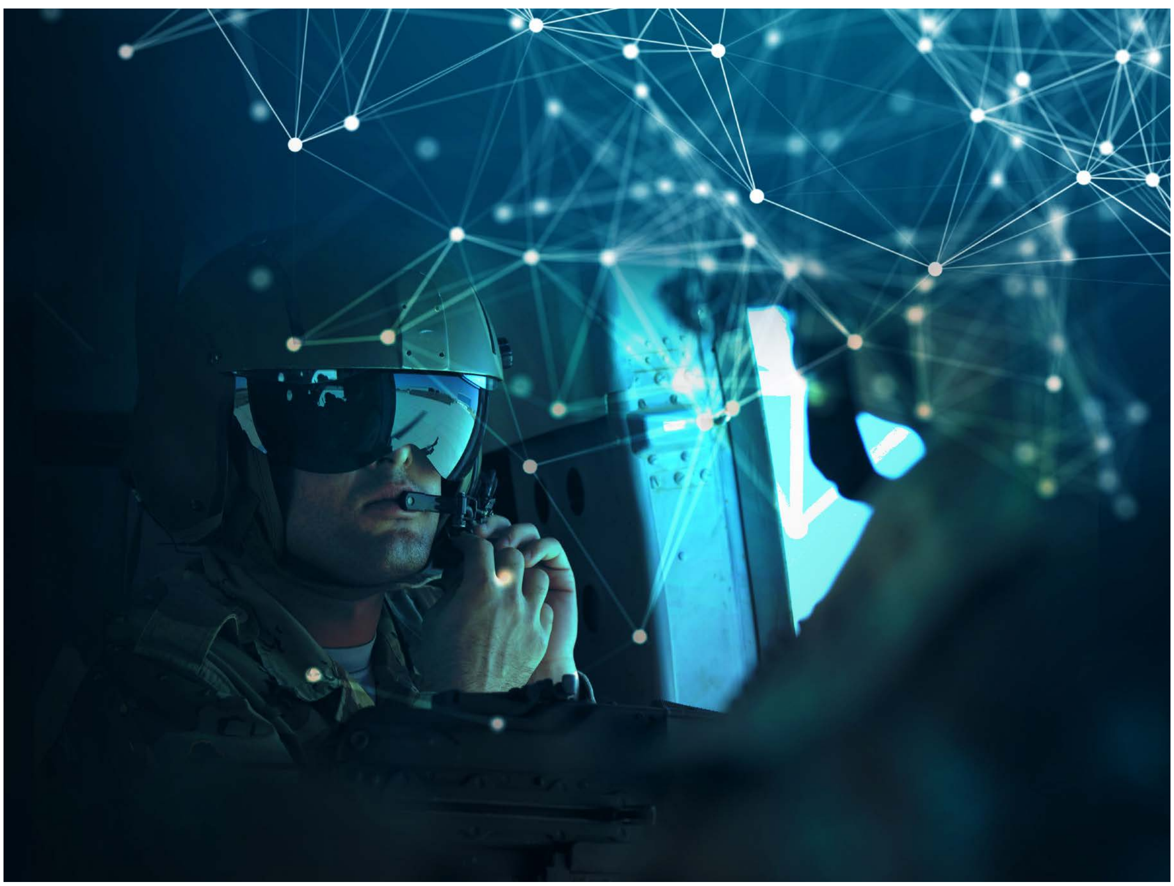
Research Report

KRISTIN WARREN, LANCE MENTHE, BRIDGET R. KANE, SYDNEY KESSLER, MARY LEE, DOMENIQUE LUMPKIN, DON SNYDER, LI ANG ZHANG, ALEKSANDR ESPARZA HARTUNIAN, SHAWN KONETZKI

Improving Sense-Making with Artificial Intelligence



For more information on this publication, visit **www.rand.org/t/RRA3152-1**.

Published by the RAND Corporation, Santa Monica, Calif.

Library of Congress Cataloging-in-Publication Data is available for this publication.

ISBN: 978-1-9774-1499-1

Cover: Composite design by Carol Ponce adapted from images by Your123/Adobe Stock and Julianne Showalter/332d Air Expeditionary Wing.

About This Report

This report addresses how key Department of the Air Force (DAF) sense-making units—including the Air Force Distributed Common Ground System (AF DCGS); the Air Operations Center Intelligence, Surveillance, and Reconnaissance (ISR) Division; and U.S. Space Force (USSF) intelligence elements—can leverage artificial intelligence (AI) capabilities to address pressing sense-making problems. The report is intended specifically for these organizations but may also be of interest to parallel organizations in the other services and in the wider intelligence community.

The research reported here was commissioned by the Air Force Chief Data and AI Office (SAF/CND) and conducted within the Force Modernization and Employment Program of RAND Project AIR FORCE as part of a fiscal year 2024 project, "Enabling Effective Modernization of Sense-Making for Key Operation Missions."

RAND Project AIR FORCE

RAND Project AIR FORCE (PAF), a division of RAND, is the Department of the Air Force's (DAF's) federally funded research and development center for studies and analyses, supporting both the United States Air Force and the United States Space Force. PAF provides the DAF with independent analyses of policy alternatives affecting the development, employment, combat readiness, and support of current and future air, space, and cyber forces. Research is conducted in four programs: Strategy and Doctrine; Force Modernization and Employment; Resource Management; and Workforce, Development, and Health. The research reported here was prepared under contract FA7014-22-D-0001.

Additional information about PAF is available on our website:
www.rand.org/paf/

This report documents work originally shared with the DAF on September 10, 2024. The draft report, dated December 2024, was reviewed by formal peer reviewers and DAF subject-matter experts.

Acknowledgments

We are grateful to our sponsor, Chandra L. Donelson, Space Chief Data and Artificial Intelligence Officer and Acting Air Force Chief Data and AI Officer, and her predecessor in her Air Force role, Eileen M. Vidrine, for their assistance in supporting and guiding this research. We also thank Lynne Graves and Lt Col Joseph O. Chapa, from the same office, for all their help, without which this would not have been possible.

We also owe a debt of gratitude to the many experts who made time to speak with us to discuss DAF sense-making processes and challenges. At the USSF, we thank Lt Col Daniel P. Beall, AF/AF5/7DI, and Lt Col Laura A. Carter, HAF/A2/6, for discussing ISR capabilities. At the USSF, we thank Maj Brandon Poje, USSF SpOC Space Delta 7/DO; Maj Arthur J. Thibert III, SpOC HQ/S2Z; Maj Cameron L. Ross, USSF SpOC HQ/S2X; Maj Christopher A. Sargent, SpOC HQ/S2T; Kenneth D. "Bama" Burnett, USSF SpOC Space Delta 7/TD; and many others for sharing their understanding of USSF issues. At USAF Air Combat Command (ACC), we thank Shane P. Hamilton, ACC/A2/A-A2; Kristofer W. Gifford, ACC A2/A23; James M. Giesken, ACC A2/A22; Michael W. Kennedy, ACC/A2/A28; Cristina M. Stone, ACC/A2/A29; and many others for sharing their knowledge of future plans and systems. At the AF DCGS in Virginia, we thank Capt Tyler A. Troesch; Lt Col Christopher W. Potts, 480 ISRW/A5/8; TSgt David E. Schiller; and many others for all their time and support. At USAF in Hawaii, we thank Lt Col Billy R. Huffman, PACAF/A51X; Col Sean M. Callahan, PACAF 613 AOC/A32; and Col Dennis M. Borrman, 692 ISRG/CC, for their generous support throughout.

We also thank many RAND colleagues for sharing their expertise with us in this work: Jin Kim, Ashwin Acharya, Aisha Nájera, Maj Elyse Crimm, Brandon De Bruhl, Henry Hargrove, Ajay Kochhar, Minami Makino, Neeti Pokhriyal, Morgan Sandler, Éder Sousa, David Stebbins, and Daniel Tapia. Many thanks to Dulani Woods for his thorough remarks on the draft and Chad Heitzenrater for his thoughtful feedback on the draft, as well as his sage advice throughout the research process. We are thankful to these and many others who made this work possible. Any errors that remain are ours alone.

Summary

Issue

As the Department of the Air Force (DAF) pivots toward addressing peer threats, DAF sense-making processes must be able to handle the rapidly increasing proliferation of sensors and targets. We identified 20 challenges associated with scaling DAF sense-making processes in five key areas—collection orchestration, data access and sharing, data fusion and analysis, model management, and skills and training—and indicated where and how five major categories of current artificial intelligence (AI) capabilities could be used to address them.

Approach

To identify sense-making challenges, we conducted site visits with DAF sense-making organizations, including Air Operations Centers (AOCs); the DAF intelligence, surveillance, and reconnaissance (ISR) wings and groups that operate the DAF Distributed Common Ground System; and their U.S. Space Force (USSF) counterparts. To understand current AI capabilities, we interviewed Air Force Research Laboratory researchers and AI experts. To identify how these capabilities could be applied to these challenges, we conducted a Delphi elicitation exercise.

Key Findings

Table S.1 shows how AI capabilities can be combined to meet DAF sense-making challenges in the five key areas described earlier. These approaches are independent of mission threads and primarily deal with common processes of the intelligence cycle; we believe the DAF should take a mission-independent approach as much as possible. In addition, the following findings regarding the use of data and algorithms cut across all AI applications described in Table S.1.

- **Datasets and knowledge representations need to be carefully curated.** High-quality datasets must be built with care and must also be associated with the right metadata to support object-based production (OBP) and subsequent algorithm development.
- **Analysts can and should anticipate AI failure modes.** Understanding the limits of an AI system's training data or knowledge representation will help analysts anticipate errors and use the AI systems more effectively.
- **Expert systems (ESs) can play an important role.** The older forms of AI remain relevant.
- **The DAF should pave the way for disruptive adoption.** Disruptive AI will be needed later, but early adoption of nondisruptive AI can help prepare the DAF for greater change.

Table S.1. Summary of Major Findings by Sense-Making Challenge Type

Challenge Area	Major Findings	CV	NLP	Plan	P/C	ES
Collection orchestration	NLP combined with ESs to elicit requirements and rephrase them into standard formats		●			●
	Planning systems to improve both deliberate and dynamic collections across multiple domains	○		●		
	CV to screen collections incapable of providing the required EEI	●				
Data access and sharing	Text classification combined with ESs to propose or confirm classification markings to assist in data transfer		●			●
	Multimodal system to assist in OBP/ontology development	○	●		○	○
Data fusion and analysis	NLP in an ESs framework to clean and condition processed data		●			●
	CV and NLP to assist tracking DOF across multiple sensor modalities and through chat/radio reports	●	○			
	P/C and Planning, with CV assistance, to anticipate future adversary movement	○		●	●	
Model management	NLP with ESs to parse code and manage adherence with cybersecurity regulations		●			●
Skills and training	ES to support customized training programs		○			●
	NLP and ESs to support knowledge management and assist knowledge transfer between units, shifts, and personnel		●			●

NOTE: CV = computer vision; DOF = disposition of forces; EEI = essential elements of information; NLP = natural language processing; P/C = prediction/classification; Plan = planning; ● = major/likely applicability; ○ = minor/possible applicability.

Recommendations

- **Follow a shared road map for developing sense-making capabilities.** DAF ISR wings and their USSF counterparts should work with DAF Chief Data and AI Officer (CDAO) to develop a set of shared priorities for AI integration into sense-making based on Table S.1.
- **Anticipate risks early.** All DAF sense-making organizations should conduct risk assessments, such as social, technological, operational, political, economic, and sustainability analysis, for AI tools they propose. The DAF Chief Information Officer should ensure responsible AI-related tasks are executed for the sense-making domain.
- **Respect tool fatigue sentiments.** DAF ISR wings and AOCs should be selective in developing and adopting AI-powered tools for sense-making. They should prioritize those that fit into existing workflows and require them to provide less training support.

- **Mitigate skill atrophy.** The DAF CDAO should develop a plan to mitigate potential atrophy of sense-making skills resulting from AI adoption, which could include developing datasets to help train analysts and help them recognize useful data in the wild.

AI holds great promise for improving sense-making, and we believe that the approaches identified in this report can help guide investments in this area. In doing so, policymakers should consider the adoption and implementation issues identified in this report and elsewhere.

Contents

Figures and Tables

Figures

Tables

Chapter 1

Introduction

It is widely expected that artificial intelligence (AI) will play a critical role in future military operations. As the 2024 Commission on the National Defense Strategy stated, "New operational concepts must also incorporate emerging technology, such as artificial intelligence (AI) and autonomous systems, as these technologies are fundamentally changing modern warfare."[1] The Department of the Air Force (DAF) states similarly that "[d]ecision advantage across the ROMO [range of military operations] will also hinge on a professional analytical workforce, teamed with AI-enabled processing, to rapidly push actionable information to warfighters" and that "the Sense-Making element is a human/machine teamed function enabled by automation, AI/ML [artificial intelligence/machine learning]."[2] More specifically, DAF plans to

> leverage AI and human-machine teaming with data analytics to analyze extensive amounts of raw data and turn it into usable information, in order to associate identifiable indications and warnings, anticipate adversary actions, prioritize and task sensors/algorithms, and rapidly provide high-quality information to warfighters and partners.[3]

To help realize this vision, the DAF Chief Information Officer (CIO) seeks to "[l]everage data, AI/ML, and their emerging technologies to solve DAF challenges related to DAF business enterprise efficiency, mission operations integration, and greater warfighting capacity."[4] As part of this effort, the DAF Chief Data and AI Officer (CDAO) asked RAND to identify the data, technologies, processes, and policies that DAF will need to enable effective sense-making in the next decade. We address this by advancing the understanding of how these elements intersect with the current state of technology by identifying challenges in the current sense-making processes and opportunities to overcome them. The effort was to focus on how sense-making occurs—where, with what, and by whom—with a particular emphasis on how information from multiple intelligence domains can be fused to find, fix, and track targets.[5]

In this report, we identify the most significant sense-making challenges facing DAF and assess how AI capabilities could address these challenges. We also provide insights for adoption through a

---

[1] Jane Harman, Eric Edelman, John M. Keane, Thomas G. Mahnken, Mara Rudman, Mariah Sixkiller, Alissa Starzak, and Roger Zakheim, *Commission on the National Defense Strategy*, RAND Corporation, MS-A3057-4, 2024, p. 34.

[2] U.S. Air Force (USAF), *Sensing Grid: Operational Framework*, June 2020, pp. 7, 15.

[3] USAF, 2020, p. 5.

[4] DAF, *Chief Information Officer Public Strategy: FY2023–FY 2028*, September 30, 2022, p. 10.

[5] *Fusion* is defined as "combining pieces of information to produce higher-quality information, knowledge, and understanding" (Christopher G. Pernin, Louis R. Moore, and Katherine Comanor, *The Knowledge Matrix Approach to Intelligence Fusion*, RAND Corporation, TR-416-A, 2007).

comparative AI adoption schema, and we conduct a systematic examination of risk to decide which capabilities to adopt and considerations on how best to implement them. AI capabilities and DAF sense-making processes are not simple: Syncing these processes requires careful consideration of how decisionmakers will use these methods and how they are integrated into the larger intelligence cycle.

What Do We Mean by Sense-Making?

Sense-making is considered by some to be an art form by which analysts structure the unknown to understand how best to operate within it;[6] in the military context, it is a "process that meaningfully translates relevant data into usable information."[7] For DAF, that sense-making structure consists of collecting, organizing, and transforming data into knowledge about an operational environment that will provide near-real-time situational awareness along with planning and decisionmaking support across the force for a particular mission.

Sense-making entails synthesizing data that address the basic analytic questions: *who*, *what*, *when*, and *where*. This includes positively identifying, geolocating, and tracking actors in an operational environment; processing, correlating, and fusing single-intelligence domain (INT) collections into a multi-INT picture; and supporting time-dominant analysis. Sense-making also includes advanced analytic questions: *why*, *how*, and *what next*. With advanced analytic questions, sense-making seeks to understand capabilities and predict actions of all actors in the operational environment, fuses multi-INT data with all-source products and foundational intelligence, and allows for content-driven analysis.[8]

Because sense-making is a natural extension of the intelligence process, the five phases of the intelligence cycle provide a framework to understand how data are obtained, processed, disseminated, and integrated for sense-making. These phases are

1. **Planning and direction:** Collection and operational priorities (or requirements) are synced. The definition of goals and objectives, establishment of intelligence requirements, and allocation of resources occurs within this phase.
2. **Collection:** Assets are tasked to collect data. This includes data from human intelligence (HUMINT), signals intelligence (SIGINT), geospatial intelligence (GEOINT), measurement and signatures intelligence (MASINT), and open-source intelligence (OSINT). Intelligence teams validate the data to ensure that they are accurate and reliable.
3. **Processing and exploitation:** Useful intelligence is sifted from all the collected data and repackaged into accessible forms. Intelligence teams ensure that the data remain protected from unauthorized users.
4. **Analysis and production:** Sifted intelligence is analyzed, fused, and transformed into meaningful, actionable intelligence that is timely and relevant. Analysts look for patterns and

---

[6] Deborah Ancona, "Sensemaking: Framing and Acting in the Unknown," in Scott Snook, Nitin Nohria, and Rakesh Khurana, eds., *The Handbook for Teaching Leadership: Knowing, Doing, and Being*, SAGE Publications, 2011, p. 3.

[7] USAF, 2020.

[8] For a discussion of time-dominant versus content-driven analysis, see Jason M. Brown and David Vernal, "Time-Dominant Fusion in a Complex World," *Trajectory*, November 11, 2014.

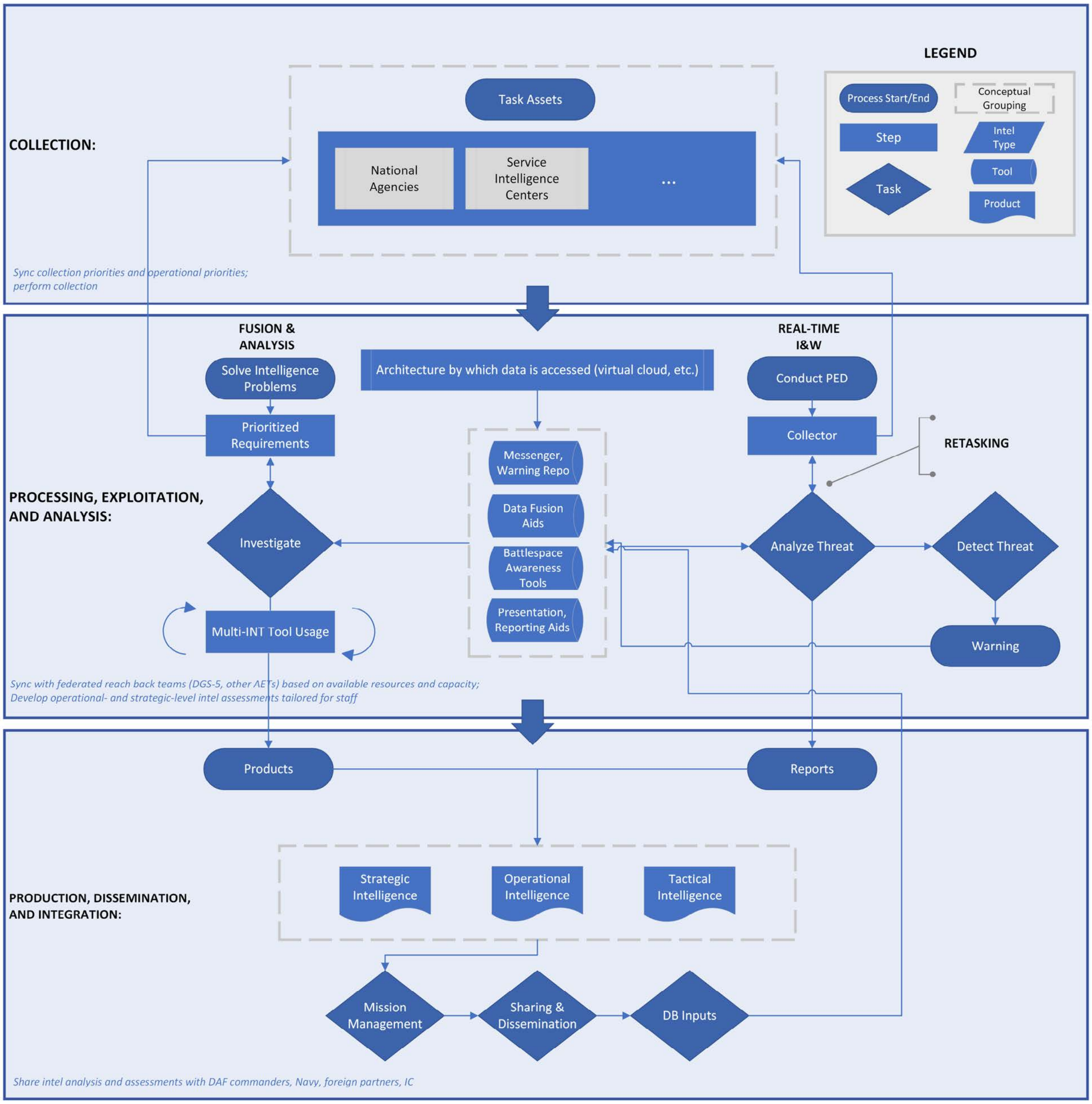
trends and consider how that affects national security and policy interests at certain degree of confidence. That intelligence is shaped into a report for decisionmakers that is quick to read and easy to understand.

5. **Dissemination and integration:** Intelligence is securely shared with stakeholders and is incorporated into operational plans or considered when making decisions. Producers of intelligence products must consider ways to customize reports based on the consumers, allowing the most important information to be easily absorbed.[9]

As a result of literature reviews and subject-matter expert (SME) interviews, we were able to map key DAF sense-making processes, including the organizations and specific tools involved. Figure 1.1 depicts a high-level process flow for how intelligence is generated, collected, and disseminated through the sense-making process. The data flow in this figure is roughly vertical, from the assets and agencies that collect raw data at the top; through the organizations that analyze, exploit, and fuse the data in the middle; to the production of reports and databases and distribution to commanders and partners on the bottom. The outputs of this data flow are ultimately used as inputs to task the next iteration of the intelligence cycle.

