize-from-feedback/blob/56b6bb613a1b58a8aa7a5e29266f65c7b980ee48/model_card.md.

GitHub. “Toxicity,” March 9, 2021. https://github.com/conversationai/perspectiveapi/blob/a51953fab94c34be2901c2ecfe4320c4d3b7039a/model-cards/English/toxicity.md.

GitHub. “XGLM Multilingual Model,” February 1, 2022. https://github.com/facebookresearch/fairseq/blob/main/examples/xglm/model_card.md.

Global Engagement Center Technology Engagement Team. “Disinfo Cloud.” U.S. Department of State, 2022. https://www.state.gov/wp-content/uploads/2021/04/Disinfo-Cloud-flyer-April-042921.pdf.

Google Cloud Model Cards. “Face Detection.” Accessed August 12, 2022. https://modelcards.withgoogle.com/face-detection.

Google Cloud Model Cards. “Object Detection.” Accessed August 12, 2022. https://modelcards.withgoogle.com/object-detection.

Google Developers. “The Size and Quality of a Data Set | Machine Learning.” Machine Learning, July 18, 2022. https://developers.google.com/machine-learning/data-prep/construct/collect/data-size-quality.

Google Developers. “Classification: ROC Curve and AUC | Machine Learning.” Machine Learning, July 18, 2022. https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc.

Google. “Model Card Hand Tracking (Lite_Full).” Google MediaPipe, October 2021. https://drive.google.com/file/d/1-rmIgTfuCbBPW_IFHkh3f0-U_lnGrWpg/preview?usp=embed_facebook.

Gourville, John T. “Eager Sellers and Stony Buyers: Understanding the Psychology of New-Product Adoption.” *Harvard Business Review*, June 1, 2006. https://hbr.org/2006/06/eager-sellers-and-stony-buyers-understanding-the-psychology-of-new-product-adoption.

Govindaraj, Vishnuvarthanan, Arunprasath Thiyagarajan, Pallikonda Rajasekaran, Yudong Zhang, and Rajesh Krishnasamy. "Automated Unsupervised Learning-Based Clustering Approach for Effective Anomaly Detection in Brain Magnetic Resonance Imaging (MRI)." *IET Image Processing* 14, no. 14 (2020): 3516–26. https://doi.org/10.1049/iet-ipr.2020.0597.

Graf-Vlachy, Lorenz, Katharina Buhtz, and Andreas König. "Social Influence in Technology Adoption: Taking Stock and Moving Forward." *Management Review Quarterly* 68, no. 1 (February 2018): 37–76. https://doi.org/10.1007/s11301-017-0133-3.

Granovetter, Mark S. *Society and Economy: Framework and Principles*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press, 2017.

Graphika, Stanford Internet Observatory. *Unheard Voice: Evaluating Five Years of pro-Western Covert Influence Operations*. Stanford, CA: Stanford Cyber Policy Center, 2022.

Green, Nekesha, Chavez Procope, Adeel Cheema, and Adekunle Adediji. "System Cards, a New Resource for Understanding How AI Systems Work." Meta AI, February 23, 2022. https://ai.facebook.com/blog/system-cards-a-new-resource-for-understanding-how-ai-systems-work/.

Grishchenko, Ivan, Artsiom Ablavatski, and Yury Kartynnik. "Model Card MediaPipe Face Mesh." Google MediaPipe, October 10, 2018. https://drive.google.com/file/d/1QvwWNfFoweGVjsXF3DXzcrCnz-mx-Lha/preview?usp=embed_facebook.

Gunning, David, Eric Vorm, Jennifer Yunyan Wang, and Matt Turek. "DARPA's Explainable AI (XAI) Program: A Retrospective." *Applied AI Letters* 2, no. 4 (2021). https://doi.org/10.1002/ail2.61.

Hailu, Tulu Tilahun, Junqing Yu, and Tessfu Geteye Fantaye. "Intrinsic and Extrinsic Automatic Evaluation Strategies for Paraphrase Generation Systems." *Journal of Computer and Communications* 8, no. 2 (2020): 1–16. https://doi.org/10.4236/jcc.2020.82001.

Hall, Patrick. "On the Art and Science of Machine Learning Explanations." In *2019 KDD XAI Workshop*. Anchorage, AK: arXiv, 2020. http://arxiv.org/abs/1810.02909.

Hangl, Johannes, Viktoria Joy Behrens, and Simon Krause. "Barriers, Drivers, and Social Considerations for AI Adoption in Supply Chain Management: A Tertiary Study." *Logistics* 6, no. 3 (2022). https://doi.org/10.3390/logistics6030063.

Hanif, Muhammad Abdullah, and Muhammad Shafique. "Dependable Deep Learning: Towards Cost-Efficient Resilience of Deep Neural Network Accelerators against Soft Errors and Permanent Faults." In *2020 IEEE 26th International Symposium on On-Line Testing and Robust System Design (IOLTS)*, 1–4, 2020. https://doi.org/10.1109/IOLTS50870.2020.9159734.

Hardin, Russell. *Trust and Trustworthiness*. New York: Russell Sage Foundation, 2002.

Harnessing Artificial Intelligence 2.0 - Human Machine Interaction AI, 2021. https://www.youtube.com/watch?v=eMx2_gZqs7U.

Harris, Mark. *Five Came Back*. New York, NY: Penguin Books, 2014.

Hartwick, John, and Henri Barki. "Explaining the Role of User Participation in Information System Use." *Management Science* 40, no. 4 (1994): 440–65. http://www.jstor.org/stable/2632752.

Hasija, Abhinav, and Terry L. Esper. "In Artificial Intelligence (AI) We Trust: A Qualitative Investigation of AI Technology Acceptance." *Journal of Business Logistics* 43, no. 3 (February 1, 2022): 388–412. https://doi.org/10.1111/jbl.12301.

Hawley, Katherine. *TRUST: A Very Short Introduction*. First edition. Oxford, UK: Oxford University Press, 2012.

Hershcovich, Daniel, Nicolas Webersinke, Mathias Kraus, Julia Anna Bingler, and Markus Leippold. "Towards Climate Awareness in NLP Research." In *2022 Conference on Empirical Methods in Natural Language Processing*. Abu Dhabi: arXiv, 2022. http://arxiv.org/abs/2205.05071.

Hertzum, Morten. "Usability Testing: A Practitioner's Guide to Evaluating the User Experience." *Synthesis Lectures on Human-Centered Informatics* 1, no. 1 (March 9, 2020): i–105. https://doi.org/10.2200/S00987ED1V01Y202001HCI045.

Hicks, Kathleen H. "Implementing Responsible Artificial Intelligence in the Department of Defense." Official memorandum, Washington, DC: Department of Defense, 2021.

Hill, Gladwin. "The American Films' Program in Occupied Germany." In *A Psychological Warfare Casebook*, edited by William E. Daugherty and Morris Janowitz, 574–79. Baltimore, MD: The Johns Hopkins Press, 1958.

Hind, Michael, Stephanie Houde, Jacquelyn Martino, Aleksandra Mojsilovic, David Piorkowski, John Richards, and Kush R. Varshney. "Experiences with Improving the Transparency of AI Models and Services." In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–8. CHI EA '20. New York, NY: Association for Computing Machinery, 2020. https://doi.org/10.1145/3334480.3383051.

Hinen, Anthony L., David J. Slattery, J. Matthew Venhaus, Brian E. Rathgeb, Michael W. Fulk, and John R. Woodhill. *Data Guide for Special Operations Forces (SOF) Military Information Support Operations (MISO)*. Tampa, FL: MITRE, 2020.

Hoo, Zhe Hui, Jane Candlish, and Dawn Teare. "What Is an ROC Curve?" *Emergency Medicine Journal* 34, no. 6 (June 2017): 357–59. https://doi.org/10.1136/emermed-2017-206735.

Horowitz, Michael C., Lauren Kahn, and Laura Resnick Samotin. "A Force for the Future." *Foreign Affairs*, June 2022. https://www.foreignaffairs.com/articles/united-states/2022-04-19/force-future.

