2021-12

# EVALUATING ARTIFICIAL INTELLIGENCE METHODS FOR USE IN KILL CHAIN FUNCTIONS

Burns, Gregory R.; Collier, Ryan T.; Cornish, Richard J.; Curley, Kyle J.; Freeman, Allan; Spears, Jared

Monterey, CA; Naval Postgraduate School


# NAVAL POSTGRADUATE SCHOOL

**MONTEREY, CALIFORNIA**

# SYSTEMS ENGINEERING CAPSTONE REPORT

**EVALUATING ARTIFICIAL INTELLIGENCE METHODS FOR USE IN KILL CHAIN FUNCTIONS**

by

Gregory R. Burns, Ryan T. Collier, Richard J. Cornish,
Kyle J. Curley, Allan Freeman, and Jared Spears

December 2021

Advisor: Bonnie W. Johnson
Co-Advisor: John M. Green

THIS PAGE INTENTIONALLY LEFT BLANK

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



| **1. AGENCY USE ONLY** *(Leave blank)* | **2. REPORT DATE** December 2021 | **3. REPORT TYPE AND DATES COVERED** Systems Engineering Capstone Report |
|---|---|---|

| **4. TITLE AND SUBTITLE** EVALUATING ARTIFICIAL INTELLIGENCE METHODS FOR USE IN KILL CHAIN FUNCTIONS | **5. FUNDING NUMBERS** |
|---|---|
| **6. AUTHOR(S)** Gregory R. Burns, Ryan T. Collier, Richard J. Cornish, Kyle J. Curley, Allan Freeman, and Jared Spears | |
| **7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)** Naval Postgraduate School Monterey, CA 93943-5000 | **8. PERFORMING ORGANIZATION REPORT NUMBER** |
| **9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)** N/A | **10. SPONSORING / MONITORING AGENCY REPORT NUMBER** |

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

| **12a. DISTRIBUTION / AVAILABILITY STATEMENT** Approved for public release. Distribution is unlimited. | **12b. DISTRIBUTION CODE** A |
|---|---|

**13. ABSTRACT (maximum 200 words)**

Current naval operations require sailors to make time-critical and high-stakes decisions based on uncertain situational knowledge in dynamic operational environments. Recent tragic events have resulted in unnecessary casualties, and they represent the decision complexity involved in naval operations and specifically highlight challenges within the OODA loop (Observe, Orient, Decide, and Assess). Kill chain decisions involving the use of weapon systems are a particularly stressing category within the OODA loop—with unexpected threats that are difficult to identify with certainty, shortened decision reaction times, and lethal consequences. An effective kill chain requires the proper setup and employment of shipboard sensors; the identification and classification of unknown contacts; the analysis of contact intentions based on kinematics and intelligence; an awareness of the environment; and decision analysis and resource selection. This project explored the use of automation and artificial intelligence (AI) to improve naval kill chain decisions. The team studied naval kill chain functions and developed specific evaluation criteria for each function for determining the efficacy of specific AI methods. The team identified and studied AI methods and applied the evaluation criteria to map specific AI methods to specific kill chain functions.

| **14. SUBJECT TERMS** kill chain, OODA, mapping, artificial intelligence, AI, machine learning, ML, decision aids, air and missile defense, AMD, model-based systems engineering, MBSE | **15. NUMBER OF PAGES** 219 |
|---|---|
| | **16. PRICE CODE** |

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

NSN 7540-01-280-5500

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

THIS PAGE INTENTIONALLY LEFT BLANK

**Approved for public release. Distribution is unlimited.**

# EVALUATING ARTIFICIAL INTELLIGENCE METHODS FOR USE IN KILL CHAIN FUNCTIONS

Gregory R. Burns, Ryan T. Collier, Richard J. Cornish,
Capt Kyle J. Curley (USMC), Allan Freeman, and Jared Spears

Submitted in partial fulfillment of the
requirements for the degrees of

**MASTER OF SCIENCE IN SYSTEMS ENGINEERING**

and

**MASTER OF SCIENCE IN ENGINEERING SYSTEMS**

from the

**NAVAL POSTGRADUATE SCHOOL**
**December 2021**

Lead Editor: Jared Spears

Reviewed by:
Bonnie W. Johnson
Advisor

John M. Green
Co-Advisor

Accepted by:
Oleg A. Yakimenko
Chair, Department of Systems Engineering

THIS PAGE INTENTIONALLY LEFT BLANK

# ABSTRACT

Current naval operations require sailors to make time-critical and high-stakes decisions based on uncertain situational knowledge in dynamic operational environments. Recent tragic events have resulted in unnecessary casualties, and they represent the decision complexity involved in naval operations and specifically highlight challenges within the OODA loop (Observe, Orient, Decide, and Assess). Kill chain decisions involving the use of weapon systems are a particularly stressing category within the OODA loop—with unexpected threats that are difficult to identify with certainty, shortened decision reaction times, and lethal consequences. An effective kill chain requires the proper setup and employment of shipboard sensors; the identification and classification of unknown contacts; the analysis of contact intentions based on kinematics and intelligence; an awareness of the environment; and decision analysis and resource selection. This project explored the use of automation and artificial intelligence (AI) to improve naval kill chain decisions. The team studied naval kill chain functions and developed specific evaluation criteria for each function for determining the efficacy of specific AI methods. The team identified and studied AI methods and applied the evaluation criteria to map specific AI methods to specific kill chain functions.

THIS PAGE INTENTIONALLY LEFT BLANK

# TABLE OF CONTENTS

# LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

# LIST OF TABLES

# LIST OF ACRONYMS AND ABBREVIATIONS

| | |
|---|---|
| AI | artificial intelligence |
| ALSA | air land sea application |
| AMD | air and missile defense |
| AN | Army Navy |
| ATC | air traffic control |
| ATD | automatic target detection |
| AW | air warfare |
| BDA | battle damage assessment |
| C2 | command and control |
| C3 | command, control, and communications |
| CART | classification and regression tree |
| CBS | Columbia Broadcasting System |
| CDE | collateral damage estimate |
| CIC | combat information center |
| CJCS | Chairman of the Joint Chiefs of Staff |
| COA | courses of action |
| DARPA | Defense Advanced Research Projects Agency |
| DNA | Dam Neck Activity |
| DOD | Department of Defense |
| DODAF | Department of Defense architecture framework |
| DODD | Department of Defense Directive |
| e.g. | exempli gratia, meaning “for example” |
| ELINT | electronic intelligence |
| et al. | Latin for “and others” |
| etc. | et cetera; meaning “and so on and so forth” |
| F2T2EA | find, fix, track, target, engage, assess |
| FN | false negative |
| FP | false positive |

| | |
|---|---|
| GAN | generative adversarial network |
| HDD | hard drive disk |
| HED | hierarchical event descriptor |
| HIL | human-in-the-loop |
| HPT | high-payoff target |
| HVT | high value target |
| IBM | International Business Machines Corporation |
| ID | identification |
| IoT | internet of things |
| IPR | In-progress review |
| ISEA | In-Service Engineering Agent |
| JAIC | Joint Artificial Intelligence Center |
| JCS | Joint Chiefs of Staff |
| JFC | Joint Forces Commander |
| JIPTL | joint integrated prioritized target list |
| JP | joint publication |
| JTCB | Joint Targeting Coordination Board |
| JTL | joint task and/or targeting list |
| KNN | K-nearest neighbors |
| LAWS | lethal autonomous weapon systems |
| LOF | lack of friends |
| M&S | modeling and simulation |
| MBSE | model-based systems engineering |
| MCC | Matthew's correlation coefficient |
| MCTS | Monte Carlo tree search |
| METOC | meteorological and oceanographic |
| ML | machine learning |
| MSE | mean squared error |
| MV ACX | multi-view advanced combat experimental |
| n.d. | no date |

| | |
|---|---|
| NATO | North Atlantic Treaty Organization |
| NATO | North Atlantic Treaty Organization |
| NCTI/R | non-cooperative target identification and/or recognition |
| NN | neural networks |
| NPS | Naval Postgraduate School |
| NSWCDD | NAVSEA Warfare Center Division Dahlgren |
| NTSB | National Transportation Safety Board |
| NWSC | Naval Surface Warfare Center |
| OODA | observe, orient, decide, and assess |
| OTA | other transaction authority |
| OV-1 | operation viewpoint |
| OV-5a | operational activity decomposition tree |
| OV-6c | event trace description |
| PCA | principal component analysis |
| PDS | passive detection system |
| PID | positive or procedural identification |
| POAM | plan of action and milestones |
| PPLI | precise participant location and identification |
| PTL | prioritized target list |
| RAI | responsible artificial intelligence |
| RCA | root cause analysis |
| ROE | rules of engagement |
| SCP | Strategic Computing Program |
| SE | systems engineering |
| SI3-CMD | serial interactions in imperfect games applied to complex military Decision-making |
| SIF | selective identification feature |
| SIGINT | signal intelligence |
| SoS | system of systems |
| SSDS | ship self-defense system |
| SVM | support vector machines |

| | |
|---|---|
| TA | target acquisition |
| TACSITs | tactical situations |
| TAO | Tactical Action Officer |
| TBM | theater ballistic missile |
| TN | true negative |
| TP | true positive |
| TPO | thesis processing office |
| TST | time-sensitive target |
| TTP | tactics, techniques, and procedures |
| TV | television |
| UAVs | unmanned aerial vehicles |
| USN | United States Navy |
| USS | United States ship |
| V&V | verification and validation |
| VID | visual identification |
| w.r.t. | with respect to |
| WTPs | weapon target pairs |
| XAI | explainable artificial intelligence |

# EXECUTIVE SUMMARY

Current naval operations are typically fast-paced, critical, and have high-stakes decisions to be made which are, at times, based on uncertain knowledge in very dynamic theaters of operation. Many examples highlight the need for increased effectiveness in decision making and a need to lessen the load of watch teams. Examples of where the above was lacking include the USS *Fitzgerald* (DDG 62) and MV ACX Crystal collision in 2017 to the grounding of the USS Port Royal (CG 73) in 2009. Some root causes were inexperience, fatigue, and stress of personnel involved.

The above accidents showcase the difficulty of military operations and demonstrate the challenges within the OODA (Observe, Orient, Decide, and Assess) loop (Jones et al. 2020). Human error, human cognitive limits, and the inherent decision complexity of naval operations lead to challenges in the OODA loop and more specifically in the kill chain process.

The modern battle space is composed of a large amount of data from multiple domains such as conventional land, air, and sea but also from space and cyberspace. Decision makers have many considerations to account for ranging from rules of engagement (ROE), weapons to be used, sensors, and evaluation of intent. The find, fix, track, target, engage, assess (F2T2EA) kill chain model alleviates some of the difficulty with this process (Joint Chiefs of Staff 2013). Artificial intelligence (AI) and machine learning (ML) can aid naval kill chain decisions in the tactical domain by mapping AI methods to the kill chain functions via an analysis of alternatives and the use of evaluation criteria. This is done in three phases spread throughout five chapters in this capstone report.

This report utilized hundreds of sources and primarily leveraged previous research conducted by the Naval Postgraduate School AI-OODA team in their Capstone Report (2020), "Leveraging Artificial Intelligence (AI) for Air and Missile Defense (AMD): An Outcome-Oriented Decision Aid." They combined their work with John Boyd's Observe, Orient, Decide and Act decision-making framework. As a preliminary step to their analysis, the AI-OODA team explicitly and tightly coupled specific OODA functions to specific

F2T2EA functions. This report, however, asserts that the OODA loop is a decision cycle that is nested within each of the functions of the kill-chain rather than mapping specifically to one or more kill-chain functions in a high or low stress scenario. This capstone team developed a set of 28 kill chain functions based on the F2T2EA model.

A good decision is hard to determine in developing evaluation criteria for mapping AI methods to the kill chain and is critical in decision assessment. In evaluating a decision, one must consider the state of knowledge awareness at the time of action selection along with explain-ability. Several methods of scoring a decision are used ranging from defining and prioritizing weapon-target pairs of interest to developing scoring criteria and reporting assessment findings for others to review.

Currently, the state of AI is vast and must be explained in order to understand AI applicability to functions in the kill chain. A high-level overview of selected AI methods is discussed in this report with a portion of the most popular methods highlighted. First, AI is difficult to define with no commonly accepted definition. Next, there is a difference in AI versus machine learning (ML). ML allows incremental gains in accuracy and predictability; AI takes in data and provides an output via an algorithm. The history of AI ranges from Alan Turing's enciphering machine in the 1940s to U.S. government use in the 1980s within the Strategic Computing Program to today in the Joint Artificial Intelligence Center (JAIC) with their five pillars for AI strategy ranging from a leading AI workforce to safety and ethics. The Defense Advanced Research Projects Agency (DARPA) has described the direction of AI in a 3-wave framework categorized by Handcrafted Knowledge (Wave 1), Statistical Learning (Wave 2), and Contextual Reasoning (Wave 3) within 1–4 dimensions of intelligence parameters' attributes (Launchbury 2017). These attributes include perceiving, reasoning, abstracting, and learning.

Artificial intelligence can involve supervised learning which can predict a result based on input values. There are several techniques for learning with supervised learning. Examples include linear regression and classification. Also, many numeric methods can analyze the effectiveness of the learning that took place such as F-score and Accuracy score. Artificial intelligence can also use unsupervised learning, which is a type of machine

learning that uses algorithms to discover data patterns or groupings in unlabeled/untagged data sets. Unsupervised learning is beneficial when analyzing unknown (the y's) responses to reveal patterns in the labeled (the x's) data. A famous example in the data analysis community is the Iris flower data set. Using only the labeled data, one can see that the responses cluster together and can determine that patterns exist in the response (the species of flower). Methods of unsupervised learning include clustering and K-means, but there are other methods. Reinforcement learning has an agent able to receive feedback from the environment and understand the basic goal. Also, there is a trade-off between exploration and exploitation as Sutton and Barto in (2018) explain. Finally, the Generational Adversarial Network (GAN) utilizes unsupervised learning and reinforcement learning and is typically used in Neural Networks (NN). Neural networks are an excellent source of machine learning algorithms that have an extreme number of inputs which, in turn, make for a lot of computation. NN's are good to use in simulations, natural language processing, game theory, and computer vision. NN's are simply a way to map inputs to outputs that allow for learning along the way. However, NN's can be described as a "black box" learning technique since it is hard to explain what is going on and an explainable AI (XAI) technique is often needed. The three main components of XAI are Explainable Models, the Explanation Interface, and the Psychology of Explanation (Gunning 2019). Data security must be considered along with "Big Data" which refers to unstructured, complex and large data sets and is characterized by the five v's: volume, velocity (the increase of the amount of data changing over time), variety, veracity, and value. Other theories include decision theory, fuzzy logic, and utility functions.

Using the above literature reviews, the team developed a framework for mapping AI/ML to the AMD (Air Missile Defense) kill chain. Four steps were taken: 1) establish the modeling framework, 2) identify decision points, 3) apply AI/ML methodologies, and 4) analyze findings. The team identified the following AI/ML methods for the kill chain mapping analysis: Linear Regression, Logistic Regression, Clustering, Association, Random Forest, Neural Networks, GAN, and Naïve Bayes. The evaluation criteria were called "decision points" and were posed as four questions: (1) what is the type of the required output, (2) what is the type of learning required, (3) what level of explainability

(XAI) is required, and (4) how many predictors are required? The team performed the mapping by evaluating each method for each kill chain function based on a set of decision points and a scoring process. A score of +1 was given for a method that was considered well suited to a task, 0 if the method was suited but sub optimal, and –1 if the method was not suited for the task.

The team conducted the mapping analysis, analyzing the AI methods according to the evaluation criteria (decision points) in relation to each of the 28 functions of the kill chain. The team used the scoring method to determine the best overall AI/ML score for each kill chain function. The team's mapping is shown in 0.

Table 1. Map of AI/ML Methods to the Kill Chain

| Step | Number | Function | AL/ML Method |
|---|---|---|---|
| Find | 1.1.1 | Initial Detection | Clustering |
| | 1.1.2 | Battle Damage Assessment (BDA) Detection | Clustering |
| | 1.1.3 | Re-Task Detection | Clustering |
| Fix | 1.2.1 | Define Target/Threat | Association |
| | 1.2.2 | Characterize | Clustering |
| | 1.2.3 | Classify | Logistic Regression, Association |
| | 1.2.4 | Identify | Logistic Regression, Association |
| | 1.2.5 | Locate | Clustering |
| | 1.2.6 | Disseminate Target /Threat | Association |
| Track | 1.3.1 | Generate / Update Track | Clustering |
| | 1.3.2 | Sort | Linear Regression |
| | 1.3.3 | Determine Target / Threat Urgency | Linear Regression |
| | 1.3.4 | Assess Blue Force Proximity | Association |
| | 1.3.5 | Validate Target / Threat | Association |
| Target | 1.4.1 | Nominate Engagement Option | Logistic Regression, Association |
| | 1.4.2 | Prioritize Target / Threat | Linear Regression |
| | 1.4.3 | Determine Time Available | Linear Regression |
| | 1.4.4 | Maintain Track | Clustering |
| | 1.4.5 | Select Attack Option | Logistic Regression, Association |
| | 1.4.6 | Verify Rules of Engagement (ROE) | Association |
| Engage | 1.5.1 | Issue Order | Association |
| | 1.5.2 | Attack Target / Threat | Linear Regression |
| | 1.5.3 | Track Weapon | Clustering |
| | 1.5.4 | Confirm Impact | Clustering |
| | 1.5.5 | Task Re-Attack | Linear Regression |
| Assess | 1.6.1 | Conduct Dynamic Assessment | Clustering |
| | 1.6.2 | Evaluate | Clustering |

The team's AI/ML mapping to kill chain functions provides two critical benefits to the DOD and Navy. First, the map, itself, is an important starting point and foundation for the design and development of AI-enabled tactical decision aids to support kill chain decisions. Secondly, the team's analytical process for mapping AI methods to the kill chain can be used for understanding the application of AI to many other military and non-military domains. The process of identifying appropriate AI methods, developing evaluation criteria and a scoring process, and laying out the functions of a process to conduct an analytical mapping, has far-reaching potential for supporting the engineering of many different AI-enabled systems.