[9] Office of the Director of National Intelligence, "How the IC Works," webpage, undated.

Figure 1.1. Sense-Making Process Flow



NOTE: AET = Analysis and Exploration Team; IC = intelligence community; I&W = indicators and warnings; PED = processing, exploitation, and dissemination.

While often open to interpretation and expertise, sense-making relies heavily on technology to supply analysts with information, the tools to process that information, and the ability to share that information with stakeholders, allowing for consideration of more data than ever before, which can be both a blessing and a curse. Analysts increasingly have access to a plethora of data, and while that bounty is great for having insights coming in from different sources, it is also a challenge for the same reason. The ability to collect and accumulate great amounts of data continues to exponentially increase while the fragmentation of defense organizations and national assets remains. As the Air Force

Deputy Chief of Staff for Intelligence, Surveillance, and Reconnaissance (ISR) Lt Gen David A. Deptula famously warned in 2009, without new technologies and processes, the USAF would soon be "swimming in sensors and drowning in data."[10] This primes sense-making as a top candidate for technical solutions to help analysts work with large amounts of data, understand their implications, and forward the resulting information to decisionmakers.

There are many organizations within DAF that conduct sense-making activities. For this report, we looked primarily at the Air Force Distributed Common Ground System (AF DCGS) operated by the 480th ISR Wing;[11] the ISR Divisions (ISRDs) within Air Operations Centers (AOCs); and U.S. Space Force (USSF) Space Delta 7.[12] These organizations reflect the lion's share of the sense-making activity within DAF, and we believe that the sense-making challenges identified in this report are representative of the whole. Because of time constraints, we were unable to include other important sense-making organizations, such as the 55th Wing, the 70th ISR Wing, other Space Deltas that can play a role in sense-making, and unit-level intelligence squadrons. Furthering this work by including those organizations would provide a more complete picture.

What Do We Mean by Artificial Intelligence?

There is no universally accepted definition of AI.[13] In this report, we follow a line of RAND work that defines *AI* broadly as "the use of computers to carry out tasks that previously required human intelligence."[14] This statement is an updated version of AI pioneer Marvin Minsky's original definition from 1968,[15] which is consistent with the working definition from the 2018 U.S. Department of Defense (DoD) AI Strategy: "AI refers to the ability of machines to perform tasks that normally

---

[10] Stew Magnuson, "Military 'Swimming in Sensors and Drowning in Data," *National Defense*, January 1, 2010; Isaac R. Porche III, Bradley Wilson, Erin-Elizabeth Johnson, Shane Tierney, and Evan Saltzman, *Data Flood: Helping the Navy Address the Rising Tide of Sensor Information*, RAND Corporation, RR-315-NAVY, 2014.

[11] AF DCGS refers to the AN/GSQ-272 SENTINEL weapon system.

[12] Space Delta 7 is "the operational Intelligence, Surveillance and Reconnaissance (ISR) element of the U.S. Space Force" (USSF Peterson and Schriever Space Force Base, "Space Delta 7 Fact Sheet," fact sheet, October 2023).

[13] An extensive survey of different AI definitions explains, "It is not surprising that AI is so difficult to define clearly. It is, after all, an imitation or simulation of something we do not yet fully understand ourselves: human intelligence" (Haroon Sheikh, Corien Prins, and Erik Schrijvers, *Mission AI: The New System Technology*, Springer Cham, 2023, pp. 15–41).

[14] Lance Menthe, Dahlia Anne Goldfeld, Abbie Tingstad, Sherrill Lingel, Edward Geist, Donald Brunk, Amanda Wicker, Sarah Soliman, Balys Gintautas, Anne Stickells, and Amado Cordova, *Technology Innovation and the Future of Air Force Intelligence Analysis*: Vol. 2, *Technical Analysis and Supporting Material*, RAND Corporation, RR-A341-2, 2021b, p. 46; and Lance Menthe, Li Ang Zhang, Edward Geist, Joshua Steier, Aaron B. Frank, Erik Van Hegewald, Gary J. Briggs, Keller Scholl, Yusuf Ashpari, and Anthony Jacques, *Understanding the Limits of Artificial Intelligence for Warfighters*: Vol. 1, *Summary*, RAND Corporation, RR-A1722-1, 2024, p. 2.

[15] Minsky's original definition of AI was "the science of making machines do things that would require intelligence if done by men" (Marvin Minsky, ed., *Semantic Information Processing*, MIT Press, 1969, p. v).

require human intelligence."[16] Following the same line of RAND work, we further divide current AI capabilities into the five general areas and subtypes as shown in Table 1.1.[17]

Table 1.1. Current Artificial Intelligence Capabilities Framework

AI Capability Areas	Definition	Subtypes	Examples
Computer vision (CV)	Detection and classification of objects in visual media	• Object detection • Object recognition • Object tracking • Image/video generation	Apple Face ID, You Only Look Once (YOLO) model
Natural language processing (NLP)	Recognition and translation of speech and text	• Translation • Transcription • Text classification • Text generation	Amazon Alexa, large language models (LLMs) (e.g., GPT-4)
Planning	Systems that use models to find a sequence of actions that lead to a prescribed goal	—	DeepMind's AlphaGO
Prediction and classification (P/C)[a]	Discriminative models based on experience extracted from past data	—	Credit card fraud detection
Expert systems (ESs)	Rules-based systems created from expert knowledge and general heuristics	—	Aviation autopilot

SOURCE: Features information from Menthe et al., 2024.
[a] Excluding visual media and text, which are included as separate categories.

In this report, we consider only the AI capability areas listed in Table 1.1. While we believe this reasonably represents the foreseeable development of these capabilities, such as higher-performing algorithms or specialized applications built on larger datasets, we do not consider new future AI capability areas or artificial general intelligence—the as-yet hypothetical ability to achieve human-like cognition.

We take an evolutionary or "crawl-walk-run" approach to AI insertion: Instead of seeking an all-in-one solution to sense-making, we envision chaining together a series of narrow AI applications within the sense-making workflow as soon as is feasible, adding improvements and adjusting the

[16] DoD, *Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance Our Security and Prosperity*, 2019, p. 5.

[17] These groupings were originally derived from Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed., Pearson, 2021. They first appeared in Li Ang Zhang, Lance Menthe, Ian Fleischmann, Sale Lilly, Joshua Kerrigan, Michael J. Gaines, and Gregory A. Schumacher, *Incorporating Artificial Intelligence into Army Intelligence Processes*, RAND Corporation, 2021, Not available to the general public. For this report, we separated the "generative learning" category by media type because of the extensive growth of those methods.

workflow over time.[18] We focus on the potential for incremental change on the theory that "highly ambitious moon shots are less likely to be successful than 'low-hanging fruit' projects that enhance business processes."[19] This is also consistent with DoD's *Data, Analytics, and Artificial Intelligence Adoption Strategy*, which notes that components should adopt an agile approach to "incrementally reduce risk" through iterative tool releases.[20] However, more-ambitious projects are still important; we consider the potential for disruptive adoption of AI in sense-making processes in Chapter 3.

Approach

What data, technologies, processes, and policies will DAF need to enable effective sense-making in the next decade? How can AI capabilities improve the sense-making process as tools become more advanced and data lakes grow? To answer these questions, we mapped the current sense-making processes, identified challenges associated with them, and selected mission sets to help evaluate AI applicability using mission improvement metrics and expert solicitation.

Selecting the Missions

Through an iterative process, we developed a list of eight DAF missions (described in Table 1.2) as starting options from which two would ultimately be selected to guide the analysis.[21] These missions were identified based on discussions with SMEs and reviews of DAF documents as the most likely to stress sense-making processes: to test the find, fix, track, target, engage, assess (F2T2EA) kill chain; to cover both global and functional mission sets; and to span the spectrum of content-driven and time-dominant intelligence.[22] Each mission is applicable to the Indo-Pacific or European regions.

To select the two final missions, a group of SMEs scored each mission based on the breadth and depth of intelligence challenges posed and a quick initial assessment of likely AI applicability. They provided qualitative scores of 1, 2, or 3 (low, medium, or high, respectively) for each of the five major categories and subcategories in the framework. We then computed a weighted sum of all scores to obtain a single score for each mission. All categories were weighted equally except for DAF criticality, as noted in the following list:

1. Initial assessment of potential for AI benefits within the mission: CV, NLP, planning, P/C, and ES.
2. Diversity of intelligence domains associated with the mission: GEOINT, SIGINT, MASINT, HUMINT, and OSINT.

[18] See Zhang, et al., 2021, p. 11.

[19] Thomas H. Davenport and Rajeev Ronanki, "Artificial Intelligence for the Real World," *Harvard Business Review*, January 1, 2018.

[20] DoD, *Data, Analytics, and Artificial Intelligence Adoption Strategy: Accelerating Decision Advantage*, June 27, 2023.

[21] We started with 12 missions, but the counterterrorism and nuclear-related missions were removed based on sponsor guidance, and gray zone and hypersonic defense missions were removed because we were concerned about limited access to data associated with these missions.

[22] Content-driven intelligence is that for which accuracy and completeness are of paramount importance; time-dominant intelligence is that for which speed is of paramount importance. See Brown and Vernal, 2014.

3. Diversity of intelligence methodologies and products associated with the mission: Joint Intelligence Preparation of the Environment, warning intelligence, current intelligence, general military intelligence, target intelligence, scientific and technical intelligence, counterintelligence, estimative intelligence, and identity intelligence.
4. DAF criticality (i.e., broad applicability of the mission to DAF strategic goals): This category was given only half the weight of the others because we anticipated and later received more direct guidance from the DAF CDAO concerning their strategic priorities. This weighting did not affect ranking results because all but two of them received the same score.
5. Applicability to DAF stakeholders.

Table 1.2. Initial Mission Set with Final Selections Identified

Analytic Focus	Mission	Description
Content-driven	Disposition of forces (DOF)	Tracking adversary DOF across phases of conflict
	Target tracking and identification with proliferated ISR	**Target tracking and identification using low-cost, proliferated, attritable small unmanned aircraft systems (e.g., Replicator)**
	Anti-satellite (ASAT) defense	Defend against ASAT capabilities in a space domain conflict
	Combat search and rescue	Rescue of a downed pilot
	Support for humanitarian aid/ disaster relief	Provide support following a natural disaster
	Suppression of enemy air defenses	Provide support for target engagement
	Over-the-horizon targeting (OTHT)	**Provide support for target engagements at extended range, including maintaining chain of custody**
Time-dominant	Missile warning/missile tracking/missile defense	Defend against a conventional missile attack on U.S. or allied territory

NOTE: Bold text indicates missions ultimately selected for in-depth analysis. Replicator refers to DoD's initiative to field thousands of low-cost, attritable, uncrewed systems by August 2025 (Kelley M. Sayler, *DOD Replicator Initiative: Background and Issues for Congress*, Congressional Research Service, IF12611, updated March 22, 2024).

In addition, we considered the presumed classification levels and relevant intelligence teams involved (e.g., AF DCGS AETs) when compiling the list to ensure that the necessary data and personnel would be accessible.

We ultimately selected *target tracking and identification with proliferated ISR* and *OTHT* as the two missions for which to conduct in-depth analysis. Although the ASAT mission initially rated higher, it was not chosen because it overlapped with concurrent RAND efforts.[23] The proliferated ISR mission was also deemed of particular interest because the operational concept "does not naturally fit into the

[23] For further discussion on the ASAT mission, see Alexander Fiore, "Deterrent and Defensive Applications of Orbital Antisatellite Weapons," *Æther: A Journal of Strategic Airpower & Spacepower*, Vol. 2, Winter 2023.

air targeting cycle (much as today's unmanned aircraft initially did not) and raises complicated issues with rules of engagement, airspace deconfliction, and the DT [dynamic targeting] process."[24] It has also long been recognized as fertile ground for AI application.[25]

Mapping Current Sense-Making Processes

To understand DAF sense-making processes as they work today, we visited Space Delta 7, the 480th ISR Wing, the 497th ISR Group (Distributed Ground Station[DGS]-1), the 8th Intelligence Squadron (DGS-5), the 613 AOC ISRD, and the Air Combat Command (ACC) Intelligence Directorate (ACC/A2). Through these site visits, as well as separate discussions with the Air Force Research Laboratory and other SMEs, we carried out approximately 20 semistructured interviews focused on eliciting information regarding current sense-making processes, understanding how those processes supported our mission areas, and identifying how AI might (or might not) aid in those processes. The interviews coalesced around the following three guiding questions:

1. What are your most significant challenges regarding data tools, fusion tools, human workflows, and other processes used in your day-to-day mission responsibilities for sense-making?
2. What (data, tools, processes, etc.) would make your work more effective, more efficient, or more resilient?
3. Where could AI be best applied in your sense-making work and what would be needed to implement those applications?

Site visit hosts briefed us on organizational responsibilities, challenges, and progress. When feasible, we were also granted the chance to observe live operational processes and to ask follow-up questions with individuals actively performing sense-making tasks. This mixed-methods approach to elicitation allowed for a natural evolution of project structure and research outputs. While we set out to map the technical details of current sense-making processes (see Figure 1.1), outputs from discussions with stakeholders compelled us to shift focus from primarily considering AI solutions to specific technical elements of the sense-making process to considering the process more holistically and understanding where AI capabilities apply across a spectrum of technical and organizational steps, pain points, and workflows.

We extracted more than 50 interrelated challenges from those stakeholder discussions and grouped them into five challenge types: *collection orchestration*, *data access and sharing*, *data fusion and analysis*, *model management*, and *skills and training*. The first three challenge types roughly align to the intelligence cycle described earlier. The last two are crosscutting, focusing on systems and people, respectively. We prioritized challenges for consideration that came up repeatedly with different

[24] Sherrill Lingel, Jeff Hagen, Eric Hastings, Mary Lee, Matthew Sargent, Matthew Walsh, Li Ang Zhang, and David Blancett, *Joint All-Domain Command and Control for Modern Warfare: An Analytic Framework for Identifying and Developing Artificial Intelligence Applications*, RAND Corporation, RR-4408/1-AF, 2020, p. 27.

[25] For example, "[c]oordination of groups of remotely piloted vehicles (RPVs) in a surveillance mission . . . is particularly suited to distributed artificial intelligence (DAI) techniques" (Randall Steeb, Stephanie Cammarata, Sanjai Narain, Jeff Rothenberg, and William Giarla, *Cooperative Intelligence for Remotely Piloted Vehicle Fleet Control: Analysis and Simulation*, RAND Corporation, R-3408-ARPA, 1986, p. v.

stakeholders, that affected many parts of the sense-making process, or that were cited as significant pain points in at least one area. We ultimately consolidated them into the list of challenges described in Chapter 2.

Identifying Artificial Intelligence Applicability

To identify applications of various AI capabilities for the selected missions, we employed a modified Delphi method with a panel of experts that included machine learning specialists, data scientists, AI policy researchers, and sense-making and military operations SMEs.[26] Each panelist first completed a pre-workshop questionnaire scoring AI capabilities on their potential to address sense-making challenges. Following this, the panel engaged in a virtual workshop, in which the research team had timed discussions on each challenge and AI capability, after which the panel had the opportunity to revise their answers and potentially converge on a group response. Additional details for the team's specific Delphi approach are provided in Appendix A.

Following previous RAND work, we considered that AI may yield mission improvements in four broad improvement measures: *efficiency*, *effectiveness*, *use of human capital*, and *agility*.[27]

- Improving efficiency involves increasing the quantity of output per unit time and unit input—doing things faster and doing more with less.
- Improving effectiveness is about improving output quality—enhancing product accuracy, completeness, and timeliness.[28]
- Improving the use of human capital involves improving working conditions and leveraging the higher cognitive functions in humans that AI cannot yet match.
- Improving agility or resilience means improving an organization's ability to adapt to new situations and to perform under stress.

The improvement areas and sample metrics are described further in Table 1.3.

---

[26] The Delphi method is a structured method developed by RAND researchers in the 1950s for eliciting consensus expert judgment. The Delphi method traditionally works as follows: Experts provide their opinions on a question; an anonymized summary of those opinions is provided to the group; the experts then reconsider their decisions and can change their opinions; the panel reiterates until consensus has been reached or the desired number of iterations has been reached (Olaf Helmer, *Analysis of the Future: The Delphi Method*, RAND Corporation, P-3558, 1967). Note that in the pre-workshop survey and during the June workshop, participant identities were not anonymous. The panelists backgrounds are briefly discussed in Appendix A. A larger panel with a broader set of skillsets and backgrounds could yield different insights.

[27] Different sources disagree on how to define some of these terms. We follow the framework described in Lance Menthe, Dahlia Anne Goldfeld, Abbie Tingstad, Sherrill Lingel, Edward Geist, Donald Brunk, Amanda Wicker, Sarah Soliman, Balys Gintautas, Anne Stickells, and Amado Cordova, *Technology Innovation and the Future of Air Force Intelligence Analysis: Volume 1, Findings and Recommendations*, RAND Corporation, RR-A341-1, 2021a.

[28] Timeliness and speed are closely related but not the same. A timely process produces results when they are needed. A rapid process produces results quickly. By allocating resources judiciously, a process may operate slower in some instances and faster in others while remaining timely.

Table 1.3. Sense-Making Improvement Measures

Improvement Area	Description	Example Metrics
Efficiency	Increasing the production rate, production capacity, or combined throughput; decreasing associated resource costs	Speed, data throughput, person-hours, cost
Effectiveness	Improving the accuracy, completeness, or timeliness of a process; providing a more thorough analysis of intelligence questions	Probability of detection, false alarm rate, priority of intelligence questions addressed
Human capital	Making better use of humans in human-machine teaming; engaging higher cognitive functions	Task workload, job satisfaction, mental health, skills inventory
Agility	Improving the ability to adapt to changing requirements; continuing to perform under stress	Change in performance under surge conditions; ability to pivot to new areas

SOURCE: Features information from Menthe et al., 2021a, pp. 6–7.

AI and automation are often considered only in terms of their potential impact on efficiency, and it is important to remember that a wider variety of benefits are possible for sense-making. If AI does not significantly speed up the sense-making process but can help analysts build products that are more accurate and more complete, it adds value. If AI does not improve efficiency or effectiveness but can improve working conditions and free analysts to work on other problems, it adds value. If AI makes no improvements to the process but adds agility—allowing the sense-making enterprise to adapt fluidly to unexpected changes in the battlespace or surges in demand—it adds value. Planners should consider all four of these areas when deciding where and how to incorporate AI into sense-making processes.

How to Read This Report

This report uses the aforementioned methodology to highlight the challenges and potential solutions necessary to modernize sense-making through the adoption of AI capabilities. Chapter 2 highlights the opportunities for enhancement—or sense-making challenges—along with metrics for improvement and the AI tools and capabilities that could assist the warfighter in the five challenge types: collection orchestration, data access and sharing, data fusion and analysis, model management, and skills and training. Chapter 3 discusses different AI adoption strategies that go beyond the crawl-walk-run model discussed in Chapter 2. Chapter 4 describes how to assess risks invoking AI adoption in a standardized way following DoD guidance. Chapter 5 concludes with a summary of the major findings and a discussion of future research efforts that are still needed. Appendix A describes the Delphi workshop used to help assess which AI capabilities were most likely applicable to which sense-making challenges.

Chapter 2

Opportunities for Enhancing Sense-Making

The ways in which AI capabilities are likely to affect future military sense-making are not yet fully understood. Some experts argue that "[t]he modern battlefield is growing progressively more transparent because of the proliferation of advanced technologies—smart devices, sensors, emitters, etc.—as well as the emergence of hyperconnected global communications and social media."[29]

But others warn that

> [m]any of the hopes about how artificial intelligence will affect military applications flow from the assumption that computers will be able to take different kinds of data from multiple sensors and use it to track multiple targets at the same time . . . [but] [i]n some cases, additional sensors can make things worse by introducing new reasons to doubt evidence you should have trusted . . . . In the worst case these challenges can combine to make additional sensors worse than useless.[30]

For this reason, we identified challenges to the sense-making process without assuming that AI will help us meet them. We are mindful that making sense of the world is ultimately a human problem. As a recent RAND study on AI and the intelligence preparation of the battlefield reminds us, "events do not interpret themselves."[31] We join those who caution that advanced technology alone "will not magically alleviate the knowledge quality problems at the heart of our strategic dilemma . . . . If we fail to ask the right questions, neither quantum nor AI can save us."[32]

In this chapter, through a combination of literature reviews, interviews, and a capstone Delphi exercise, we identify the major challenges facing the DAF sense-making enterprise and assess how likely it is that AI could (or could not) assist in addressing them. We first describe the full set of sense-making challenges and the process by which we assessed AI applicability, then we walk through the challenges by type and describe specific findings. We emphasize that not all these challenges are difficult for DAF to do today; DAF can do some quite well, at least under current conditions. Rather,

---

[29] U.S. Army Training and Doctrine Command, *The Operational Environment 2024–2034: Large-Scale Combat Operations*, TRADOC Pamphlet 525-92, December 2024, p. 14.

[30] Edward Geist, *Deterrence Under Uncertainty: Artificial Intelligence and Nuclear Warfare*, Oxford University Press, 2023, pp. 131–135.

[31] David Stebbins, Richard S. Girven, Timothy Parker, Thomas Deen, Brandon F. De Bruhl, James Ryseff, Jessica Welburn Paige, Annie Yu Kleiman, Sunny D. Bhatt, Éder M. Sousa, Marta Kepe, and Matthew Fay, *Exploring Artificial Intelligence Use to Mitigate Potential Human Bias Within U.S. Army Intelligence Preparation of the Battlefield Processes*, RAND Corporation, RR-A2763-1, 2024, p. 1.

[32] Geist, 2023, p. 167.

they are processes that we believe are likely to become increasingly more challenging to complete successfully on a greater scale in the future.

Overview of Sense-Making Challenges

To assess how specific capabilities might be used to address specific needs, it is highly advantageous, if not strictly necessary, to distill the overall demand into a manageable set of individual challenges. To assemble a puzzle, one must first gather the pieces. However, reducing such a complex process as sense-making into a series of specific challenges is difficult because any reduction will necessarily omit parts of the process, and any division will be somewhat artificial. For this reason, we took an open-ended, iterative approach to discovering these challenges. We constructed an initial list of challenges based on prior research to shape our interview protocol. With each site visit and interview, we extended and refined the list of challenges based on loose lists of concerns and issues. Through some distillation and clustering, we ultimately arrived at the final list of 20 sense-making challenges in five broad challenge types, as shown in Table 2.1.