Horowitz, Michael, Paul Scharre, Gregory C. Allen, Kara Frederick, Anthony Cho, and Edoardo Saravalle. *Artificial Intelligence and International Security*. Washington, DC: Center for a New American Security, 2018. https://www.cnas.org/publications/reports/artificial-intelligence-and-international-security.

HuggingFace. "Bert-Base-Multilingual-Uncased · Hugging Face." Accessed September 23, 2022. https://huggingface.co/bert-base-multilingual-uncased.

HuggingFace. "Dalle-Mini/Dalle-Mini · Hugging Face," 2021. https://huggingface.co/dalle-mini/dalle-mini.

Hurlburt, George. "How Much to Trust Artificial Intelligence?" *IT Professional* 19, no. 4 (2017): 7–11. https://doi.org/10.1109/MITP.2017.3051326.

Hutchinson, Ben, Negar Rostamzadeh, Christina Greer, Katherine Heller, and Vinodkumar Prabhakaran. "Evaluation Gaps in Machine Learning Practice." In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. Seoul, Republic of Korea: ACM, 2022. https://doi.org/10.1145/3531146.3533233.

IBM Research. "AI FactSheets 360." IBM Research AI FactSheets 360. Accessed October 2, 2022. https://aifs360.mybluemix.net/.

IBM Research. "Weather Forecaster FactSheet." IBM Research AI FactSheets 360. Accessed September 29, 2022. https://aifs360.mybluemix.net/examples/max_weather_forecaster.

IBM Watson. *IBM Global AI Adoption Index 2022*. Armonk, NY: IBM Corporation, 2022. https://www.ibm.com/watson/resources/ai-adoption.

Ilachinski, Andy, and David Broyles. “Face/Off.” Produced by CNA. In *AI with AI*, November 19. 2021. Podcast, MP3 audio, 40:13. https://www.cna.org/our-media/podcasts/ai-with-ai/season-5/5-3.

IMDb. “2001: A Space Odyssey (1968).” Accessed November 13, 2022. http://www.imdb.com/title/tt0062622/characters/nm0706937.

International Organization for Standardization. “ISO 9241–11 Ergonomics of Human-System Interaction — Part 11: Usability: Definitions and Concepts.” ISO Online Browsing Platform (OBP), 2018. https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en.

International Skin Imaging Collaboration (ISIC). “Dataset Nutrition Label: 2020 SIIM-ISIC Melanoma Classification Challenge Dataset [Draft].” Data Nutrition Project, 2020. https://datanutrition.org/labels/isic-2020/.

Ish, Daniel, Jared Ettinger, and Christopher Ferris. *Evaluating the Effectiveness of Artificial Intelligence Systems in Intelligence Analysis*. Santa Monica, CA: RAND Corporation, 2021. https://www.rand.org/pubs/research_reports/RRA464-1.html.

IST Research. “Pulse,” 2014. https://web.archive.org/web/20151127082326/http://istresearch.com/pulse/.

Jain, Ruchika, Naval Garg, and Shikha N. Khera. “Adoption of AI-Enabled Tools in Social Development Organizations in India: An Extension of UTAUT Model.” *Frontiers in Psychology* 13 (June 20, 2022). https://doi.org/10.3389/fpsyg.2022.893691.

Jöhnk, Jan, Malte Weißert, and Katrin Wyrtki. “Ready or Not, AI Comes— An Interview Study of Organizational AI Readiness Factors.” *Business & Information Systems Engineering* 63, no. 1 (February 2021): 5–20. https://doi.org/10.1007/s12599-020-00676-7.

Johnson, Bradley. “How Marketers Got Back in the Game: Ad Spending Surged Last Year with the Biggest Increase Seen since the 1970s. But Budgets Could Come under Pressure amid Signs of a Looming Recession.” *Advertising Age* 93, no. 10 (June 27, 2022): 14–14.

Joint Artificial Intelligence Center. “Test and Evaluation Framework for AI-Enabled Systems.” Department of Defense, Washington, DC, April 27, 2022. https://testscience.org/wp-content/uploads/formidable/13/Final-change-1-20220427-DataWORKS-Session-1A-TE-for-AI-Enabled-Systems-Lt-Col-Woolley.pdf.

Joint Artificial Intelligence Center. "JAIC T&E Workshop." Presentation, Department of Defense, October 28, 2021.

Joint Artificial Intelligence Center. "Responsible Artificial Intelligence Test & Evaluation (RAITE)." Washington, DC: Department of Defense, April 11, 2022. https://www.ai.mil/docs/raite_slick_220411.pdf.

Joint Artificial Intelligence Center. *Responsible Artificial Intelligence (RAI): Transforming the Department of Defense Through AI*. Washington, DC: Department of Defense, 2022. https://www.ai.mil/docs/rai_slick_sheet-dopsr_appvd_220317.pdf.

Joint Chiefs of Staff. "Joint Doctrine Publications." Accessed October 10, 2022. https://www.jcs.mil/Doctrine/Joint-Doctine-Pubs/.

Joint Chiefs of Staff. *Civil-Military Operations*. JP 3-57. Washington, DC: Department of Defense, 2018. https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_57.pdf.

Joint Chiefs of Staff. *Information in Joint Operations*. JP 3-04. Washington, DC: Department of Defense, 2022.

Joint Chiefs of Staff. *Information Operations*. JP 3-13. Washington, DC: Department of Defense, 2014. https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_13.pdf.

Joint Chiefs of Staff. *Joint Intelligence Preparation of the Operational Environment*. Washington, DC: Department of Defense, 2014. https://irp.fas.org/doddir/dod/jp2-01-3.pdf.

Joint Chiefs of Staff. *Joint Operations*. JP 3-0. Washington, DC: Department of Defense, 2022. https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_13.pdf.

Joint Chiefs of Staff. *Joint Planning*. JP 5-0. Washington, DC: Department of Defense, 2020. https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_13.pdf.

Joint Chiefs of Staff. *Military Information Support Operations*. JP 3-13.2. Washington, DC: Department of Defense, 2014. https://irp.fas.org/doddir/dod/jp3-13-2.pdf.

Joint Chiefs of Staff. *Public Affairs*. JP 3-61. Washington, DC: Department of Defense, 2016. https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_61.pdf.

Jones, Harry. "Going Beyond Reliability to Robustness and Resilience in Space Life Support Systems." In *50th International Conference on Environmental Systems*. Lisbon, Portugal: Texas Tech University, 2021. https://ttu-ir.tdl.org/handle/2346/87122.

Kaplan, Alexandra D., Theresa T. Kessler, J. Christopher Brill, and P. A. Hancock. "Trust in Artificial Intelligence: Meta-Analytic Findings." *Human Factors*, May 28, 2021. https://doi.org/10.1177/00187208211013988.

Karahanna, Elena, Ritu Agarwal, and Corey M. Angst. "Reconceptualizing Compatibility Beliefs in Technology Acceptance Research." *MIS Quarterly* 30, no. 4 (2006): 781–804. https://doi.org/10.2307/25148754.

Karahasanovic, Amela, Erik Gausta Nilsson, Giorgio Grani, Kjell Fredrik Pettersen, Leo Karabeg, and Patrick Schittekat. "User Involvement in the Design of ML-Infused Systems." In *CHI Greece 2021: 1st International Conference of the ACM Greek SIGCHI Chapter*. Athens, Greece: ACM, 2021. https://doi.org/10.1145/3489410.3489421.

Kartynnik, Yury. "Model Card MediaPipe BlazeFace Full Range." Google MediaPipe, January 19, 2021. https://drive.google.com/file/d/1jpQt8TB1nMFQ49VSSBKdNEdQOygNRvCP/preview?usp=embed_facebook.

Kaur, Davinder, Suleyman Uslu, Kaley J. Rittichier, and Arjan Durresi. "Trustworthy Artificial Intelligence: A Review." *ACM Computing Surveys* 55, no. 2 (January 18, 2022): 39:1-39:38. https://doi.org/10.1145/3491209.