## References


Gunning, David, and David Aha. 2019. "DARPA's Explainable Artificial Intelligence (XAI) Program." *AI Magazine* 40 (2): 44–58.

Joint Chiefs of Staff. 2013. *Joint Targeting (JP 3-60)*. https://www.justsecurity.org/wp-content/uploads/2015/06/Joint_Chiefs-Joint_Targeting_20130131.pdf.

———. 2019. *Joint Fire Support (JP 3-09)*. https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_09.pdf.

Jones, Julian I., II, Russell Kress, William J. Newmeyer Jr., and Adam I. Rahman. 2020. "Leveraging Artificial Intelligence (AI) for Air and Missile Defense (AMD): An Outcome-Oriented Decision Aid." Systems engineering capstone report, Naval Postgraduate School. http://hdl.handle.net/10945/66088.

Launchbury, John. 2017. "A DARPA Perspective on Artificial Intelligence." https://www.darpa.mil/about-us/darpa-perspective-on-ai.

Sutton, Richard S., and Andrew G. Barto. 2018. *Reinforcement Learning, Second Edition: An Introduction*. MIT Press.

THIS PAGE INTENTIONALLY LEFT BLANK

# I. INTRODUCTION

## A. STATEMENT OF NEED

Current naval operations require sailors to make time critical and high-stakes decisions based on uncertain situational knowledge in dynamic operational environments. Recent tragic events have resulted in unnecessary casualties. They highlight the need to increase the efficacy of naval decision processes, while lessening the burden on watch standers and watch teams in the Combat Information Center (CIC) and the Bridge. The following four events illustrate the need for improvements to naval operational decision processes.

### 1. USS Fitzgerald and MV ACX Crystal Collision

**Incident:** On June 17, 2017, a United States Navy destroyer (USS Fitzgerald, DDG 62) was involved in a collision with a Philippine-flagged container ship (MV ACX Crystal) off the coast of Japan as shown in Figure 1. The resulting collision led to the flooding of berthing areas of the ship leading to seven deaths, multiple Injuries and approximately $523 million in repair costs (Mizokami 2020).

Figure 1. Damage from the USS Fitzgerald Collision with ACX Crystal Container Ship. Source: USNI (2017).

**Root Cause:** The National Transportation Safety Board (NTSB) Marine Accident Report (2020, sec. Abstract) on the accident identified the following "safety issues: Fitzgerald crew's insufficient training, Fitzgerald crew's fatigue, U.S. Navy practice of not broadcasting automatic identification signals, Fitzgerald commanding officer not augmenting bridge watch standing personnel with a more experienced officer while the vessel was crossing busy coastal traffic route and Fitzgerald commanding officer not adequately assessing the hazard presented by the vessel's intended transit."

### 2. USS John S. McCain and Alnic MC Collision

**Incident:** On August 21, 2017, a United Stated Navy destroyer (USS John S. McCain, DDG 56) was involved in a collision with a Liberian-flagged tanker (Alnic MC) of the coast of Singapore and Malaysia as shown in Figure 2. The collision resulted in the flooding of crew berthing, communications, and machinery compartments, which led to deaths of 10 sailors and multiple reported injuries with repair costs estimated at approximately $223 million (Werner 2017).

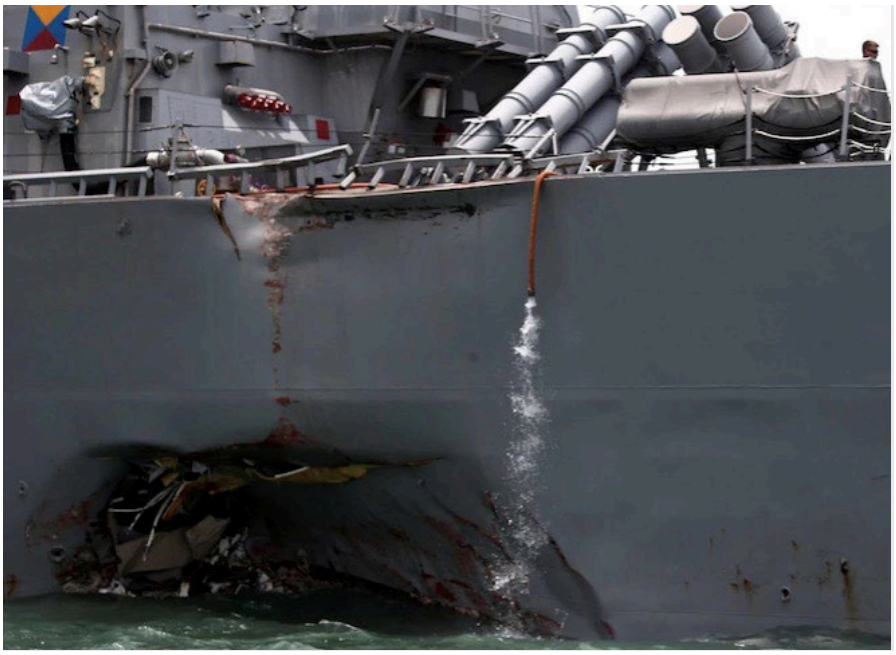


Figure 2. Hull Damage from the USS John S. McCain and Alnic MC Collision. Source: Reuters (2017).

**Root Cause:** The NTSB Marine Accident Report (NTSB 2019) on the accident identified multiple safety issues including the following: lack of sleep by those operating the bridge, the design and training of the bridge equipment to include supporting geolocation systems equipment, and "the procedures for the transfers of steering and thrust control on board the John S McCain (NTSB 2019, viii)"

### 3. USS Port Royal Grounding

**Incident:** On February 5, 2009, a United States guided missile cruiser (USS Port Royal, CG 73) ran aground on a coral reef off the coast of Oahu, Hawaii as shown in Figure 3. The incident resulted in significant damage to the ship and the coral reef costing approximately $40 million in repairs (Konrad 2009).


Figure 3. USS Port Royal Grounded off the Coast of Oahu, Hawaii. Source: USNI (n.d.).

**Root Cause:** When the investigation was completed, Naval investigators determined that the USS Port Royal grounding was causes by several factors. Equipment was determined to be both faulty and not correctly utilized by a crew who lacked sufficient

training on the ship's equipment. Additionally, it was determined that the commanding officer was overly tired due to lack of sleep (Cole 2009).

### 4. USS Vincennes and Iran Air Flight 655 Incident

**Incident:** On July 3, 1998, an Airbus A300 (Iran Air Flight 655) was destroyed by a surface-to-air missile shot by a United States Navy guided missile cruiser [USS Vincennes, CG 49 (Figure 4)] while the flight was in transit from Tehran to Dubai. All 290 occupants on board the flight were killed (Pasley 2020).


Figure 4. USS Vincennes Launching Missile from its Deck. Source: CBS News (n.d.).

**Root Cause:** A Department of Defense (DOD) investigation report (DOD 1988) found that the incident was caused by a misidentification of the civilian airliner as a military aircraft, communications failure, inexperience, and heightened levels of crew stress due to recent events in the area.

### 5. Problem Description: Kill Chain Decision Complexity

These incidents represent the decision complexity involved in naval operations and specifically highlight challenges within the OODA loop (Observe, Orient, Decide, and Assess). Kill chain decisions involving the use of weapon systems are a particularly stressing category within the OODA loop—with unexpected threats that are difficult to identify with certainty, shortened decision reaction times, and lethal consequences. An effective kill chain requires the proper setup and employment of shipboard sensors, the identification and classification of unknown contacts, the analysis of contact intentions based on kinematics and intelligence, awareness of the environment, and decision analysis and resource selection.

Limits to human cognition and human-induced issues can also contribute to the decision complexity of the operational naval OODA loop and kill chain. Examples include:

- watch-stander fatigue
- sensor misuse/misinterpretation
- lack of operator training
- general lack of situational awareness
- deficiencies in situational assessments

The combination of human error, human cognitive limits, and the inherent decision complexity of naval operations lead to challenges in the OODA loop and more specifically in the kill chain.

## B. BACKGROUND AND PROBLEM MOTIVATION

The modern battle space is composed of multiple domains: space, cyberspace, air, land, sea, and underwater. Naval sailors and tactical watch officers face a complex set of tasks and responsibilities during normal at-sea operations that only increase in complexity during threat conditions. Air, surface, and subsurface watch-standers are responsible for the safe navigation and employment of the ship. This includes identifying, classifying, and

tracking all air, surface, and subsurface contacts in addition to navigating hazardous environments. Officers must gain situational awareness of their environment, assess possible threats, and conduct multiple missions while complying with rules of engagement (ROE), and managing a diverse set of weapons, sensors, and communication systems. This requires the aggregation of many data sources and the continuous analysis of contact movement and intent. Figure 5 depicts the find, fix, track, target, engage, and assess (F2T2EA) processes that represent tactical operations. Much of the normal at-sea operations are focused on the "find," "fix," and "track" processes. However, in threat conditions, operations also include "target," "engage," and "assess" functions.

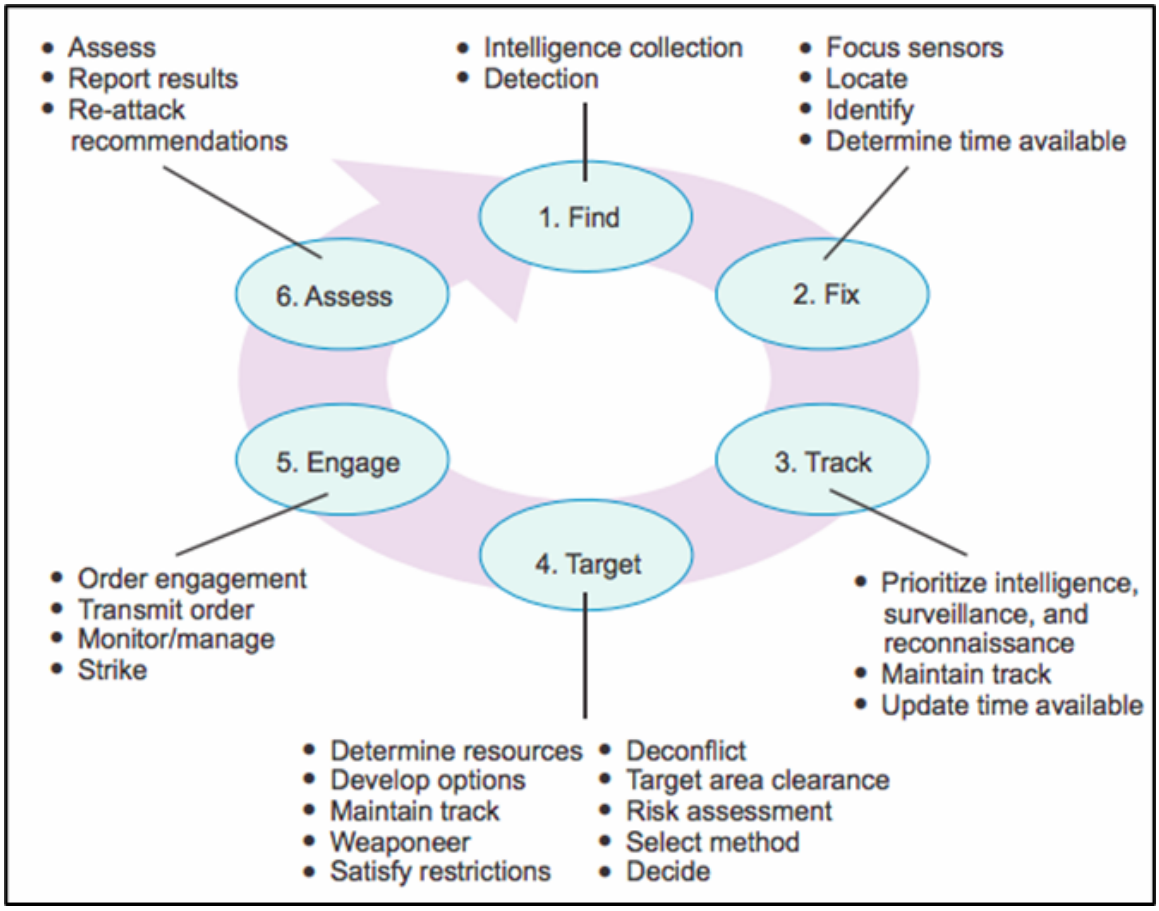



Figure 5. F2T2EA Targeting Cycle. Source: Joint Chiefs of Staff (2013).

The Navy and DOD have recently focused on AI methods as a capability that can improve tactical operations including the F2T2EA kill chain processes. John Hammerer, Naval Postgraduate School (NPS) Chair of Integrated Air and Missile Defense (2021, 1), wrote that the Navy "needs to have a much, much better common tactical picture and then very fast, sophisticated, and automated decision-making." In 2019, the DOD released the Defense Artificial Intelligence Strategy that directs the DOD to "accelerate the adoption of AI and the creation of a force fit for our time" (Joint Chiefs of Staff 2013, 4). DOD also established the Joint Artificial Intelligence Center (JAIC) with the charter to "accelerate the delivery of AI-enabled capabilities, scale the Department-wide impact of AI, and synchronize DOD AI activities to expand Joint Force advantages" (DOD 2019, 9). The DOD plans to apply AI methods to address tactical decision complexity and process the large amounts of tactical data that can be characterized by the three Vs—volume, velocity, and variety. The AI will not replace the human element, but instead support human-machine teaming to enhance situational awareness and decision-making (Alpaydin 2010).

AI is the logical next step in the evolution of the kill chain. Brose (2020, chap. 8) writes that "this is why we must view the emergence of intelligent machines not merely as a way to optimize our existing battle networks but also as an opportunity to break from our present model and reimagine it for the future." DOD and the Navy must explore the use of AI for the kill chain and for many other applications—for their potential benefits and because our peer competitor nations are doing the same. Brose (2020) also determined that China is major producer of AI research in the world by the fact that they produce the majority of the cited published research papers. AI is an emerging frontier of technology innovation that will define Naval dominance in the 21st century and beyond.

## C. PROJECT OBJECTIVES

This capstone project explored the use of automation in concert with artificial intelligence (AI) to improve naval kill chain decisions. The project studied the naval kill chain functions and developed specific evaluation criteria for each function for determining the efficacy of specific AI methods. The project identified and studied AI methods and

applied the evaluation criteria to map specific AI methods to specific kill chain functions. The following research questions guided the project approach:

1. How can AI methods be evaluated for their effective application to the naval kill chain?
2. What criteria need to be established to provide transparent and useful feedback for future AI implementation?
3. What are the evaluation criteria for each kill chain function?
4. What are current applicable AI methods?
5. Which AI methods best map to which functions within the kill chain?

## D. PROJECT OVERVIEW, SCOPE, AND DEFINITIONS

This capstone project contributes to the Navy's knowledge of how AI can apply in the tactical domain. Utilizing the F2T2EA kill chain processes, the team explored the use of automation and artificial intelligence to improve naval kill chain decisions. Figure 6 contains an operational view (OV-1) of this capstone project.


Figure 6. OV-1: Operational Concept for Improving the Naval Kill Chain by Leveraging Artificial Intelligence

This capstone project focused on the naval kill chain functions and development of specific evaluation criteria for each function for determining the efficacy of specific AI methods. The scope was limited to current kill chain processes and the tactical battlespace domain was limited to air and missile defense (AMD). The project was also limited to using only unclassified sources and information throughout the research.

Several key terms are repeated throughout this report. The team has defined the following key terms utilizing multiple sources for the reader's reference.

- Artificial Intelligence (AI) "refers to the ability of machines to perform tasks that normally require human intelligence – for example, recognizing patterns, learning from experience, drawing conclusions, making predictions, or taking action – whether digitally or as the smart software behind autonomous physical systems." (DOD 2019, 5)
- Kill Chain "consists of mission tasks or functions required to successfully employ a specific weapon against a specific threat and the platforms that could provide the required functionality (e.g., target detection could be done by an aircraft or a surface ship). In addition, the kill chain includes the major decision nodes (e.g., the decision to commit an aircraft to visually identify a tracked object) as well as the communication links required to transmit information between and

within units." (Clawson et al. 2015, sec. Developing Kill Chains and TACSITs)

## E. PROJECT APPROACH

This project applied a systems engineering analysis approach to study the application of AI methods to the tactical kill chain. The capstone team developed system models and operational scenarios models and the development of evaluation criteria. The team studied solution concepts involving the application of AI methods to the kill chain. The project conducted a mapping of AI methods to kill chain functions through analysis of alternatives and the use of the evaluation criteria and a system behavioral study using the kill chain scenarios. The project was conducted in three phases, as illustrated in Figure 7.