Table 2.1. List of Sense-Making Challenges

Challenge Type	Challenges
Collection management	• Eliciting clear, well-defined collection requirements • Optimizing collection plans across multiple domains • Dynamically recuing and retasking collection assets in a timely manner • Supporting the collection manager pipeline
Data access and sharing	• Transferring data across classification barriers in a timely manner • Working between separate Title 10/Title 50 authorities • Developing taxonomies to support object-based production (OBP) • Sharing data to support multiple common intelligence picture (CIP)/common operational picture (COP) tools and DOF dashboards
Data fusion and analysis	• Fusing data from different sources with inconsistent formatting or conditioning • Finding the “unknown unknowns” • Predicting movements and activity of adversaries and civilians • Maintaining custody of large numbers of targets
Model management	• Obtaining authorization to operate (ATO) and cybersecurity accreditation for new tools • Maintaining common software model repositories • Modifying software code as needed • Integrating new tools into workflows and adapting workflows to new capabilities
Skills and training	• Working with different tools and methods at different AF DCGS sites • Maintaining continuity of effort across different work shifts and time zones • Training for new software and methods (e.g., “tool fatigue”) • Building and retaining human skills and knowledge

This set of challenges is not exhaustive. With further site visits and interviews, it is likely that we could refine the list further.[33] Nevertheless, we believe it to be sufficiently concise and complete to guide our analysis. During this long process of narrowing down challenges, we set aside certain issues as beyond our scope or as less pressing.

Collection orchestration: We excluded from our list the challenge of obtaining highly specialized sensor data in a timely manner. Although it is important, the scarcity of exquisite collection assets lies beyond the scope of our work, and integrating AI into the sense-making workflow would not address that lack of assets. To the extent that the issue is partly one of prioritization and coordination, we address it within the challenges of optimizing collection plans and dynamic retasking.

Data access and sharing: We did not include the basic challenge of storing and transferring ever-increasing amounts of data. Although communications latency is a perennial concern—and this challenge will no doubt continue to grow alongside the growth in data—the DAF and DoD are engaging in significant efforts to address these issues at higher levels, including the use of cloud computing and greater integration of AOC and AF DCGS networks.[34] Furthermore, while we observed some latency with specific systems, most SMEs felt that policy issues constrained data management more than any technical limitations.

We also did not include the difficulty of spreading awareness of what AF DCGS can offer to the wider IC and integrating the organization more fully into it. This appears to have become less of a concern as AETs have built relationships with intelligence squadrons, targeting cells, and SMEs at service intelligence centers. This is an encouraging trend: At the end of the day, sense-making is a team sport.

Fusion and analysis: Although we initially expected a lack of standardized sense-making operating procedures across different theaters to be a significant issue as well, we believe that many of these issues have been resolved except in terms of the challenges of building CIP/COP dashboards and reporting on DOF, which we describe. We also did not look at the important question of using AI to help address human biases in analysis; this was already being addressed in a comprehensive RAND report.[35]

Model management: We were initially concerned that sense-makers lacked awareness of the variety of tools available and that there was insufficient common understanding of AI capabilities and responsibilities to know how they might be used. While these concerns may be relevant in some areas, they did not emerge as significant factors in our visits or interviews.

Overview of Artificial Intelligence Assessments

Once we compiled the set of 11 AI major capabilities and subtypes (see Chapter 1) and the final list of 20 sense-making challenges, we built a large matrix to guide the process of assessing which capabilities were applicable to which challenges. After a dry run of the exercise that led us to

[33] It is largely a happy accident that each of the challenge types has precisely four challenges.

[34] For example, see the projected $9 billion DoD investment in the Joint Warfighting Cloud Capability (Justin Doubleday, "DoD CIO Looks Ahead to 'JWCC 2.0' and Next Steps for Cloud in 2024," Federal News Network, December 14, 2023.

[35] Stebbins, et al., 2024.

reorganize some of the items into their final forms as they appear here, we conducted a two-round Delphi exercise to tap expert opinion and achieve our final scores.

For each of the 220 pairings, SMEs rated whether the AI capability was a good fit to the sense-making challenge, a poor fit, or somewhere in between. We are grateful to all the participants for their patience in walking through these pairings; the details of the exercise can be found in Appendix A. We considered many ways to analyze the resulting sets of scores to find concurrences and disagreements, including averages and other statistics,[36] but in the end, we settled on a simple rule: If a majority of SMEs on the Delphi panel considered the pairing to be a good fit (i.e., they rated the fit as 4 or 5 out of a possible 5), then we marked it as a potential good fit. Table 2.2 shows this summary of the ratings of the panel.

[36] For example, see Elizabeth A. Holey, Jennifer L. Feely, John Dixon, and Vicki J. Whittaker, "An Exploration of the Use of Simple Statistics to Measure Consensus and Stability in Delphi Studies," *BMC Medical Research Methodology*, Vol. 7, No. 52, 2007.

Table 2.2. Matrix of Artificial Intelligence Capabilities and Sense-Making Challenges

		AI Capabilities										
		CV				NLP				Other		
Sense-Making Category	Sense-Making Challenge	Object Detection	Object Recognition	Object Tracking	Image/Video Generation	Translation	Transcription	Text Classification	Text Generation	Planning	P/C	ESs
Collection orchestration	Eliciting clear, well-defined collection requirements											
	Optimizing collection plans across multiple domains											
	Dynamically recuing and retasking collection assets in a timely manner											
	Supporting the collection manager pipeline											
Data access and sharing	Transferring data across classification barriers in a timely manner											
	Working between separate Title 10/Title 50 authorities											
	Developing taxonomies to support OBP											
	Sharing data to support multiple CIP/COP tools and DOF dashboards											
Data fusion and analysis	Fusing data from different sources with inconsistent formatting or conditioning											

		AI Capabilities										
	Finding the “unknown unknowns"											
	Predicting movements and activity of adversaries and civilians											
	Maintaining custody of large numbers of targets											
Model management	Obtaining ATO and cybersecurity accreditation for new tools											
	Maintaining common model repositories											
	Modifying software code as needed											
	Integrating new tools into workflows and adapting workflows to new capabilities											
Skills and training	Working with different tools and methods at different AF DCGS sites											
	Maintaining continuity of effort across different work shifts and time zones											
	Training for new software and methods (e.g., “tool fatigue”)											
	Building and retaining human skills and knowledge											

NOTE: A gray-filled block indicates that a majority of the Delphi expert panel rated the pairing as a potential good fit.

A few trends emerged from this analysis. First, there was a surprisingly strong showing for ES, the oldest type of AI. This includes such systems as Mycin, an early ES used to identify bacteria and diagnose and treat patients by recommending appropriate antibiotics;[37] it is also the only type that does not primarily employ the kinds of deep learning neural networks that have revolutionized ML over the past decade, although other forms of ML are also used. ESs were rated as the most widely applicable AI solution method of all we considered: Some form of ES, albeit a hypothetical system that has not yet been developed, appeared likely to be applicable to most of the sense-making challenges, including at least one challenge of every type. Their more modern cousins, planning systems, were not far behind. As one Delphi panel member put it: "ESs continue to play a role because they can capture military TTPs [tactics, techniques, and procedures] to generate guardrails . . . I think people are fixated on AI models driven by big data."[38]

Second, generative AI involving text and visual media was found to have wide applicability to sense-making challenges. Together, the two generative capabilities covered almost as many sense-making challenges as ESs. Notably, most SMEs in the Delphi panel believed that generative AI could be used to help address all the challenges associated with skills and training in some way. To this point, while hailing AI as a "game-changer,"[39] the USSF recently curtailed the procurement of generative AI systems, including LLMs, for important safety reasons.[40] As these safety issues are resolved, the USSF should consider how these generative AI capabilities may be applied broadly to sense-making needs.

Finally, it is striking that the most familiar AI capabilities—CV and NLP, minus their generative subtypes—filled the matrix only sparsely. It is easy to see how automated target recognition and automatic transcription can assist analysts in performing specific analysis tasks, but these are only one part of the sense-making process, and the associated challenges go well beyond answering the basic intelligence questions. Most subtypes of CV and NLP applied to only three or four of the challenges, and if these were the only AI capabilities available, seven of the 20 challenges could not be addressed at all. When considering how AI and automation might assist in resolving future sense-making challenges, it is therefore important to cast a wide net.

The reader will observe that nearly all these sense-making challenges deal with processes that are common across multiple intelligence disciplines and do not depend on individual mission threads. This was an unexpected finding that shaped how we addressed these challenges in the project. When addressing sense-making challenges, DAF should take a mission-independent approach as much as possible. In the next sections, we walk through the results of our assessment of AI capabilities for each of the five types of sense-making challenges.

---

[37] Edward H. Shortliffe, "Mycin: A Knowledge-Based Computer Program Applied to Infectious Diseases," Annual Meeting of the Society of Computer Medicine, November 10, 1977.

[38] Remarks at RAND workshop, June 27, 2024.

[39] Unshin Lee Harpley, "Space Force CTIO: AI Will Be 'Game-Changer' for Operational Space," *Air & Space Forces Magazine*, November 14, 2023.

[40] Lisa A. Costa, "Responsible Adoption of Generative AI (GenAI) and Large Language Models (LLMs) Within the United States Space Force (USSF)," memorandum to the Guardian workforce, U.S. Department of the Air Force, September 29, 2023.

Collection Orchestration

Collection orchestration includes collection requirements management (CRM), "the generation of tasking requests to collection management authorities," and collection operations management, "the direction, scheduling, and control of specific collection operations."[41] Because this report concerns sense-making rather than sensor employment, we look primarily at CRM, except to the extent that dynamic retasking involves sensor operations. CRM is a critical part of the sense-making processes because it closes the loop in the intelligence cycle: After they have analyzed the available data and disseminated what they have learned, analysts submit sensor tasking requests to collect additional data and begin the cycle anew.

The CRM community is small, and their processes are ripe for automation. Despite the implementation of software to manage collection requests for specific assets, theater CRM remains primarily the realm of spreadsheets, whiteboards, and frantic telephone calls. AI capabilities can help automate several important parts of this process.

This is why, despite its relatively small size, we call out the field of CRM for potential insertion of AI capabilities to improve processes.

Eliciting Clear, Well-Defined Collection Requirements

The CRM process can stumble on its very first step: articulating the requests to be managed. When the requests are unclear, planners are uncertain how to meet them and the resulting taskings may be insufficient. USAF doctrine on collection requirements underscores this point: "To make the planning process more efficient, information requesters should clearly articulate collection requirements. Precise requirements allow collection managers and operations planners to determine the best way to meet requirements."[42]

Collection managers spend much of their time making calls, sending emails, and attending meetings with requestors to try to understand what they need. This is because requirements may come from different people in different formats with different understanding of what assets can provide, and the timeline for submitting requests is increasingly short. Although the CRM process can be effective today, handling tasking requests on a case-by-case basis requires a great deal of time, as well as investment in human relationships, a process that is unlikely to remain effective when demand surges. As one USAF analyst we interviewed put it, "bro-level doesn't scale."

The Delphi panel concluded that a combination of NLP capabilities and ESs could help elicit these requirements. Such a hybrid system could reformulate a verbal or written collection request into a standard format, based on the rules it understands regarding collection assets, and repeat back to the requestor an improved statement of their needs, which they may then accept or amend. The key would be to specify the essential elements of information (EEI) needed to ensure that the request describes what information they need, rather than overspecifying the specific collection assets or sensor modalities they believe can obtain it because they might not understand the full capabilities available, or their limitations. Such an AI solution would be expected to improve effectiveness and, to

[41] Joint Chiefs of Staff, *DoD Dictionary of Military and Associated Terms*, U.S. Department of Defense, November 2021, p. 39.

[42] Air Force Doctrine Publication 2-0, *Intelligence*, U.S. Air Force, June 1, 2023, p. 18.

a lesser extent, improve the use of human capital by preserving collection managers' time for the most pressing needs.

Building on this, a second step might be to use AI planning capabilities to review the historical outputs of this kind of request to illustrate what the output might be, which could assist the requestor in making a more informed choice. A vocal minority of the Delphi panel also stated that CV could be used to assess existing imagery to assist in this process for GEOINT collections when moving toward the "walk" or "run" part of the AI insertion process.

Optimizing Collection Plans Across Multiple Domains

Collection assets gather data in all warfighting domains—air, land, sea, space, etc.—and are typically managed by agencies tasked to manage those domains, not the sense-making organizations that we focus on in this report (e.g., AOC and AF DCGS). Many of those agencies maintain software programs to help them optimize collection within their domain. A concern here is optimizing collection plans across domains, which effectively means that the sense-making organizations should find a way to avoid duplication and coordinate requests before they go out to the sensor management agencies.

The panel unanimously stated that an AI planning tool that could help "game out" how collections were requested from different sources could help develop a more unified collection strategy for the sense-making organizations—which was one of the only unanimous outcomes—but beyond that, the results were mixed, and no other capability rose to the top. Several argued that NLP could be used to "identify redundancies or priority gaps based on existing collection plans . . . [and] deconflict redundant collection priorities,"[43] while others felt that CV tools were needed to take advantage of high-quality historical products to better inform future collection plans. Both the planning technologies and potential NLP technologies could improve process efficiency.

Dynamically Recuing and Retasking of Collection Assets in a Timely Manner

The previous two challenges concern deliberate planning and typically entail submitting requirements in time for the appropriate regular operations cycle for the relevant collection assets (e.g., the traditional 72-hour cycle for the Air Tasking Order). However, these timelines can be long, and the increasing speed of warfare often demands dynamic reassignment—recuing and retasking—of assets to collect on targets of opportunity or to re-collect when a previous attempt proved insufficient to deliver the required EEI.

The ability to pivot on the fly is increasingly important as a "degraded/denied environment may prevent effective execution of collection operations management over ISR assets."[44] Moreover, even

[43] Remarks at RAND workshop, June 27, 2024.

[44] Melissa Sidwell-Bowron and Matthew Winot, "The Intelligence, Surveillance, Reconnaissance Liaison Officer: A Critical Intelligence Node in Agile Combat Operations," Air Land Sea Space Application Center, February 1, 2023.

where the threat environment is permissive, target tracks and identification can be dropped or lost because of insufficient contact, so sensors must be retasked to maintain or regain custody of targets.

Here, the panel made another unanimous call for an AI planning capability to help, in large part because the work for dynamic retargeting is similar to the deliberate planning problem—it is effectively the same asset/target pairing problem, just with considerably less time and many more constraints. This is similar to the problem of devising an air attack plan, which is "currently approached in an almost entirely manual fashion," and for which an "AI system . . . would greatly accelerate the planning process, improve plan quality, and free up significant human capital."[45]

The panel also saw an important role for CV to screen images to determine if they were capable of providing EEI or if re-collection was necessary. (Note that this is not the same as determining whether the image actually *does* provide EEI. It is comparatively easy to determine the presence of haze or clouds that can render an image useless.) Some also noted that NLP could be used to transcribe retasking requests from radio, telephone calls, chat, or emails, similar to the process described above for eliciting clear collection requests. We would expect the application of AI tools to improve effectiveness by allowing for additional, timely collections that otherwise might not be scheduled while the platforms are still in the area.

Supporting the Collection Manager Pipeline

The last of this challenge type arguably belongs in the skills and training category, but we carve it out for special attention because of its importance, fragility, and specific application. Collection management accreditation certainly exists, but practical on-the-job training is limited at best.[46] The proposed AI methods to assist are (1) a hybrid of NLP capabilities and (2) an ES to create a training environment that would help collection managers learn the capabilities of the systems they manage and understand how collection requests are received.

The Delphi panel also considered that NLP, particularly generative tools, could assist the collection manager in altering collection plans on the fly when the scale of collections increase beyond what is seen at present. As one panelist explained, "Say you have a thousand images of an air view. You can't draw information from that manually, so some CV algorithm [is needed] to identify a priority.

---

[45] Matthew Walsh, Lance Menthe, Edward Geist, Eric Hastings, Joshua Kerrigan, Jasmin Léveillé, Joshua Margolis, Nicholas Martin, and Brian P. Donnelly, *Exploring the Feasibility and Utility of Machine Learning-Assisted Command and Control*: Vol. 2, *Supporting Technical Analysis*, RAND Corporation, RR-A263-2, 2021b, p. 53.

[46] DoD Intelligence and Security Professional Certification, "Collection Management Professional Certification," webpage, undated. One article from 2017, which is still relevant considering little has changed in this process over the past eight years, said,

> At the tactical level, the supported commanders' ISR professionals must be able to understand and help employ the full scope of joint capabilities. While service-centric training is inadequate to this task, 'joint' training, sadly, is far worse. It often consists of either PowerPoint slides which outline collection platforms' capabilities or a 'how-to' pamphlet. Slides and pamphlets are poor substitutes for rigorous training programs that emphasize the practical application of combat ISR capabilities (Jaylan M. Haley, "Putting the Right Man in the Loop: Views Intelligence, Surveillance, and Reconnaissance Tactical Controllers," *Air & Space Power Journal*, Vol. 31, No. 1, Spring 2017, p. 42).

Which ones are useful? What types of aircraft? Like a RAG [retrieval-augmented generation] where you ask a specific question for a specific system."[47]

We would expect a tool utilizing these AI capabilities to improve the use of human capital and agility, allowing collection managers to pivot more quickly to new demands and to handle a larger number of requests more smoothly. If any of the other applications described in this section are adopted, there would be additional need for training with those as well.

Data Access and Sharing

We identified four challenges related to data access and sharing (Table 2.1). Challenges in this section can greatly affect the remainder of the sense-making process, because assembling the right data at the right place and the right time is essential to provide decision advantage. Efficiency and effectiveness are essential. A warning from a RAND report in 2016 remains true today: "As the decision loop of our adversaries shrinks, so too does the window for which the data we collect are timely and relevant."[48]

Transferring Data Across Classification Barriers in a Timely Manner

Sensor data and the intelligence products built from them can be classified for many reasons, including sources and methods, the processes by which they are analyzed, the systems on which they are stored, and the communications channels used to transmit them. Transferring data from one system to another—or one group of analysts to another—may run into technical and policy barriers associated with different classification restrictions. Fusing different sets of data can entail similar issues. These can severely limit how quickly data are transferred and can slow data movement through the sense-making process in general. AI solutions for this challenge could add efficiency while also making better use of human capital whose time may otherwise be dedicated to tackling the additional hurdles created by classification challenges.

P/C was identified as the AI capability most likely to lend itself to overcoming this challenge, followed by text classification and ESs. These capabilities would be able to summarize text into forms that may be less informative but would be adequate for the receiving entity, which can traverse classification barriers. Furthermore, these tools may be able to find more data sources within classified networks for summarization, correctly label classified information, and flag when data increase in classification when merging different sources based on classification guidance (i.e., contained in, revealed by, and compilation). Part of the challenge is that classification guidance is not always clear or

---

[47] Remarks at RAND workshop, June 27, 2024. "Retrieval-Augmented Generation (RAG) is the process of optimizing the output of a large language model, so it references an authoritative knowledge base outside of its training data sources before generating a response" (Amazon Web Services, "What Is RAG (Retrieval-Augmented Generation)?" webpage, undated).

[48] Brien Alkire, Abbie Tingstad, Dale Benedetti, Amado Cordova, Irina Danescu, William Fry, D. Scott George, Lawrence M. Hanser, Lance Menthe, Erik Nemeth, David Ochmanek, Julia Pollak, Jessie Riposo, Timothy Smith, and Alexander Stephenson, *Leveraging the Past to Prepare for the Future of Air Force Intelligence Analysis*, RAND Corporation, RR-1330-AF, 2016, p. 42.

fully consistent, so review and deconfliction may be needed as part of the development process; this is not a simple challenge.

P/C and ESs may also be able to identify potentially interested users of the data. Semantic segmentation could help determine not only where to focus but also where to remove object identifications that are at a different classification level than the one used by the person seeking the information. As one SME on the Delphi panel explained:

> Think of image segmentation. You have an image of multiple things. You can have the algorithm take out the classification material and bring it to a lower classification without material. I don't think that the current systems are able to do this. But I can envision a future where you can get info like classified name/location classified at different level. So, an algorithm that combines them, it can know the classification has changed.[49]

As with any AI capability, there are some concerns surrounding protecting the transfer of sensitive data learned by the AI that may be alleviated with proper scaffolding and ensembling.[50] In sharing with partner nations, the United States must consider interoperability with their development of AI systems. Allowing AI to detect objects or summarize large volumes of text will still require a human in the loop for validation purposes.

Working Between Separate Title 10/Title 50 Authorities

Title 10 of the U.S. Code authorizes DoD operations, whereas Title 50 authorizes intelligence agencies to collect data. The convergence of intelligence and military resources for ISR have complicated information sharing procedures across organizations. This can result in legal barriers to data access.[51] When the DAF relies on national assets, this challenge poses complications not only to sense-making but to collection across various intelligence domains.

The only AI capability that was rated as a potential solution by the Delphi panel is Ess; the panelists viewed Title 10/Title 50 authorities as more of a policy issue than one that AI could help solve. ESs might be able to encode the complex rules and procedures required to allow for the sharing of information in different circumstances. They could also help a user navigate the process better. If an AI capability could assist with working between Title 10 and Title 50, improvements could be realized in efficiency, agility, and human capital. Perhaps in the process of building an ES, one might flag additional policy issues that could be amended to allow for better data access and sharing.

---

[49] Remarks at RAND workshop, June 27, 2024.

[50] *Scaffolding* refers to creating a "basic framework on which you can provide create, read, update and delete (CRUD) functions to allow access to a database through a web application" (IDERA, "Scaffolding," webpage, undated). *Ensembling* is an ML technique that combines forecasts from different models to improve prediction performance, or accuracy, of predictive models (Jacob Murel and Eda Kavlakoglu, "What Is Ensemble Learning?" webpage, IBM, March 18, 2024).