Kelman, Herbert C. "Compliance, Identification, and Internalization Three Processes of Attitude Change." *Journal of Conflict Resolution* 2, no. 1 (March 1958): 51–60. https://doi.org/10.1177/002200275800200106.

Kemp, Simon. "Digital 2022: Global Overview Report." DataReportal – Global Digital Insights, January 26, 2022. https://datareportal.com/reports/digital-2022-global-overview-report.

Kettler, Brian. "Influence Campaign Awareness and Sensemaking." Defense Advanced Research Projects Agency. Accessed August 29, 2022. https://www.darpa.mil/program/influence-campaign-awareness-and-sensemaking.

Khan, Muhammad Zohaib. "Comparing the Performance of NLP Toolkits and Evaluation Measures in Legal Tech." Master's thesis, Universität Passau, 2021. http://arxiv.org/abs/2103.11792.

Khanijahani, Ahmad, Shabnam Iezadi, Sage Dudley, Megan Goettler, Peter Kroetsch, and Jama Wise. "Organizational, Professional, and Patient Characteristics Associated with Artificial Intelligence Adoption in Healthcare: A Systematic Review." *Health Policy and Technology* 11, no. 1 (March 1, 2022). https://doi.org/10.1016/j.hlpt.2022.100602.

Kinder-Kurlanda, Katharina E., and Katrin Weller. “Perspective: Acknowledging Data Work in the Social Media Research Life cycle.” *Frontiers in Big Data* 3 (December 2020). https://doi.org/10.3389/fdata.2020.509954.

Kiron, David, and Michael Schrage. “Strategy For and With AI.” *MIT Sloan Management Review* 60, no. 4 (Summer 2019): 30–35. http://www.proquest.com/docview/2273705001/abstract/9489BC4062544CF0PQ/1.

Kiron, David. “AI Can Change How You Measure - and How You Manage.” *MIT Sloan Management Review* 63, no. 3 (Spring 2022): 24–28. https://www.proquest.com/docview/2655629383/abstract/5D3E21D5094F4D31PQ/1.

Knight, Will. “AI’s Smarts Now Come With a Big Price Tag.” *Wired*, October 14, 2021. https://www.wired.com/story/ai-smarts-big-price-tag/.

Konaev, Margarita, Tina Huang, and Husanjot Chahal. *Trusted Partners: Human-Machine Teaming and the Future of Military AI*. Washington, DC: Center for Security and Emerging Technology, 2021. https://cset.georgetown.edu/publication/trusted-partners/.

Kulkarni, Ajay, Deri Chong, and Feras A. Batarseh. “Foundations of Data Imbalance and Solutions for a Data Democracy.” In *Data Democracy*, 83–106. Amsterdam, Netherlands: Elsevier, 2020. https://doi.org/10.1016/B978-0-12-818366-3.00005-8.

Kumar, Anuj, and Anjali Kalse. “Usage and Adoption of Artificial Intelligence in SMEs.” *Materials Today: Proceedings*, February 26, 2021. https://doi.org/10.1016/j.matpr.2021.01.595.

Lamb, Christopher J. *Review of Psychological Operations Lessons Learned from Recent Operational Experience*. Washington, DC: National Defense University, 2005. https://apps.dtic.mil/sti/citations/ADA445151.

Launchbury, John. “A DARPA Perspective on Artificial Intelligence.” Presentation, DARPA, Arlington County, VA, February 15, 2017. https://www.darpa.mil/attachments/AIFull.pdf.

Lavigne, Mathieu. “Strengthening Ties: The Influence of Microtargeting on Partisan Attitudes and the Vote.” *Party Politics* 27, no. 5 (September 1, 2021): 965–76. https://doi.org/10.1177/1354068820918387.

Law, Effie Lai-Chong, Virpi Roto, Marc Hassenzahl, Arnold P.O.S. Vermeeren, and Joke Kort. “Understanding, Scoping and Defining User Experience: A Survey Approach.” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 719–28. New York, NY: Association for Computing Machinery, 2009. https://doi.org/10.1145/1518701.1518813.

Lee, J., M. Ghaffari, and S. Elmeligy. “Self-Maintenance and Engineering Immune Systems: Towards Smarter Machines and Manufacturing Systems.” *Annual Reviews in Control* 35, no. 1 (April 2011): 111–22. https://doi.org/10.1016/j.arcontrol.2011.03.007.

Leeuwen, Rik van, and Ger Koole. “Data-Driven Market Segmentation in Hospitality Using Unsupervised Machine Learning.” *Machine Learning with Applications* 10 (December 15, 2022). https://doi.org/10.1016/j.mlwa.2022.100414.

Li, Haoran, Junnan Zhu, Cong Ma, Jiajun Zhang, and Chengqing Zong. “Read, Watch, Listen, and Summarize: Multi-Modal Summarization for Asynchronous Text, Image, Audio and Video.” *IEEE Transactions on Knowledge and Data Engineering* 31, no. 5 (May 2019): 996–1009. https://doi.org/10.1109/TKDE.2018.2848260.

Liu, Ming-Yu, Xun Huang, Jiahui Yu, Ting-Chun Wang, and Arun Mallya. “Generative Adversarial Networks for Image and Video Synthesis: Algorithms and Applications.” *Proceedings of the IEEE* 109, no. 5 (November 30, 2020): 839–62. https://doi.org/10.1109/JPROC.2021.3049196.

Liveley, Genevieve, and Sam Thomas. “Homer’s Intelligent Machines: AI in Antiquity.” In *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*, edited by Stephen Cave, Kanta Dihal, and Sarah Dillon, 25–48. Oxford, UK: Oxford University Press, 2020.

Loukides, Mike. “AI Adoption in the Enterprise 2022.” O’Reilly Media, March 31, 2022. https://www.oreilly.com/radar/ai-adoption-in-the-enterprise-2022/.

MacVaugh, Jason, and Francesco Schiavone. “Limits to the Diffusion of Innovation: A Literature Review and Integrative Model.” *European Journal of Innovation Management* 13, no. 2 (2010): 197–221. https://doi.org/10.1108/14601061011040258.

Maharana, Adyasha. “Storydalle.” GitHub, September 15, 2022. https://github.com/adymaharana/storydalle/blob/615881ace2dd0a5f5757cfeb4bdd1fb2c08f3337/MODEL_CARD.MD.

Makino, Sharma, M. Shrivastava, and B. Agarwal. “Denial-of-Service and Botnet Analysis, Detection, and Mitigation.” In *Forensic Investigations and Risk Management in Mobile and Wireless Communications*. Hershey, PA: IGI Global, 114AD.

Marcellino, William, Madeline Magnuson, Anne Stickells, Benjamin Boudreaux, Todd Helmus, Edward Geist, and Zev Winkelman. *Counter-Radicalization Bot Research: Using Social Bots to Fight Violent Extremism*. RR2705. Santa Monica, CA: RAND Corporation, 2020. https://doi.org/10.7249/RR2705.

Marikyan, Davit, and Savvas Papagiannidis. "Unified Theory of Acceptance and Use of Technology: A Review." In *TheoryHub Book*, edited by Savvas Papagiannidis, 16, 2021. https://open.ncl.ac.uk/theories/2/unified-theory-of-acceptance-and-use-of-technology/.

Markiewicz, Anne, and Ian Patrick. *Developing Monitoring and Evaluation Frameworks*. Thousand Oaks, CA: SAGE Publications, Inc, 2016. https://doi.org/10.4135/9781071878774.

Marsh, Alton K. *Guide to Defense & Aerospace Expert Systems*. Arlington, VA: Pasha Publications Inc., 1986. https://apps.dtic.mil/sti/pdfs/ADA636820.pdf.

Martinez, Dave, Nick Malyska, Bill Streilein, Rajmonda Caceres, William Cambell, Charlie Dagli, Vijay Tran et al. *Artificial Intelligence: Short History, Present Developments, and Future Outlook Final Report*. Cambridge, MA: Massachusetts Institute of Technology, 2019.

MAXQDA. "Document Map: Arranging Documents According to Similarity." MAXQDA, 2022. https://www.maxqda.com/help-mx20/visual-tools/document-map-arranging-documents-according-to-similarity.