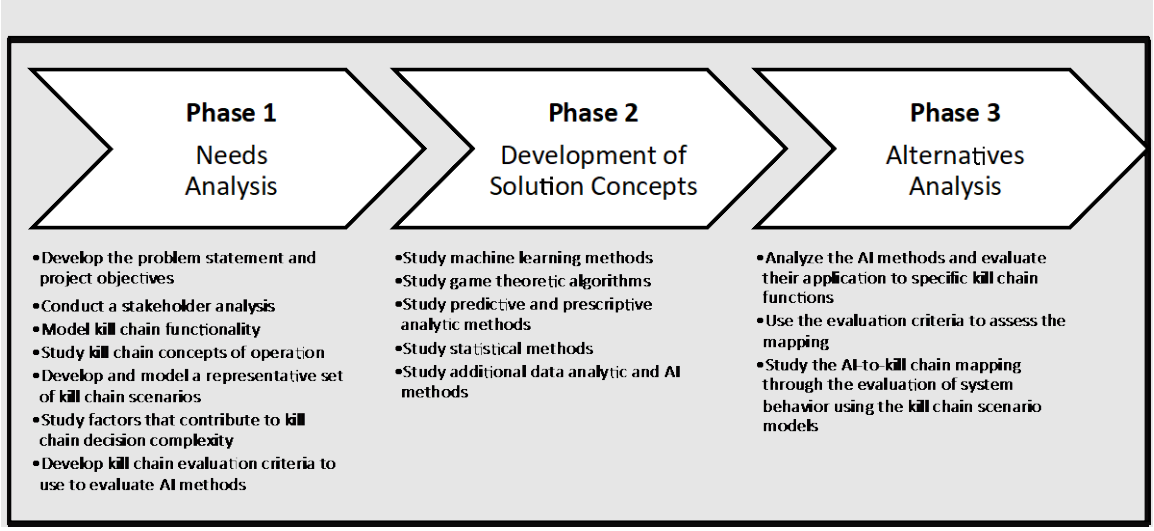



Figure 7. Project Phases

Phase 1 involved a comprehensive needs analysis with a focus on the development of initial project deliverables and analyses of kill chain functionality and AI methodologies. As the foundation of this study, a thorough review was conducted on "Leveraging Artificial Intelligence (AI) for Air and Missile Defense (AMD): An Outcome Oriented Decision Aid" authored by NPS AI-OODA Team from a previous cohort which included a high-level analysis aimed at proving the effectiveness of automation within the kill chain in

AMD (Jones et al. 2020). The outputs of Phase 1 included the project framework and Capstone Proposal.

The transition to Phase 2 represented a shift from high-level analysis to detailed literature reviews on kill chain and AI resources. This included looking at machine learning models, analytic and statistical methods, and theoretical algorithms in addition to the development of functional block diagrams and use cases for initial mapping. Further emphasis was placed on the development of evaluation criteria for refinement in Phase 3 with all outputs of Phase 2 being research framework.

AI Mapping was finalized in Phase 3 along with the application of evaluation criteria to assess behavior of the model in scenario-based employment. AI to kill chain mapping was refined based on scenario results and all project work was written out into five report chapters spanning across two project reviews with advisors and key stakeholders. Major outputs of this phase included the Final Report and brief to all project stakeholders.

## F. TEAM STRUCTURE

The capstone team, "AI 6," was comprised of the six team members represented in Table 1. The team represented a diverse set of organizational backgrounds across several disciplines and geographic locations. Each team member was assigned a main area of focus; however, each member also contributed to all aspects of the project.

Table 1. Team Structure

| TEAM MEMBER | POSITION | ROLES | ORGANIZATIONAL BACKGROUND |
|---|---|---|---|
| Gregory Burns | Systems Engineering Lead | -Direct system testing and evaluation requirements<br>-Conduct functional analysis to translate project requirements into verified deliverables | Systems Engineer/Lab Manager in support of DDS/ASDS. Systems include AN/SPQ-12/ 14/15 and AN/SPA-25G/H. Part of ISEA team installing HW/SW on Radars on DDG/CG/CVN/ LHA/LPDs for DDS/ASDS working from NSWCDD DNA. |
| Todd Collier | Simulation Lead | -Verify modeling and simulation software in support of project requirements<br>-Validate support activities using related tools and software | Meteorological and Oceanographic (METOC) System Engineer supporting atmospheric characterization systems utilized for Air Traffic Control (ATC) weather forecasting for Naval air stations. |
| Richard Cornish | Modeling Lead | -Execute modeling and analysis of project systems<br>-Provide MBSE guidance in support of project requirements | Software Development Lead at NSWCDD DNA for the Aegis and SSDS ATD Integrated Training System. |
| Kyle Curley | Team Lead | -Lead team meetings and organize execution of project deliverables<br>-Provide updates to Capstone Advisors and project stakeholders | Air Support Control Officer at United States Marine Corps, Marine Tactical Air Command Squadron 38. |
| Allan Freeman | Lead Analyst | -Oversee all collections and analytics requirements<br>- Directs the research of historical project related data | Test and Evaluation engineer supporting DT for the USN SPY-6 family of radars. |
| Jared Spears | Chief Editor | -Quality control and configuration management<br>-Report and documentation editing | Digital Systems lab team lead at the Office of Naval Intelligence for the Navy Foreign Materiel Exploitation program. |

Figure 8 shows how the team was organized with the team lead coordinating with the Capstone advisors and the other five team members reporting to the lead.

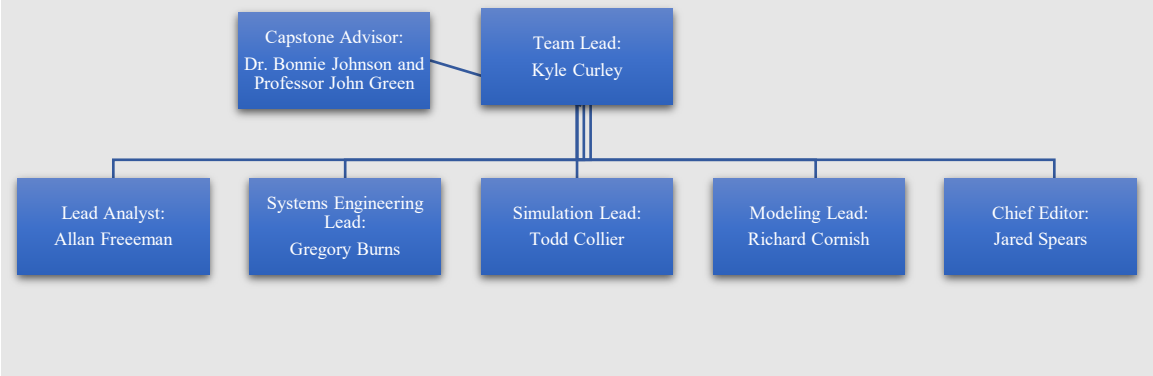



Figure 8. Team Roles

## G. BENEFITS OF THE STUDY

This project supports Navy and DOD stakeholders by providing a deeper understanding of how to evaluate the application of AI methods to the kill chain. The project provides a critical analysis of specific AI methods to evaluate their suitability for specific functions and decisions within the tactical kill chain. The project delivers a mapping of AI methods to the kill chain based on a structured needs analysis, the development of evaluation criteria, a study of AI methods, and an analysis that includes the development and use of MBSE artifacts to understand both the problem space and solution alternatives. These insights provide a knowledge foundation for naval engineers and AI developers as they design and build AI systems for tactical operations. The results of this project will eventually support operational warfare officers as future AI-enabled systems are built and deployed.

## H. THESIS OUTLINE

This capstone report contains five chapters. Chapter I introduces the project and describes a statement of need, the project background and motivation, project objectives, scope, and definitions, the project approach, the team structure, and an overview of the benefits of the study. Chapter II contains a literature review of the tactical kill chain—

describing the kill chain functions, prior work on analyzing the application of AI for the kill chain, and prior work on developing criteria to evaluate kill chains. Chapter III contains a literature review of AI—providing a short history of AI and its applications in the military domain and describing a variety of AI methods. Chapter IV contains the team's analysis—describing the team's evaluation of AI methods for specific kill chain functions. Finally, Chapter V summarizes the project results, presents the team's recommendations, and discusses benefits of the study and future work.

# II. KILL CHAIN: LITERATURE REVIEW AND ANALYTICAL FRAMEWORK

This chapter covers the capstone team's work on establishing a kill chain analytical framework. To conduct a mapping of AI methods onto the kill chain, the team first had to establish a set of kill chain functions that were representative of the tactical decisions and processes involved in gaining situational awareness of a battlespace and making weapon engagement decisions. The team conducted a literature review of relevant source material to inform the development of this study's set of kill chain functions. The team's literature review heavily leveraged a recent NPS capstone study's findings on the tactical kill chain (Jones et al. 2020) and Joint Publication (JP) 3–60: Joint Targeting (2013), which describes the kill chain F2T2EA model as developed by the Joint Chiefs of Staff (JCS).

The capstone team also developed a set of evaluation criteria as part of the kill chain analytical framework. The team reviewed literature sources on the use of different evaluation criteria to assess kill chain performance and effectiveness. The team leveraged these findings to develop a set of evaluation criteria to support the project's analytical framework. This chapter presents the team's evaluation criteria and the literature review findings that supported it.

This chapter is organized into two sections: section A contains the team's literature review findings and section B contains the kill chain part of the analytical framework that the team developed for this project.

## A. KILL CHAIN LITERATURE REVIEW

### 1. Kill Chain Functions

A recent NPS capstone team (the AI-OODA team) studied the use of AI for the AMD kill chain (Jones et al. 2020). As part of their study, the team conducted an in-depth review of kill chain functionality and developed a set of 17 kill chain functions as shown in Table 2. The table shows how the 17 kill chain functions are related to the OODA kill chain model as well as the F2T2EA kill chain model. The previous team's 17 kill chain functions were scoped to just include AMD functions.

Table 2. 17 Kill Chain Functions for AMD. Source: Jones et al. (2020, 44).

| OODA | F2T2EA | Functions |
|---|---|---|
| Observe | Find | Collects Data |
| | | Initial Detection |
| | | Identifies Emerging Target |
| | Fix | Request Further Information |
| | | Classifies Target |
| | | Locates Target |
| | | Validates Detection |
| Orient | Track | Update Target Track |
| | | Validates Target |
| | | Assess Blue Proximity |
| Decide | Target | Nominate Engagement Options |
| | | Prioritize Target |
| | | Select Attack Option |
| Act | Engage | Issue Orders |
| | | Attack Target |
| | Assess | Assess Status of Target |
| | | Authorize Re-attack |

JP 3-60 introduces the F2T2EA concept and breaks down the model into individual process diagrams for each step of the kill chain. Figures 9–13 show functions process diagrams from JP 3-60. Since the F2T2EA kill chain was not designed for AMD alone, several planning and battle management steps essential to joint targeting are incorporated into Figures 9–13 that are not shown in the list of 17 kill chain functions in Table 2.

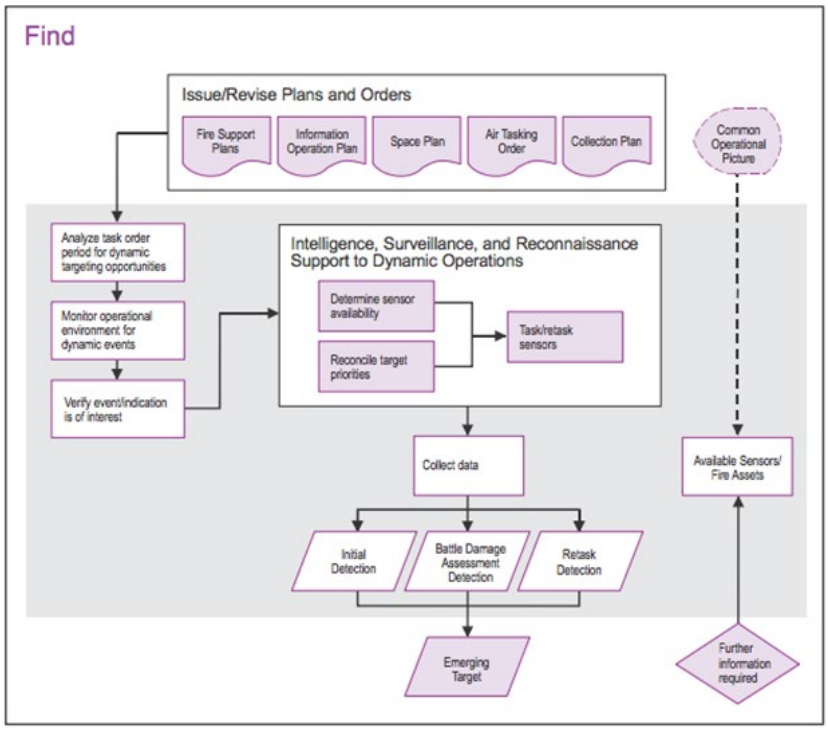



Figure 9. Find: F2T2EA Process Diagram. Adapted from Joint Chiefs of Staff (2013, fig. II–11).

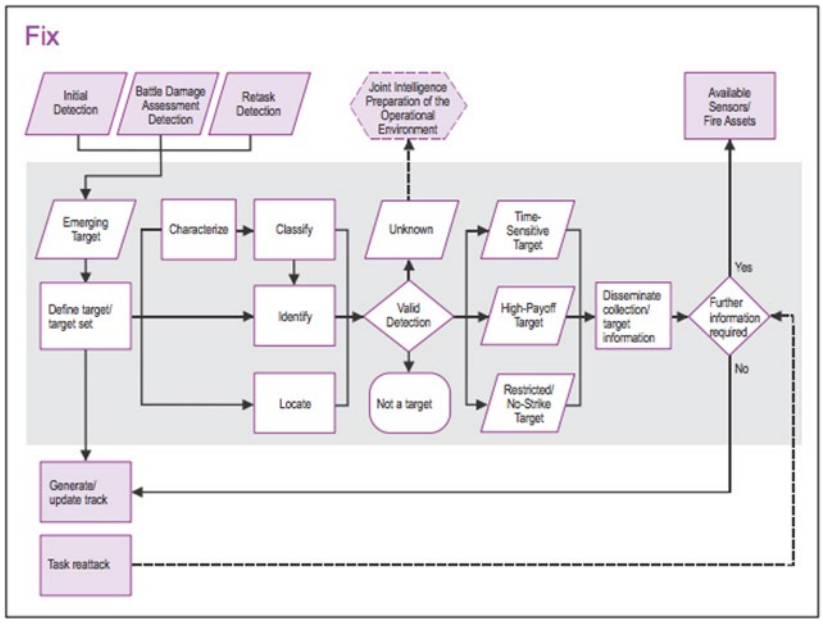



Figure 10. Fix: F2T2EA Process Diagram. Adapted from Joint Chiefs of Staff (2013, fig. II–12).

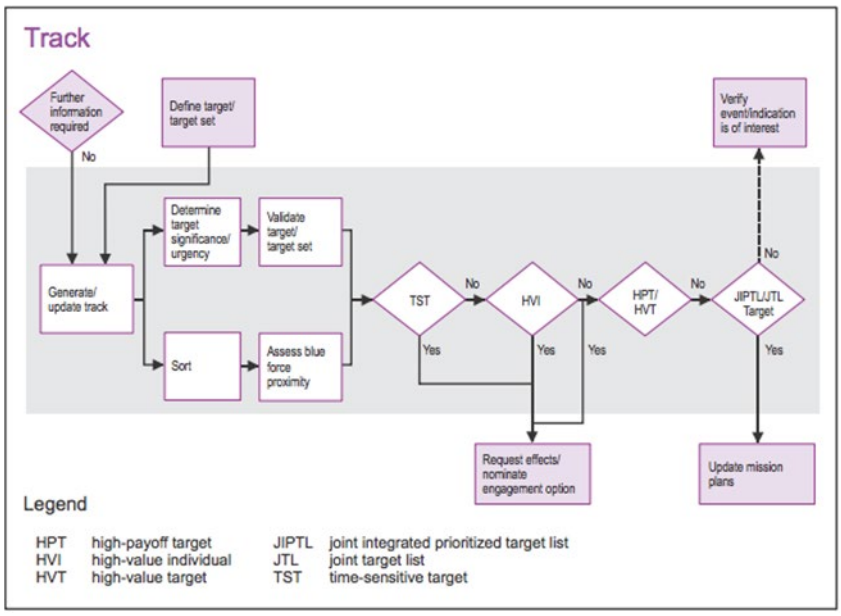



Figure 11. Track: F2T2EA Process Diagram. Adapted from Joint Chiefs of Staff (2013, fig. II–13).

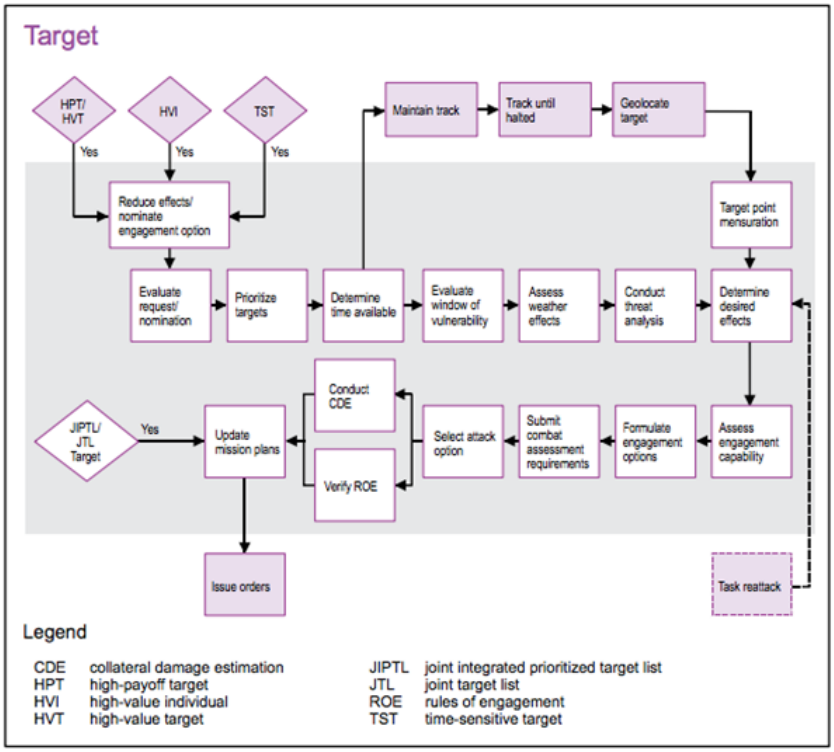



Figure 12. Target: F2T2EA Process Diagram. Adapted from Joint Chiefs of Staff (2013, fig. II–14).