[51] These are longstanding issues in the ISR community. See, for example, Andru E. Wall, "Demystifying the Title 10-Title 50 Debate: Distinguishing Military Operations, Intelligence Activities & Covert Action," *Harvard Law School National Security Journal*, Vol. 3, 2011.

Developing Taxonomies to Support Object-Based Production

OBP is a cross-agency analytic effort that "creates a conceptual 'object' for people, places, and things and then uses that object as a 'bucket' to store all information and intelligence produced about those people, places, and things."[52] This is meant to replace the current mix of overlapping systems that store data based on the sensor that collected it, the sensor modality (e.g., infrared images versus radar returns), the geolocation data associated with the collection, the enemy's order of battle, and other organizational constructs. The goal of OBP is to increase information integration across the IC and DoD and to help ensure that analysts do not omit pertinent information or mistakenly assume a knowledge gap exists because their particular group or organization had not previously examined the target or area. This metaphorical bucket becomes the single repository of and the ultimate starting point to find all information collected on the object by the IC. Developing a taxonomy to support OBP would help alleviate the efficiency, effectiveness, and human capital issues associated with

- information requesters not knowing if the data are available or how to search for them
- being able to access the data but requiring someone (or software) to "discover" or make sense of it
- associating data with the object of interest.

Building OBP requires an ontology, "a formal representation of a domain of knowledge. It is comprised of a taxonomy as an integral part, with an underlying vocabulary including definitions of terms representing universals, defined classes, and axioms from which rational arguments can be made."[53] The IC and DoD have been working to build an ontology for sense-making since 2015, and the effort continues. To assist in developing an ontology, a multimodal AI approach may be beneficial because object tracking, text classification, P/C, ESs, and image/video generation were all identified as AI capabilities that could aid in developing them for OBP.

Other AI capabilities could also assist with OBP and are multimodal. Once an ontology is created, these "buckets" must be filled. Object tracking would be able to identify objects within an image and use that in combination with text classification to place items of interest in the appropriate "bucket." Similarly, P/C and ESs could do the same thing based on the processes and data they were trained on. Another AI capability that could aid in OBP is image/video generation. It was identified as a possible solution for this challenge. Some experts stated that it could be useful when sense-makers would like to see more than text excerpts. Depending on the modalities selected, training of analysts would be paramount in ensuring that objects are labeled consistently across agencies. Perhaps this would be another benefit of using the aforementioned AI capabilities—to present potential links and options for proper tags and classification based on existing buckets.

[52] Catherine Johnston, Elmo C. Wright, Jr., Jessica Bice, Jennifer Almendarez, and Linwood Creekmore, "Transforming Defense Analysis," *Joint Force Quarterly*, Vol. 79, October 2015.

[53] Air Combat Command Manual 14-422, *Intelligence Data Governance*, Air Combat Command, October 24, 2023.

Sharing Data to Support Multiple Common Intelligence Picture and Common Operational Picture Tools and Disposition of Forces Dashboards

A *COP* is "a single identical display of relevant information shared by more than one command that facilitates collaborative planning and assists all echelons to achieve situational awareness," with the goal of real-time situational awareness across echelons.[54] Similarly, a *CIP* is a continuously updated display that includes critical information about an adversary's operating state. Information from the COP/CIP helps to build the DOF dashboards. To keep the tools and dashboards continuously updated, data from multiple sources (e.g., satellite or aircraft) and multiple sources themselves need to be fed into COP/CIP tools and DOF dashboards. A further complication is that these tools and dashboards are often built by individual organizations to the specifications of that organization, so they tend to be siloed; this situation ultimately leads to the use of different taxonomies. The challenges with feeding data directly into the tools and dashboards are that

- COPs are not being updated because the data may have different classification levels, the communications links do not exist, or data may be otherwise difficult to share
- COP tools do not interact with each other, leading to inconsistencies or missing information.

Should AI capabilities be implemented, areas of improvement would include efficiency in getting relevant data to the tools and dashboards; effectiveness by doing so quickly with accuracy and completeness; human capital, because less time would be spent verifying COP/CIP/DOF completeness; and agility, because as battlefield circumstances change, the tools and dashboards must be able to capture them to ensure mission success. Text classification and P/C were identified as the most promising AI capabilities to address this challenge. While it would be beneficial to have all the relevant data, having too much data—especially data that does not succinctly tie into the requirements—could lead to a lake of data or information that analysts could drown in and would go against DoD's desire for a rapid transition from sensing to deciding. Text classification and P/C capabilities could be able to filter the available data to make sure the user's needs are precisely met.

Data Fusion and Analysis

Four challenges relate to data fusion and analysis. The proliferation of sensor data since the early 2000s has contributed to an environment in which making sense of massive amounts of information is increasingly critical to thorough analysis and effective completion of the F2T2EA kill chain. Solutions to challenges in this section are foundational to model and tool creation, whose outputs feed back into effective collection practices.

[54] Joint Chiefs of Staff, 2021, p. 42.

Fusing Data from Different Sources with Inconsistent Formatting or Conditioning

Sensing data come from a variety of sources. Making effective use of these inputs requires imposing a level of standardization not currently exhibited across the full spectrum of sense-making processes. Examples include standardization (or lack thereof) of text fields and freeform chat, the way in which numbers are captured, and in second degree applications of data, such as the creation of reports specific to one organization that are then leveraged as inputs by another. Examples of this challenge include target location data that are provided in multiple databases but are labeled "latitude" and "longitude" in one database, and "lat" and "lon" in another. Additionally, in one spreadsheet file, an entire sheet might be marked classified; in another file, individual columns are marked as classified. These inconsistencies can be further exacerbated by processes that require "significant cutting and pasting . . . of metadata from one system to another, and manual preparation of many products. For high-altitude imagery, most of the analyst's time is spent formatting, not analyzing."[55]

We note that standardization of data does not necessarily require standardized procedures for data collection and dissemination. Indeed, attempting too much standardization can be unhelpful. As a recent RAND report explains,

> [T]he standardization of data, which occurred in the case of STANAG [Standardization Agreement] 4607, yields evolution and innovation, while the standardization of transport and link layers, which occurred in Link 16, has inhibited innovation. We believe this is because data are a resource that can be exploited, while transports and links are constraints that must be overcome.[56]

Robust formatting and conditioning practices are critical to ensuring accuracy, compatibility, integration, and scalability of analytic applications, but they are difficult to employ, given the ubiquity of the assets and processes that contribute to this challenge. In some circumstances, it may even be difficult to identify the data or system owner to begin making these changes. A lack of these capabilities or standards in place poses challenges to data fusion, which would enable easier data processing, enhanced accuracy, ease of interpretation, and, eventually, analytic and/or algorithmic outputs.

The Delphi panel suggested that the use of NLP to extract, clean, and standardize text—both in general and particularly from chat messages—would provide benefit here. In one example, a participant described the possibility of using CV for such a task as AI-driven geographic information system cleaning. The most applicable AI capabilities for this challenge include text classification and text generation followed by ESs. In terms of sense-making improvement measures, text classification and generation increase effectiveness, human capital, and agility by enabling automation and fusion processes.

Policy implications for this challenge include an emphasis on good data management strategies baked into new system development and data capture/governance processes. For those existing systems, policies could be updated to reflect standardization needs—perhaps incorporating some AI-

---

[55] Menthe et al., 2021b.

[56] Jon Schmid, Bonnie L. Triezenberg, James Dimarogonas, and Samuel Absher, *The Role of Standards in Fostering Capability Evolution: Does Design Matter? Insights from Interoperability Standards,* RAND Corporation, RR-A1576-1, 2022, p. 40.

enabled tool intended to account for less structured inputs. From the training perspective, it is critical to espouse the importance of capturing good inputs (i.e., meaningful inputs to unstructured text fields) at all levels of the organization. It would be unwise to rely on AI-enabled tools to fix all issues related to formatting and conditioning. This is especially true with the possibility that these AI tools may alter the data as they try to reformat them; therefore, safeguards should be put in place. To the extent that AI technologies can assist at the operational level—perhaps by describing the importance of meaningful inputs and how they affect the organization's ability to fuse data and provide follow-on analysis—the organization should seek to decipher technology usefulness, particularly as the process of introducing new tools or requesting changes through the data owner could take a significant amount of time.

Finding the "Unknown Unknowns"

Given the proliferation of sensor data in recent years, analysts can harness more information than ever before. The individuals that the team spoke with were excited about this opportunity, and voiced a desire to find unique ways to identify things they could not have known to "go after" previously. *Unknown unknowns*—as described to the research team—include factors or variables relevant to an issue that are particularly challenging.[57] This is because they may be beyond the organization's current awareness or understanding, such as the effects that a new sensing capability might have on a group's ability to operate during a covert mission. Finding the unknown unknowns affects the ability to do further analysis and is compounded by issues related to data access (foundationally), model management (to enable big data analysis), and training (to foster a culture of inquiry). An example of this challenge includes patterns that are difficult to find and interpret, given enormous amounts of data and existing methodologies.

This challenge can be addressed, in part, by leveraging big data analytics and exploratory data analysis to uncover patterns and anomalies that might suggest unknown factors or relationships. It could also be possible to use simulation techniques to test various hypotheses and scenarios, to reveal unforeseen variables or issues, or to leverage modeling capabilities to predict outcomes under different conditions.

The Delphi panel results for this challenge were relatively neutral, with participants agreeing that many AI capabilities relate to this challenge, but without strong inputs for particular capabilities. The exception to this is P/C, which was a notable outlier. One participant described how it is possible to look for patterns even without previous training data. For example, in the case of enemy submarines, an analyst might not know to look for a precise pattern, but they can look for "holes in the ocean" or other abnormalities to guide their search. Another participant provided a solution to this challenge in the context of satellite data; by using object recognition to ask a system for a description and

---

[57] The phrase was popularized in the military context by former U.S. Secretary of Defense Donald Rumsfeld:

> There are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns—the ones we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tend to be the difficult ones (Donald H. Rumsfeld, "DoD News Briefing," transcript of news briefing delivered at the Pentagon, February 12, 2002).

comparison with reference objects that we already have access to, we can ask, "What does it seem most similar to? What about it is novel?"

This issue is particularly challenging given the nature of unknown unknowns, but steps could be taken to enact strategies and policies friendly to systems that support big data analysis—such as those that promote strong data standards—including the use and development of P/C algorithms and other simulation and modeling tools. This primarily offers improvements in the use of human capital and agility, in the sense that analysts may be better enabled to quickly pivot to new areas of interest or concern. In terms of training, fostering a culture of inquiry might include the encouragement of analytical hypothesis testing (perhaps an addendum to training related to the IC's Analysis of Competing Hypotheses),[58] or training addendums to other existing methodologies.

Predicting Movements and Activity of Adversaries and Civilians

Predicting the movements and activity of adversaries and civilians is a longstanding challenge that could benefit significantly from the application of P/C methods and techniques. An important part of this challenge is the need to leverage large amounts of information to enable predictions. Like other challenges of this type, the prediction of movements and activity of adversaries and civilians naturally affects and is affected by model management issues. An example of this challenge includes the act of predicting Red defensive posturing and response actions.

The Delphi participants confirmed the AI capability relevance to this challenge, specifically P/C, and noted the potential use of generative AI to help simulate such things as aircraft trajectories and vehicle movements. That said, participants indicated that much of this work was still experimental and other applications could provide more near-term value, such as using CV to provide near-real-time correlations, particularly in cases in which the analyst might be experiencing a significant number of sensing data returns; this then feeds into the subsequent challenge of maintaining custody of large numbers of targets.

Workshop participants took this challenge at a rather granular level from the AI capability perspective and focused on such things as the potential to use real-time data in the generation of natural movements in the area, such as individuals on sidewalks. However, as one participant noted, reinforcement learning and similar simulation methods could play a significant role in helping Blue forces to predict an adversary's next move—both literally and strategically—based on previous knowledge. All facets of CV (i.e., object detection, object recognition, object tracking, image/video generation) fared well for this challenge.

Planning also scored particularly high for this challenge, likely because of the predictive element of the description. As described, this could refer to the near-real-time planning of granular elements of a mission (say, prediction of an aircraft's flight path) or the more complex development of courses of action to respond to potential threat activity. Tactically, the prediction of movements and activity increases efficiency and human capital by enabling faster processing of significant amounts of sensor data, allowing those sensors to be quickly retasked to capture more useful data as the target moves.

[58] Richards J. Heuer, Jr., *Psychology of Intelligence Analysis*, Center for the Study of Intelligence, Central Intelligence Agency, 1999.

Strategically, this task improves the use of human capital and agility by freeing up mental capacity to consider alternate courses of action based on near-real-time adversarial maneuvers.

Policy implications for operations and near-term missions include crisis management, contingency planning, and the generation of response strategies for potential crises or sudden escalations. Policy implications at the strategic or campaign level help to inform such things as force deployments and resource allocation. This challenge also heavily affects allies, partnerships, and legal and ethical considerations related to civilian involvement in force operations. Training considerations follow a similar delineation to that of policy implications; analysts could benefit from training regarding the use of AI-enabled tools to support critical mission operations, and commanders and planners could benefit from understanding the value that AI-enabled capabilities could provide at the strategic level to encourage a greater understanding of adversarial activity and follow-on effects.

Maintaining Custody of Large Numbers of Targets

The availability of significant amounts of sensing data is immensely valuable for data fusion and analysis. However, with a crowded field of sensors, it becomes increasingly difficult to wade through the noise of surrounding data and track targets of interest. Multi-INT analysis helps in this regard, but timeliness and data access issues—as well as a lack of well-integrated multi-INT analysis capabilities—make this a persistent challenge. This challenge is closely tied to other challenges in both data access and sharing and data fusion and analysis; it is both hindered and enabled by complexities and successes, respectively, regarding the fusion of data from different sources and the prediction of movements or activity. Technical challenges involve sensor capabilities (data fusion, as described), data processing (filtering and noise reduction), and communication systems (bandwidth and cybersecurity). Operational challenges include multidomain operations, in which targets span different domains, requiring seamless integration and coordination between systems; resource allocation (prioritization and dynamic reallocation); and human factors related to decisionmaking, training, and readiness. One example of this challenge is how the DAF must maintain effective chain of custody as targets are handed across multiple platforms over long distances.

Not surprisingly, the Delphi panel participants rated object tracking very highly for this challenge, followed by such related CV tasks as object detection and recognition. Discussions among the participants mirrored this sentiment, noting that "if you're keeping track of multiple things, [CV] plays well for object detection, recognition, [and] tracking in real time. In general, this is an area where CV could be used extensively."[59] Capabilities related to planning and P/C also fared well for this challenge. Multidomain operations also play a significant role in retaining custody of large numbers of targets and can benefit significantly from multi-INT fusion and classification techniques.

Policy implications include the multifaceted issue that the challenge presents: Maintaining custody of large numbers of targets requires advanced analytic techniques and integrated operations. Training should emphasize coordination and integration of resources, as well as ensuring that personnel are trained to operate advanced systems and respond to rapidly evolving threats and circumstances.

---

[59] Remarks at RAND workshop, June 27, 2024.

Model Management

These challenges describe the difficulties of managing computer systems, including AI models, within the DAF sense-making enterprise. A combination of strict governance and lack of centralized programming resources can hinder the development and maintenance of new analytic tools and software models. This category includes management of all computing tools used for sense-making, not only the AI applications described in this report.

Obtaining Authorization to Operate and Cybersecurity Accreditation for New Tools

Current DoD policy requires that all "DoD systems (e.g., weapons systems, stand-alone systems, control systems, or any other type of systems with digital capabilities) must receive and maintain a valid authorization before beginning operations."[60] Any new software or capability intended to run on a system requires an ATO, which verifies its ability to be integrated into existing infrastructure networks.[61] This process of obtaining an ATO requires careful consideration of risk management and can take many months, if not longer, while incurring significant costs to the resourcing point, potentially creating a misalignment with the operational realities of combat operations.[62] Efforts are underway to respond to "complaints from industry officials about how the ATO process is hindering rapid technology and software innovation,"[63] but the recurring nature of these issues suggests that persistent efforts may be needed to prevent the process from ballooning out of control.

The Delphi panel indicated that ESs are the best capability for this challenge. In practice, obtaining an ATO requires satisfying many requirements and checklists; any automation tool used for this purpose will likely need heuristics to navigate the ATO process and perform deterministic steps via if-then logic. The panel also suggested text generation to create the necessary artifacts for ATO and cybersecurity accreditation. The panelists recognized a general need to streamline the accreditation process and suggested that leadership consider reevaluating ATO processes to better accommodate AI capabilities across DoD.[64] Efficiency would be the major improvement metric for improvements in the ATO process, with rates of both identification and production increasing

---

[60] Department of Defense Instruction 8510.01, *Risk Management Framework for DoD Systems*, U.S. Department of Defense, July 19, 2022, p. 13.

[61] DoD points to the National Institute of Standards and Technology for its definition of an *ATO*: "The official management decision given by a senior organizational official to authorize operation of an information system and to explicitly accept the risk to organizational operations (including mission, functions, image, or reputation), organizational assets, individuals, other organizations, and the Nation based on the implementation of an agreed-upon set of security controls" (National Institute of Standards and Technology, *Managing Information Security Risk: Organization, Mission, and Information System View*, U.S. Department of Commerce, NIST Special Publication 800-39, March 2011.

[62] "ATOs across government have traditionally taken 6–18 months, with a lot of slow back-and-forth between system owners and the assessors" (Aiden Feldman, "Taking the ATO Process from 6 Months to 30 Days," 18F, July 19, 2018).

[63] Brandi Vincent, "Pentagon Issues New Guidance to Address Industry Gripes About ATO Process," DefenseScoop, May 8, 2024.

[64] The relatively new "continuous ATO" concept is one example (David W. McKeown, "Continuous Authorization To Operate (cATO)," memorandum to senior Pentagon leadership and Defense Agency and DoD Field Activity Directors, Office of the Secretary of Defense, February 3, 2022).

significantly. However, these issues apply far beyond DAF sense-making tools, with implications for almost every part of DoD. Any solutions are likely to require coordination and efforts well beyond the scope of this report.

Maintaining Common Model Repositories

The issue of common software model repositories extends well beyond AI systems. It involves the maintenance of the numerous custom scripts for geographic information systems, Visual Basic for Microsoft Excel spreadsheets, and Python code to pipe data between systems that are routinely created locally to enhance the sensemaking process, as well as many other examples. It has special relevance to AI capabilities, which depend on models tailored for specific anticipated scenarios, which may in turn require frequent updates or user selection of appropriate models for specific situations. All these issues would be well served by a common repository for operators and analysts to select models, customize them, and share them. Currently, the USAF lacks such a repository, and operators may be unaware of the benefits of the various tools and frameworks that exist across the intelligence enterprise. Additionally, asking operators to learn such industry-standard tools as GitHub could add significantly to their mental workload without additional support.[65]

The Delphi panel agreed that text generation is a promising solution for addressing this challenge. They highlighted the benefits of using chatbots to raise awareness and answer queries when navigating model repositories. Such an application could generate model documentation creation and conduct code inspection to facilitate their use. Finally, the panel suggested containerization techniques to streamline deployment of models from a repository like GitHub on the classified side.

The policy implications are twofold. First, policies regarding practices around maintaining and distributing AI models and documentation within classified and unclassified networks must be updated to accommodate the ever-changing computing landscape. Second, codifying policies for training personnel on the use of AI for generating documentation and querying information is imperative.

Modifying Software Code as Needed

The dynamic and fast-changing battlefield environment presents a significant challenge for AI capabilities at both the enterprise and warfighter levels. Many AI systems are not trained to handle emerging threats, such as cyber threats, and cannot be easily updated to do so. Therefore, modifying AI software through code or model updating is crucial. The Delphi panel identified NLP and ESs as potential AI solutions for this challenge.

NLP can be used to clean, document, explain, and standardize code. It can also facilitate the conversion of code from one programming language to another. NLP could potentially convert spoken

---

[65] "Git, created by Linus Torvalds in 2005, reigns as the most popular Distributed Version Control System (DVCS) globally. . . . GitHub is often thought of as just a repository service . . . . However, GitHub now offers a platform that covers the entire development life cycle, from writing applications to building and releasing" (Yuki Hattori and Isabel Drost-Fromm, *DevOps Unleashed with Git and GitHub: Automate, Collaborate, and Innovate to Enhance Your DevOps Workflow and Development Experience*, Packt Publishing, 2024).

commands or intentions directly into code. However, some panelists expressed skepticism about AI's ability to generate code in this manner in highly specialized contexts outside the code it was trained on—which is largely open-source for current models.[66] Additionally, NLP could be used to analyze code to detect cybersecurity vulnerabilities, explain code failures, and perhaps even identify tampered code. Another possible application could be the generation of cryptologic code, making it difficult for adversaries to decipher commands—although that likely remains experimental at this stage.

ESs were also recognized for their potential to manage code and ensure adherence to cybersecurity standards. For example, Python Enhancement Proposals provide rule-based coding conventions for best practices in Python, and ES tools have long been established to help users adhere to these conventions.

The major policy implication for these AI solutions is the need to enable and incorporate AI tools into coding and cybersecurity practices. Facilitating the use of AI tools for code management and modification will allow various USAF organizations to test, train, and develop appropriate procedures. This will also provide leaders with the opportunity to evaluate the implications of integrating AI into their organizations.

The need to modify AI software specifically is also important in its own right because of the separate and growing phenomenon of AI model drift or model decay:

> In most real-world application scenarios, the machine learning model's performance deteriorates in production and consistently degrades as the systems evolve. . . . The accuracy of machine learning systems is prone to drop . . . [when] the system environment is dynamic and progressively subject to changes, making it difficult for a single model to provide accurate predictions.[67]

This is a general feature of AI models that are trained on static data sets but must perform in dynamic environments. An AI workflow that separates the training phase from the inference or working phase "relies on the implicit assumption that the training data is indeed representative for the target task in the working phase. . . . However, in many practical tasks and relevant real world scenarios, the assumed separation of training and working phase appears artificial and cannot be justified."[68] In cybersecurity, for example, the threat landscape changes steadily, and, thus, AI models trained on older threats can become measurably less effective in a matter of months or even days.[69] The ability to retrain or model software code in general is therefore of particular importance when the software in question is an AI application.