MAXQDA. "Similarity Analysis for Documents." MAXQDA, 2022. https://www.maxqda.com/help-mx20/mixed-methods-functions/similarity-analysis-for-documents.

McCarthy, John. "What Is Artificial Intelligence?," November 12, 2007. http://jmc.stanford.edu/articles/whatisai/whatisai.pdf.

McKendrick, Joe. "AI Adoption Skyrocketed Over the Last 18 Months." *Harvard Business Review*, September 27, 2021. https://hbr.org/2021/09/ai-adoption-skyrocketed-over-the-last-18-months.

Medsker, Larry R. *Hybrid Neural Network and Expert Systems*. Boston, MA: Springer U.S., 1994. https://doi.org/10.1007/978-1-4615-2726-8.

Meegahapola, Lakmal, Vijini Mallawaarachchi, Eranga Heshan, Roshan Madusanka, Dulani Meedeniya, and Sampath Jayarathne. "Random Forest Classifier Based Scheduler Optimization for Search Engine Web Crawlers," 2018. https://doi.org/10.1145/3185089.3185103.

Menke, Joshua, and Tony Martinez. "Improving Supervised Learning By Adapting the Problem to the Learner." *International Journal of Neural Systems* 19, no. 1 (2009): 1–9. https://doi.org/10.1142/S0129065709001793.

Meta AI. "Instagram Feed Ranking System Card," February 23, 2022. https://ai.facebook.com/tools/system-cards/instagram-feed-ranking.

Metcalf, Alexander L., Justin W. Angle, Conor N. Phelan, B. Allyson Muth, and James C. Finley. "More 'Bank' for the Buck: Microtargeting and Normative Appeals to Increase Social Marketing Efficiency." *Social Marketing Quarterly* 25, no. 1 (March 1, 2019): 26–39. https://doi.org/10.1177/1524500418818063.

Miller, Ann. "Expert Systems: The Structure, History, and Future of Successful AI Applications." *IEEE Potentials* 5, no. 3 (October 1986): 12–15. https://doi.org/10.1109/MP.1986.6500801.

Min, Bonan, Hayley Ross, Elior Sulem, Amir Pouran Ben Veyseh, Thien Huu Nguyen, Oscar Sainz, Eneko Agirre, Ilana Heinz, and Dan Roth. "Recent Advances in Natural Language Processing via Large Pre-Trained Language Models: A Survey," November 1, 2021. http://arxiv.org/abs/2111.01243.

Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. "Model Cards for Model Reporting." In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–29. Atlanta GA: ACM, 2019. https://doi.org/10.1145/3287560.3287596.

ML Kit. "MLKit Selfie Segmentation," September 23, 2022. https://developers.google.com/ml-kit/images/vision/selfie-segmentation/selfie-model-card.pdf.

Mohan, Neeraj, Ruchi Singla, Priyanka Kaushal, and Seifedine Kadry, eds. *Artificial Intelligence, Machine Learning, and Data Science Technologies: Future Impact and Well-Being for Society 5.0*. Boca Raton, FL: CRC Press, 2021. https://doi.org/10.1201/9781003153405.

Mohanty, Sachi Nandan, Jyotir Moy Chatterjee, Sarika Jain, Ahmed A. Elngar, and Priya Gupta. *Recommender System with Machine Learning and Artificial Intelligence: Practical Tools and Applications in Medical, Agricultural and Other Industries*. Newark, NJ: John Wiley & Sons, Incorporated, 2020. http://ebookcentral.proquest.com/lib/ebook-nps/detail.action?docID=6225842.

Momani, Alaa. "The Unified Theory of Acceptance and Use of Technology: A New Approach in Technology Acceptance." *International Journal of Sociotechnology and Knowledge Development* 12 (July 1, 2020): 79–98. https://doi.org/10.4018/IJSKD.2020070105.

Moore, Geoffrey A. *Crossing the Chasm: Marketing and Selling Technology Project*. HarperCollins e-Books, 2014.

Mutasa, Simukayi, Shawn Sun, and Richard Ha. "Understanding Artificial Intelligence Based Radiology Studies: What Is Overfitting?" *Clinical Imaging* 65 (September 2020): 96–99. https://doi.org/10.1016/j.clinimag.2020.04.025.

Naja, Iman, Milan Markovic, Peter Edwards, Wei Pang, Caitlin Cottrill, and Rebecca Williams. "Using Knowledge Graphs to Unlock Practical Collection, Integration, and Audit of AI Accountability Information." *IEEE Access* 10 (2022): 74383–411. https://doi.org/10.1109/ACCESS.2022.3188967.

Nakashima, Ellen. "Pentagon Opens Sweeping Review of Clandestine Psychological Operations." *The Washington Post*, September 19, 2022. https://www.washingtonpost.com/national-security/2022/09/19/pentagon-psychological-operations-facebook-twitter/.

National Security Commission on Artificial Intelligence. *National Security Commission on Artificial Intelligence Final Report*. Washington, DC: National Security Commission on Artificial Intelligence, 2021. https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf.

Neumann, Oliver, Katharina Guirguis, and Reto Steiner. "Exploring Artificial Intelligence Adoption in Public Organizations: A Comparative Case Study." *Public Management Review*, March 20, 2022, 1–27. https://doi.org/10.1080/14719037.2022.2048685.

Nielsen, Michael. *Neural Networks and Deep Learning*. Determination Press, 2015. http://neuralnetworksanddeeplearning.com/.

Nightingale, Sophie J., and Hany Farid. "AI-Synthesized Faces Are Indistinguishable from Real Faces and More Trustworthy." *Proceedings of the National Academy of Sciences* 119, no. 8 (February 22, 2022). https://doi.org/10.1073/pnas.2120481119.

Nvidia NGC Catalog. "FaceDetect," January 13, 2022. https://catalog.ngc.nvidia.com/orgs/nvidia/teams/tao/models/facenet.

Nvidia NGC Catalog. "LPDNet," May 25, 2022. https://catalog.ngc.nvidia.com/orgs/nvidia/teams/tao/models/lpdnet.

Nvidia NGC Catalog. "NeMo - Text to Speech," September 22, 2022. https://catalog.ngc.nvidia.com/orgs/nvidia/collections/nemo_tts.

Nvidia NGC Catalog. "PeopleNet," May 26, 2022. https://catalog.ngc.nvidia.com/orgs/nvidia/teams/tao/models/peoplenet.

Nvidia NGC Catalog. "PeopleSegNet," August 22, 2022. https://catalog.ngc.nvidia.com/orgs/nvidia/teams/tao/models/peoplesegnet.

Nvidia NGC Catalog. "STT En Conformer-CTC Large," June 23, 2022. https://catalog.ngc.nvidia.com/orgs/nvidia/teams/nemo/models/stt_en_conformer_ctc_large.

Nvidia NGC Catalog. "TrafficCamNet," May 25, 2022. https://catalog.ngc.nvidia.com/orgs/nvidia/teams/tao/models/trafficcamnet.

Nvidia NGC Catalog. "VehicleTypeNet," November 7, 2021. https://catalog.ngc.nvidia.com/orgs/nvidia/teams/tao/models/vehicletypenet.

Nyczepir, Dave. "SOCOM Looks to Combat Disinformation in Africa on New Governmentwide Contract." *FedScoop*, July 27, 2020. https://www.fedscoop.com/socafrica-disinformation-ist-research/.

Ó Fathaigh, Ronan, Tom Dobber, Frederik Zuiderveen Borgesius, and James Shires. "Microtargeted Propaganda by Foreign Actors: An Interdisciplinary Exploration." *Maastricht Journal of European and Comparative Law* 28, no. 6 (December 1, 2021): 856–77. https://doi.org/10.1177/1023263X211042471.

O'Neill, Onora. *A Question of Trust*. Cambridge, UK: Cambridge University Press, 2003.

O'Sullivan, Patrick B, and Caleb T Carr. "Masspersonal Communication: A Model Bridging the Mass-Interpersonal Divide." *New Media & Society* 20, no. 3 (March 1, 2018): 1161–80. https://doi.org/10.1177/1461444816686104.