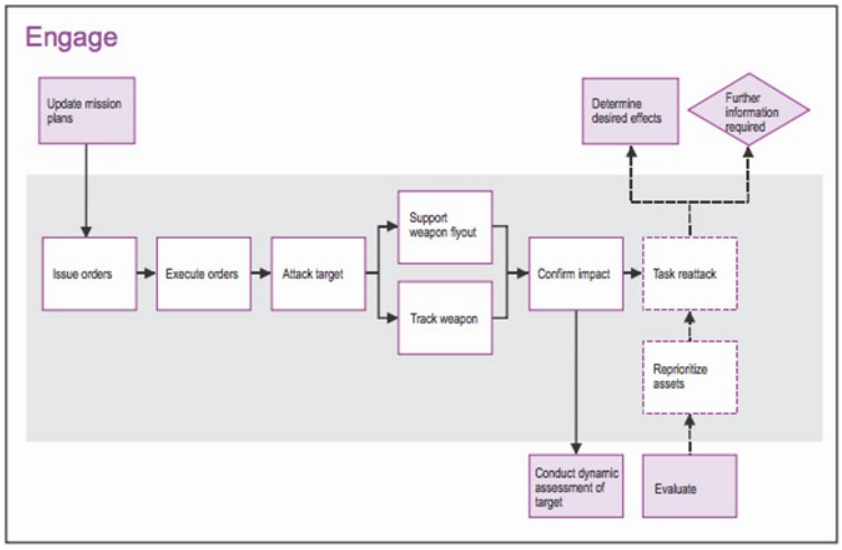



Figure 13. Engage: F2T2EA Process Diagram. Adapted from Joint Chiefs of Staff (2013, fig. II–15).

### 2. Analysis of Automating AMD Kill Chain

This section summarizes the previous NPS capstone analysis by Jones et. al. (2020) (called the AI-OODA team) that studied how the AMD kill chain could be automated and evaluated the decision risk involved in automating the specific functions of the AMD kill chain. Jones et al. (2020) assessed the risk associated with differing levels of automation, studied the feasibility of AI-enabled decision support aids to speed up the execution of the kill-chain, and categorized AMD threat scenarios based on the speed of the incoming threat which would affect the amount of decision time available for performing the kill chain functions.

The AI-OODA team's analysis was based on the ten levels of automation (Parasuraman, Sheridan, and Wickens 2000) shown in Figure 14. The team used the ten levels of automation as a basis for characterizing and analyzing different possible human-machine interactions for the functions of the AMD kill chain.

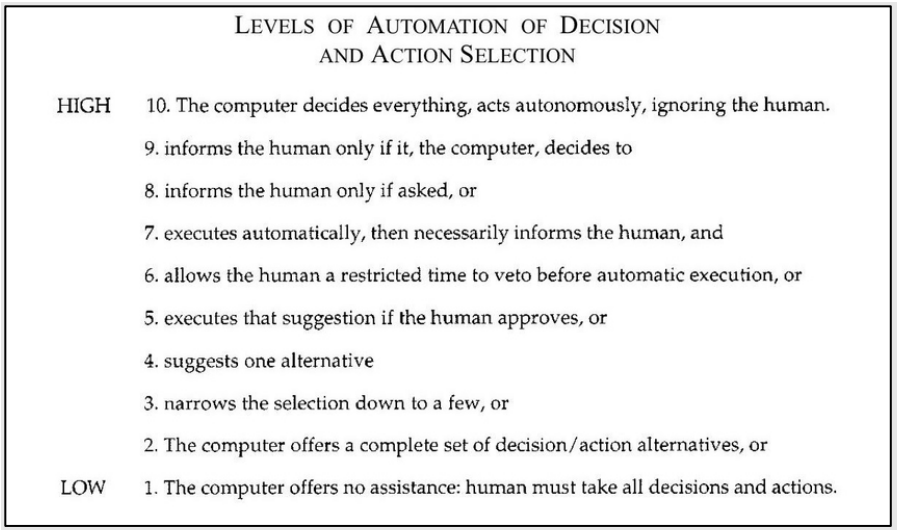



Figure 14. Levels of Automation. Source: Parasuraman, Sheridan, and Wickens (2000, 287).

The AI-OODA team identified three categories of AMD threats based on the time that is available for the kill chain response—effectively the level of decision stress place on the blue force defense. The three scenarios were classified as low, moderate and high stress as shown in Table 3. The AI-OODA team calculated red force munitions speeds by using unclassified maximum operational ranges and velocities. These time values are annotated in the bottom half of the table under column heading "Time to Respond to Red Threat (minutes)." The dashed circles represent the three defined scenario timelines.

Table 3. AI-OODA Blue Force Threat vs. Red Force Threat Timelines. Source: Jones et al. (2020, fig. 39).

| | Red Threat: | DF-21D (China) | DF-17 Hypersonic (China) | WZ-8 Drone (China) | P-270 Moskit (Russia) | BrahMos-11 (Russia) | 3M-54 Kalibr (Russia) | Khalij Fars (Iran) | Shahed-129 Drone (Iran) |
|---|---|---|---|---|---|---|---|---|---|
| | Maximum Operational Range (km) | 2000 | 2500 | 6000 | 250 | 600 | 50 | 300 | 170 |
| | Velocity (kpm) | 205.80 | 205.80 | 82.32 | 61.75 | 144.07 | 16.08 | 61.73 | 2.92 |
| | Time to engage at Maximum Operational Range (minutes) | 9.72 | 12.15 | 72.89 | 4.05 | 4.16 | 3.11 | 4.86 | 58.29 |
| | Blue Force Detection Assets | Detection Range (Km) | Time to Respond to Red Threat (minutes) | | | | | | |
| Navy | Aegis Weapon System (AWS) / AN/SPY-1 Radar | 310 | 1.5 | 1.5 | 3.8 | 4.0 | 2.2 | 3.1 | 4.9 | 58.3 |
| Army | Sentinel Radar | 75 | 0.4 | 0.4 | 0.9 | 1.2 | 0.5 | 3.1 | 1.2 | 25.7 |
| | AN/TPQ-53 (replacement for AN/TPQ-36) | 20 | 0.1 | 0.1 | 0.2 | 0.3 | 0.1 | 1.2 | 0.3 | 6.9 |
| | AN/TPQ-50 Light-weight Counter Mortar Radar (LCMR) | 10 | 0.0 | 0.0 | 0.1 | 0.2 | 0.1 | 0.6 | 0.2 | 3.4 |
| | Patriot system | 100 | 0.5 | 0.5 | 1.2 | 1.6 | 0.7 | 3.1 | 1.6 | 34.3 |
| | Terminal High Altitude Area Defense (THAAD) | 1000 | 4.9 | 4.9 | 12.1 | 4.0 | 4.2 | 3.1 | 4.9 | 58.3 |
| Missile Defense Agency | Upgraded Early Warning Radars (UEWR) | 4828 | 9.7 | 12.1 | 58.7 | 4.0 | 4.2 | 3.1 | 4.9 | 58.3 |
| | Cobra Dane radar | 3218 | 9.7 | 12.1 | 39.1 | 4.0 | 4.2 | 3.1 | 4.9 | 58.3 |
| | Sea-based X-band radar | 4023 | 9.7 | 12.1 | 48.9 | 4.0 | 4.2 | 3.1 | 4.9 | 58.3 |

Next, the AI-OODA team allocated time to each task in both the low stress and high stress scenarios. The low stress scenario represented a fully human-driven scenario with no automation and a threat time to engage of just under 59 minutes. The high-stress scenario represented a fully automated scenario against a threat with a time to engage of 1.5 minutes. For both the high and low stress scenarios, timelines were evenly distributed to each of the 17 critical tasks. The moderate stress scenario represented a human-machine team against a threat with a time to engage of 9.7 minutes. To determine levels of automation, the human-machine team classified each of the 17 critical tasks in terms of perceived risk involved. This risk analysis contributed to the generation of an automation utility curve that aided in the allocation of time to each function in the moderate stress scenario.

A primary outcome of the AI-OODA team's study was that the amount of automation required for the kill chain depends on the threat situation. For low stress scenarios, there may be enough reaction time to perform the kill chain functions manually with minimal aid from an automated decision aid. For high-stress scenarios, the reaction time will be so small, that the kill chain functions will have to be fully automated. And in

moderate stress scenarios, the kill chain functions can be performed by a hybrid combination of manual decisions supported by automated functions.

Another key finding of the AI-OODA team's study was that higher levels of decision risk will have to be accepted in the high stress threat scenarios. The team's simulation analysis showed that success against moderate stress scenario threats was heavily dependent on risk acceptance or accepting more automation in the kill chain functions. For the moderate scenario simulation runs, initial Monte Carlo runs were unsuccessful 100% of the time. After the team revisited timeline allocation and added more automation requiring the acceptance of more decision risk, the simulation runs demonstrated 100% success. The high stress scenario simulations were run with full automation and demonstrated an 83.5% success rate.

### 3. Decision-Making

This section presents the team's findings on decision-making as it relates to tactical kill chain decisions. The team explored work done on this topic to better understand decision risk and challenges that warfighters may face in making tactical decisions. The previous AI-OODA capstone team analyzed the kill chain by studying the time available for making decisions. This study sought to expand the analysis to also address the quality of the kill chain decisions as well.

During the literature review, the team identified an exemplary concept for this study - bounded rationality coined by Herbert Simon (as cited in Campitelli and Gobet 2010). Simon presented the bounded rationality concept as an argument against the assumption of a perfectly rational thinker. Human decision makers often employ a decision-making concept called satisficing—or searching through a set of available alternatives until finding one that meets an acceptability criterion. Humans often use this method to select from available courses of action. However, this method does not guarantee an ideal selection, but rather the decision-maker often settles on the first such option that meets the required threshold for acceptability. Some "complications" often lead to the use of the satisficing method rather than seeking ideal options (or perfect rationality). These include:

- inability to fully process all available information
- lack of pertinent information
- limited knowledge to act confidently
- risk acceptance
- personal experience(s)
- information overload
- analysis paralysis ("overthinking")

These "complications" are evident in each of the naval mishaps that were described in Chapter I. Shipboard watch-standers have varying levels of knowledge, proficiencies with systems, and experiences. They are often flooded and overwhelmed with information that demand swift processing and action. Figure 15 exemplifies the amount of information that must be processed simultaneously and continuously by Tactical Action Officers (TAO) on surface combatants to gain and maintain situational awareness.

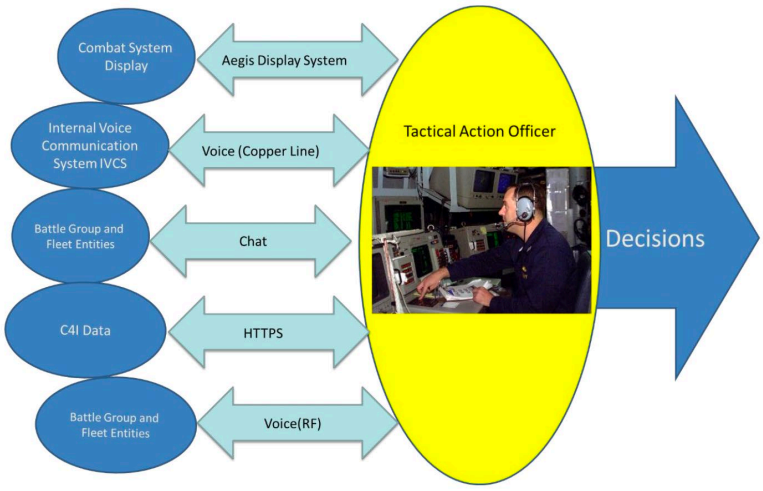



Figure 15. Tactical Action Officer Information Flow. Source: Iversen and DiVita (2019, fig. 2).

A critical factor in developing evaluation criteria for the mapping of AI methods to kill chain functions is decision assessment. How do we classify a decision as "good"? It is important to highlight that decision theory dictates that *decision* quality must be distinguished from *outcome* quality (Horvitz, Breese, and Henrion 1988). Decision assessment must consider the state of knowledge awareness at the time of action selection—or the quality of the information that the decision is based on. The old adage that "garbage in equals garbage out" will apply to the quality of decisions. In some circumstances good decisions can lead to unfavorable outcomes and vice versa. Thus, decision quality must be continually evaluated to ensure that decision processes provide good outcomes.

Conditional probability theory is an important foundation to decision science. According to Horvitz, Breese and Henrion (1988, 249):

> Probability provides a language for making statements about uncertainty and thus makes explicit the notion of partial belief and incomplete information. Decision theory extends this language to allow us to make statements about what alternative actions are and how alternative outcomes (the results of actions) are valued relative to another.

Research on decision theory highlights three types of decisions – decisions under certainty, decisions under uncertainty, and decisions under conflict. The relevance here is obvious as the kill chain represents decision making under all three types. This capstone project noted the importance of considering these factors in the development of evaluation criteria for assessing kill chain decision-making methods.

The literature review also identified the importance of interpretability in automated and AI systems and role of human interaction with AI systems. To ensure the success of tactical operations, human decision makers must understand how automated decision aids arrive at recommendations. This will not only guarantee operator understanding but will also provide confidence in suggested courses of action.

## 4. Assessing Kill Chains

To fully understand and develop appropriate evaluation criteria for AI methods, the AI-6 team studied research completed by Naval Surface Warfare Center (NSWC) Dahlgren

Division in their evaluation of test-fires in tactical situations (TACSITs) (Clawson et al. 2015). The research team developed an assessment process to evaluate kill chains based on the steps described in the following subsections (Clawson et al. 2015).

a. Defining and Prioritizing Weapon-Target Pairs of Interest
b. Developing Kill Chains and Tactical Situations (TACSITs)
c. Developing Scoring Criteria
d. Scoring Kill Chains
e. Performing an Integrated Kill-Chain Assessment
f. Reporting Assessment Findings for Future Commanders

#### *a. Defining and Prioritizing Weapon-Target Pairs of Interest*

The NSWC Dahlgren Division describes the first process as "developing a fleet-prioritized list of weapon-target pairs (WTPs) that identify a specific weapon for use against a specific target" (Clawson et al. 2015). These prioritized lists are provided by operating forces and weapons assignments are based on assessed effectiveness against the target within prescribed TACSITs.

#### *b. Developing Kill Chains and Tactical Situations (TACSITs)*

The development of kill chains and tactical situations is dependent on the specific mission set. This includes tasking and the associated command, control, and communications (C3) components which can change as each step is executed (Clawson et al. 2015). NSWC Dahlgren Division uses the AW-3 Detect task as an example in Figure 16 to show the complexity of the kill chain pathway.

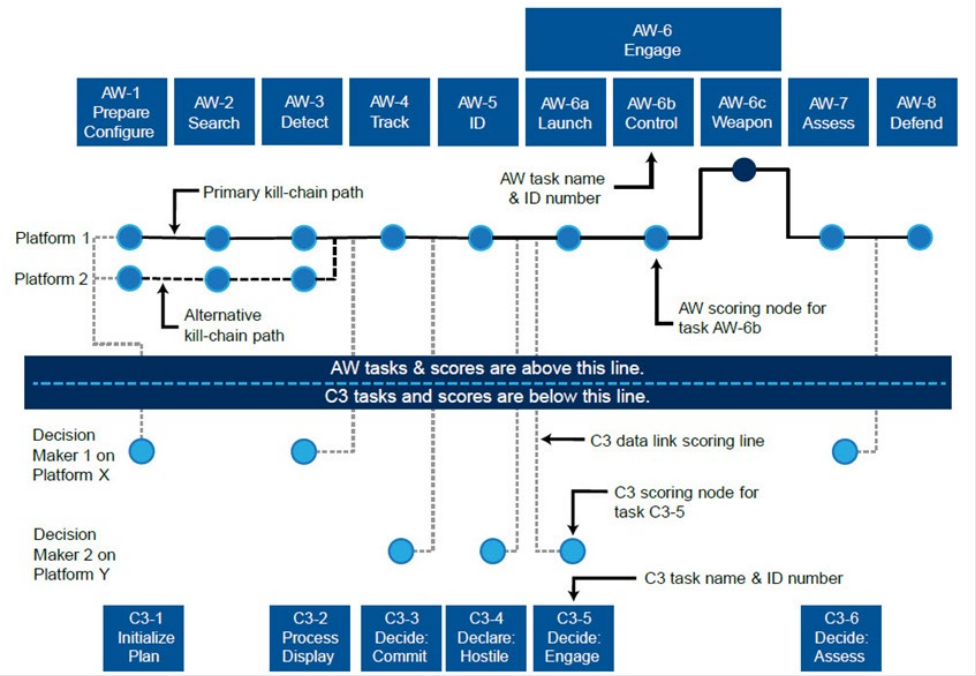



Figure 16. Notional Unscored Kill Chain for the Air Warfare Mission. Source: Clawson et al. (n.d., fig. 2).