---

[66] Since the panel discussion, there have been many new releases of LLM-based systems, including AlphaDev and Devin.ai, that have the potential to out-code humans in a variety of contexts in the near future.

[67] Firas Bayram, Bestoun S. Ahmed, and Andreas Kassler, "From Concept Drift to Model Degradation: An Overview on Performance-Aware Drift Detectors," *Knowledge-Based Systems*, Vol. 245, June 7, 2022.

[68] Michiel Straat, Fthi Abadi, Zhuoyun Kan, Christina Göpfert, Barbara Hammer, and Michael Biehl "Supervised Learning in the Presence of Concept Drift: A Modelling Framework," *Neural Computing and Applications*, Vol. 34, 2022, pp. 101–118.

[69] Joshua Steier, Erik Van Hegewald, Anthony Jacques, Gavin S. Hartnett, and Lance Menthe, *Understanding the Limits of Artificial Intelligence for Warfighters*: Vol. 2, *Distributional Shift in Cybersecurity Datasets*, RAND Corporation, RR-A1722-2, 2024, pp. 10–12.

Integrating New Tools into Workflows or Adapting Workflows to Capabilities

The final challenge within the model management section concerns integration. DAF has long-established workflows for various tasks, which may vary between locations. Cultural inertia often leads to resistance to change, and integrating AI can disrupt existing workflows. Moreover, many current AI tools are optimized for cloud compute and are not containerized or standalone but may house capabilities that would need to run in an on-premises, disconnected environment to aid in sense-making workflow. Analysts cannot be expected to continually retrain AI algorithms onsite.

The Delphi panel explored possible AI solutions to help workflows adapt to new AI tools, focusing on two main possibilities: NLP and AI-driven workflow design. NLP can generate instructions and best practices for implementing new tools. Additionally, it can monitor workflows, user queries, and other text data to evaluate the impact of new tools and identify common problems with specific tools or tasks. An AI-driven workflow can determine the most relevant tools needed for each step in a workflow and suggest solutions or tools for improvement.

The policy implications primarily involve leadership and education. Leadership should understand the force-multiplying potential of AI tools, including AI-driven workflow enhancements. Personnel may need to be trained in interacting with AI systems to identify and address workflow issues. As noted in the next chapter, the strategy chosen to adopt AI tools in the near term can have long-term implications.

Skills and Training

The efficient and effective use of manpower is a ripe opportunity area for implementing AI in the sense-making process. The final four challenges focus on the use of AI for improving skills and training.

USAF employment of AI to manage manpower is underway and, perhaps, can expand into the sense-making process. For example, the USAF Manpower Analysis Agency employs AI through Project HIPPOPATMUS. Project HIPPOPATMUS is a toolset that supports strategic planning and complex manpower spending decisions while optimizing personnel mix and considering training limitations.[70] Our work seeks to address the structural DAF dynamics that affect sense-making personnel development, effort continuity, and skill level, resulting in limited manpower availability, nonstandardized processes, and variable tool usage. As one USAF commander we interviewed noted, "Training and resources are the biggest concern for the upcoming fight."[71] Through interviews and the expert panel, we learned about the difficulties in maintaining continuity of effort across different work shifts and time zones, difficulties with adopting new tools and methods at different DGS sites, and complications in building and retaining human skills and knowledge effects on the sense-making process.

---

[70] Air Force Manpower Analysis Agency, homepage, undated.

[71] Remarks at RAND workshop, June 27, 2024.

Working with Different Tools and Methods at Different DGS Sites

Sense-making centers in different theaters sometimes use different tools and methods to perform the same types of analysis or report the same types of information. One airman noted that an AOC or DGS site might seek authorization from their associated combatant command to use their independently developed protocol or system, but the combatant command might ask them to pause until other units catch up—which can lead to an impasse when each theater is waiting for the other to move. To that end, tools across USAF functions may be either unknown or incompatible, disrupting the cross-functional flow of information between different AOCs, across the AF DCGS enterprise, and for those personnel moving between sites.

Analysis of the Delphi panel results showed that ESs could help support data, tool, and method integration across DGS sites allowing for improved use of human capital. A combination of ESs and NLP can help translate data from one format to another structure, enabling integration. ESs would codify hard rules and NLP would codify soft matching, soft rules, and classification. Additionally, machine translation NLP for personnel at a new site can help with having different tools at sites and adjusting as necessary.[72]

Maintaining Continuity of Effort

Maintaining continuity of effort can be challenging given work shift changes, working time zones, and divisional rotations. A USAF analyst characterized this opportunity area as "24/7 operational sense-making demands are not met with 24/7 airmen availability."[73] For example, the number of available personnel is constrained because they rotate and are shared across the division. Additionally, senior intelligence domain officers operate on a 24/7 watch function with a very limited number of personnel, many of whom are frequently rotated to other functions. Also, the nature of missions spanning multiple time zones can disrupt effort continuity because the OTHT mission analysts work across various time zones and rotate on eight-hour shifts for the mission length, which can take days. Similarly, communication and information flow may disrupt the continuity of effort because of the lack of co-location. NLP can help maintain effort continuity between airmen, civilians, and contractors under these organizational conditions.

Analysis of the Delphi panel results showed that NLP tools, including LLMs, could support agile knowledge transfer and operational resilience amid personnel and time zone changes. For example, transcription and generative text NLP can be used to generate shift reports or summaries, enabling airmen to document accomplished tasks, concerns, and important open items using speech-to-text. AI-selected images based on text descriptions from the previous analyst could facilitate knowledge transfer for visual learners. Additionally, predictive models could compile information from previous shifts and suggest next steps for the incoming shift. Translation capabilities would be beneficial for

[72] In addition to on-boarding new tools, it is important to off-board them in a timely manner as well. As an earlier RAND report advised, "All software is ultimately transient, and all tools will need to be offboarded eventually. Doing so efficiently, even ruthlessly, can be important to minimizing unnecessary training on obsolete systems" (Menthe et al., 2021b, p. 92).

[73] Remarks at RAND workshop, June 27, 2024.

embedded personnel exchanges with foreign allies and partners; this could have significant benefits for DGS sites that collaborate closely with allies.

Training for New Software and Methods

Analysts are often burdened with learning new requisite tools, each presenting its own learning curve and varying levels of usefulness. This process is time-consuming for data analysts who already have limited bandwidth, leading to "tool fatigue."[74] Several USAF personnel mentioned being "worn out by change."[75] Moreover, tools may not be used as intended—for example, a tool may be borrowed to provide an impromptu solution for some other need—and when used outside of its validated parameters, it can cause unexpected errors. Furthermore, enhancing technological and methodological awareness, education, usability, and applicability could greatly benefit leaders, enabling more informed advocacy for resources and facilitating communication across services. For example, new analysts would benefit from early training on OTHT mission inputs and collaboration with other services to accurately interpret and present information. Several USAF sense-making teams mentioned that many available tools remain underused because of a lack of proper training. New analysts have significant responsibilities that quicker ramp-up would support.

In general, DoD would benefit from more advanced training methods. A recent RAND report notes that "[t]he potential benefits to the DAF from adoption of more advanced training technologies might be considerable, both in terms of economizing time and resources and in terms of operational benefits."[76] In particular, AI may be able to reduce the training required to understand datasets. Text generation capabilities via NLP can improve sense-making efficiency by reducing the person-hours spent on training or searching for tools. They can also help deliver customized training experiences. The authors of a separate recent RAND report that looked specifically at LLMs for training agree:

> LLMs hold promise for adoption and delivery of adaptive training. For training delivery, LLMs have significant advantages over other forms of adaptive training content; for instance, they can efficiently generate content tailored to the unique needs of a trainee and within constraints defined by a curriculum . . . [and] help overcome some deficiencies in data quality or competency specification in training content.[77]

AI can facilitate efficient training for analysts on their job duties, thereby reducing the time required for training in highly skilled roles. Summaries of new tools, explaining their use and applications, can replace traditional manuals. Additionally, chatbots (e.g., ChatGPT) can provide

---

[74] In the context of AI specifically, such fatigue is being recognized as "a quiet revolution of weariness towards technology that, if ignored, could sabotage even the best-laid digital strategies" (Sherzod Odilov, "Here's How Leaders Can Manage AI Fatigue," *Forbes*, February 14, 2024).

[75] Remarks at RAND workshop, June 27, 2024.

[76] Emmi Yonekura, Mark Toukan, Timothy Marler, Andrea M. Abler, Henry Hargrove, Eddie Ro, Isabelle Winston, and Sankalp Kumar, *Accelerating the Transfer of Training Technologies to Support Evolving Department of the Air Force Mission Capabilities: A Framework, Lessons Learned, and Recommendations*, RAND Corporation, RR-A2326-2, 2024, p. 49.

[77] Mark Toukan, Jair Aguirre, Sean Mann, and Eddie Ro, *Lessons Learned from Integrating a Computational Cognitive Model for Personalized Linguist Training*, RAND Corporation, RR-A2454-1, 2024, p. 26.

specialized technical support that is intuitive and has a low learning curve. Moreover, CV could be used to enhance training based on performance, especially when there are few experts available in many locations.

Building and Retaining Human Skills and Knowledge

Insufficient manpower to manage target quality tracks, coupled with a shortage of SIGINT personnel to enable effective battlespace awareness, highlights significant opportunity areas for building and retaining human skills and knowledge in sense-making. This issue is exacerbated by deficient SIGINT collector manager pipelines and a lack of AOC data strategists and analysts to analyze reporting product trends. The increased volume of data inputs necessitates a greater number of analysts, resulting in backlogs. This is part of a more persistent issue of retaining technically savvy talent in the modern age. As one article recently put it,

> "The Air Force is losing the war for technical talent," [Capt Kyle] Palko said. "We are finding it increasingly difficult to compete with the Googles, Amazons, or Facebooks of the world to recruit digital talent. Instead, the Air Force is going to have to enable opportunities to build our expertise from within."[78]

However, AI may offer a solution by closing skill gaps and efficiently ramping up new personnel. For instance, ad hoc trend analysis of mission reports currently enables only vague monthly reporting. Implementing AI could standardize and analyze mission reports, allowing for more frequent and precise trend analysis. Additionally, ISR operations personnel often develop tools in-house; AI-assisted development could streamline this process, enabling airmen to focus more on operations.

Analysis of the Delphi panel results showed that a wide variety of AI capabilities, including generative AI for all media formats and ESs, could significantly enhance personnel development efficiency and effectiveness. Indeed, the potential application of ESs to knowledge management was the highest rated pairing of all 220 options that the panel considered. NLP tools could be employed to summarize and retain knowledge or to translate historical cases into learning materials. CV could be beneficial for studying historical cases in a learning environment and is extensively used for teaching new skills through various mediums, such as virtual reality headsets. For example, virtual reality can guide users through such processes as riveting a wing or identifying objects on a map.[79] CV could also improve interpersonal communication training by imparting the skills of monitoring behavior, facial expressions, and conversational cues.

Planning algorithms and ESs could be used to build scenarios for analysts to play through, facilitating learning from past experiences, and reinforcing skills during training exercises. Generative text NLP could explain complex concepts, generate summaries, write code or analyses, and produce plots as learning supplements. Comprehensive training applications, such as conducting interviews, could benefit from the integration of multimodal AI, including CV, NLP, and predictive systems.

---

[78] Jordyn Fetter, "Project NEXUS: Empowering the Air Force's Digital Talent," Joint Base San Antonio, November 19, 2019.

[79] D. Mourtzis, V. Zogopoulos, and E. Vlachou, "Augmented Reality Application to Support Remote Maintenance as a Service in the Robotics Industry," *Procedia CIRP*, Vol. 63, 2017.

Chapter 3

Choosing the Right Artificial Intelligence Adoption Strategy

The discussion in Chapter 2 considers an evolutionary approach to AI insertion. The applications identified by the panel are intended to be inserted into current workflows with relative ease. In this chapter, we consider the potential for AI to induce more disruptive changes in the workflow.

Adoption Strategies

Adopting AI presents costs and challenges, as well as opportunities and benefits. Tools must be developed and sustained. Training data need to be acquired and curated where necessary for the AI technology. Staff must be trained. Policies need to be adjusted. If successful, human effort could be redirected to different tasks, and some tasks might be performed more quickly, completely, or accurately. AI adoption strategies can be broadly divided into two categories:

- *Nondisruptive adoption*, in which the AI supports a well-defined task within a workflow without disturbing the overall workflow. This is the strategy considered in Chapter 2.
- *Disruptive adoption*, in which AI fundamentally changes a larger workflow, redistributing tasks and responsibilities.

An example of nondisruptive adoption would be the use of AI to assist an airman in target recognition. The AI system essentially replicates or assists the airman in the same task the airman already performs; it does not significantly change the workflow upstream or downstream of it. An example of disruptive adoption would be the use of AI at the tactical edge to process, edit, and make sense of raw data at the sensor without human intervention. Such a capability could overhaul the entire sense-making workflow, including the reallocating of data ownership, roles, and responsibilities. Real world examples can, of course, fall between these end member cases; it is possible to disrupt a meaningful part of a workflow without disturbing the whole.[80]

In general, nondisruptive adoption of AI has lower implementation costs and lower risk of failure, other factors being equal. However, its benefits are also limited to increasing the efficiency,

---

[80] Ajay Agrawal, Joshua S. Gans, and Avi Goldfarb, "Artificial Intelligence Adoption and System-Wide Change," *Journal of Economics and Management Strategy*, Vol. 33, No. 2, Summer 2024; Ajay Agrawal, Joshua Gans, and Avi Goldfarb, *Power and Prediction: The Disruptive Economics of Artificial Intelligence*, Harvard Business Review Press, 2022. See also Erin E. Makarius, Debmalya Mukherjee, Joseph D. Fox, and Alexa K. Fox, "Rising with the Machines: A Sociotechnical Framework for Bringing Artificial Intelligence into the Organization," *Journal of Business Research*, Vol. 120, November 2020; Aizhan Tursunbayeva and Hila Chalutz-Ben Gal, "Adoption of Artificial Intelligence: A TOP Framework-Based Checklist for Digital Leaders," *Business Horizons*, Vol. 6, No. 4, July–August 2024.

effectiveness, use of human capital, and agility in performing the isolated tasks to which it is applied. Disruptive adoption often comes at a higher price—the reworking of a larger workflow—but if successful, it can yield a higher return on investment.[81] An example of disruptive adoption is the introduction of AI into the vehicle-for-hire business. This business was once dominated by taxi services, and taxi drivers needed to acquire an encyclopedic knowledge of a city; the barrier to entry was high.[82] New platforms, such as rideshare applications, have allowed independent individuals to easily identify optimal routes without specialized city knowledge and to engage customers through a peer-to-peer booking system.[83] AI is used by rideshare applications to

1. improve user experience by matching passengers to drivers based on distance, traffic, user preferences, etc.
2. increase efficiency by predicting estimated times of arrival, updating maps in real time to improve navigation and destination prediction, optimizing routes with convenient pickup spots, and engaging surge or dynamic pricing
3. enhance safety by using NLP and CV to detect emergencies, such as reckless driving, near-miss incidents, and sudden stops.[84]

Lessons from Nondisruptive Adoption

AI-assisted diagnostic imaging is an example of nondisruptive adoption that is similar enough to sense-making processes to provide useful lessons. It is one of the more intense areas of investigation of the use of AI in medicine because there can be many diagnostic images for each patient, these images are generally digitized (unlike some other areas of medicine), and CV is a productive and active area of AI research.[85]

Figure 3.1 compares AI-assisted image classification in radiology with AI-assisted target recognition in sense-making. The figure shows many parallels. In radiology, misreading an image can lead to inappropriate treatments and harm to patients. In sense-making, an error can lead to missing the target, to unacceptable collateral damage, and even to violations of the laws of armed conflict that undermine strategic objectives.

---

[81] Carlos J. Pérez and Carlos J. Ponce, "Disruption Costs, Learning by Doing, and Technology Adoption," *International Journal of Industrial Organization*, Vol. 41, July 2015.

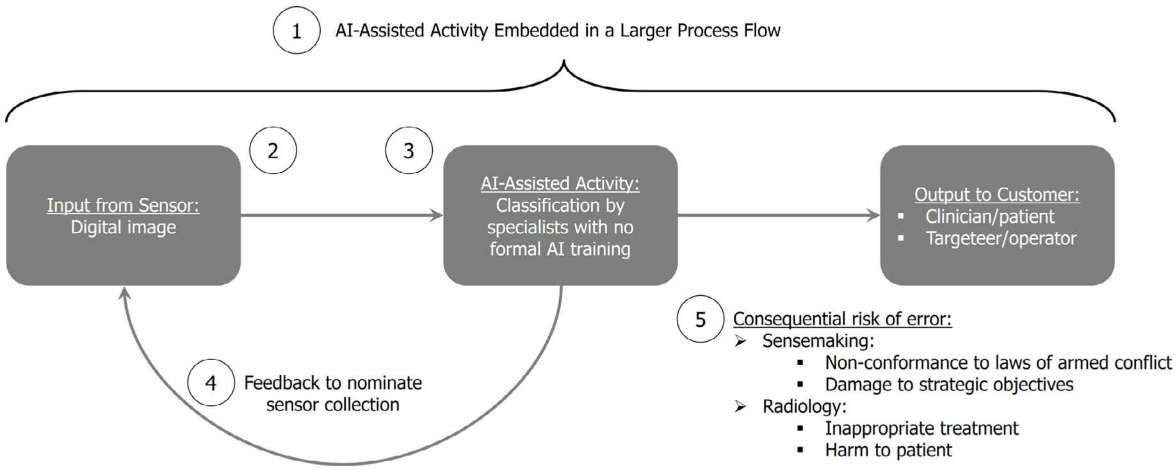
[82] "Between 2002 and 2014, the price of a [taxi] medallion rose to more than $1 million from $200,000, even though city records showed that driver incomes barely changed. About 4,000 drivers bought medallions in that period, records show" (Brian M. Rosenthal, "They Were Conned': How Reckless Loans Devastated a Generation of Taxi Drivers," *New York Times*, May 19, 2019).

[83] Agrawal et al., 2022; Agrawal et al., 2024.

[84] Anshika Mathews, "How Uber's Predictive Machine Learning Is Changing User Experience?" AIM Research, July 8, 2024; Sarah Conlisk, "How Lyft Uses AI to Get You Where You Want to Go, Faster," Lyft, August 10, 2023; LENS Corporation, "How AI Is Optimizing Your Taxi Rides," Stack the Tech Newsletter, LinkedIn, May 24, 2024; Shripal Gandhi, "Case Study: How Uber Uses AI to Optimize Surge Pricing," Hyperscale Business Newsletter, LinkedIn, June 10, 2024.

[85] Julie Sogani, Bibb Allen, Jr., Keith Dreyer, and Geraldine McGinty, "Artificial Intelligence in Radiology: The Ecosystem Essential to Improving Patient Care," *Clinical Imaging*, Vol. 59, No. 1, January 2020.

Figure 3.1. Comparison of Artificial Intelligence-Assisted Target Recognition and Radiology



SOURCES: Features information from Sogani et al., 2020; Marta N. Flory, Sandy Napel, and Emily B. Tsai, "Artificial Intelligence in Radiology: Opportunities and Challenges," *Seminars in Ultrasound, CT and MRI*, Vol. 45, No. 2, April 2024.

NOTE: This figure shows many parallels between AI-assisted image classification in radiology and AI-assisted target recognition in sense-making. These parallels are numbered as follows: (1) both are embedded in a larger workflow that is not changed; (2) both receive digitized images from sensors; (3) both assist, but do not replace, specialists who are not specifically trained in AI; (4) in both cases, assessments by these specialists can be used to alter choices of what future images to take to clarify decisionmaking; and (5) in both cases, errors in classification can have significant consequences. One difference, however, is that the "ground truth" is known in diagnostic imagery datasets. This is not always the case with sense-making, which can make curating a training set more time-consuming.

Using this radiology analogy, we can explore how adoption decisions for AI technologies could be made. Decisions about whether to adopt AI in radiology vary from country to country and institution to institution, yet there are common elements. A key attribute that has significantly influenced how AI has been adopted in radiology is the liability associated with errors. This attribute has driven a cautious approach to adoption in clinical settings, leading to the following practices that have clear parallels for DAF sense-making processes:

- Develop high-quality data curated with a desire to remove known biases.[86]
- Verify and validate algorithms to reduce both false positives and false negatives. False negatives can be deadly, and false positives can lead to more work—not less—for practitioners and incalculable pain and suffering for patients.[87]

---

[86] Sogani et al., 2020. In an example from radiology, imaging machines put metadata on the images that AI can read, such as in the case of portable imaging machines' metadata. Portable imaging machines are often used in intensive care units, and because of the *portability* of this metadata, AI may conclude that the patient has a malignant tumor. To have high-quality datasets, they must be cleaned of printed metadata.

[87] Lea Strohm, Charisma Hehakaya, Erik R. Ranschaert, Wouter P. C. Boon, and Ellen H. M. Moors, "Implementation of Artificial Intelligence (AI) Applications in Radiology: Hindering and Facilitating Factors," *European Radiology*, Vol. 30, No. 10, October 2020; Flory et al., 2024.

- Understand the limits of the training dataset—what it does and does not cover—and be cautious about using AI algorithms for cases that extrapolate beyond those limits.[88]
- Be cognizant that algorithms can make simple errors that human are unlikely to make, such as deducing that a tumor is malignant based on the presence of a chest tube.[89]
- Reserve decisionmaking for a specialist and do not surrender final classification judgments to the algorithm.[90]

Successful adoption of AI in radiology likewise has been found to correlate with the following implementation practices, which are reinforced by observations from other economic sectors and have clear parallels to the adoption of AI applications for DAF sense-making:[91]

- establishing a close relationship between the AI developer and the AI user that extends into the sustainment phase[92]
- training practitioners in enough of the details of the AI model to understand its strengths and weakness, but not more than is required. Surveys indicate that practitioners do not request explainable AI, per se—they do not feel that they need to know how the model works. They are satisfied with verified and validated models built on datasets whose underlying biases they know.[93]
- having a local champion of the AI application who is a respected practitioner and who is also knowledgeable about AI.[94] As a recent RAND report noted, "Advocates often underestimate the challenge and importance of socializing the value proposition for a technology program. Pursuing this earlier in the program can help overcome cultural resistance to new technologies."[95]

---

[88] Scott Monteith, Tash Glenn, John R. Geddes, Eric D. Achtyes, Peter C. Whybrow, and Michael Bauer, "Differences Between Human and Artificial/Augmented Intelligence in Medicine," *Computers in Human Behavior: Artificial Humans*, Vol. 2, No. 2, August–December 2024.