Office of the Director of National. "Artificial Intelligence Ethics Framework for the Intelligence Community." INTEL.gov, June 2020. https://www.intelligence.gov/artificial-intelligence-ethics-framework-for-the-intelligence-community.

Office of the Under Secretary of Defense (Comptroller)/Chief Financial Officer. *Operation and Maintenance Overview: Fiscal Year 2022 Budget Request*. Washington, DC: Department of Defense, 2021. https://comptroller.defense.gov/Portals/45/Documents/defbudget/FY2022/FY2022_OM_Overview.pdf.

Office of the Under Secretary of Defense for Acquisition and Sustainment. *DOD Instruction 5000.02 Operation of the Adaptive Acquisition Framework*. Washington, DC: Department of Defense, 2022. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500002p.pdf.

Ognyanova, Katherine. “Multistep Flow of Communication: Network Effects.” In *The International Encyclopedia of Media Effects*, edited by Patrick Rössler, Cynthia A. Hoffner, and Liesbet Zoonen, 1st ed., 1–10. Wiley, 2017. https://doi.org/10.1002/9781118783764.wbieme0056.

Oliveira, Tiago, and Maria Fraga Martins. “Literature Review of Information Technology Adoption Models at Firm Level.” *Electronic Journal of Information Systems Evaluation* 14, no. 1 (January 1, 2011): 110-121. https://academic-publishing.org/index.php/ejise/article/view/389.

Omaar, Hodan. “NSF Data Shows AI Adoption in the United States Remains Low But Big Companies Are Leading the Way.” *Center for Data Innovation* (blog), March 17, 2022. https://datainnovation.org/2022/03/nsf-data-shows-ai-adoption-in-the-united-states-remains-low-but-big-companies-are-leading-the-way/.

OpenAI. “DALL·E 2,” 2022. https://openai.com/dall-e-2/.

OpenAI. “OpenAI API.” Accessed November 21, 2022. https://beta.openai.com.

Ozkaya, Ipek. “The Behavioral Science of Software Engineering and Human–Machine Teaming.” *IEEE Software* 37, no. 6 (November 2020): 3–6. https://doi.org/10.1109/MS.2020.3019190.

Pak, Richard, Nicole Fink, Margaux Price, Brock Bass, and Lindsay Sturre. “Decision Support Aids with Anthropomorphic Characteristics Influence Trust and Performance in Younger and Older Adults.” *Ergonomics* 55, no. 9 (September 1, 2012): 1059–72. https://doi.org/10.1080/00140139.2012.691554.

Paterson, Ryan. “DARPA Funds Development of Pulse Platform — IST Research.” IST Research, September 13, 2013. https://web.archive.org/web/20160125090234/http://istresearch.com/news/2013/9/6/darpa-funds-development-of-pulse-platform.

Patil, DJ, Hilary Mason, and Mike Loukides. “Of Oaths and Checklists.” O’Reilly Media, July 17, 2018. https://www.oreilly.com/radar/of-oaths-and-checklists/.

Patton, Michael Quinn. *Essentials of Utilization-Focused Evaluation*. Los Angeles, CA: SAGE, 2012.

Paul, Christopher, Colin Clarke, Bonnie Triezenberg, David Manheim, and Bradley Wilson. *Improving C2 and Situational Awareness for Operations in and Through the Information Environment*. Santa Monica, CA: RAND Corporation, 2018. https://doi.org/10.7249/RR2489.

Paul, Christopher, Jessica Yeats, Colin P. Clarke, Miriam Matthews, and Lauren Skrabala. *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Handbook for Practitioners*. Santa Monica, CA: RAND, 2015.

Pillai, Rajasshrie, and Brijesh Sivathanu. "Adoption of AI-Based Chatbots for Hospitality and Tourism." *International Journal of Contemporary Hospitality Management* 32, no. 10 (2020): 3199–3226. https://doi.org/10.1108/IJCHM-04-2020-0259.

Piorkowski, David, John Richards, and Michael Hind. "Evaluating a Methodology for Increasing AI Transparency: A Case Study." *ArXiv*, January 24, 2022. http://arxiv.org/abs/2201.13224.

Pomerleau, Mark. "Pentagon's AI Center to Field New Psychological Operations Tool." *C4ISRNet*, September 11, 2020, sec. Artificial Intelligence. https://www.c4isrnet.com/artificial-intelligence/2020/09/11/pentagons-ai-center-to-field-new-psychological-operations-tool/.

Potts, Christopher. "Model Card for DynaSent Model 0 and Model 1." GitHub, January 4, 2021. https://github.com/cgpotts/dynasent/blob/a771f0d344ffa6d21716644ef8766c0ee743fc1b/dynasent_modelcard.md.

Procope, Chavez, Adeel Cheema, David Adkins, Bilal Alsallakh, Nekesha Green, Emily Mcreynolds, Grace Pehl, Erin Wang, and Polina Zvyagina. *System-Level Transparency of Machine Learning*. Menlo Park, CA: Meta AI, 2022. https://ai.facebook.com/research/publications/system-level-transparency-of-machine-learning.

Pumplun, Luisa, Christoph Tauchert, and Margareta Heidt. "A New Organizational Chassis for Artificial Intelligence-Exploring Organizational Readiness Factors." In *27th European Conference on Information Systems (ECIS)*, 16. Stockholm, Sweden: ResearchGate, 2019. https://aisel.aisnet.org/ecis2019_rp/106.

Punn, Narinder, and Sonali Agarwal. "Automated Diagnosis of COVID-19 with Limited Posteroanterior Chest X-Ray Images Using Fine-Tuned Deep Neural Networks." *Applied Intelligence* 51 (May 1, 2021): 1–14. https://doi.org/10.1007/s10489-020-01900-3.

Qian, Ming, and Davis Qian. "Defining a Human-Machine Teaming Model for AI-Powered Human-Centered Machine Translation Agent by Learning from Human-Human Group Discussion: Dialog Categories and Dialog Moves." In *Artificial Intelligence in HCI*, edited by Helmut Degen and Lauren Reinerman-Jones, 70–81. Lecture Notes in Computer Science. Cham, Denmark: Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-50334-5_5.

Qin, Li, Yongbeom Kim, Jeffrey Hsu, and Xin Tan. "The Effects of Social Influence on User Acceptance of Online Social Networks." *International Journal of Human-Computer Interaction* 27, no. 9 (September 2011): 885–99. https://doi.org/10.1080/10447318.2011.555311.

Qiu, Lingyun, and Izak Benbasat. "Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems." *Journal of Management Information Systems* 25, no. 4 (2009): 145–81. https://www.jstor.org/stable/40398956.

Rae, Jack W., Sebastian Borgeaud, Trevor Cai, Katie Millican, Jordan Hoffmann, Francis Song, John Aslanides et al. "Scaling Language Models: Methods, Analysis & Insights from Training Gopher." arXiv, January 21, 2022. https://doi.org/10.48550/arXiv.2112.11446.

Rafail, Patrick, and Isaac Freitas. "Natural Language Processing." In *SAGE Research Methods Foundations*, edited by Paul Atkinson, Sara Delamont, Alexandru Cernat, Joseph W Sakshaug, and Richard A Williams. London, UK: SAGE Publications Ltd, 2020. https://dx.doi.org/9781529749120.

Ram, Sundaresan, and Jagdish N. Sheth. "Consumer Resistance To Innovations: The Marketing Problem A." *The Journal of Consumer Marketing* 6, no. 2 (Spring 1989). https://www.proquest.com/docview/220132547/abstract/725F9F68A45C4184PQ/1.

Ramírez, Ricardo, and Dal Brodhead. *Utilization Focused Evaluation: A Primer for Evaluators.* Penang, Malaysia: Southbound, 2013.

Reddy, Y, Viswanath Pulabaigari, and Eswara B. "Semi-Supervised Learning: A Brief Review." *International Journal of Engineering & Technology* 7 (February 9, 2018): 81. https://doi.org/10.14419/ijet.v7i1.8.9977.