The research team offers the following explanation (Clawson et al. 2015):

> In increment 3, the AW kill chains consisted of eight mission tasks plus three subtasks and six C3 nodes...Given the number of combinations of mission tasks and platforms and/or performers that could carry out those tasks, there could be many "paths" through a kill chain. Many of these paths will have "broken" links, some will have "weak" but not broken links, and (hopefully) some will be comprised of all "strong" links. Since assessing all possible paths through a kill chain is not generally possible, WCB focuses on assessing the primary path, which represents the Navy's preferred path, based on current doctrine, training, etc., and possibly one alternative path that may circumvent weak or broken links in the preferred path.

Due to the complexity of kill chain pathways, NSWC Dahlgren Division states that the TACSIT is the key element that provides context in development of scoring criteria (Clawson et al. 2015).

### *c. Developing Scoring Criteria*

The next step is the development of scoring criteria. NSWC Dahlgren Division developed the following color scheme to assess the projected performance of a naval platform or system (Clawson et al. 2015):

- Green: The platform provides the full level of performance required by the TACSIT.
- Yellow: The platform provides a partial or degraded level of performance in the constraints defined by the TACSIT. For instance, a yellow score would be applied to performance that occurred between a desired and a minimum threshold, if both were defined. In some circumstances, the desired and minimum threshold are the same, and a yellow score does not exist, i.e., performance is either Green or Red.
- Red: The platform fails to provide the minimum level of performance required by the TACSIT.
- White: Test data does not exist or is insufficient to accurately score system performance. The PI responsible for scoring a given system or function makes the decision as to whether there is sufficient data to score system performance. In some cases, accredited M&S is used to supplement test results. If the M&S supports the trend observed in the test data, then the function will be scored; otherwise, the White score remains in effect.

### *d. Scoring Kill Chains*

In this step, the research team assessed system components through a combination of data and scoring criteria. Table 4 provides an example of scoring a weapon system with the lowest score assignment being responsible for the overall score of the system. In this example, three critical measures were evaluated: whether the correct weapon had been selected, whether the prelaunch data had been provided to the weapon, and whether the launcher was able to fire all requested rounds for the engagement. In this example, an issue was identified with hardware interfaces preventing the launch for some of the attempts. This criterion was scored as “yellow,” and subsequently, the overall task score was given a “yellow” rating.

Table 4. Example Assessment. Source: Clawson et al. (2015, tab. - 1).

| Critical Measures | Notional Scoring Criteria | Notional Score and Rationale | Overall Task Score |
|---|---|---|---|
| Launch_1 | Is the correct weapon selected? | **GREEN:** No issues observed in testing | **YELLOW** |
| Launch_2 | Is all prelaunch data provided to the weapon? | **YELLOW:** Hardware interface issues prevented launch in x of y attempts | |
| Launch_3 | Is the launcher able to fire all requested rounds for the engagement? | **GREEN:** No issues observed in testing | |

### *e. Performing an Integrated Kill-Chain Assessment*

The assessment step represents a critique of scoring for all kill chain measures. A key element of this step included ensuring that the TACSIT scenario and capabilities of all related components were accurately captured (Clawson et al. 2015).

### *f. Reporting Assessment Findings for Future Commanders*

The final step develops a report of the assessment findings, with the main purpose being the identification of deficiencies within the kill chain and anything that would negate the success of the weapon system against a particular target (Clawson et al. 2015). This analysis provided the AI6 Team with a structural process for development of criteria that will be expanded on in Chapter IV. It was decided that the research would focus on the time aspect covered by the AI-OODA team while incorporating the concepts of quality of decision-making and accuracy. Further assessment was provided after development of the model.

## B. KILL CHAIN ANALYTICAL FRAMEWORK

This section presents the kill chain analytical framework developed by the capstone team for this project. The team developed a set of kill chain functions, based on the F2T2EA kill chain model and the kill chain analysis conducted by the AI-OODA team. This capstone project assessed the F2T2EA processes for AI mapping potential—the resulting capstone kill chain model serves as the basis for the functional methodology and analysis presented in this capstone report in Chapter IV.

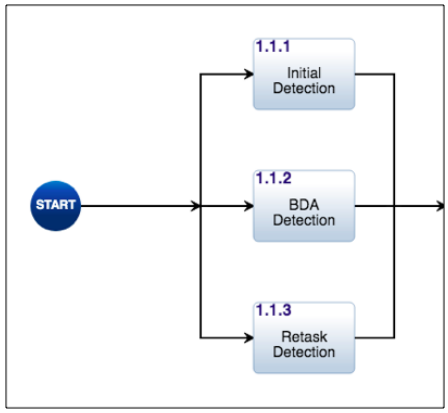
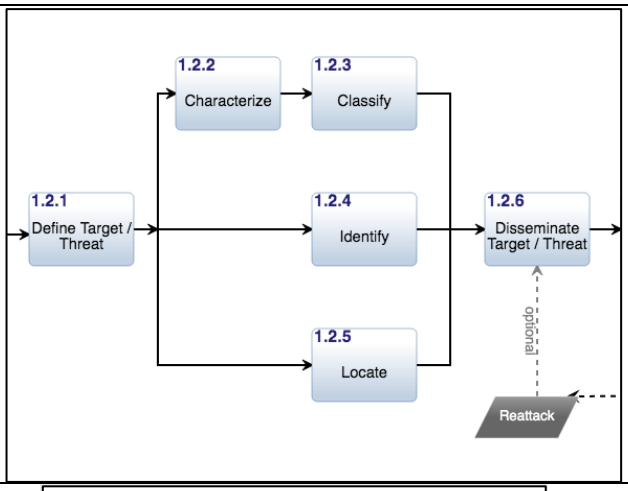
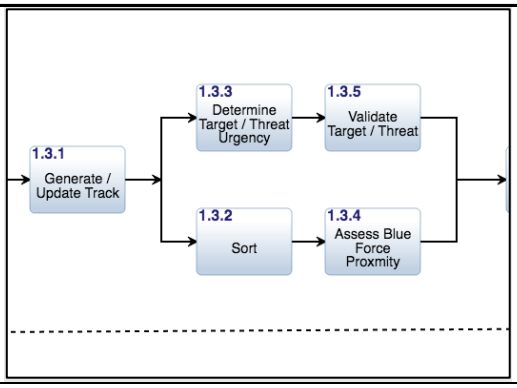
This capstone team interpreted the kill chain slightly differently from the AI-OODA team's interpretation. The AI-OODA team explicitly and tightly coupled specific OODA functions to specific F2T2EA functions. This capstone team asserts that the OODA loop is a decision cycle that is nested within each function of the kill-chain rather than a mapping to one of more kill chain functions. Therefore, this capstone team settled on a model of the kill chain functions that is more closely based on the F2T2EA model.

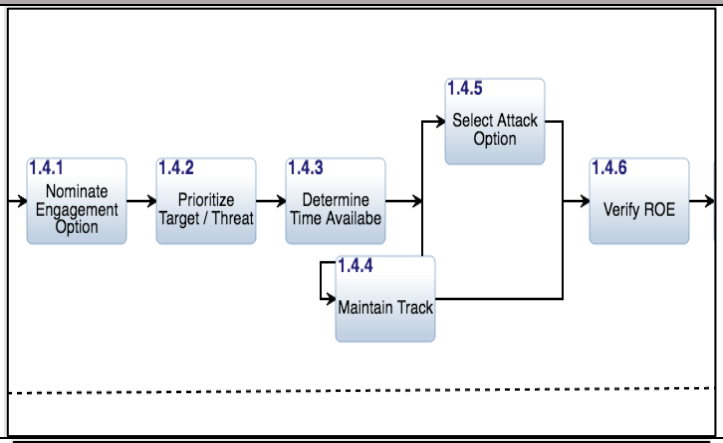
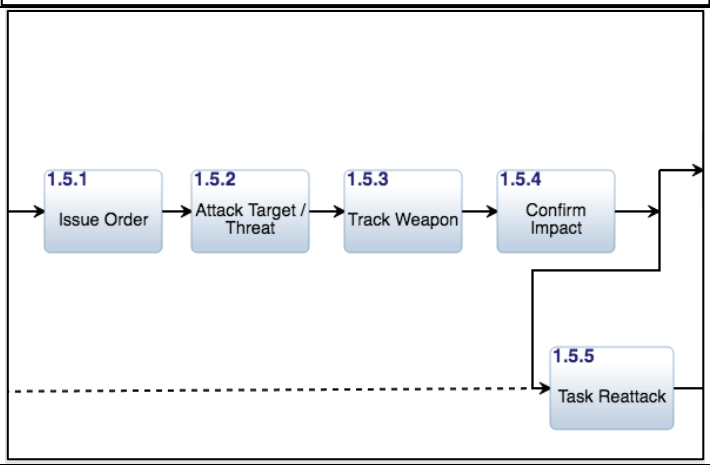
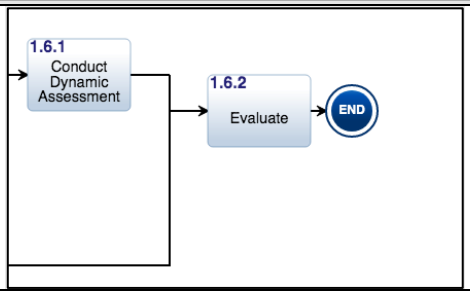
During the capstone team's review of kill chain related literature, the team reviewed methods for evaluating kill chains. The team identified a set of evaluation factors that they incorporated into the analytical framework. This section contains a description of those factors.

### 1. Kill Chain Functional Block Diagram Architecture

The team began by developing a functional block diagram (FBD) based on the functions of the kill chain. The team identified kill chain functions for each of the primary steps in the F2T2EA kill chain model. Table 5 contains the team's FBD kill chain architecture.

Table 5. Initial AMD Functional Block Diagram. Adapted from (Joint Chiefs of Staff 2013, fig. II-11 to II-16).

| Kill Chain Step | Functional Block Diagram |
|---|---|
| Find |  |
| Fix |  |
| Track |  |

| Kill Chain Step | Functional Block Diagram |
|---|---|
| Target |  |
| Engage |  |
| Assess |  |

## 2. Kill Chain Model with 28 Functions

The team identified a kill chain model for this study that is based on the F2T2EA model and has 28 functions. Table 6 lists the 28 kill chain functions and groups them according to the primary F2T2EA functions and to the FBD architecture.

Table 6. Kill Chain Model with 28 Functions

| Kill Chain Step | FBD # | # | Function |
|---|---|---|---|
| Find | 1.1.1 | 1 | Initial Detection |
| | 1.1.2 | 2 | Battle Damage Assessment (BDA) Detection |
| | 1.1.3 | 3 | Re-Task Detection |
| Fix | 1.2.1 | 4 | Define Target/Threat |
| | 1.2.2 | 5 | Characterize |
| | 1.2.3 | 6 | Classify |
| | 1.2.4 | 7 | Identify |
| | 1.2.5 | 8 | Locate |
| | 1.2.6 | 9 | Validate Detection |
| | 1.2.7 | 10 | Disseminate Target /Threat Information |
| Track | 1.3.1 | 11 | Generate / Update Track |
| | 1.3.2 | 12 | Sort |
| | 1.3.3 | 13 | Determine Target / Threat Urgency |
| | 1.3.4 | 14 | Assess Blue Force Proximity |
| | 1.3.5 | 15 | Validate Target / Threat |
| Target | 1.4.1 | 16 | Nominate Engagement Option |
| | 1.4.2 | 17 | Prioritize Target / Threat |
| | 1.4.3 | 18 | Determine Time Available |
| | 1.4.4 | 19 | Maintain Track |
| | 1.4.5 | 20 | Select Attack Option |
| | 1.4.6 | 21 | Verify Rules of Engagement (ROE) |
| Engage | 1.5.1 | 22 | Issue Order |
| | 1.5.2 | 23 | Attack Target / Threat |
| | 1.5.3 | 24 | Track Weapon |
| | 1.5.4 | 25 | Confirm Impact |
| | 1.5.5 | 26 | Task Re-Attack |
| Assess | 1.6.1 | 27 | Conduct Dynamic Assessment |
| | 1.6.2 | 28 | Evaluate |

The kill chain model begins with detection of the threat via multiple sources of collection and sensors. The process then goes through steps to identify and validate track information of the threat, concluding with weapon to target matching requirements and attack. Finally, inherent to any kill chain process, there is an opportunity to assess and re-engage the threat, as operational contingencies necessitate.

### 3. Operational Viewpoints

The team developed some operational viewpoint models to accompany the FBD and 28-function kill chain model. The team developed the OV-1 high-level operational illustration that was shown in Chapter I in Figure 6. The OV-1 is a holistic view of the operational concept based on AMD contingency requirements while demonstrating the integration of multiple systems and networking to all key stakeholders the OV-1 highlights aspects and systems in tactical operations that this study focused on: blue forces, red forces, weapon systems, sensors, and C2 systems.

#### *a. OV-5a: Operational Activity Decomposition Tree (Initial)*

The capstone team developed an OV-5a, operational activity decomposition tree model. This model highlighted a top-level view of all functional components within the kill chain model and is shown in Figure 17. The OV-5a placed the 28 designated functions in a hierarchical structure and clearly delineated lines of responsibility associated within the kill chain and assisted in identifying any redundancies during the team's analysis.

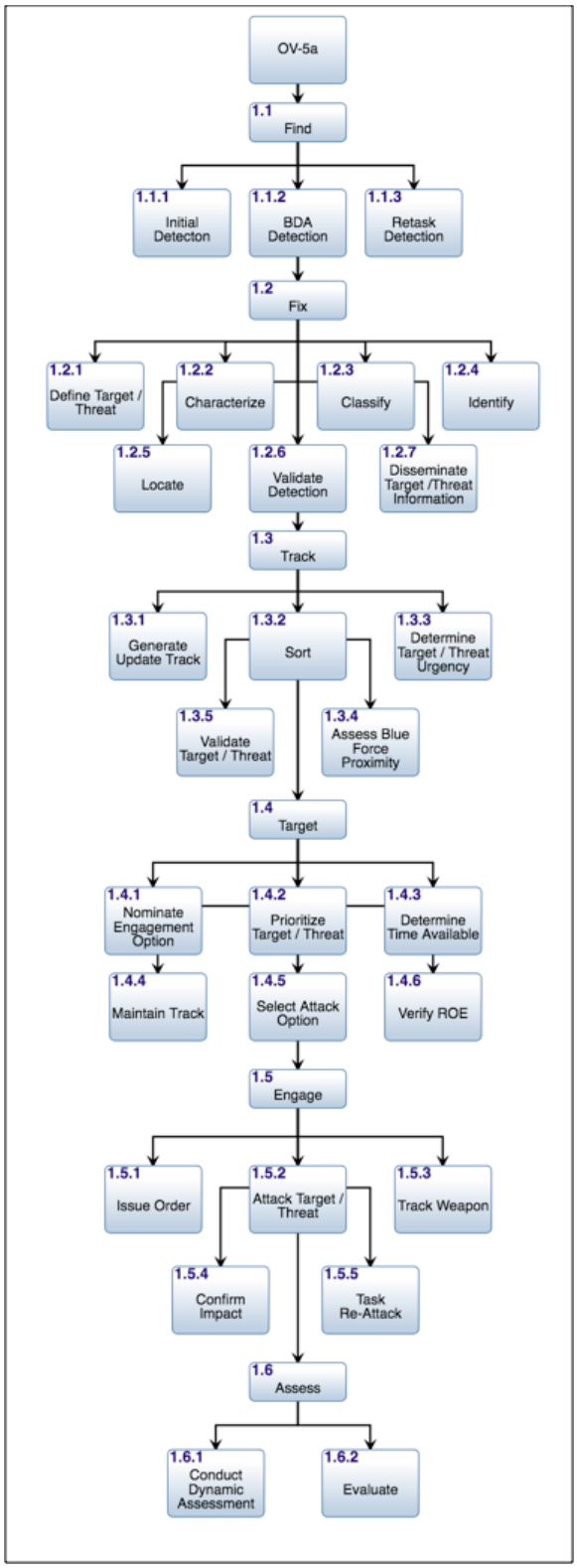



Figure 17. OV-5a: Initial Operational Activity Decomposition Tree

### *b. OV-6c: Event Trace (Initial)*

The capstone team modeled an OV-6c event trace description based on a common AMD scenario of a theater ballistic missile (TBM) attack. The OV-6c demonstrated scenario progression through sequence of events across all relevant systems and is depicted in Figure 18.