[89] Imon Banerjee, Kamanasish Bhattacharjee, John L. Burns, Hari Trivedi, Saptarshi Purkayastha, Laleh Seyyed-Kalantari, Bhavik N. Patel, Rakesh Shiradkar, and Judy Gichoya, "'Shortcuts' Causing Bias in Radiology Artificial Intelligence: Causes, Evaluation, and Mitigation," *Journal of the American College of Radiology*, Vol. 20, No. 9, September 2023. This concurs with other work that indicates that "intelligence soldiers [need to] have the skills and training necessary to identify different kinds of failures and to be able work around them as needed, to ensure the critical Army intelligence workflows can proceed even under difficult conditions" (Zhang et al., 2021).

[90] Charlene Liew, "The Future of Radiology Augmented with Artificial Intelligence: A Strategy for Success," *European Journal of Radiology*, Vol. 102, May 2018.

[91] Marija Cubric, "Drivers, Barriers and Social Considerations for AI Adoption in Business and Management: A Tertiary Study," *Technology in Society*, Vol. 62, August 2020.

[92] Luis Marco-Ruiz, Miguel Ángel Tejedor Hernández, Phuong Dinh Ngo, Alexandra Makhlysheva, Therese Olsen Svenning, Kari Dyb, Taridzo Chomutare, Carlos Fernández Llatas, Jorge Muñoz-Gama, and Maryam Tayefi, "A Multinational Study on Artificial Intelligence Adoption: Clinical Implementers' Perspectives," *International Journal of Medical Informatics*, Vol. 184, April 2024; Flory et al., 2024; Sam Solaimani and Lucas Swaak, "Critical Success Factors in a Multi-Stage Adoption of Artificial Intelligence: A Necessary Condition Analysis," *Journal of Engineering and Technology Management*, Vol. 69, July–September 2023.

[93] Marco-Ruiz et al., 2024.

[94] Strohm et al., 2020.

[95] Yonekura et al., 2024, p. 44.

- establishing a structured implementation process and ensuring that any AI tools integrate well with other digital tools to minimize manual inputs and other workflow disruptions[96]
- evaluating and adjusting for how the AI adoption affects task execution and workflow, not just evaluating success based on how accurate the AI models are.[97]

Adopting AI to perform tasks can also lead to atrophy of human skills related to that task because humans will perform that task less frequently. This kind of atrophy is a topic of concern in diagnostic imaging.[98] Human skill atrophy can lead to two additional consequences. First, if the task is performed by a single AI tool, the risk of a "monoculture" of solutions can arise, meaning that the task outputs lose the diversity of thought that humans provide.[99] Second, without human expertise adding to a training data base, AI might train on AI-generated data, which can in some circumstances lead to what has been termed *model collapse*, which is the regression of solutions toward the median, to the exclusion of outliers.[100]

Even nondisruptive adoption of AI will require some adjustments by workers in how they perform their duties, and success will depend on their ability to adapt to the new technology.[101]

A Lesson from Disruptive Adoption

It is worth noting that the use of AI on diagnostic imagery was not a guaranteed success. Early attempts to adopt such a system famously failed:

> In 2013, the MD Anderson Cancer Center launched a "moon shot" project: diagnose and recommend treatment plans for certain forms of cancer using IBM's Watson cognitive system. But in 2017, the project was put on hold after costs topped $62 million—and the system had yet to be used on patients. At the same time, the cancer center's IT group was experimenting with using cognitive technologies to do much less ambitious jobs . . . which contributed to . . . a decline in time spent on tedious data entry by the hospital's care managers.[102]

Disruptive adoption of AI has the potential to revolutionize large-scale processes, thereby bestowing benefits beyond marginal improvements in task execution, but it also brings additional risk of failure if the organization struggles to implement the requisite workflow changes. When large-scale changes are made to an organization's workflow processes, the ability of the organization to successfully adapt is a change management challenge.

---

[96] Strohm et al., 2020; Makarius et al., 2020.

[97] Marco-Ruiz et al., 2024. See also Makarius et al., 2020.

[98] Flory et al., 2024.

[99] Lisa Messeri and M. J. Crockett, "Artificial Intelligence and Illusions of Understanding in Scientific Research," *Nature*, Vol. 627, March 7, 2024.

[100] Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson, and Yarin Gal, "AI Models Collapse When Trained on Recursively Generated Data," *Nature*, Vol. 631, July 25, 2024.

[101] Makarius et al., 2020.

[102] Davenport and Ronanki, 2018.

The disruptive adoption of AI is a special case of change management involving the need to ingest new technical capabilities from outside the organization.[103] The ability of an organization to acquire, assimilate, transform, and exploit new technical knowledge to competitive advantage is called an organization's *absorptive capacity*.[104] An organization's absorptive capacity in a technical area is a function of its knowledge of that area, which can be expanded.[105] In the case of AI, that would mean expanding knowledge of AI, its limitations, and its applications. Developing a sufficiently receptive environment to grow this internal knowledge can require changes to organizational culture.[106] A recent RAND report concurs that "increasing absorptive capacity is within control of the DAF because absorptive capacity characterizes the ability to discover, understand, and accept technology transfer from external organizations."[107]

We have fewer lessons to draw from disruptive adoption of AI by industry in part because the types of disruptions can be very different, so they can be difficult to compare—in part because they may be less common or less publicized, and in part because such disruption can also, as in the case of taxi services, be fatal to the industries that are disrupted. It is clear, however, that early adoption of rapidly evolving technologies can enhance absorptive capacity and reduce the likelihood of future "lockout."[108] Lockout results from waiting too long to enter a technical field, at which point the field has evolved to a level of sophistication that prevents the organization from achieving a level of proficiency on par with competitors.[109]

Failure to invest early in AI could lead to low future absorptive capacity for AI incorporation, which in turn could lead to failure to maintain competitive advantage against adversaries. For this reason, nondisruptive AI adoption and disruptive AI adoption can be complementary: Nondisruptive AI adoption is one relatively low-risk way to expand absorptive capacity and pave the way for greater change.

Moving Forward

In this report, we primarily describe how AI can be deployed nondisruptively to improve DAF sense-making processes by combining narrow applications built with existing AI capabilities. We advocate such evolutionary adoption in the near term for two main reasons. First, it enables insertion of AI into sense-making workflows more rapidly, which allows the promised improvements to be

---

[103] Tursunbayeva and Chalutz-Ben Gal, 2024.

[104] Wesley M. Cohen and Daniel A. Levinthal, "Absorptive Capacity: A New Perspective on Learning and Innovation," *Administrative Science Quarterly*, Vol. 35, No. 1, March 1990; Shaker A. Zahra and Gerard George, "Absorptive Capacity: A Review, Reconceptualization, and Extension," *Academy of Management Review*, Vol. 27, No. 2, April 2002.

[105] Cohen and Levinthal, 1990.

[106] Chinho Lin, Bertram Tan, and Shofang Chang, "The Critical Factors for Technology Absorptive Capacity," *Industrial Management & Data Systems*, Vol. 102, No. 6, 2002; Jean-Pierre Noblet, Eric Simon, and Robert Parent, "Absorptive Capacity: A Proposed Operationalization," *Knowledge Management Research & Practice*, Vol. 9, No. 4, 2011.

[107] Yonekura et al., 2024, p. 17.

[108] Cohen and Levinthal, 1990.

[109] Something similar happened to Blockbuster, which famously chose not to invest in Netflix and was ultimately unable to catch up when streaming services replaced physical media, despite many early advantages (Greg Satell, "A Look Back at Why Blockbuster Really Failed and Why It Didn't Have To," *Forbes*, September 5, 2014).

realized sooner while growing absorptive capacity within DAF sense-making enterprise necessary for further successful AI adoption. Second, it attempts to mitigate the substantial problems described in Chapter 2 of analysts feeling "worn out by change" by minimizing the apparent change to the workflow and reducing the additional training needed to adopt these AI capabilities.

However, sense-making workflows have changed significantly over time and must continue to change in response to changing demands and the new opportunities afforded by technology. In 2019, AF DCGS shifted from its single-INT assembly line approach to the current AET system. Instead of focusing "only on the specific piece in the processing chain for which they were responsible," today's airmen now work to produce "multisource, fused intelligence . . . augmented by collection requirements developed by the AETs themselves."[110] It is essential that the successful insertion of AI tools into the current workflow not be allowed to lock DAF into that workflow. In other words, the AI adoption program we describe in this report should be understood as a series of steps to help balance the risks of change while preparing for more to come.

In Chapter 4, we look at how to consider these risks in a more systematic way for individual AI applications.

[110] Kelly Borukhovich and Tyler Morton, "DCGS Next Generation: Accelerating Change to Deliver Decision Advantage," *Over the Horizon*, September 26, 2020.

Chapter 4

Mitigating Artificial Intelligence Risk

In Chapter 3, we championed the idea that AI must be adopted with an understanding of its effect on the larger workflow, but the broader ecosystem considerations go beyond that. Much attention has been directed as of late to the larger need for AI systems to be trustworthy, safe, and responsible. DoD has initiated several efforts to adopt a Responsible AI (RAI) approach to integrating AI into DoD processes. *RAI* is "a dynamic approach to the design, development, deployment, and use of artificial intelligence systems that implements the DoD AI Ethical Principles to advance the trustworthiness of such systems."[111] Part of these efforts include the development of an RAI Toolkit: a multitude of frameworks, worksheets, and tools to help developers and maintainers incorporate the tenets of RAI into their practices.[112]

One tool in the RAI Toolkit is the Defense AI Guide on Risk (DAGR), whose purpose is to "holistically guide risk evaluation, provide abstracted risk models to manage risk, and suggest an approach to quantify the holistic risk of AI capabilities."[113] In this chapter, we demonstrate how such a risk analysis would be conducted for a notional new AI tool using the Social, Technological, Operational, Political, Economic, and Sustainability (STOPES) framework presented in DAGR.[114] First, we introduce a fictional AI application based on the identified sense-making challenges, and then we demonstrate how to analyze such a system by imagining risks that could apply to that AI application in the context of each of the six STOPES factors. This type of analysis, or something similar, should be conducted for any proposed AI application to assist in sense-making.

Example Artificial Intelligence Application: CoordClass

The F2T2EA process can be highly time sensitive and may require rapid coordination and data-sharing among collection managers, intelligence analysts, mission planners, and targeteers.

---

111 DoD Responsible AI Working Council, *U.S. Department of Defense Responsible Artificial Intelligence Strategy and Implementation Pathway*, June 2022, p. 41. The DoD Ethical AI Principles are that AI be responsible, equitable, traceable, reliable, and governable (U.S. Department of Defense, "DOD Adopts Ethical Principles for Artificial Intelligence," press release, February 24, 2020).

112 Among other aids, the RAI Toolkit provides the SHIELD assessment, a thorough guide for implementing AI systems, a list of external tools to aid that assessment, and a template for data and model cards for AI systems. SHIELD is an abbreviation that stands for Set Foundations, Hone Operationalizations, Improve and Innovate, Evaluate Status, Log for Traceability, and Detect via continuous Monitoring (M. K. Johnson, Michael M. Hanna, M. V. Clemens-Sewall, and D. P. Staheli, "RAI Toolkit Executive Summary," webpage, *Responsible AI Toolkit (RAI Toolkit 1.0)*, Responsible AI, U.S. Department of Defense, undated-b).

113 M. K. Johnson, Michael M. Hanna, M. V. Clemens-Sewall, and D. P. Staheli, *Responsible AI Toolkit (RAI Toolkit 1.0)*, Responsible AI, U.S. Department of Defense, undated-c.

114 The STOPES analysis is described in Section 6 ("STOPES AI Risk Considerations") of Johnson et al., undated-c.

Coordinates for particular targets are often communicated through a chat system that requires each message be tagged by the user to indicate its classification level; other users will be able to see only the messages that are tagged at the classification levels they are authorized to see. However, while the basic target coordinates by themselves might be classified at one level, more advanced information about the target, and the sources and methods by which that intelligence was derived, might be classified at a higher level. This can create an overclassification problem: Personnel who might need the target coordinates may not be able to see them, thereby impeding the timeline.[115]

"CoordClass" is a notional AI application that we conceptualize as part of a larger AI system that addresses general classification issues using the combination of NLP and ESs described in Chapter 2. CoordClass helps minimize the overclassification of coordinates in chat communications by acting as a plug-in on chat platforms, ingesting message drafts before they are sent and flagging the user via a pop-up message if it detects the presence of coordinates in a message with a higher classification tag. The pop-up message would recommend that they be tagged and propose an appropriate classification level. If the user accepts, CoordClass will send two separate messages: one at the original classification level and one at the lower classification level with the coordinate information only.

Despite its modest aims, CoordClass is not a simple tool. It is an AI ensemble that uses three different algorithms to detect coordinates within a message: an ES that matches coordinates to regular expressions, an open-source LLM to parse the text, and an NLP algorithm that applies basic semantic analysis. The pop-up message is triggered only if at least two of the three algorithms agree about the presence of coordinates.

Social, Technological, Operational, Political, Economic, and Sustainability Analysis

To illustrate the application of the DAGR methodology, we conducted a STOPES analysis on the notional CoordClass tool. Our analysis focused on risks incurred by the DAF only. The STOPES analysis lists a series of questions that are meant to guide the analysis but are not all-inclusive or prescriptive. The STOPES analysis has six factors:

- **Social**: factors related to community, social support, income, education, race and ethnicity, employment, and social perceptions
- **Technological**: factors related to the organizational affects of a technological capability being inoperable, compromised, or operating incorrectly, and appropriate supply chain risks related to technology and security
- **Operational**: factors that may result in adverse change in resources resulting from operational events, such as military (combat and noncombat) operations; operations inoperability or incorrectness of internal processes, systems, or controls; external events; and appropriate supply chain risks related to operations. Operational factors also include reputation, legal

[115] While mentioned in this notional example, this is the actual process of how coordinates are shared today, which poses a real problem for which there does not exist many real solutions.

factors, ethical factors, and human-machine interaction and the corresponding feedback loop of this interaction.

- **Political**: factors related to government policy, changes in legislation, political climate, and international relations
- **Economic**: factors that may influence the organization, such as access to funding, acquisition processes and vehicles, labor costs and workforce skill, market conditions, and appropriate supply chain risks related to economics
- **Sustainability**: factors related to human, environmental, social, and economic sustainability. The intersection and balance of environment, economy, and social equity support sustainability initiatives.[116]

We applied this framework to our notional CoordClass application and list the potential risks in Table 4.1. Thinking through these potential issues ahead of time is important to present a strong proposal for adoption and help fortify the implementation process against foreseeable failures.

Table 4.1. STOPES Analysis of CoordClass

Factor	Risks
Social	• Personnel might see CoordClass as an insult to their own abilities to correctly classify data and feel they are being coddled. This may cause tension between the users of the tool and the management that implemented it.
Technological	• Test and evaluation would have to be done to confirm that CoordClass is performing within the specified technical parameters. Test and validation criteria would need to be developed for this tool, including obtaining permissions to test it on classified data. • Each different AI algorithm has different considerations: – ES: If coordinates are manually typed into the chat window, there is a chance for human error. CoordClass may not be able to catch this error and might not trigger a pop-up when it should. – LLM: To the extent that the algorithm is vulnerable to adversarial attacks, there could be greater issues if the LLM is open-source or if its algorithm or training data are unknown to test and evaluation teams. • The expanded attack surface or vulnerability introduced by such a system must be evaluated, and risks must be mitigated.

[116] Social and economic sustainability factors are specialized topics within the aforementioned economic and social categories. Factors include concepts related to climate change, the environment, energy usage, social responsibility, human security, and appropriate supply chain risks related to sustainability.

Factor	Risks
Operational	• Having to click through many pop-up messages could slow transfer of many targets. • Users may resend pop-up messages if they appear too frequently. This may cause them to turn off the plug-in entirely or cause them to click through the prompt without thinking, potentially releasing misclassified information. • To avoid triggering the pop-up, users might begin to format coordinates in a way that CoordClass does not detect, which would render CoordClass useless and require extra time from personnel on the receiving end to decode the coordinates. • If even one instance of misclassification occurs, users might start to mistrust the system. This could cause users to abandon a system that improved their process. • If CoordClass mistakenly changes information in the messages it sends—such as sending the wrong coordinates—the larger mission that information is meant to inform could fail if that error is not caught.
Political	• CoordClass, or the larger AI model to which it belongs, would need a model card and data card to be consistent with USAF CDAO and DoD CDAO guidelines. • Changes in classification guidance could require a major update to CoordClass. • Concerns regarding worker displacement and ownership of the tool's development and implementation could alienate some and possibly lose support for the tool.
Economic	• Concerns depend on how the application that CoordClass sits within is acquired – Built in-house: CoordClass would require a significant amount of personnel time to develop and test it, which may take away from their other duties. This may be especially problematic for any understaffed DGS. – Bought off the shelf: If CoordClass was initially built for another service (or another purpose) and was acquired by the DAF, significant time and cost might be required to adapt and integrate CoordClass into DAF networks. If it was not built with DAF requirements in mind, additional effort may be needed to modify CoordClass to fit within DAF policies and practices. – Created with an industry partner: Soliciting a partner to build this app may prove costly in terms of both the initial labor and capital required to create it and maintain it, which might include hosting personnel from an industry partner to maintain the system on-site. • The total costs associated with getting CoordClass up and running and to sustain it may not be worth the benefit it provides.

Factor	Risks
Sustainability	• Without occasional checks, it may be easy to miss any misclassifications that CoordClass might make. One way to perform the verification would be to record its input and output and have a human or algorithm validate it. Storing that data could create new complications and would require extra human or machine power to perform those validation checks. However, if CoordClass' performance over time is not measured, users may be unaware when its performance begins to degrade. • If there are any changes to the standard coordinate configuration or if any new coordinate conventions begin to be used, CoordClass will have to be updated. Without these changes, CoordClass might miss new coordinate data, resulting in a return to the norm of coordinate data being overclassified. • Using CoordClass would require personnel to update and maintain it. If CoordClass was built in-house, this could pose problems if the personnel who built CoordClass are transferred to another unit where they no longer have access to or no longer have the incentive to maintain CoordClass. • If users are not properly trained on how CoordClass works or the system lacks clarity in its design or messaging, users might misunderstand how CoordClass works. They might believe that it is an authoritative source and may accept CoordClass' recommendations thoughtlessly, or, conversely, they might believe that CoordClass is only a recommendation and ignore its messages. • Running CoordClass expends energy, thereby increasing energy costs for as long as it runs. This includes the compute power to run and retrain the AI models used in CoordClass, as well as the compute power to run each individual query.

SOURCE: Authors' analysis of Johnson et al., undated-c.

Although each of the outcomes of these risk considerations affects the STOPES domains differently, these potential risks have a common set of the following outcomes:

- CoordClass could be underused by personnel, rendering the investment moot.
- CoordClass prompts might be automatically accepted, which could result in a classification error.
- CoordClass could require an unacceptable amount of personnel or resources to operate.
- CoordClass could create tension between different units that need to operate closely.
- CoordClass could give incorrect information.

Risk Mitigation

The ultimate purpose of identifying risk considerations is to aid in identifying what steps to take for risk mitigation. Some of the risks listed in Table 4.1 will be highly unlikely, while others will be much more likely, though identifying the level of likelihood is outside the scope of this analysis. Beyond likelihood, steps for mitigating these risks depend on the level of tolerance for each of the risks. To illustrate how the final part of this analysis process would go, Table 4.2 lists a few key mitigation steps and related guidance from the RAI Toolkit.

Table 4.2. Risk Mitigations for CoordClass with RAI Toolkit Guidance

Risk Mitigation	Related SHIELD Guidance	Tools for Enacting Guidance
Develop a comprehensive test, evaluation, verification, and validation (TEVV) criteria to ensure CoordClass works as expected	• Step 5: TEVV	• Baseline datasets • Robustness tools • Drift tools
Record incidents when a misclassification incident occurs	• Step 7: Use • Step 7.4: Record Lessons Learned	• RAI use case repository • AI incident repository
Assign a clear maintainer of this system, with a clear hierarchy of who has what responsibility	• Step 2: Ideation • Step 2.5: Accountability, Responsibility, Access Flows and Governance	• Appendix 5: Responsibility Flows Questionnaire • Appendix 7: Personas List and Descriptions
Fill out a model and data card for CoordClass	• Step 3: Assessment • Step 3.2: Exploratory Data Analysis	• Appendix 10: Data and Model Card Templates and Guidance

SOURCE: Authors' analysis of Johnson et al., undated-c.
NOTE: The steps in the middle column are the steps of the SHIELD process.

The first column of Table 4.2 lists a specific risk mitigation, and the second and third columns list guidance from the RAI Toolkit for performing that risk mitigation. Part of the RAI Toolkit is the SHIELD assessment, which outlines seven stages of the AI product life cycle and the considerations that should be taken for each step to identify and mitigate risk. The second column points to the associated section of the SHIELD assessment that should be completed to enact that risk mitigation. The third column lists tools and other guidance from the RAI Toolkit that can be used to complete the SHIELD assessment steps. For example, if reducing the risk of CoordClass giving incorrect information is a high concern, one way to mitigate risk would be to develop a comprehensive TEVV criteria. Step 5 (TEVV) of the SHIELD assessment walks stakeholders through key questions and actions needed to perform TEVV and links to additional tools to accomplish this. For example, Question 2 of Step 5 asks, "Have there been unit tests of each component in isolation? Have there been integration tests to understand how the components interact with one another within the overall system?" and links to a set of baseline datasets that can be used to perform the unit tests and integration tests.[117] The risk mitigation steps listed here are merely fictional suggestions for a fictional tool, created by our research team to illustrate what risk mitigations could look like. But deciding which risks are tolerated, at what level, and the mitigation steps to address intolerable risks is a responsibility that rests on the unit using those tools.