Regona, Massimo, Tan Yigitcanlar, Bo Xia, and Rita Yi Man Li. "Opportunities and Adoption Challenges of AI in the Construction Industry: A PRISMA Review." *Journal of Open Innovation: Technology, Market, and Complexity* 8, no. 1 (March 2022). https://doi.org/10.3390/joitmc8010045.

Ribeiro, Filipe N., Koustuv Saha, Mahmoudreza Babaei, Lucas Henrique, Johnnatan Messias, Fabricio Benevenuto, Oana Goga, Krishna P. Gummadi, and Elissa M. Redmiles. "On Microtargeting Socially Divisive Ads: A Case Study of Russia-Linked Ad Campaigns on Facebook." In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 140–49. Atlanta, GA: ACM, 2019. https://doi.org/10.1145/3287560.3287580.

Richards, John, David Piorkowski, Michael Hind, Stephanie Houde, Aleksandra Mojsilovic, and Kush R Varshney. “A Human-Centered Methodology for Creating AI FactSheets.” *Bulletin of the Technical Committee on Data Engineering*, December 2021.

Ridout, Travis N., Erika Franklin Fowler, and Michael M. Franz. “Spending Fast and Furious: Political Advertising in 2020.” *The Forum* 18, no. 4 (March 1, 2021): 465–92. https://doi.org/10.1515/for-2020-2109.

Rogers, Everett M. *Diffusion of Innovations*. 4th ed. New York, NY: The Free Press, 1995.

Roh, Yuji, Geon Heo, and Steven Euijong Whang. “A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective.” *IEEE Transactions on Knowledge and Data Engineering* 33, no. 4 (April 2021): 1328–47. https://doi.org/10.1109/TKDE.2019.2946162.

Rossi, Francesca. “Building Trust in Artificial Intelligence.” *Journal of International Affairs* 72, no. 1 (2018): 127–34. http://www.jstor.org/stable/26588348.

Rousseau, Denise, Sim Sitkin, Ronald Burt, and Colin Camerer. “Not So Different After All: A Cross-Discipline View of Trust.” *Academy of Management Review* 23 (July 1, 1998). https://doi.org/10.5465/AMR.1998.926617.

Rowe, Neil C. “Algorithms for Artificial Intelligence.” *Computer* 55, no. 7 (July 2022): 97–102. https://doi.org/10.1109/MC.2022.3169360.

Saenz, Maria Jesus, Elena Revilla, and Cristina Simón. “Designing AI Systems With Human-Machine Teams.” *MIT Sloan Management Review* 61, no. 3 (Spring 2020): 1–5. https://www.proquest.com/docview/2392464050/abstract/2BE117B7C37847A0PQ/1.

Salesforce AI Research. “Einstein OCR Model Card.” Einstein Vision and Language, February 2022. https://developer.salesforce.com/docs/analytics/einstein-vision-language/guide/einstein-ocr-model-card.html.

Salesforce Marketing Cloud Einstein. “Einstein Engagement Frequency Model Details.” Salesforce Help, May 2020. https://help.salesforce.com/s/articleView?id=sf.mc_anb_einstein_engagement_frequency_model_details.htm&type=5.

Salesforce Marketing Cloud Einstein. “Einstein Engagement Scoring for Mobile Model Card.” Salesforce Help, August 2020. https://help.salesforce.com/s/articleView?id=sf.mc_anb_einstein_engagement_scoring_for_mobile_model_card.htm&type=5.

Salesforce Marketing Cloud Einstein. “Einstein Messaging Insights Model Card.” Salesforce Help, May 2020. https://help.salesforce.com/s/articleView?id=sf.mc_anb_einstein_messaging_insights_model_card.htm&type=5.

Salman, Fatima M., and Samy S. Abu-Naser. “Expert System for COVID-19 Diagnosis,” 2020. http://dspace.alazhar.edu.ps/xmlui/handle/123456789/588.

Schaefer, Cindy, Kristina Lemmer, Kret Samy Kret, Maija Ylinen, Patrick Mikalef, and Bjoern Niehaves. “Truth or Dare? – How Can We Influence the Adoption of Artificial Intelligence in Municipalities?” In *Hawaii International Conference on System Sciences*. Honolulu, HI: University of Hawaiʻi at Mānoa, 2021. https://doi.org/10.24251/HICSS.2021.286.

Schäwel, Johanna, Regine Frener, and Sabine Trepte. “Political Microtargeting and Online Privacy: A Theoretical Approach to Understanding Users’ Privacy Behaviors.” *Media and Communication* 9, no. 4 (2021): 158–69. http://dx.doi.org.libproxy.nps.edu/10.17645/mac.v9i4.4085.

Schirmer, Peter, Amber Jaycocks, Sean Mann, William Marcellino, Luke J. Matthews, John David Parsons, and David Schulker. *Natural Language Processing: Security- and Defense-Related Lessons Learned*. Santa Monica, CA: RAND Corporation, 2021. https://www.rand.org/pubs/perspectives/PEA926-1.html.

Schmidhuber, Juergen. “Deep Learning in Neural Networks: An Overview.” *Neural Networks* 61 (January 2015): 85–117. https://doi.org/10.1016/j.neunet.2014.09.003.

Schramm, Wilbur. “How Communication Works.” In *Readings in Psychological Operations*, 1–24. ST 33–151. Fort Bragg, NC: U.S. Army Special Warfare School, 1963.

Schroden, Jonathan. “WHY OPERATIONS ASSESSMENTS FAIL: It’s Not Just the Metrics.” *Naval War College Review* 64, no. 4 (2011): 89–102. https://www.jstor.org/stable/26397245.

Schwille, Michael, Anthony Atler, Jonathan Welch, Christopher Paul, and Richard C. Baffa. *Intelligence Support for Operations in the Information Environment: Dividing Roles and Responsibilities Between Intelligence and Information Professionals*. Santa Monica, CA: RAND Corporation, 2020. https://doi.org/10.7249/RR3161.

Sethumadhavan, Arathi. “Trust in Artificial Intelligence.” *Ergonomics in Design* 27, no. 2 (April 1, 2019): 34–34. https://doi.org/10.1177/1064804618818592.

Sharma, Shavneet, Gurmeet Singh, Nazrul Islam, and Amandeep Dhir. "Why Do SMEs Adopt Artificial Intelligence-Based Chatbots?" *IEEE Transactions on Engineering Management*, 2022, 1–14. https://doi.org/10.1109/TEM.2022.3203469.

Shen, Hong, Wesley H. Deng, Aditi Chattopadhyay, Zhiwei Steven Wu, Xu Wang, and Haiyi Zhu. "Value Cards: An Educational Toolkit for Teaching Social Impacts of Machine Learning through Deliberation." In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 850–61. Virtual Event Canada: ACM, 2021. https://doi.org/10.1145/3442188.3445971.

Sheppard, Lindsey. *Accelerating the Defense Department's AI Adoption*. New York, NY: Council on Foreign Relations, 2020. http://www.jstor.org/stable/resrep29909.

Sheshadri, Chatterjee, and Kalyan Kumar Bhattacharjee. "Adoption of Artificial Intelligence in Higher Education: A Quantitative Analysis Using Structural Equation Modelling." *Education and Information Technologies* 25, no. 5 (September 2020): 3443–63. https://doi.org/10.1007/s10639-020-10159-7.

Shi, Si, Yuhuang Gong, and Dogan Gursoy. "Antecedents of Trust and Adoption Intention toward Artificially Intelligent Recommendation Systems in Travel Planning: A Heuristic–Systematic Model." *Journal of Travel Research* 60, no. 8 (November 1, 2021): 1714–34. https://doi.org/10.1177/0047287520966395.

Shi, Yong, Yuanchun Zheng, Kun Guo, Luyao Zhu, and Yi Qu. "Intrinsic or Extrinsic Evaluation: An Overview of Word Embedding Evaluation." In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, 1255–62, 2018. https://doi.org/10.1109/ICDMW.2018.00179.