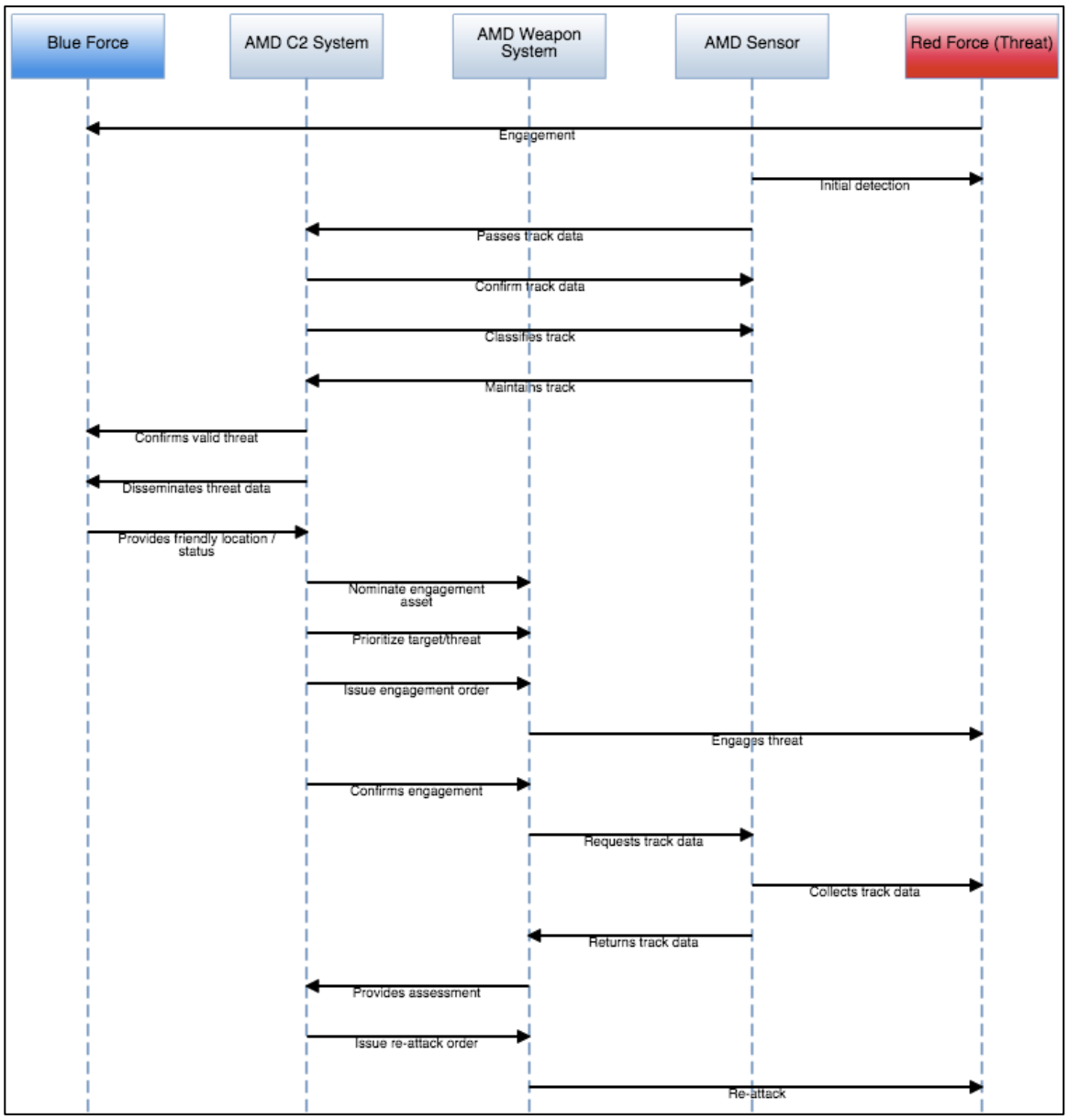



Figure 18. OV-6c: Initial Event Trace Description for a TBM Launch

The sequence begins with an engagement by a hostile force in the form of a TBM launch and progresses through the identified kill chain steps as referenced earlier. The scenario concludes with engagement and termination of the threat through subsequent re-attack of the TBM. Of note, the OV-6c highlighted several events dependent on critical decision-making across multiple systems. The model also identified the need to execute further analysis of the behavioral aspect of steps within the kill chain, as noted earlier in the section covering development of evaluation criteria.

### 4. Kill Chain Evaluation Factors

The final section of this chapter involved identifying kill chain evaluation factors that could be used as part of the mapping process in the analytical portion of this report. Based on the principles outlined by NSWC Dahlgren Division and the kill chain literature review, the team established the following preliminary elements as areas suitable for research as evaluation factors:

- assumptions and assessment of the tactical situation
- development of decision criteria
- development of a set scoring process

These factors, combined with the AI literature review, would be the basis for forming the evaluation criteria and methodology introduced in Chapter IV.

# III. ARTIFICIAL INTELLIGENCE: LITERATURE REVIEW AND ANALYTICAL FRAMEWORK

This chapter contains an overview of AI-related information that the capstone team gathered through literature review. The team explored topics in AI to gain an understanding of AI methods and establish a foundation for the mapping of AI methods to the kill chain. Due to the breadth of the AI knowledge space, only a subset of the most prominent methods is presented in this chapter. The intent is to provide a high-level overview of each of the selected AI methods.

This chapter is organized into two parts: (1) a literature review of AI topics relevant to this project and (2) a description of the team's analytical framework based on the AI literature review. The AI literature review includes: (1) an overview of the history of AI and its application to DOD and the Navy, (2) an overview of the three waves of AI, and (3) descriptions of specialized topics in AI that relate to this project. The analytical framework discussion in this chapter describes the set of AI methods and introduces the evaluation criteria that the team decided to use for the mapping analysis. Figure 19 illustrates the organization and flow of this chapter.

## A. AI LITERATURE REVIEW

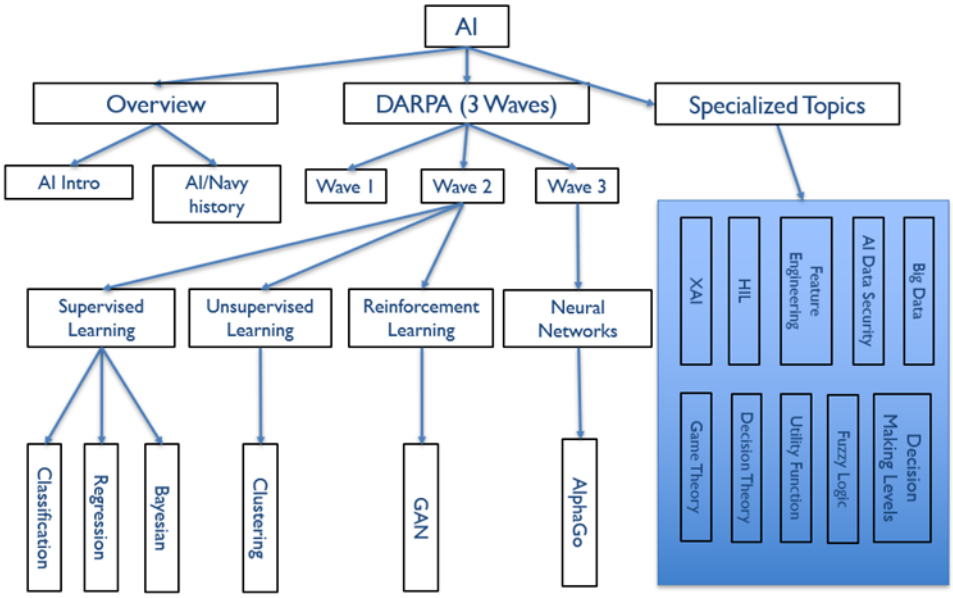



Figure 19. Layout of AI Topics in the Literature Review

### 1. AI Overview

This section provides an introduction and overview of AI. The section begins with a discussion about how AI, as a field, lacks a concrete universally agreed upon definition. Various definitions of AI are provided to highlight this statement. Next, this section explains the difference between hand-crafted knowledge AI and machine learning AI and discusses that ML is a subset the AI discipline. Finally, this section presents a short history of how AI got started and discusses the current way that DOD and the Navy is organized to develop and implement AI for military applications.

#### *a. Definitions of AI*

A recent Congressional Research Service report sums up the difficulty of trying to define AI by noting that "almost all academic studies in artificial intelligence acknowledge that no commonly accepted definition of AI exists, in part because of the diverse approaches to research in the field" (Tarraf et al. 2019, 1). This subsection presents some

different definitions of AI to show how academics and practitioners are thinking about AI as a discipline.

This capstone study used the DOD definition of AI (2019, 5) in Chapter I: "AI refers to the ability of machines to perform tasks that normally require human intelligence—for example, recognizing patterns, learning from experience, drawing conclusions, making predictions, or taking action—whether digitally or as the smart software behind autonomous physical systems."

Some of the foremost academics describe AI as the process of determining the best outcome or expected outcome from a given set of inputs. AI is the building of "agents" that can make these intelligent outcome determinations (Russell and Norvig 2021, 37). "Artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment" (Nilsson 2010, 13). See Figure 20 below for the AI definitions from the FY 2019 National Defense Authorization Act defines AI (2018, sec. 1051).

(f) DEFINITION OF ARTIFICIAL INTELLIGENCE.—In this section, the term "artificial intelligence" includes each of the following:

(1) Any artificial system that performs tasks under varying and unpredictable circumstances without significant human oversight, or that can learn from experience and improve performance when exposed to data sets.

(2) An artificial system developed in computer software, physical hardware, or other context that solves tasks requiring human-like perception, cognition, planning, learning, communication, or physical action.

(3) An artificial system designed to think or act like a human, including cognitive architectures and neural networks.

(4) A set of techniques, including machine learning that is designed to approximate a cognitive task.

(5) An artificial system designed to act rationally, including an intelligent software agent or embodied robot that achieves goals using perception, planning, reasoning, learning, communicating, decision-making, and acting.

Figure 20. NDDA 2018 AI Definition. Source: Office of the Federal Register (2018, sec. 1051).

### *b. Handcrafted Knowledge AI vs. Machine Learning AI*

AI is generally categorized as either handcrafted knowledge or machine learning. Figure 21 depicts the differences between these categories.

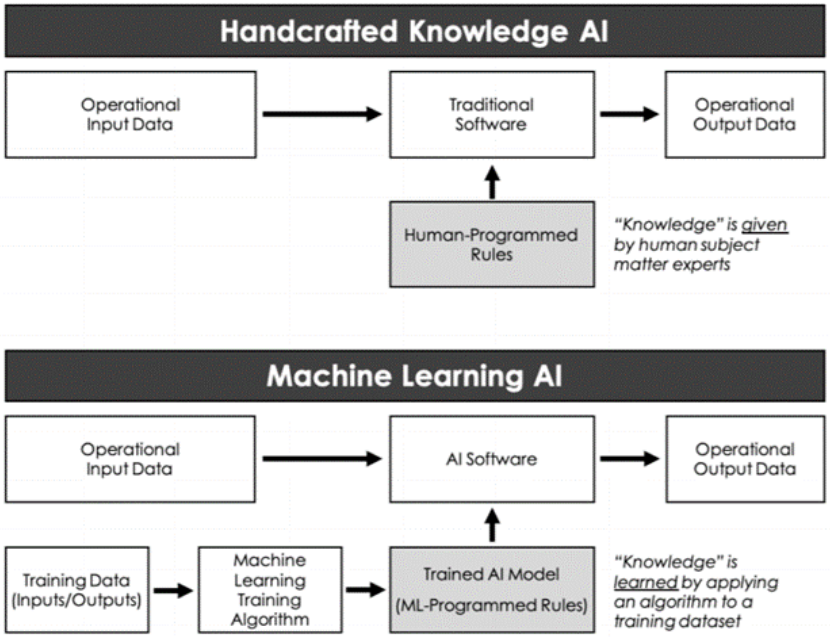



Figure 21. Simplified Diagram of AI Approaches. Source: Allen (2020, fig. 2).

Handcrafted knowledge AI systems use traditional software to process operational input data and produce operational output data. The software is developed by humans—it is "handcrafted" and is based on human-programmed rules developed by human subject matter experts. Allen (2020, 6) describes handcrafted knowledge AI as being based on human-programmed rules such that if given "x input," the system will produce "y output."

In contrast, machine learning AI is based on a process of training an AI model using training data and training algorithms. The AI model being trained "learns" based on many iterations of running the system using the training data. This training process produces the AI software that is then used operationally to produce operational outputs given operational data inputs. ML methods develop a model that learns from a given set of inputs (data) that in turn makes an ML prediction (Theodoridis 2015). As shown in Figure 22, AI is the

development of a machine that can sense, reason, act, and adapt as human might where a machine learning machine without being explicitly programmed to learn to improve performance as more data becomes available. ML is a subset of AI—and comprises machines that improve over time and are based on learning methods rather than explicitly programmed methods.

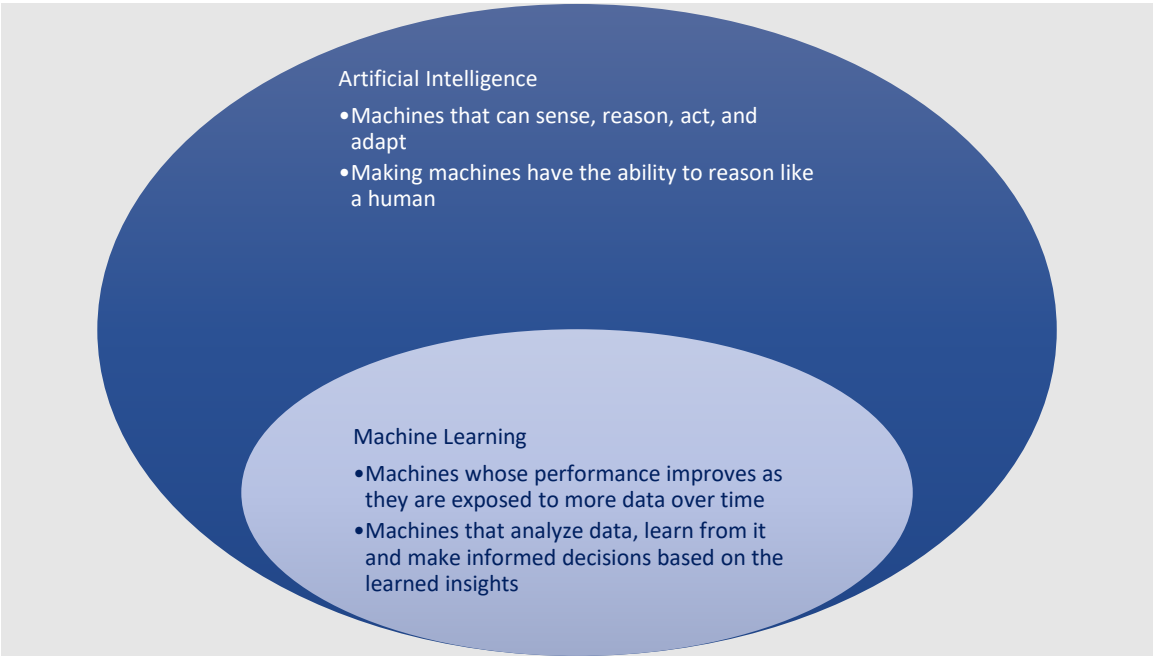



Figure 22. ML as a Subset of AI. Adapted from (Singh 2018, fig. Cousins of AI; Oppermann 2019, fig. AI vs. ML vs. DL).

### *c. AI within the DOD and Navy*

One of the first significant uses of AI was by Alan Turing to crack the 'Enigma' code on the German Enigma enciphering machine during World War II in the 1940s. Turing developed an electro-mechanical machine known as the Bombe that used logic to decipher the encrypted codes of the Enigma machine (Cox 2018, sec. Cracking the Code) .Turing was one of the first to attempt to define AI by saying: "What we want is a machine that can learn from experience," and that the "possibility of letting the machine alter its own instructions provides the mechanism for this"(Copeland 2020, sec. Theoretical work). The Advanced Research Projects Agency (ARPA) later named DARPA emerged as the

primary DOD funder of AI research focusing on natural language processing, facial recognition, and target detection in the 1950s until the mid-1970s. However, due to disappointing results, AI funding was limited until 1983 when DARPA launched the Strategic Computing Program (SCP). The SCP objective was to provide "a broad base of machine intelligence technology" (Roland and Shiman 2002, 76) but never quite met the expectations of stakeholders and closed shop around 1993. The DOD interest and funding increased again around 2010 "due to the convergence of three enabling developments: (1) the availability of "'big data'" sources, (2) improvements to machine learning approaches, and (3) increases in computer processing power" (Sayler 2019, 2).

In 2018 the DOD established the Joint Artificial Intelligence Center (JAIC) with a mission "to transform the DOD by accelerating the delivery and adoption of AI to achieve mission impact at scale" and vision to "transform the DOD through Artificial Intelligence." (DoDCIO 2021) The JAIC is currently working to achieve this mission and vision. The JAIC has established five pillars as the DOD strategy for developing and implementing AI. The pillars are shown in Figure 23.

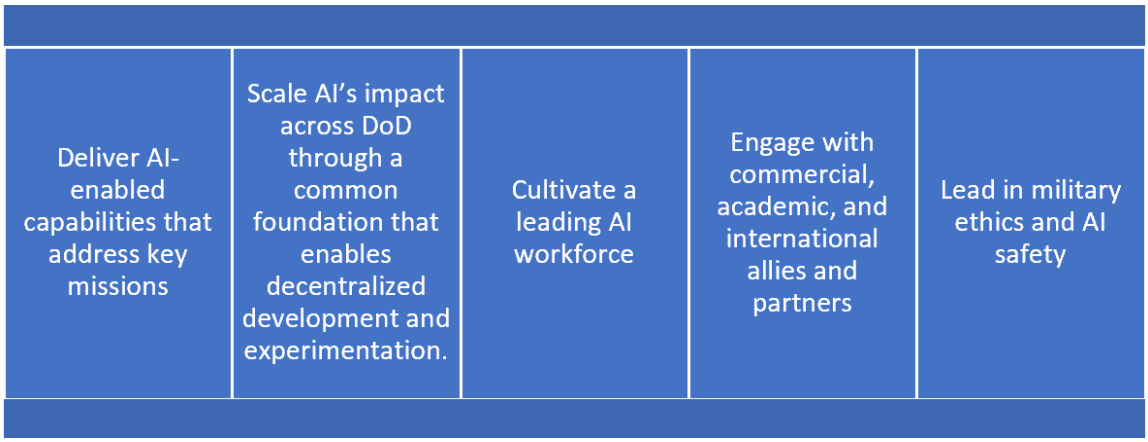



Figure 23. DOD AI Strategy organized in Five Pillars. Source: DoDCIO (2021, sec. Overview).