The RAI Toolkit defines responsibilities for different roles in the RAI approach; however, many of these roles involve AI expertise. For instance, an AI ethics and risk specialist, a role in the Defense Cyber Workforce Framework, is "responsible for tracking consistency with the DoD AI Ethical

[117] Johnson et al., undated-c.

Principles and RAI practices."[118] In the sense-making domain, it is unclear who is meant to fulfill these responsibilities. Additional clarification on whose responsibility it is to perform RAI-related tasks—such as a risk analysis like the one presented here—in the sense-making domain would help ensure that AI tools that the DAF uses are indeed following the DoD Ethical AI Principles and the tenets of RAI.

[118] M. K. Johnson, Michael M. Hanna, M. V. Clemens-Sewall, and D. P. Staheli, "Appendix 7. Personas List and Descriptions," webpage, *Responsible AI Toolkit (RAI Toolkit 1.0)*, Responsible AI, U.S. Department of Defense, undated-a.

Chapter 5

Conclusions

In this chapter, we summarize the proposed AI applications from all sense-making types, along with the lessons learned from considering adoption strategies and risk mitigation analysis. We conclude with a discussion of topics for future work.

Summary of Findings

From our site visits, our interviews with SMEs, and the Delphi elicitation exercise, roughly a dozen findings emerged as to how near-term AI capabilities could be applied to address DAF sense-making challenges. Table 5.1 compiles these use cases from Chapter 2 and identifies which of the five general types of AI capability are indicated. The AI capabilities marked as having major or likely application are those that are central to the proposed use case. The AI capabilities marked as having minor or potential application are those that could be used to extend the use case in subsequent development.

The proposals in Table 5.1 are a mix of the novel and the familiar. Together they represent a portfolio of development options for USAF and USSF sense-making organizations independent of the specific systems used today.

Table 5.1. Summary of Use Cases by Sense-Making Challenge

Challenge Area	Major Findings	CV	NLP	Plan	P/C	ES
Collection orchestration	NLP combined with ES to elicit requirements and rephrase them into standard formats		●			●
	Planning systems to improve both deliberate and dynamic collections across multiple domains	○		●		
	CV to screen collections incapable of providing the required EEI	●				
Data access and sharing	Text classification combined with ES to propose or confirm classification markings to assist in data transfer		●			●
	Multimodal system to assist in OBP and ontology development	○	●		○	○
Data fusion and analysis	NLP in an ES framework to clean and condition processed data		●			●
	CV and NLP to assist tracking DOF across multiple sensor modalities and through chat and radio reports	●	○			

Challenge Area	Major Findings	CV	NLP	Plan	P/C	ES
	P/C and planning with CV assistance to anticipate future adversary movement	○		●	●	
Model management	NLP with ES to parse code and manage adherence with cybersecurity regulations		●			●
Skills and training	ES to support customized training programs		○			●
	NLP and ES to support knowledge management and assist knowledge transfer between units, shifts, and personnel		●			●

NOTE: ● = major/likely applicability; ○ = minor/possible applicability.

In addition to these proposals, we draw a handful of crosscutting findings from the discussion of AI adoption strategies in Chapter 3 and the discussions of risk in Chapter 4.

- **Datasets and knowledge representations need to be carefully curated.** By now it is well recognized that AI is only as good as the data on which it is trained, or, in the case of ESs, the knowledge representations on which it is built.[119] For AI capabilities to be effectively applied to sense-making problems, high quality datasets must be built with care—not scraped randomly off the internet—and must also be associated with the right metadata to support OBP and enable subsequent algorithm development. However effective an AI system might be today in performing a specific task, if it has been constructed using an impoverished dataset or an isolated knowledge representation, it ultimately will become an obstacle to the broader adoption of more integrated sense-making applications in the future.
- **Analysts can and should anticipate AI failure modes.** AI systems work best when they are assigned tasks for which the output can unambiguously be deemed correct or incorrect.[120] However, real-world sense-making does not always lend itself to such certainty, especially when the ground truth is not known. Understanding the limits of an AI system's training data or knowledge representation will help analysts anticipate the types of errors it can make and be cautious about using AI for cases that extrapolate beyond those limits.[121] Analysts should also be trained to anticipate unusual errors. AI is known to make simple errors that humans are unlikely to make, which humans therefore may have difficulty anticipating. AI algorithms are also vulnerable to adversarial attacks designed to exploit these weaknesses, as well as attacks against its design and implementation—as is any computing system.[122] This does not,

[119] "For AI systems, what 'exists' is that which can be represented. When the knowledge of a domain is represented in a declarative formalism, the set of objects . . . and the describable relationships among them, are reflected in the representational vocabulary with which a knowledge-based program represents knowledge" (Thomas R. Gruber, "Toward Principles for the Design of Ontologies Used for Knowledge Sharing," *International Journal Human-Computer Studies*, Vol. 43, Nos. 5–6, November 1995).

[120] Nora Young, "What AI Can and Can't Do," *Spark* podcast, November 10, 2022.

[121] As a recent RAND report put it, "AI classification algorithms cannot be relied upon to learn what they are not taught" (Menthe et al., 2024, p. 22).

[122] For more on this, see John Matsumura, Lance Menthe, Henry Hargrove, Casidhe Hutchison, Bridget R. Kane, Philip Song, Joshua Steier, Anton Wu, and Elie Alhajjar, *Mitigating the Impact of Future Adversarial Attacks on Artificial Intelligence (AI) Applications for the Army*, RAND Corporation, RR-A2432-1, 2024, Not available to the general public.

however, require that users understand the inner workings of the model, but the AI must be "explainable" in the more limited sense that users can comprehend and trust the model based on DAF validation processes and recognize its underlying biases.[123] Users should also understand the basis for the algorithm's outputs and have the ability to investigate unexpected behavior.

- **ESs can play an important role.** The older forms of AI remain relevant. Policymakers should not look only to neural networks to help build AI solutions. Although these are the oldest forms of AI, ESs remain relevant today because they are not probabilistic, as most other trained forms of AI are. Instead, they are deterministic and, therefore, predictable and relatively easy to understand and explain.
- **The DAF should pave the way for disruptive adoption.** Disruptive AI will be needed later, but early adoption of nondisruptive AI can help prepare the DAF for greater change by building its absorptive capacity. It should also be expected that learning new systems takes time and productivity can decline during the learning period, especially when these systems are disruptive.[124]

Recommendations

Consideration of the enablers and constraints associated with adopting these AI technologies leads to four crosscutting recommendations.

Follow a Shared Road Map for Developing Sense-Making Capabilities

To organize integration efforts, USAF sense-making wings (e.g., 480th ISR Wing) and their USSF counterparts should work with the USAF CDAO to develop a set of shared priorities for AI integration based on the capabilities summarized in Table 5.1, starting with the least disruptive AI tools that easily drop into the existing sense-making workflow and require limited additional training. While we believe that the data fusion and analysis tools are of particularly high importance, the prioritization needs to consider all relevant factors at play. Beginning with the nondisruptive tools would pave the way for necessary disruptive AI adoption in the future.

Anticipate Risks Early

Because some of these processes are currently unclear and would benefit from codification, the USAF CIO should take ownership of ensuring that RAI-related tasks are executed for the sense-

---

[123] Like most things related to AI, explainability is an ill-defined and evolving concept. One survey concludes that it is now about achieving a "good interpretability-accuracy tradeoff" rather than requiring everything to be understood (Sajid Ali, Tamer Abuhmed, Shaker El-Sappagh, Khan Muhammad, Jose M. Alonso-Moral, Roberto Confalonieri, Riccardo Guidotti, Javier Del Ser, Natalia Díaz-Rodríguez, and Francisco Herrera, "Explainable Artificial Intelligence (XAI): What We Know and What Is Left to Attain Trustworthy Artificial Intelligence," *Information Fusion*, Vol. 99, November 2023).

[124] "Productivity often temporarily declines after introduction of a new technology . . . improvements are needed to lay the groundwork for AI/ML to come" (Menthe et al., 2021b, p. 98).

making domain to guarantee that AI tools that the DAF uses are following the DoD Ethical AI Principles and the tenets of RAI. Furthermore, the DAF sense-making organization that proposes a new AI tool—whether it be an AOC, a wing, or a group that operates the AF DCGS, ACC, or another ISR or targeting wing or group—should perform a standard risk assessment process, such as the STOPES analysis described here, for that proposal. These risks can and should be considered early in the development process. Addressing them early will smooth adoption later.

Respect Tool Fatigue Sentiments

DAF sense-makers are exhausted by the seemingly endless parade of new tools presented for their use. Having knowledgeable, local champions to help design and develop AI applications can help combat this fatigue, but it must not overly burden the DAF sense-making community. Too much inclusion in the development process can become its own burden if a software tool is not close enough to what is needed and significant reshaping is required.

The need for units to supply their own training to support these tools also adds to the sense of tool fatigue. As one DAF commander explained, "if we have to pay the training tax," then they will be reluctant to adopt new tools.[125] This is a significant change from even a few years ago when RAND researchers found that airmen were eager to adopt new tools and "AF DCGS ops floor[s] are fertile ground for sowing new tools, technologies, and processes."[126] We can only speculate on the reasons for the change, but overpromising and underdelivering on recent programs may have contributed.[127] Tool fatigue cannot be wished away; at this juncture, the DAF should be selective in adopting AI-powered tools and prioritizing those that require less training and fit well into the existing workflow.

Mitigate Skill Atrophy

As noted in the case of diagnostic imaging, adoption of AI to perform tasks can lead to atrophy of human skills related to that task because humans will perform that task less frequently. This can lead to a lack of diversity in thinking about the task area, misunderstandings of the process, and less innovation in areas that may need to evolve. Skill atrophy could also make it more difficult to improve the dataset if humans lack the expertise to recognize useful data. The USAF CDAO should develop a mitigation plan for skill atrophy as more AI tools are integrated into the sense-making process. As part of such a plan, the USAF CDAO should consider the development of training datasets that can not only be used to train and validate AI tools but will also provide analysts data to train on and help them recognize useful data in the wild. Analysts could occasionally practice in exercises and training events without AI tool assistance. This could also aid in analysts' propensity to recognize bad data or anomalies, capacity to investigate questionable AI outputs, and ability to make recommendations on how to improve AI tools.

---

[125] Remarks at RAND workshop, June 27, 2024.

[126] Menthe et al., 2021b, p. 93.

[127] The implementation issues with Project Maven were mentioned by some interviewees and documented in another RAND report (Chad Heitzenrater, Bradley Wilson, Sarah W. Denton, James Ryseff, and Jeffery Broughton, *Lessons Learned from the Algorithmic Warfare Cross Functional Team: Project Maven*, RAND Corporation, 2024, Not available to the general public).

Conclusions

The DAF sense-making enterprise slewed the bulk of its efforts over the first two decades of the 21st century toward supporting operations in relatively permissive environments, such as Afghanistan, Iraq, and Syria. As the focus pivots back toward peer threats, the astonishing proliferation of sensors and targets—and the need to make sense of all the data—have made scalability perhaps the most critical concern for sense-making processes. As we look to where AI can assist, we are mindful that the goal should be to support intelligence analysis, not to replace it. As stated in a previous RAND report, "Analysis is not a burdensome step in the intelligence cycle that should be eliminated but a critical step that should be strengthened."[128]

In this report, we identified 20 broad challenges to performing sense-making at scale and suggested roughly a dozen of those challenges where judicious application of AI capabilities are likely to be useful. Most of these recommendations involve chaining together combinations of AI capabilities, because sense-making problems are complex. As a RAND report on AI for command and control noted, "Problem characteristics call for multiple solution capabilities, some of which are hard to achieve together . . . hybrid approaches are often needed to deal with the range of characteristics."[129]

We also identified several crosscutting findings on data, algorithms, and implementation to help guide AI adoption into the sense-making process. Most of these findings dovetail with findings from other reports, but the importance of tool fatigue is a new finding for this area.[130] We originally sought to look for specific lessons for proliferated ISR and for specific technical details, but following from discussions with stakeholders, we ultimately shifted toward the more holistic view of how AI capabilities might apply that is described here. We are mindful as well that the sense-making enterprise lacks the resources to revolutionize every area at once.[131] We intend that the discussion here helps elucidate certain challenges for attention.

One area that we did not consider in this analysis is the need for assessments of the sense-making process to measure if any of the AI systems suggested here succeed when considering all success metrics, including costs, ease of implementation and use, achieving desired outcomes, and accuracy. This is a problem for ISR in general: "There is no common assessment approach between (or even within) USAF airborne ISR organizations; very limited availability of reliable, accurate data; a lack of common terminology and data standards; and, in many cases, lack of either feedback from end users or access to contextual information needed for ISR specialists to make assessments."[132] Relatedly, we also did not look at the potential application of AI to tactical assessments, such as battle damage

---

[128] Menthe et al., 2021a, p. 23.

[129] Matthew Walsh, Lance Menthe, Edward Geist, Eric Hastings, Joshua Kerrigan, Jasmin Léveillé, Joshua Margolis, Nicholas Martin, and Brian P. Donnelly, *Exploring the Feasibility and Utility of Machine Learning-Assisted Command and Control*: Vol. 1, *Findings and Recommendations*, RAND Corporation, RR-A263-1, 2021a, pp. 66-67.

[130] When asked what would not be helpful from a RAND study, one interviewee candidly remarked, "Please don't recommend another hundred new tools" (remarks at RAND workshop, June 27, 2024).

[131] "The AF DCGS lacks the resources to make the kinds of large-scale investments that will be required to achieve AI/ML breakthroughs in all relevant collection disciplines" (Menthe et al., 2021a, p. 22).

[132] Abbie Tingstad, Dahlia Anne Goldfeld, Lance Menthe, Robert A. Guffey, Zachary Haldeman, Krista Langeland, Amado Cordova, Elizabeth M. Waina, and Balys Gintautas, *Assessing the Value of Intelligence Collected by U.S. Air Force Airborne Intelligence, Surveillance, and Reconnaissance Platforms*, RAND Corporation, RR-2742-AF, 2021.

assessment. Finally, the study was scoped to look at the early stages of the targeting cycle only. The application of AI to targeting more generally is an important area of research. Future research could investigate those areas that were not in our scope and include an assessment of the DAF's ability to support technical innovation by recommending capabilities, tools, and organizational frameworks to support model development and machine learning operations at all levels of command.

As the DAF looks to integrate AI into sense-making, we are mindful that the human part of the equation remains paramount and is likely to be so for the foreseeable future. As a recent analysis of DAF liaison officers noted, "Data and net-centric warfare require human based relationships."[133] AI holds great promise for improving sense-making, and we believe that the approaches identified in this report can help guide investments in this area, but policymakers should consider the adoption and implementation issues identified in this report and elsewhere. Employment of AI must not become complete dependence. A warning from a 1949 RAND report remains as true in 2025 as it was then:

> No intelligence technique has been developed to the point where its performance even approximates its "inherent" potentialities consistently . . . . There would be little justification at present, and probably for some time to come, for policymakers to grant any intelligence technique an exclusive field of operations.[134]

[133] Sidwell-Bowron and Winot, 2023.

[134] Alexander L. George, *The Intelligence Value of Content Analysis: A Preliminary Progress Report*, RAND Corporation, RM-116, February 15, 1949, p. 102.

Appendix A

Sense-Making Delphi Workshop

The Delphi method is a structured method, developed by RAND researchers in the 1950s, for eliciting consensus expert judgment.[135] It works as follows: A panel of experts provide their opinions on a question; an anonymized summary of those opinions is provided to the group; the experts then reconsider their opinions and iterate for several rounds or until consensus has been reached.[136]

The research team distributed an initial survey and then held a virtual four-hour workshop in June 2024 using a Delphi exercise with nine RAND SMEs. The goal of the workshop was to identify which AI capabilities (if any) hold potential to address sense-making challenges. The background of the SMEs included intermediate to expert knowledge of AI applications in a military context. Many SMEs also had at least intermediate knowledge of sense-making operations, and most had expert knowledge of general AI algorithms and applications. Military service among SMEs ranged from none to 18 years.

Prior to the Delphi workshop, the research team identified 20 sense-making challenges (see Table 2.1) and identified 11 AI capabilities (including subtypes) as potential solutions to those challenges (see Table 1.1).[137] On the initial survey, which served as the first round of the Delphi process, workshop participants were asked to score each AI capability against its potential to overcome each sense-making challenge on a scale from 1 to 5:

1. Inapplicable or very poor fit
2. Poor fit
3. Neither poor nor good fit
4. Possible fit
5. A likely good fit

During the workshop, the research team defined each challenge in detail, discussed concrete sense-making examples of each of those challenges, and encouraged the experts to share their rationales for their scores. Participants were then asked to rescore each capability and challenge pairing using the same 1 to 5 scale. (Participants could choose not to score a particular pairing if they felt that they did not have sufficient expertise.) Following the workshop, the team used the final score sets along with analysis of DAF documents and SME interviews to identify how AI capabilities could be applied to sense-making challenges.

[135] Helmer-Hirschberg, 1967.

[136] The workshop was conducted in a virtual meeting space. Participant identities and inputs were not anonymous.

[137] The initial list did not call out object detection separately. This was added as a capability for the workshop and this report following the pre-workshop discussion.

Abbreviations

ACC	Air Combat Command
AET	Analysis and Exploitation Team
AF DCGS	Air Force Distributed Common Ground System
AI	artificial intelligence
AOC	Air Operations Center
ASAT	anti-satellite
ATO	authorization to operate
CDAO	Chief Data and AI Officer
CIO	Chief Information Officer
CIP	common intelligence picture
COP	common operational picture
CRM	collection requirements management
CV	computer vision
DAF	Department of the Air Force
DAGR	Defense AI Guide on Risk
DGS	Distributed Ground Station
DoD	U.S. Department of Defense
DOF	disposition of forces
EEI	essential elements of information
ES	expert system
F2T2EA	find, fix, track, target, engage, assess
GEOINT	geospatial intelligence
HUMINT	human intelligence
IC	intelligence community
INT	intelligence domain
ISR	intelligence, surveillance, and reconnaissance
ISRD	Intelligence, Surveillance, and Reconnaissance Division
LLM	large language model
MASINT	measurement and signatures intelligence
ML	machine learning
NLP	natural language processing
OBP	object-based production
OSINT	open-source intelligence
OTHT	over-the-horizon targeting

P/C	prediction/classification
RAI	responsible artificial intelligence
SHIELD	Set Foundations, Hone Operationalizations, Improve and Innovate, Evaluate Status, Log for Traceability, and Detect via continuous Monitoring
SIGINT	signals intelligence
SME	subject-matter expert
STOPES	Social, Technological, Operational, Political, Economic, and Sustainability
TEVV	test, evaluation, verification, and validation
USAF	U.S. Air Force
USSF	U.S. Space Force

References

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb, *Power and Prediction: The Disruptive Economics of Artificial Intelligence*, Harvard Business Review Press, 2022.

Agrawal, Ajay, Joshua S. Gans, and Avi Goldfarb, "Artificial Intelligence Adoption and System-Wide Change," *Journal of Economics and Management Strategy*, Vol. 33, No. 2, Summer 2024.

Air Combat Command Manual 14-422, *Intelligence Data Governance*, Air Combat Command, October 24, 2023.

Air Force Doctrine Publication 2-0, *Intelligence*, U.S. Air Force, June 1, 2023.

Air Force Manpower Analysis Agency, homepage, undated. As of January 9, 2025: https://www.afmaa.af.mil

Ali, Sajid, Tamer Abuhmed, Shaker El-Sappagh, Khan Muhammad, Jose M. Alonso-Moral, Roberto Confalonieri, Riccardo Guidotti, Javier Del Ser, Natalia Díaz-Rodríguez, and Francisco Herrera, "Explainable Artificial Intelligence (XAI): What We Know and What Is Left to Attain Trustworthy Artificial Intelligence," *Information Fusion*, Vol. 99, November 2023.

Alkire, Brien, Abbie Tingstad, Dale Benedetti, Amado Cordova, Irina Danescu, William Fry, D. Scott George, Lawrence M. Hanser, Lance Menthe, Erik Nemeth, David Ochmanek, Julia Pollak, Jessie Riposo, Timothy Smith, and Alexander Stephenson, *Leveraging the Past to Prepare for the Future of Air Force Intelligence Analysis*, RAND Corporation, RR-1330-AF, 2016. As of January 6, 2025: https://www.rand.org/pubs/research_reports/RR1330.html

Amazon Web Services, "What Is RAG (Retrieval-Augmented Generation)?" webpage, undated. As of August 11, 2024: https://aws.amazon.com/what-is/retrieval-augmented-generation/

Ancona, Deborah, "Sensemaking: Framing and Acting in the Unknown," in Scott Snook, Nitin Nohria, and Rakesh Khurana, eds., *The Handbook for Teaching Leadership: Knowing, Doing, and Being*, SAGE Publications, 2011.

Banerjee, Imon, Kamanasish Bhattacharjee, John L. Burns, Hari Trivedi, Saptarshi Purkayastha, Laleh Seyyed-Kalantari, Bhavik N. Patel, Rakesh Shiradkar, and Judy Gichoya, "'Shortcuts' Causing Bias in Radiology Artificial Intelligence: Causes, Evaluation, and Mitigation," *Journal of the American College of Radiology*, Vol. 20, No. 9, September 2023.

Bayram, Firas, Bestoun S. Ahmed, and Andreas Kassler, "From Concept Drift to Model Degradation: An Overview on Performance-Aware Drift Detectors," *Knowledge-Based Systems*, Vol. 245, June 7, 2022.

Borukhovich, Kelly, and Tyler Morton, "DCGS Next Generation: Accelerating Change to Deliver Decision Advantage," *Over the Horizon*, September 26, 2020.

Brown, Jason M., and David Vernal, "Time-Dominant Fusion in a Complex World," *Trajectory*, November 11, 2014.

Cohen, Wesley M., and Daniel A. Levinthal, "Absorptive Capacity: A New Perspective on Learning and Innovation," *Administrative Science Quarterly*, Vol. 35, No. 1, March 1990.

Conlisk, Sarah, "How Lyft Uses AI to Get You Where You Want to Go, Faster," Lyft, August 10, 2023.