SIBR-STTR. "IST Research Corp." SIBR-STTR: America's Seed Fund. Accessed November 6, 2022. https://www.sbir.gov/sbc/ist-research-llc.

Simon, Herbert A. "Bounded Rationality." In *Utility and Probability*, edited by John Eatwell, Murray Milgate, and Peter Newman, 15–18. London, UK: Palgrave Macmillan, 1990. https://doi.org/10.1007/978-1-349-20568-4_5.

Skinner, Burrhus Frederic. *Contingencies of Reinforcement: A Theoretical Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1969.

Smith, Carol J. "Designing Trustworthy AI: A Human-Machine Teaming Framework to Guide Development." In *AAAI FSS-19: Artificial Intelligence in Government and Public Sector*. Arlington, Virginia: arXiv, 2019. https://doi.org/10.48550/arXiv.1910.03515.

Smith, Carol. "Designing Trustworthy AI: A User Experience (UX) Framework." In *RSA Conference 2020*. Pittsburgh, PA: Carnegie Mellon University, 2020. https://doi.org/10.1184/R1/12198321.v1.

Sofaer, Helen R., Jennifer A. Hoeting, and Catherine S. Jarnevich. "The Area under the Precision-Recall Curve as a Performance Metric for Rare Binary Events." *Methods in Ecology and Evolution* 10, no. 4 (2019): 565–77. https://doi.org/10.1111/2041-210X.13140.

spaCy. "English · SpaCy Models Documentation," October 19, 2022. https://spacy.io/models/en#en_core_web_trf.

Steinbeck, John. *The Moon Is Down*. New York, NY: Penguin Books, 1995.

Stirbu, Vlad, Tuomas Granlund, and Tommi Mikkonen. "Continuous Design Control for Machine Learning in Certified Medical Systems." *Software Qual J*, September 13, 2022. https://doi.org/10.1007/s11219-022-09601-5.

Stone, Merlin, Eleni Aravopoulou, Yuksel Ekinci, Geraint Evans, Matt Hobbs, Ashraf Labib, Paul Laughlin, Jon Machtynger, and Liz Machtynger. "Artificial Intelligence (AI) in Strategic Marketing Decision-Making: A Research Agenda." *The Bottom Line* 33, no. 2 (2020): 183–200. https://doi.org/10.1108/BL-03-2020-0022.

Strauss, Christine, Gabriele Kotsis, A Min Tjoa, and Ismail Khalil, eds. *Database and Expert Systems Applications: 32nd International Conference, DEXA 2021, Virtual Event, September 27–30, 2021, Proceedings, Part I*. Vol. 12923. Lecture Notes in Computer Science. Cham, Denmark: Springer International Publishing, 2021. https://doi.org/10.1007/978-3-030-86472-9.

Suominen, Hanna. "Performance Evaluation Measures for Text Mining." In *Handbook of Research on Text and Web Mining Technologies*, edited by Min Song and Yi-Fang Brook Wu, 724–47. Hershey, PA: IGI Publishing, 2009. 10.4018/978-1-59904-990-8.ch041.

Sutton, Richard S., and Andrew G. Barto. *Reinforcement Learning: An Introduction*. Second edition. Cambridge, Massachusetts: The MIT Press, 2018.

Tagliabue, Jacopo, Ville Tuulos, Ciro Greco, and Valay Dave. "DAG Card Is the New Model Card." In *Thirty-Fifth Conference on Neural Information Processing Systems*. arXiv, 2021. https://doi.org/10.48550/arXiv.2110.13601.

Teasdale, Rebecca M. "Evaluative Criteria: An Integrated Model of Domains and Sources." *American Journal of Evaluation* 42, no. 3 (September 1, 2021): 354–76. https://doi.org/10.1177/1098214020955226.

Telley, Christopher. "The Influence Machine: Automated Information Operations as a Strategic Defeat Mechanism." *The Institute of Land Warfare*, no. 121 (October 2018).

TensorFlow. "Model Card Toolkit | Responsible AI Toolkit," February 9, 2021. https://www.tensorflow.org/responsible_ai/model_card_toolkit/guide.

Thiebes, Scott, Sebastian Lins, and Ali Sunyaev. "Trustworthy Artificial Intelligence." *Electronic Markets* 31, no. 2 (June 1, 2021): 447–64. https://doi.org/10.1007/s12525-020-00441-4.

Thompson, Joel E. "Influencing Trust in Human and Artificial Intelligence Teaming through Heuristics." Master's thesis, Naval Postgraduate School, 2021. https://calhoun.nps.edu/handle/10945/67823.

Thottan, M., and Chuanyi Ji. "Anomaly Detection in IP Networks." *IEEE Transactions on Signal Processing* 51, no. 8 (August 2003): 2191–2204. https://doi.org/10.1109/TSP.2003.814797.

Tidd, Joe. *Gaining Momentum: Managing The Diffusion Of Innovations*. Singapore: World Scientific Publishing Company, 2010. http://ebookcentral.proquest.com/lib/ebook-nps/detail.action?docID=731340.

Tkachenka, Andrei, Gregory Karpiak, Yury Kartynnik, Artsiom Ablavatski, Valentin Bazarevsky, and Siargey Pisarchyk. "Model Card Content: Hair Segmentation." Google MediaPipe, 2019. https://drive.google.com/file/d/1lPwJ8BD_-3UUor4LayQ0xpa_RIC_hoRh/preview?usp=embed_facebook.

Tornatzky, Louis G., and Mitchell Fleischer. *The Processes of Technological Innovation*. Lexington, MA: Lexington Books, 1990.

Trunzer, Emanuel, Iris Kirchen, Jens Folmer, Gennadiy Koltun, and Birgit Vogel-Heuser. "A Flexible Architecture for Data Mining from Heterogeneous Data Sources in Automated Production Systems." In *2017 IEEE International Conference on Industrial Technology (ICIT)*, 1106–11. Toronto, Canada: IEEE, 2017. https://doi.org/10.1109/ICIT.2017.7915517.

Tucker, Patrick. "How AI Will Soon Change Special Operations." *Defense One*, May 18, 2020. https://www.defenseone.com/technology/2020/05/how-ai-will-soon-change-special-operations/165487/.

Two Six Technologies. "IST Research and Two Six Labs Combine to Form High-Growth Government Technology Platform." GlobeNewswire News Room, February 1, 2021. https://www.globenewswire.com/news-release/2021/02/01/2167580/0/en/IST-Research-and-Two-Six-Labs-Combine-to-Form-High-Growth-Government-Technology-Platform.html.

Two Six Technologies. "Pulse: Enabling Data Collection and Two-Way Engagement." Two Six Technologies | Advanced Technology Solutions for Critical Missions, 2022. https://twosixtech.com/products/pulse/.

Two Six Technologies. “Two Six Technologies Announces New Office in San Antonio, TX.” Two Six Technologies | Advanced Technology Solutions for Critical Missions, August 18, 2022. https://twosixtech.com/news/two-six-technologies-announces-new-office-in-san-antonio-tx/.

Tyagi, Amit Kumar, and G. Rekha. “Challenges of Applying Deep Learning in Real-World Applications.” In *Challenges and Applications for Implementing Machine Learning in Computer Vision*, edited by Ramgopal Kashyap and A.V. Senthil Kumar, 92–118. Hershey, PA: IGI Global, 2020. http://www.igi.global.com/chapter/challenges-of-applying-deep-learning-in-real-world-applications/242103.

U.S. Government Accountability Office. *Artificial Intelligence: Status of Developing and Acquiring Capabilities for Weapon Systems*. GAO-22-104765. Washington, DC: U.S. Government Accountability Office, 2022. https://www.gao.gov/assets/gao-22-104765.pdf.

U.S. Marine Corps. *MCDP 8 Information*. PCN 142 000018 00. Washington, DC: Department of the Navy, Headquarters United States Marine Corps, 2022. https://www.marines.mil/Portals/1/Publications/MCDP%208.pdf?ver=6gIvEcD0CUuPAgTSmyDNag%3d%3d.