In order to "deliver AI-enabled capabilities that address key mission" pillars, the JAIC is utilizing AI to "improve situational awareness and decision-making, increase safety of operating equipment, implement predictive maintenance and supply, and to

streamline business processes" (DOD 2019, 7). The JAIC has gathered the expertise to lead the integration of AI across the DOD where lesson learned, tools, frameworks, and standards utilized by the JAIC will be shared across the department providing a common AI foundation. The current focus is on empowering and cultivating the workforce by providing AI technology that will make an immediate impact to the warfighter. The JAIC is also recruiting the best available talent while also providing applicable AI training across the entire workforce. To have the best practices available, the JAIC will utilize resources outside of the DOD such as commercial and academic partners.

To implement the final pillar, the JAIC will lead the military in ethics and AI safety and will use "AI to reduce the risk of civilian casualties and other collateral damage" (DOD 2019, 6). The JAIC following the Law of War will provide AI solutions to aid war fighters in their decision-making process. The JAIC is leading the way to implement the use of ethical principles as shown in Figure 24 in the DOD.


Figure 24. The DOD AI Ethical Principles. Source: DOD (2020).

These principals were recommended by former Secretary of Defense Dr. Mark T. Esper who stated that "the adoption of AI ethical principles will enhance the department's commitment to upholding the highest ethical standards as outlined in the DOD AI Strategy,

while embracing the U.S. military's strong history of applying rigorous testing and fielding standards for technology innovations" (DOD 2020, para. 3). These adopted DOD AI ethical principles are culmination of not only ethical but legal, safety, and policy frameworks. To enforce these frameworks while protecting privacy and civil liberties in all AI development and procurement, the Responsible AI (RAI) principals as listed in Figure 25 are governed by the JAIC for implementation (Polit 2021). The JAIC is also working with the procurement community to utilize an Other Transaction Authority (OTA) approach to help with AI related acquisitions. The JAIC developed the Tradewind OTA business model to incorporate a commitment to RAI into a streamlined acquisition approach that connects the warfighter to industry and academic leaders in AI ("Tradewind | An Acquisition Business Model for AI at the DOD" n.d.).


Figure 25. Responsible AI Principles. Source: Polit (2021).

## 2. Three Waves of AI

The direction of AI technology now and in the future can be categorized in a framework characterized by three waves (Launchbury 2017). These three waves, as categorized by DARPA (and shown in Figure 26), are handcrafted knowledge (Wave 1), statistical learning (Wave 2), and contextual reasoning (Wave 3).

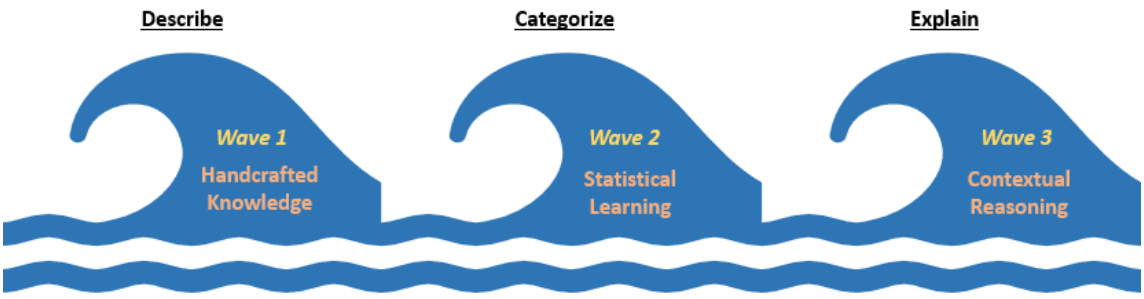



Figure 26. Three Waves of AI. Adapted from Launchbury (2017).

Each wave can be defined by the way information is processed and decisions are made describing (Wave 1), categorizing (Wave 2), and explaining (Wave 3). Additionally, each wave can be parameterized by dimensions of intelligence on a scale of 1–4: the attributes perceiving, learning, abstracting, and reasoning.

- **Perceiving** is the ability to make sense of the outside world by using input to attain an awareness or understanding.
- **Learning** is the ability to gain knowledge or skill through experience of instruction or study.
- **Abstracting** is the ability to take knowledge discovered at a certain level and apply it at another level.
- **Reasoning** is the ability to draw inferences or conclusions using reason.

The overall composition of each wave relative to perceiving, learning, abstracting, and reasoning is shown in Figure 27.

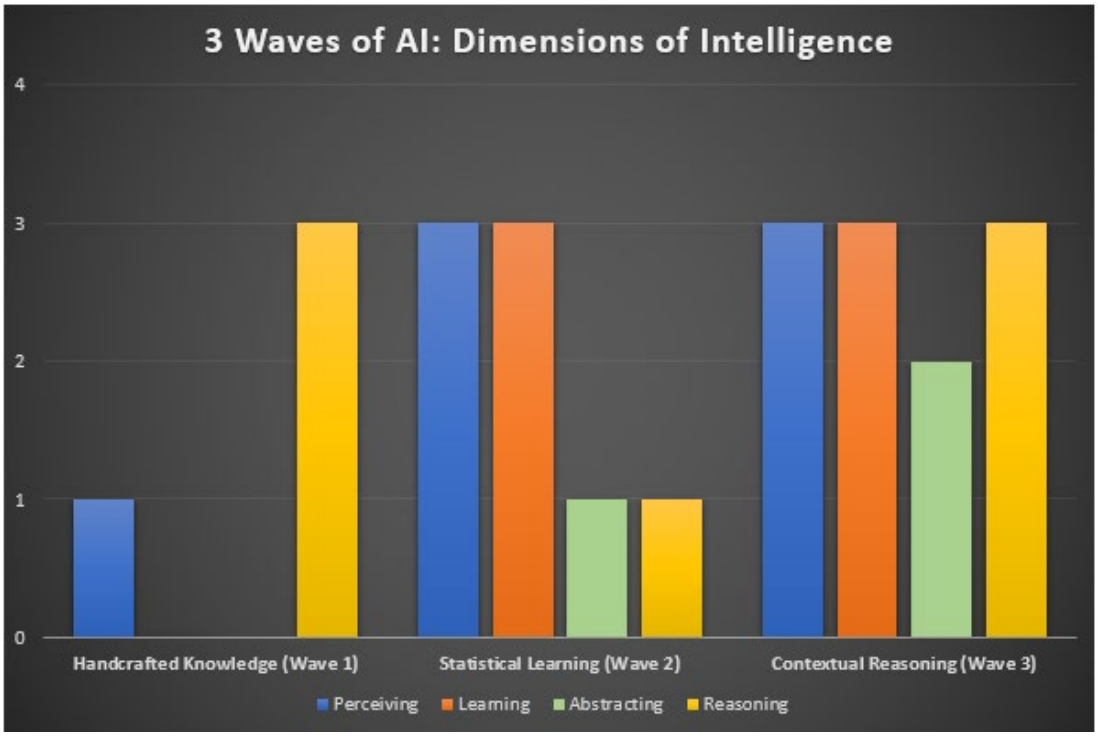



Figure 27. 3 Waves of AI: Dimensions of Intelligence. Source: Launchbury (2017).

The following three subsections provide a high-level description of the three waves of AI. More detailed information about machine learning and methods used in Waves 2 and 3 can be found in Appendix C.

### *a. Wave 1 – Handcrafted knowledge*

Knowledge about a particular domain is characterized by rules that are given to a machine. The machine uses a combination of the rules for decision making and logical reasoning based on particular facts found within a concrete situation. Reasoning is enabled but over very narrowly defined domains with "no learning capability and poor handling of uncertainty" (Launchbury 2017, 10). There are still applications for first wave handcrafted knowledge technologies which are still relevant today and some examples include logistics planning and games like chess. However, these rules-based machine learning technologies start to break down within the natural world. Handcrafted knowledge technologies have difficulties with tackling new situations and thus drive the need for probabilistic decision

making. Figure 28 shows that handcrafted knowledge is primarily based on perceiving and reasoning methods.

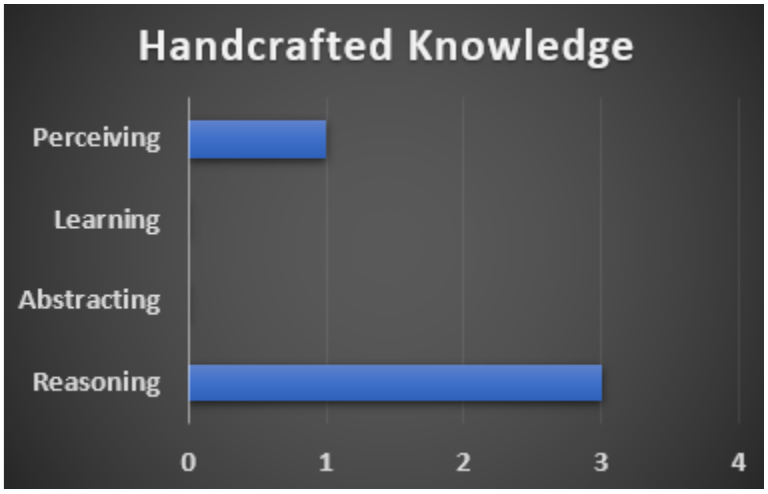



Figure 28. Handcrafted Knowledge: Dimensions of Intelligence. Source: Launchbury (2017).

### *b. Wave 2 – Statistical Learning*

Statistical learning technologies build upon first wave methods and are representative of current mainstream AI technology. Statistical learning technologies utilize big data models that characterize the problem domain to be solved and then train on those models containing specific data. As a result of systems needing to be trained, they can learn to adapt themselves to different situations. This learning is exhibited to a high degree and reflective of the integrity of the information and data provided to the system. Statistical learning systems are exceptionally good at perceiving the natural world but have "no contextual capability and minimal reasoning ability" (Launchbury 2017, 10). Additionally, these technologies exhibit no new capabilities for abstracting and taking knowledge gained in one domain and applying it to another. Examples of statistical learning technologies include voice and facial recognition applications, early neural nets, network routing, and optimizing the sharing of the electromagnetic spectrum. The tools and techniques of statistical learning employed by machine learning algorithms can be categorized as supervised learning, unsupervised learning, or reinforcement learning.

Figure 29 shows that statistical learning is based on perceiving, learning, abstracting, and reasoning.

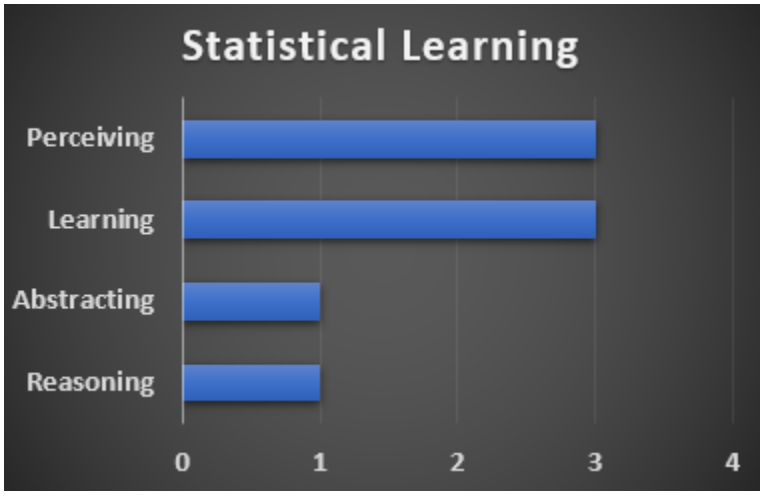



Figure 29. Statistical Learning Dimensions of Intelligence. Source: Launchbury (2017).

### c. *Wave 3* – Contextual Reasoning/Adaptation

Contextual reasoning and adaptation methods represent the future of AI. Contextual adaptation brings together handcrafted knowledge and statistical learning. These systems themselves will "construct explanatory models for classes of real-world phenomena" (Launchbury 2017, 26). For example, a system will be able to classify an image, and construct and provide an explainable model and explanation for its decision criteria. This wave is also called deep learning. Deep learning is in reference to numerous numbers of transformation layers in the model, in contrast Wave 2 model only use 1 or 2 layers.

Contextual reasoning and adaptation technologies will build around contextual models. In some instances, these systems will be constructed from only one or two examples. Over time these systems will learn about how the model should be structured, perceive the world in terms of that model, and use that model to reason. This will enable the system to make decisions and then possibly use that model to abstract and take the data further. Figure 30 shows that contextual reasoning requires perceiving, learning, abstracting, and reasoning.

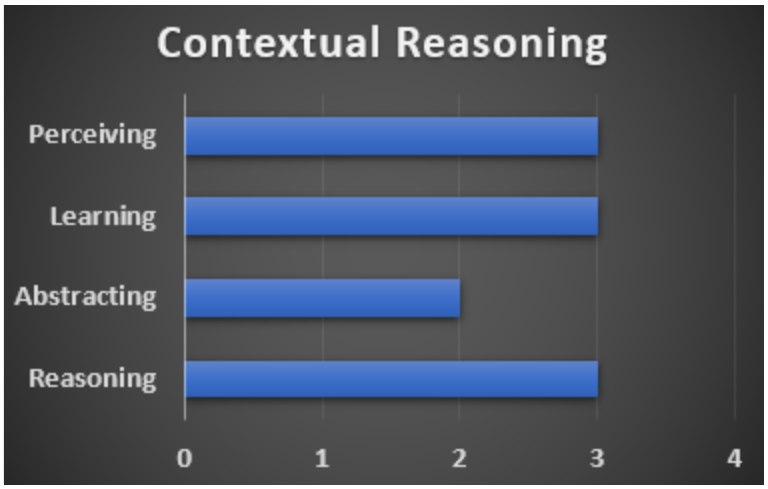



Figure 30. Contextual Reasoning (Wave 3) Dimensions of Intelligence. Source: Launchbury (2017).

### 3. Specialized Topics In AI

This section provides a summary of specialized AI topics identified by the team as areas of interest due to their relevance towards understanding how AI can be leveraged to improve tactical warfighting capabilities. If desired, the reader is encouraged to reference Appendix C for further detail on each topic area. Figure 31 depicts the selected topics; subsections summarizing these topics follow.

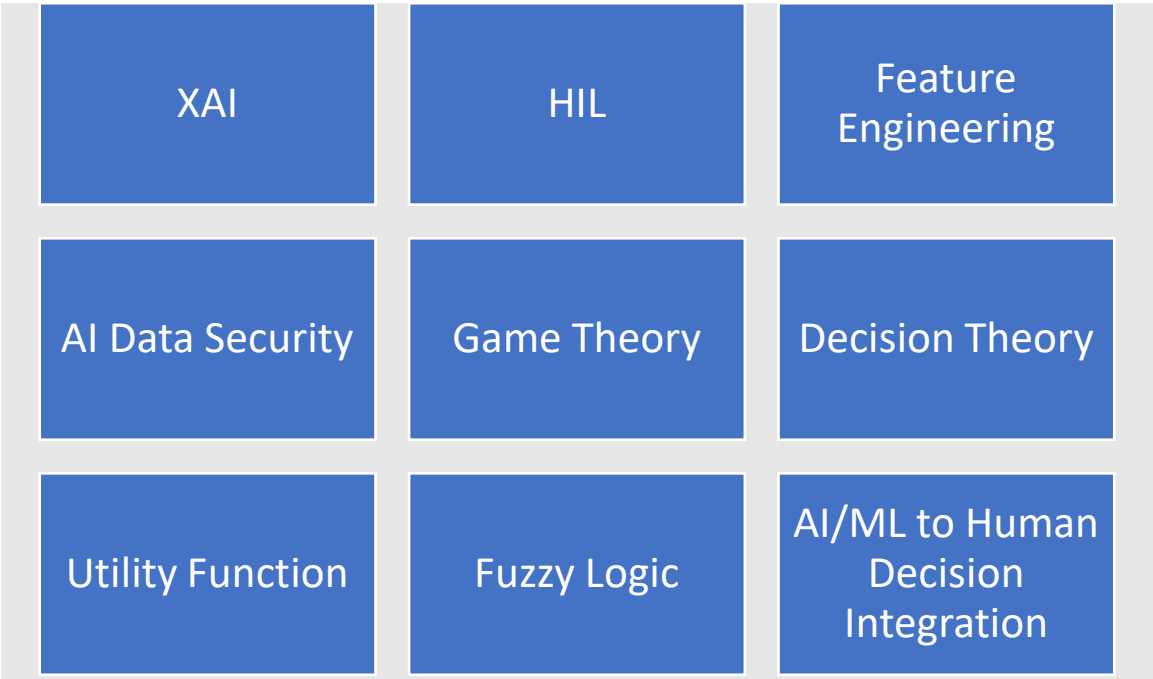



Figure 31. Additional Specialized Topics in AI for the Literature Review

### *a. XAI*

> XAI will create a suite of machine learning techniques that enables human users to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners. ("(PDF) Explainable Artificial Intelligence (XAI)" n.d., 83)

XAI seeks to overcome the black box issues inherent in AI systems by developing models that enable computers the ability to effectively reason with humans. We believe this is critical to the human-machine teaming process and is necessary to establish and increase human trust in AI enabled decision support systems. XAI methods often rely on text or visual aids. Figure 32 presents an overarching view of how XAI can be incorporated into the AI process. Particularly illustrating the addition of two primary components: explainable models and interface, which aim to provide the end user increased understanding of the AI-enabled decision-making process.