Costa, Lisa A., "Responsible Adoption of Generative AI (GenAI) and Large Language Models (LLMs) Within the United States Space Force (USSF)," memorandum to the Guardian workforce, U.S. Department of the Air Force, September 29, 2023.

Cubric, Marija, "Drivers, Barriers and Social Considerations for AI Adoption in Business and Management: A Tertiary Study," *Technology in Society*, Vol. 62, August 2020.

DAF—*See* Department of the Air Force.

Davenport, Thomas H., and Rajeev Ronanki, "Artificial Intelligence for the Real World," *Harvard Business Review*, January 1, 2018.

Department of Defense Instruction 8510.01, *Risk Management Framework for DoD Systems*, U.S. Department of Defense, July 19, 2022.

Department of the Air Force, *Chief Information Officer Public Strategy: FY2023–FY 2028*, September 30, 2022.

DoD—*See* U.S. Department of Defense.

Doubleday, Justin, "DoD CIO Looks Ahead to 'JWCC 2.0' and Next Steps for Cloud in 2024," Federal News Network, December 14, 2023.

Feldman, Aiden, "Taking the ATO Process from 6 Months to 30 Days," 18F, July 19, 2018.

Fetter, Jordyn, "Project NEXUS: Empowering the Air Force's Digital Talent," Joint Base San Antonio, November 19, 2019.

Fiore, Alexander, "Deterrent and Defensive Applications of Orbital Antisatellite Weapons," *Æther: A Journal of Strategic Airpower & Spacepower*, Vol. 2, Winter 2023.

Flory, Marta N., Sandy Napel, and Emily B. Tsai, "Artificial Intelligence in Radiology: Opportunities and Challenges," *Seminars in Ultrasound, CT and MRI*, Vol. 45, No. 2, April 2024.

Gandhi, Shripal, "Case Study: How Uber Uses AI to Optimize Surge Pricing," Hyperscale Business Newsletter, LinkedIn, June 10, 2024.

Geist, Edward, *Deterrence Under Uncertainty: Artificial Intelligence and Nuclear Warfare*, Oxford University Press, 2023.

George, Alexander L., *The Intelligence Value of Content Analysis: A Preliminary Progress Report*, RAND Corporation, RM-116, 1949. As of February 17, 2025: https://www.rand.org/pubs/research_memoranda/RM116.html

Gruber, Thomas R., "Toward Principles for the Design of Ontologies Used for Knowledge Sharing," *International Journal Human-Computer Studies*, Vol. 43, Nos. 5–6, November 1995.

Haley, Jaylan M., "Putting the Right Man in the Loop: Views Intelligence, Surveillance, and Reconnaissance Tactical Controllers," *Air & Space Power Journal*, Vol. 31, No. 1, Spring 2017.

Harman, Jane, Eric Edelman, John M. Keane, Thomas G. Mahnken, Mara Rudman, Mariah Sixkiller, Alissa Starzak, and Roger Zakheim, *Commission on the National Defense Strategy*, RAND Corporation, MS-A3057-4, 2024. As of August 9, 2024: https://www.rand.org/nsrd/projects/NDS-commission.html

Harpley, Unshin Lee, "Space Force CTIO: AI Will Be 'Game-Changer' for Operational Space," *Air & Space Forces Magazine*, November 14, 2023.

Hattori, Yuki, and Isabel Drost-Fromm, *DevOps Unleashed with Git and GitHub: Automate, Collaborate, and Innovate to Enhance Your DevOps Workflow and Development Experience*, Packt Publishing, 2024.

Heitzenrater, Chad, Bradley Wilson, Sarah W. Denton, James Ryseff, and Jeffery Broughton, *Lessons Learned from the Algorithmic Warfare Cross Functional Team: Project Maven*, RAND Corporation, 2024, Not available to the general public.

Helmer, Olaf, *Analysis of the Future: The Delphi Method*, RAND Corporation, P-3558, 1967. As of June 23, 2024: https://www.rand.org/pubs/papers/P3558.html

Heuer, Richards J., Jr., *Psychology of Intelligence Analysis*, Center for the Study of Intelligence, Central Intelligence Agency, 1999.

Holey, Elizabeth A., Jennifer L. Feely, John Dixon, and Vicki J. Whittaker, "An Exploration of the Use of Simple Statistics to Measure Consensus and Stability in Delphi Studies," *BMC Medical Research Methodology*, Vol. 7, No. 52, 2007.

IDERA, "Scaffolding," webpage, undated. As of August 21, 2024: https://www.idera.com/glossary/scaffolding/

Johnson, M. K., Michael M. Hanna, M. V. Clemens-Sewall, and D. P. Staheli, "Appendix 7. Personas List and Descriptions," webpage, *Responsible AI Toolkit (RAI Toolkit 1.0)*, Responsible AI, U.S. Department of Defense, undated-a. As of July 31, 2024: https://rai.tradewindai.com/appendix/personas-list-and-description

Johnson, M. K., Michael M. Hanna, M. V. Clemens-Sewall, and D. P. Staheli, "RAI Toolkit Executive Summary," webpage, *Responsible AI Toolkit (RAI Toolkit 1.0)*, Responsible AI, U.S. Department of Defense, undated-b. As of August 11, 2024: https://rai.tradewindai.com/executive-summary

Johnson, M. K., Michael M. Hanna, M. V. Clemens-Sewall, and D. P. Staheli, *Responsible AI Toolkit (RAI Toolkit 1.0)*, Responsible AI, U.S. Department of Defense, undated-c. As of February 3, 2025: https://rai.tradewindai.com

Johnston, Catherine, Elmo C. Wright, Jr., Jessica Bice, Jennifer Almendarez, and Linwood Creekmore, "Transforming Defense Analysis," *Joint Force Quarterly*, Vol. 79, October 2015.

Joint Chiefs of Staff, *DoD Dictionary of Military and Associated Terms*, U.S. Department of Defense, November 2021.

LENS Corporation, "How AI Is Optimizing Your Taxi Rides," Stack the Tech Newsletter, LinkedIn, May 24, 2024.

Liew, Charlene, "The Future of Radiology Augmented with Artificial Intelligence: A Strategy for Success," *European Journal of Radiology*, Vol. 102, May 2018.

Lin, Chinho, Bertram Tan, and Shofang Chang, "The Critical Factors for Technology Absorptive Capacity," *Industrial Management & Data Systems*, Vol. 102, No. 6, 2002.

Lingel, Sherrill, Jeff Hagen, Eric Hastings, Mary Lee, Matthew Sargent, Matthew Walsh, Li Ang Zhang, and David Blancett, *Joint All-Domain Command and Control for Modern Warfare: An Analytic Framework for Identifying and Developing Artificial Intelligence Applications*, RAND Corporation, RR-4408/1-AF, 2020. As of January 6, 2025: https://www.rand.org/pubs/research_reports/RR4408z1.html

Magnuson, Stew, "Military 'Swimming in Sensors and Drowning in Data,'" *National Defense*, January 1, 2010.

Makarius, Erin E., Debmalya Mukherjee, Joseph D. Fox, and Alexa K. Fox, "Rising with the Machines: A Sociotechnical Framework for Bringing Artificial Intelligence into the Organization," *Journal of Business Research*, Vol. 120, November 2020.

Marco-Ruiz, Luis, Miguel Ángel Tejedor Hernández, Phuong Dinh Ngo, Alexandra Makhlysheva, Therese Olsen Svenning, Kari Dyb, Taridzo Chomutare, Carlos Fernández Llatas, Jorge Muñoz-Gama, and Maryam Tayefi, "A Multinational Study on Artificial Intelligence Adoption: Clinical Implementers' Perspectives," *International Journal of Medical Informatics*, Vol. 184, April 2024.

Mathews, Anshika, "How Uber's Predictive Machine Learning Is Changing User Experience?" AIM Research, July 8, 2024.

Matsumura, John, Lance Menthe, Henry Hargrove, Casidhe Hutchison, Bridget R. Kane, Philip Song, Joshua Steier, Anton Wu, and Elie Alhajjar, *Mitigating the Impact of Future Adversarial Attacks on Artificial Intelligence (AI) Applications for the Army*, RAND Corporation, 2024, Not available to the general public.

McKeown, David W., "Continuous Authorization to Operate (cATO)," memorandum to senior Pentagon leadership and Defense Agency and DoD Field Activity Directors, Office of the Secretary of Defense, February 3, 2022.

Menthe, Lance, Dahlia Anne Goldfeld, Abbie Tingstad, Sherrill Lingel, Edward Geist, Donald Brunk, Amanda Wicker, Sarah Soliman, Balys Gintautas, Anne Stickells, and Amado Cordova, *Technology Innovation and the Future of Air Force Intelligence Analysis*: Vol. 1, *Findings and Recommendations*, RAND Corporation, RR-A341-1, 2021a. As of January 6, 2025: https://www.rand.org/pubs/research_reports/RRA341-1.html

Menthe, Lance, Dahlia Anne Goldfeld, Abbie Tingstad, Sherrill Lingel, Edward Geist, Donald Brunk, Amanda Wicker, Sarah Soliman, Balys Gintautas, Anne Stickells, and Amado Cordova, *Technology Innovation and the Future of Air Force Intelligence Analysis*: Vol. 2, *Technical Analysis and Supporting Material*, RAND Corporation, RR-A341-2, 2021b. As of January 6, 2025: https://www.rand.org/pubs/research_reports/RRA341-2.html

Menthe, Lance, Li Ang Zhang, Edward Geist, Joshua Steier, Aaron B. Frank, Erik Van Hegewald, Gary J. Briggs, Keller Scholl, Yusuf Ashpari, and Anthony Jacques, *Understanding the Limits of Artificial Intelligence for Warfighters*: Vol. 1, *Summary*, RAND Corporation, RR-A1722-1, 2024. As of January 6, 2025: https://www.rand.org/pubs/research_reports/RRA1722-1.html

Messeri, Lisa, and M. J. Crockett, "Artificial Intelligence and Illusions of Understanding in Scientific Research," *Nature*, Vol. 627, March 7, 2024.

Minsky, Marvin, ed., *Semantic Information Processing*, MIT Press, 1969.

Monteith, Scott, Tasha Glenn, John R. Geddes, Eric D. Achtyes, Peter C. Whybrow, and Michael Bauer, "Differences Between Human and Artificial/Augmented Intelligence in Medicine," *Computers in Human Behavior: Artificial Humans*, Vol. 2, No. 2, August–December 2024.

Mourtzis, D., V. Zogopoulos, and E. Vlachou, "Augmented Reality Application to Support Remote Maintenance as a Service in the Robotics Industry," *Procedia CIRP*, Vol. 63, 2017.

Murel, Jacob, and Eda Kavlakoglu, "What Is Ensemble Learning?" webpage, IBM, March 18, 2024. As of August 21, 2024:
https://www.ibm.com/topics/ensemble-learning

National Institute of Standards and Technology, *Managing Information Security Risk: Organization, Mission, and Information System View*, U.S. Department of Commerce, NIST Special Publication 800-39, March 2011.

Noblet, Jean-Pierre, Eric Simon, and Robert Parent, "Absorptive Capacity: A Proposed Operationalization," *Knowledge Management Research & Practice*, Vol. 9, No. 4, 2011.

Odilov, Sherzod, "Here's How Leaders Can Manage AI Fatigue," *Forbes*, February 14, 2024.

Office of the Director of National Intelligence, "How the IC Works," webpage, undated. As of January 6, 2025:
https://www.intelligence.gov/how-the-ic-works

Pérez, Carlos J., and Carlos J. Ponce, "Disruption Costs, Learning by Doing, and Technology Adoption," *International Journal of Industrial Organization*, Vol. 41, July 2015.

Pernin, Christopher G., Louis R. Moore, and Katherine Comanor, *The Knowledge Matrix Approach to Intelligence Fusion*, RAND Corporation, TR-416-A, 2007. As of January 7, 2025:
https://www.rand.org/pubs/technical_reports/TR416.html

Porche, Isaac R. III, Bradley Wilson, Erin-Elizabeth Johnson, Shane Tierney, and Evan Saltzman, *Data Flood: Helping the Navy Address the Rising Tide of Sensor Information*, RAND Corporation, RR-315-NAVY, 2014. As of October 7, 2024:
https://www.rand.org/pubs/research_reports/RR315.html

Rosenthal, Brian M., "'They Were Conned': How Reckless Loans Devastated a Generation of Taxi Drivers," *New York Times*, May 19, 2019.

Rumsfeld, Donald H., "DoD News Briefing," transcript of news briefing delivered at the Pentagon, February 12, 2002.

Russell, Stuart, and Peter Norvig, *Artificial Intelligence: A Modern Approach,* 4th ed., Pearson, 2021.

Satell, Greg, "A Look Back at Why Blockbuster Really Failed and Why It Didn't Have To," *Forbes*, September 5, 2014.

Sayler, Kelley M., *DOD Replicator Initiative: Background and Issues for Congress*, Congressional Research Service, IF12611, updated March 22, 2024.

Schmid, Jon, Bonnie L. Triezenberg, James Dimarogonas, and Samuel Absher, *The Role of Standards in Fostering Capability Evolution: Does Design Matter? Insights from Interoperability Standards*, RAND Corporation, RR-A1576-1, 2022. As of January 6, 2025:
https://www.rand.org/pubs/research_reports/RRA1576-1.html

Sheikh, Haroon, Corien Prins, and Erik Schrijvers, *Mission AI: The New System Technology*, Springer Cham, 2023.

Shortliffe, Edward H., "Mycin: A Knowledge-Based Computer Program Applied to Infectious Diseases," Annual Meeting of the Society for Computer Medicine, November 10, 1977.

Shumailov, Ilia, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson, and Yarin Gal, "AI Models Collapse When Trained on Recursively Generated Data," *Nature*, Vol. 631, July 25, 2024.

Sidwell-Bowron, Melissa, and Matthew Winot, "The Intelligence, Surveillance, Reconnaissance Liaison Officer: A Critical Intelligence Node in Agile Combat Operations," Air Land Sea Space Application Center, February 1, 2023.

Sogani, Julie, Bibb Allen, Jr., Keith Dreyer, and Geraldine McGinty, "Artificial Intelligence in Radiology: The Ecosystem Essential to Improving Patient Care," *Clinical Imaging*, Vol. 59, No. 1, January 2020.

Solaimani, Sam, and Lucas Swaak, "Critical Success Factors in a Multi-Stage Adoption of Artificial Intelligence: A Necessary Condition Analysis," *Journal of Engineering and Technology Management*, Vol. 69, July–September 2023.

Stebbins, David, Richard S. Girven, Timothy Parker, Thomas Deen, Brandon F. De Bruhl, James Ryseff, Jessica Welburn Paige, Annie Yu Kleiman, Sunny D. Bhatt, Éder M. Sousa, Marta Kepe, and Matthew Fay, *Exploring Artificial Intelligence Use to Mitigate Potential Human Bias Within U.S. Army Intelligence Preparation of the Battlefield Processes*, RAND Corporation, RR-A2763-1, 2024. As of August 9, 2024: https://www.rand.org/pubs/research_reports/RRA2763-1.html

Steeb, Randall, Stephanie Cammarata, Sanjai Narain, Jeff Rothenberg, and William Giarla, *Cooperative Intelligence for Remotely Piloted Vehicle Fleet Control: Analysis and Simulation*, RAND Corporation, R-3408-ARPA, 1986. As of January 6, 2025: https://www.rand.org/pubs/reports/R3408.html

Steier, Joshua, Erik Van Hegewald, Anthony Jacques, Gavin S. Hartnett, and Lance Menthe, *Understanding the Limits of Artificial Intelligence for Warfighters*: Vol. 2, *Distributional Shift in Cybersecurity Datasets*, RAND Corporation, RR-A1722-2, 2024. As of October 8, 2024: https://www.rand.org/pubs/research_reports/RRA1722-2.html

Straat, Michiel, Fthi Abadi, Zhuoyun Kan, Christina Göpfert, Barbara Hammer, and Michael Biehl, "Supervised Learning in the Presence of Concept Drift: A Modelling Framework," *Neural Computing and Applications*, Vol. 34, 2022.

Strohm, Lea, Charisma Hehakaya, Erik R. Ranschaert, Wouter P. C. Boon, and Ellen H. M. Moors, "Implementation of Artificial Intelligence (AI) Applications in Radiology: Hindering and Facilitating Factors," *European Radiology*, Vol. 30, No. 10, October 2020.

Tingstad, Abbie, Dahlia Anne Goldfeld, Lance Menthe, Robert A. Guffey, Zachary Haldeman, Krista Langeland, Amado Cordova, Elizabeth M. Waina, and Balys Gintautas, *Assessing the Value of Intelligence Collected by U.S. Air Force Airborne Intelligence, Surveillance, and Reconnaissance Platforms*, RAND Corporation, RR-2742-AF, 2021. As of January 6, 2025: https://www.rand.org/pubs/research_reports/RR2742.html

Toukan, Mark, Jair Aguirre, Sean Mann, and Eddie Ro, *Lessons Learned from Integrating a Computational Cognitive Model for Personalized Linguist Training*, RAND Corporation, RR-A2454-1, 2024. As of January 6, 2025: https://www.rand.org/pubs/research_reports/RRA2454-1.html

Tursunbayeva, Aizhan, and Hila Chalutz-Ben Gal, "Adoption of Artificial Intelligence: A TOP Framework-Based Checklist for Digital Leaders," *Business Horizons*, Vol. 67, No. 4, July–August 2024.

USAF—*See* U.S. Air Force.

U.S. Air Force, *Sensing Grid: Operational Framework*, June 2020.

U.S. Army Training and Doctrine Command, *The Operational Environment 2024–2034: Large-Scale Combat Operations*, TRADOC Pamphlet 525-92, December 2024.

U.S. Department of Defense, *Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance Our Security and Prosperity*, 2019.

U.S. Department of Defense, "DOD Adopts Ethical Principles for Artificial Intelligence," press release, February 24, 2020.

U.S. Department of Defense, *Data, Analytics, and Artificial Intelligence Adoption Strategy: Accelerating Decision Advantage*, June 27, 2023.

U.S. Department of Defense Intelligence and Security Professional Certification, "Collection Management Professional Certification," webpage, undated. As of August 11, 2024: https://dodcertpmo.defense.gov/CollMgmt/

U.S. Department of Defense Responsible AI Working Council, *U.S. Department of Defense Responsible Artificial Intelligence Strategy and Implementation Pathway*, June 2022.

USSF—*See* U.S. Space Force.

U.S. Space Force Peterson and Schriever Space Force Base, "Space Delta 7 Fact Sheet," fact sheet, October 2023.

Vincent, Brandi, "Pentagon Issues New Guidance to Address Industry Gripes About ATO Process," DefenseScoop, May 8, 2024.

Wall, Andru E., "Demystifying the Title 10-Title 50 Debate: Distinguishing Military Operations, Intelligence Activities & Covert Action," *Harvard Law School National Security Journal*, Vol. 3, 2011.

Walsh, Matthew, Lance Menthe, Edward Geist, Eric Hastings, Joshua Kerrigan, Jasmin Léveillé, Joshua Margolis, Nicholas Martin, and Brian P. Donnelly, *Exploring the Feasibility and Utility of Machine Learning-Assisted Command and Control*: Vol. 1, *Findings and Recommendations*, RAND Corporation, RR-A263-1, 2021a. As of January 6, 2025: https://www.rand.org/pubs/research_reports/RRA263-1.html

Walsh, Matthew, Lance Menthe, Edward Geist, Eric Hastings, Joshua Kerrigan, Jasmin Léveillé, Joshua Margolis, Nicholas Martin, and Brian P. Donnelly, *Exploring the Feasibility and Utility of Machine Learning-Assisted Command and Control*: Vol. 2, *Supporting Technical Analysis*, RAND Corporation, RR-A263-2, 2021b. As of January 7, 2025: https://www.rand.org/pubs/research_reports/RRA263-2.html

Yonekura, Emmi, Mark Toukan, Timothy Marler, Andrea M. Abler, Henry Hargrove, Eddie Ro, Isabelle Winston, and Sankalp Kumar, *Accelerating the Transfer of Training Technologies to Support Evolving Department of the Air Force Mission Capabilities: A Framework, Lessons Learned, and Recommendations*, RAND Corporation, RR-A2326-2, 2024. As of January 7, 2025: https://www.rand.org/pubs/research_reports/RRA2326-2.html

Young, Nora, "What AI Can and Can't Do," *Spark* podcast, November 10, 2022.

Zahra, Shaker A., and Gerard George, "Absorptive Capacity: A Review, Reconceptualization, and Extension," *Academy of Management Review*, Vol. 27, No. 2, April 2002.

Zhang, Li Ang, Lance Menthe, Ian Fleischmann, Sale Lilly, Joshua Kerrigan, Michael J. Gaines, and Gregory A. Schumacher, *Incorporating Artificial Intelligence into Army Intelligence Processes*, RAND Corporation, 2021, Not available to the general public.

RAND PROJECT AIR FORCE

It is widely expected that artificial intelligence (AI) will play a critical role in future military operations. As part of the Department of the Air Force (DAF)'s efforts to incorporate emerging technology into modern warfare operations, RAND researchers were tasked with studying the data, technologies, processes, and policies that the DAF will need to enable effective sense-making in the next decade. To advance the understanding of how these elements intersect with the current state of technology, RAND researchers identified challenges in the current sense-making processes and opportunities to overcome them. The effort was to focus on how sense-making occurs—where, with what, and by whom—with a particular emphasis on how information from multiple intelligence domains can be fused to find, fix, and track targets.

In this report, RAND researchers identify the most significant sense-making challenges facing the DAF and assess how AI capabilities could address these challenges. RAND researchers also provide insights for adoption through a comparative AI adoption schema and conduct a systematic examination of risk on a notional AI system, showcasing requisite considerations on how best to implement these insights. AI capabilities and DAF sense-making processes are not simple. Syncing these processes requires careful consideration of how decisionmakers will use these methods and how they are integrated into the larger intelligence cycle.

$36.00

ISBN-10 1-9774-1499-0
ISBN-13 978-1-9774-1499-1

53600

9 781977 414991

www.rand.org

RR-A3152-1