Uğur, Serap, and Gulsun Kurubacak. “Artificial Intelligence to Super Artificial Intelligence, Cyber Culture to Transhumanist Culture: Change of the Age and Human.” In *Handbook of Research on Learning in the Age of Transhumanism*, edited by Serap Sisman-Ugur and Gulsun Kurubacak, 1–16. Hershey, PA: IGI Global, 2019. https://doi.org/10.4018/978-1-5225-8431-5.

Unhelkar, Bhuvan, and Tad Gonsalves. *Artificial Intelligence for Business Optimization: Research and Applications*. Boca Raton, FL: CRC Press, 2021. https://doi.org/10.1201/9781003120926.

Usama, Muhammad, Junaid Qadir, Aunn Raza, Hunain Arif, Kok-lim Alvin Yau, Yehia Elkhatib, Amir Hussain, and Ala Al-Fuqaha. “Unsupervised Machine Learning for Networking: Techniques, Applications and Research Challenges.” *IEEE Access* 7 (2019): 65579–615. https://doi.org/10.1109/ACCESS.2019.2916648.

Vaccari, Cristian, and Andrew Chadwick. “Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News.” *Social Media + Society* 6, no. 1 (January 1, 2020). https://doi.org/10.1177/2056305120903408.

Vakunov, Andrey, Ivan Grishchenko, Artsiom Ablavatski, and Dmitry Lagun. “Model Card MediaPipe Iris.” Google MediaPipe, May 13, 2020. https://drive.google.com/file/d/1bsWbokp9AklH2ANjCfmjqEzzxO1CNbMu/preview?usp=embed_facebook.

Vallor, Shannon. *Ethics in Tech Practices: A Toolkit*. Santa Clara, CA: Markkula Center for Applied Ethics at Santa Clara University, 2018. https://www.scu.edu/ethics-in-technology-practice/ethical-toolkit/.

Vasiljeva, Tatjana, Ilmars Kreituss, and Ilze Lulle. "Artificial Intelligence: The Attitude of the Public and Representatives of Various Industries." *Journal of Risk and Financial Management* 14, no. 8 (July 21, 2021). https://doi.org/10.3390/jrfm14080339.

Venkatesh, Viswanath, and Fred D. Davis. "A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies." *Management Science* 46, no. 2 (2000): 186–204. http://www.jstor.org/stable/2634758.

Venkatesh, Viswanath, James Y. L. Thong, and Xin Xu. "Unified Theory of Acceptance and Use of Technology: A Synthesis and the Road Ahead." *Journal of the Association for Information Systems* 17, no. 5 (May 2016): 328–76. https://www.proquest.com/docview/1794948207/abstract/112891861CE74109PQ/1.

Venkatesh, Viswanath, Michael G. Morris, Gordon B. Davis, and Fred D. Davis. "User Acceptance of Information Technology: Toward a Unified View." *MIS Quarterly* 27, no. 3 (2003): 425–78. https://doi.org/10.2307/30036540.

Venkatesh, Viswanath. "Adoption and Use of AI Tools: A Research Agenda Grounded in UTAUT." *Annals of Operations Research* 308, no. 1/2 (January 2022): 641–52. https://doi.org/10.1007/s10479-020-03918-9.

Vergun, David. "Special Operations Strives to Use the Power of Artificial Intelligence." *U.S. Department of Defense News*, December 7, 2020. https://www.defense.gov/News/News-Stories/Article/Article/2438076/special-operations-strives-to-use-the-power-of-artificial-intelligence/.

Verma, Dinesh C, Archit Verma, and Utpal Mangla. "Addressing the Limitations of AI/ML in Creating Cognitive Solutions." In *2021 IEEE Third International Conference on Cognitive Machine Intelligence (CogMI)*, 189–96, 2021. https://doi.org/10.1109/CogMI52975.2021.00033.

Wang, Pei, and Ben Goertzel. "Introduction: What Is the Matter Here?" In *Theoretical Foundations of Artificial General Intelligence*, edited by Pei Wang and Ben Goertzel, 4:1–9. Atlantis Thinking Machines. Paris: Atlantis Press, 2012. https://doi.org/10.2991/978-94-91216-62-6_1.

Wang, Yabing, Guimin Huang, Jun Li, Hui Li, Ya Zhou, and Hua Jiang. "Refined Global Word Embeddings Based on Sentiment Concept for Sentiment Analysis." *IEEE Access* 9 (2021): 37075–85. https://doi.org/10.1109/ACCESS.2021.3062654.

Wang, Zhicheng. “Model Card Content: MediaPipe KNIFT,” February 5, 2020. https://drive.google.com/file/d/1RCdA83a8JDV3ZPS-mtpsV8mTj3r0F9s-/view?usp=sharing&usp=embed_facebook.

Werder, Karl, Balasubramaniam Ramesh, and Rongen (Sophia) Zhang. “Establishing Data Provenance for Responsible Artificial Intelligence Systems.” *ACM Transactions on Management Information Systems* 13, no. 2 (March 10, 2022): 22:1-22:23. https://doi.org/10.1145/3503488.

Westenberger, Jens, Kajetan Schuler, and Dennis Schlegel. “Failure of AI Projects: Understanding the Critical Factors.” *Procedia Computer Science* 196 (January 1, 2022): 69–76. https://doi.org/10.1016/j.procs.2021.11.074.

White, Andrew, and Mehrad Ansari. “Hemolytic Prediction.” peptide.bio v0.19.0, 2022. https://peptide.bio/.

Wholey, Joseph S., Harry P. Hatry, and Kathryn E. Newcomer, eds. *Handbook of Practical Program Evaluation*. Hoboken, NJ: Wiley, 2004.

Wikimedia. “Machine Learning Models/Proposed/Language Agnostic Link-Based Article Topic Model Card,” September 29, 2022. https://meta.wikimedia.org/wiki/Machine_learning_models/Proposed/Language_agnostic_link-based_article_topic_model_card.

Wilner, Alex, and Casey Babb. “New Technologies and Deterrence: Artificial Intelligence and Adversarial Behaviour.” In *NL ARMS Netherlands Annual Review of Military Studies 2020*, edited by F. Osinga and T. Sweijs, 17. The Hague: T.M.C. Asser Press, 2020. https://doi.org/10.1007/978-94-6265-419-8_21.

Witten, Ian H., Eibe Frank, Mark A. Hall, and Christopher J. Pal, eds. *Data Mining: Practical Machine Learning Tools and Techniques*. Fourth Edition. Amsterdam, Netherlands: Elsevier, 2017.

Yang, Xiangli, Zixing Song, Irwin King, and Zenglin Xu. “A Survey on Deep Semi-Supervised Learning.” arXiv, August 22, 2021. http://arxiv.org/abs/2103.00550.

Zerilli, John, Umang Bhatt, and Adrian Weller. “How Transparency Modulates Trust in Artificial Intelligence.” *Patterns* 3, no. 4 (April 8, 2022). https://doi.org/10.1016/j.patter.2022.100455.

Zhang, Liangkai, Adel Ahmadyan, Artsiom Ablavatski, Jianing Wei, and Matthias Grundmann. “Model Card MediaPipe Objectron.” Google MediaPipe, October 27, 2020. https://drive.google.com/file/d/1CMhN7Npdq0Dt2j0_z69mai2-m7oUTRKF/preview?usp=embed_facebook.

Zhang, Susan, Stephen Roller, Naman Goyal, Mikel Artetxe, Moya Chen, Shuohui Chen, Christopher Dewan et al. "OPT: Open Pre-Trained Transformer Language Models." arXiv, June 21, 2022. https://doi.org/10.48550/arXiv.2205.01068.

Zhang, Yumeng. "An Overview of the Theory and Application of Reinforcement Learning." In *ICMLCA 2021; 2nd International Conference on Machine Learning and Computer Application*, 1–4, 2021.

Zissis, Georges. "The R3 Concept: Reliability, Robustness, and Resilience [President's Message]." *IEEE Industry Applications Magazine* 25, no. 4 (July 2019): 5–6. https://doi.org/10.1109/MIAS.2019.2909374.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California