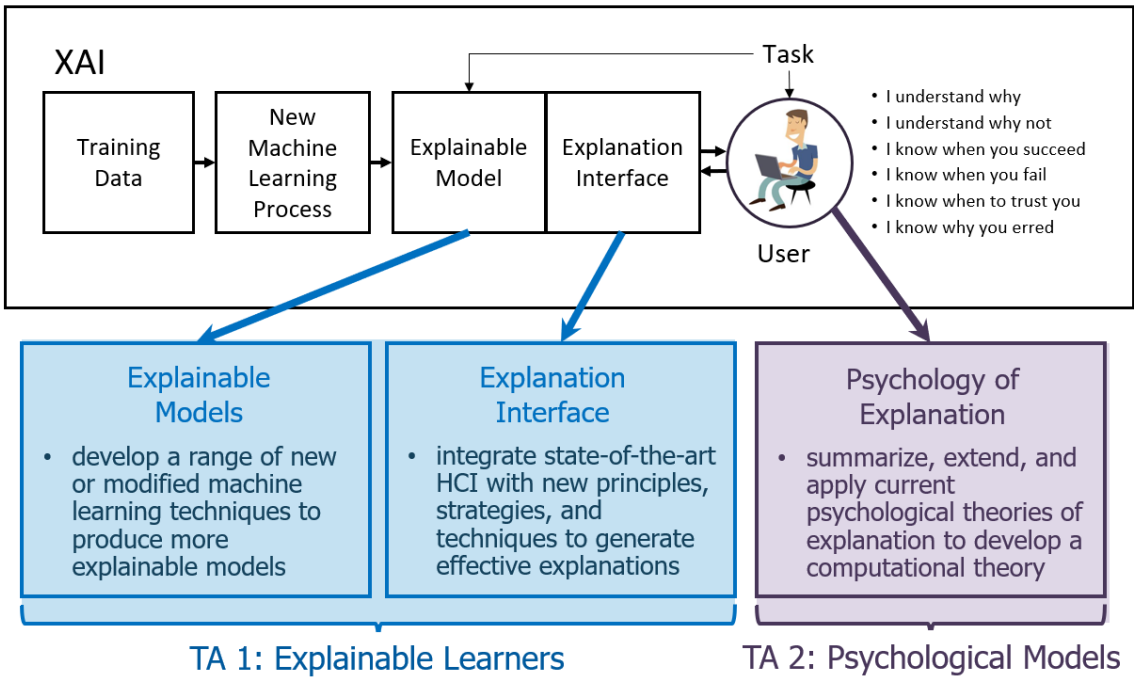



Figure 32. XAI Components. Source: Gunning and Aha (2019, slide 8).

### *b. Human-in-the-Loop Decision Making*

> Autonomy refers to a spectrum of automation in which independent decision making can be tailored for a specific mission, level of risk, and degree of human-machine teaming. (Feickert et al., n.d., 2)

The level of autonomy for an AI system is determined by the amount of human interaction or approval that is required for the system to complete its given mission. Humans are in the loop in AI in machine learning where a human labels the data, tunes the model, and validates a model. Also, humans are in the loop when humans can override decisions and actions performed by AI. The Department of Defense Directive (DODD) 3000.09 (2012) defines the United States Policy on Lethal Autonomous Weapon Systems (LAWS) which outlines the amount of human interaction with a LAWS. DODD 3000.09 defines LAWS as a "weapon system [s] that, once activated, can select and engage targets without further intervention by a human operator." This is a human out of the loop systems where other weapons systems have a human in the loop that selects the final targets. Even though LAWS have full autonomy, all weapon systems must "allow commanders and

operators to exercise appropriate levels of human judgment over the use of force" per DODD 3000.09 (Sayler 2019, 15).

The DOD has historically "stressed the need to keep a human in the loop for automated systems" (Barnett 2020, para. 5); however: there is a push for the military to move "at the speed of relevance" by transitioning to a "human on the loop" approach. This approach would allow AI in a decision-making scenario to start a course of action without human preapproval. The "human on the loop" would still have oversight but would not be the center of the decision-making process allowing the AI system to react and decreasing the time to react to incoming threats (Barnett 2020). As discussed with the USS Vincennes tragedy, too much deference was given to the system which incorrectly identified the passenger jet as a threat. The Enigma code was deciphered using the Bombe device, but it was human understanding that made the greatest advancement in the decoding (Cox 2018). Therefore, the determination of the amount of human interaction is a critical factor in AI.

### *c. Feature Engineering*

In Machine Learning a feature is an attribute of variable used to describe an aspect of an object. Informative features are the bedrock for machine learning. These features are useful for distinguishing distinct groups of objects or describing the underlying object. Note in literature and this paper "Feature" "Variable" and "attribute" are often used as synonyms. According to Dr. Jason Brownlee (2014, para. 15), "Feature engineering is the process of transforming raw data into features that better represent the underlying problem to the predictive models, resulting in improved model accuracy on unseen data."

Machine Learning algorithms are heavily dependent on the availability of relevant, high-quality data. Raw data can be presented in numerous forms and sizes; feature engineering is a necessary data processing step that is related to dimensionality reduction and will directly ensure reduced ML model error, increased learning efficiency and higher quality output.

### *d. AI Data Security*

All the methods described so far in this paper assume that the training data is correct and free of errors. This assumption highlights inherent vulnerabilities open to exploitation by our adversaries. If our enemies "deliberately influence the training data to manipulate the results of a predictive model this is" which Jagielski et al. (2018, 1) defined as "AI data poisoning." During the data poisoning the attackers inject a small number of corrupted points during the training process. This is easier to accomplish if the machine learning model is updated on a regular basis. The "field of adversarial machine learning studies the effect of such attacks against machine learning models and aims to design robust defense algorithms" (Jagielski et al. 2018, 1).

Those working in the field of adversarial machine learning are developing algorithms/models to detect and mitigate these data poisoning attacks. These algorithms fall into one of two categories Noise Resilient Regression or Adversarial-resilient regression. For Noise Resilient Regression the main idea is to identify and remove outliers from a dataset. For Adversarial-resilient regression the main idea is to assume that the data adheres to a certain distribution and detect if the distribution of the dataset falls outside of the assumption. The takeaway from this section is that your algorithm is only as good as the data you feed it and, any presumption of "clean data" must be discarded. Engineers must ensure that security by design is a foundational principle of system development.

### *e. Game Theory*

Game theory is "the study of mathematical models of conflict and cooperation between intelligent rational decision-makers" (Myerson 1997, 1). Rationality refers to the concept that every agent understands that other agents are just as rational. "Agents" is an umbrella term that is interpreted differently depending on the application of the game theoretical concepts. This team's particular interest is in game theoretic applications related to two primary disciplines:

- Strategic interactions between friendly and adversary participants (Wargaming)

- Algorithmic game theory
- In 2018, DARPA announced their AI Next campaign emphasizing a desire/need to accelerate the third wave of AI through multiple key focus areas including the exploration of the use of computational game theory in competitive and adversarial interactions. They introduced Serial Interactions in Imperfect Games Applied to Complex Military Decision-Making (SI3-CMD) program which aims to "extend current AI/game theory techniques to be effective when there are multiple interacting agents, extremely large search spaces, sequential revelation of information, use of deception, continuous resource quantities, stochastic outcomes, and the ability to learn from past iterations" (Uppal, 2020).
- Game theoretic AI-enabled systems have already been funded and fielded by numerous government agencies. Additionally, algorithm designers have utilized principles of game theory to improve AI and ML algorithms with astonishing results. Further detail and examples are provided in Appendix C; however, the key takeaway from this section is that game theoretic principles can directly result substantially more efficient and useful AI-enabled system and, thus, must be explored in greater detail to establish project specific relevance.

### *f. Decision Theory*

Decision theory is very closely related to Game Theory – the primary difference being that while game theory considers the study of how rational agents maximize expected utility in situations where there are multiple agents, decision theory is concerned with how an individual agent can maximize expected utility in situations with no other agents. Otherwise, there are many similarities, for example:

- study of how decisions are made
- how multiple decisions influence each other

- decision-making under uncertainty

There are two branches of decision theory: normative and descriptive. Descriptive decision theory focuses on how decisions are made in reality – it is concerned with observing how decisions are made and identifying potential frameworks or explanations for the way those decisions are made. Normative decision theory focuses on the perfectly rational decision makers – how decisions should be made or what decisions should be made to achieve a certain objective. The optimality of a decision is usually evaluated against some function, known as utility functions which combine concepts of probability and utility theories to determine the overall satisfaction that a decision affords an agent. Utility functions are described in detail in the following section.

### *g. Utility Functions*

Utility functions are used to facilitate an ordering of different alternatives or states by considering an agent's preferences under uncertainty and/or risk. In this regard, utility functions are strongly related to game theory – it enables the assessment of courses of actions (COAs), resulting or predicted agent states, and game outcomes. It is important to highlight the fact that utility functions will vary amongst agents in a game due to differing preferences by virtue of tailored strategies, preferences, capabilities, and perceived payoffs. To define a utility function, risk preference must be considered. There are three types of risk preference:

- **Risk averse** – Agent is reluctant to take on risk.
- **Risk seeking** – Agent is willing to take higher risks to achieve above average returns.
- **Risk neutral** – Agent does not care about the risks involved in decision making.

Military tactical experts, along with other domain experts, will be instrumental to the development of AI-enabled warfighting systems. Utility and risk acceptance will need to be addressed and incorporated within AI systems engineering efforts.

### *h. Fuzzy Logic*

"Fuzzy logic is a method of reasoning that resembles human reasoning" (Das 2020). More specifically, there is an emphasis on multi-valued reasoning that is capable of handling partial truths or degrees of truth. Fuzzy logic enables development of fuzzy inference systems, which are a form of modern AI. Fuzzy logic is based on fuzzy set theory, which assigns elements to sets based on degrees of membership vice assigning elements based on precise properties of membership.

Fuzzy logic aims to mimic more accurately the human reasoning and logic process. It highlights the need for specialized implementation that considers a computer's tendency to determine binary (true or false) outcomes with the natural concept of partial truths present in human reasoning. This is extremely relevant with regards to tactical warfighting decision-making and must be considered concurrently with utility and risk as many warfighting scenarios do not offer the time required to determine a "perfect" solution.

### *i. AI/ML to Human Decision Integration*

Decision making is becoming increasingly complex yet continually reliant on outdated methods. Organizations are working to integrate AI methods into the decision-making process. "In the next four years, 69% of what a manager currently does will be automated. In such a disruptive environment, enterprises need a reality check on how best they can integrate AI into their strategy and be ready for forthcoming disruptions" (Hippold 2020, para. 15). Integrating AI methods into an organization's decision-making process must be done, so developing a method that maps the appropriate technology into the decision-making process is ideal. "Consider what kind of data you need, what data you could exploit, what pieces of the decision making are best left to humans and what should be handled by machines. And determine the collaborations that are critical, rather than what you can manage" (Rollings 2021, para. 8). For mapping AI/ML methods to the decision-making process, research identified nine key areas that provide a framework:

1. Efficacy Characteristics
2. Decision Factors

3. Degree of AI in Decision Making

4. Human/Machine Decision and Solution Complexity

5. Accuracy and Interpretability

6. Engagement Level

7. Functional Roles

8. Event Descriptors

9. Data Characteristics

"In a recent survey, Gartner found that 65% of decisions made are more complex (involving more stakeholders or choices) than they were two years ago. The current state of decision making is unsustainable" (Rollings 2021, para. 8).

This framework is outlined in greater detail in Appendix C and should, in combination with the other identified specialized topics, be considered in future efforts to assess applicability of holistic AI approaches to solving combat challenges such as though highlighted in the Introduction section of this paper. Many of these specialized topics were outside of the scope this team's analysis, however, remain extremely critical to future work in engineering AI-enabled combat systems.

## B. AI ANALYTICAL FRAMEWORK

### 1. AI Methods for Kill Chain Mapping

The following ML methods (Figure 33) will be assessed for applicable mapping to kill chain functions in Chapter IV:

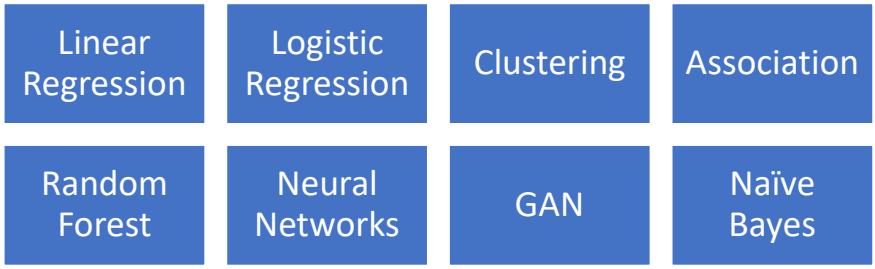



Figure 33. ML Methods

This team selected this set of methods based on the belief that they enabled a comprehensive analysis of AI benefits to kill chain functions while also bounding the scope of such analysis to ensure meaningful progress within this study. While this list is by no means fully exhaustive, it should serve as a foundational piece (at the very least) to future efforts of much more in-depth analysis of specific ML algorithms.

***a. Linear Regression***

- In this method Linear relationships are identified between input predictors (measures) (X) and output variables (Y) to develop functions that create predictive models. Response values are numerical. Discussed in greater detail in section V.D.1.a(4)

***b. Logistic Regression***

- In this method relationships and predictive models are developed based on a categorical response. Discussed in greater detail in section V.D.1.a(5)

***c. Clustering***

- This method which separates datasets into groups based on similarities in properties and/or features which potentially uncover meaningful relationships and patterns amongst samples in the dataset.

### *d. Association*

- This method examines large datasets to find relationships between variables. The output can be a series of if/then statements related to the response variable of interest.

### *e. Random Forest*

- This method uses a predictive "decision-tree" algorithm to partition the training set predictor space into multidimensional boxes. Generating many trees using random subsets of the input features Bias is minimized by growing large trees and variance is reduced by averaging the results over the trees (bagging).

### *f. Neural Networks*

- This method maps inputs (features) to outputs (responses) using complex transformations as the data traverses through the multiple layers.

### *g. GAN*

- In this method learning is accomplished using two networks facing off against one another. The results of this faceoff can inform the dynamics of future models.

### *h. Naïve Bayes*

- This method uses repeated application of bayes rule to bring to predict classifications based upon probabilistic inference.

## 2. AI Evaluation Factors

The analysis will center of the following four decision points:

1) What is the type of the required output?

- **Quantitative:** The output/response contains real number values. Infinite number of values.

- **Qualitative:** The output/response consists of categorical data. This is data that has been assigned to one of the predefined categorical values (numbers or strings).
- **Clusters:** The output/response consists of clusters of data that is grouped by strongly associated qualities. This is useful when trying to find patterns in the dataset.
- **Rules:** The output/response consists of a series of if/then rules. One Common application of this is recommender systems.

2) What is the type of learning required?
   - **Supervised:** A fully labeled dataset containing the predictors and response variables will be available to the method for training and model creation.
   - **Unsupervised:** No response variables, only predictors in the dataset used for training and model creation.
   - **Reinforcement:** Fully labeled dataset not available (partial or non-labeled available), general rules are defined such that the method can generate feedback as it learns.
3) What level of XAI is required?
   - **XAI mandatory:** The output of the method must include or allow an easy translation to an explainable output.
   - **XAI desired:** The output of the method may include or allow an easy translation to an explainable output, but it is not required.
   - **XAI not needed:** The output of the method is not required to have the ability to explain its reasoning for generating the response.
4) How many predictors? This is the number of input features.
   - 1 – 9

- 10 – 99
- 100+

The method of scoring has been formulated as a simplistic representation of three distinct criterion based on a method's suitability for achieving a kill chain function's desired output. The three criteria are:

1. method is well suited for the task
2. method is suitable, however may be suboptimal
3. method is ill suited for the task

Using the prescribed single function analysis, these decision points and scoring criteria allow the team to present an uncomplicated scorecard type of output artifact that will aid the user in assessing the applicability of a specific method to a specific kill chain function. The desire is that the scorecard can then be presented during conceptual engineering reviews and serve as a foundation for explaining AI method reasoning to an audience with various levels of technical understanding. Also, future efforts that may expand on this analysis can grow the scope of the scorecard output as applicable to user needs.

THIS PAGE INTENTIONALLY LEFT BLANK

# IV. MAPPING OF AI METHODS TO THE KILL CHAIN

This chapter presents the methodology used for developing evaluation and decision criteria of AI/ML methods for mapping them to kill chain functions for use in designing future systems. This includes establishing a modeling framework and an example single function analysis to characterize information flow through the kill chain. The team's approach to mapping AI methodology to the AMD kill chain was based on the following sequence (Figure 34):


Figure 34. AI/ML Kill Chain Mapping Methodology

This process began with an assessment of kill chain functional processes and event tracing building off what that the AI-OODA team concluded. Figure 35 represents the methodology and thought process from the perspective of engineering the system. It represents the fusion of the kill chain framework and AI/ML method application to points of integration within the kill chain decision-making process. The top of the graphic illustrates the kill chain framework. Below that it is mapped to points of decision integration along the kill chain continuum where decisions are made. These decisions are characterized by the criteria in the block. As information and data flow, at the human and machine levels, through each phase of the kill chain, decision integration criteria are continually and iteratively addressed. Because these are continually and iteratively addressed, so too are the variations of AI methods that can be used to augment, enhance, or automate these decisions. AI methods were initially characterized using the criteria in the blocks below in order to develop the final evaluation and scoring criteria needed to create a mapping schema.

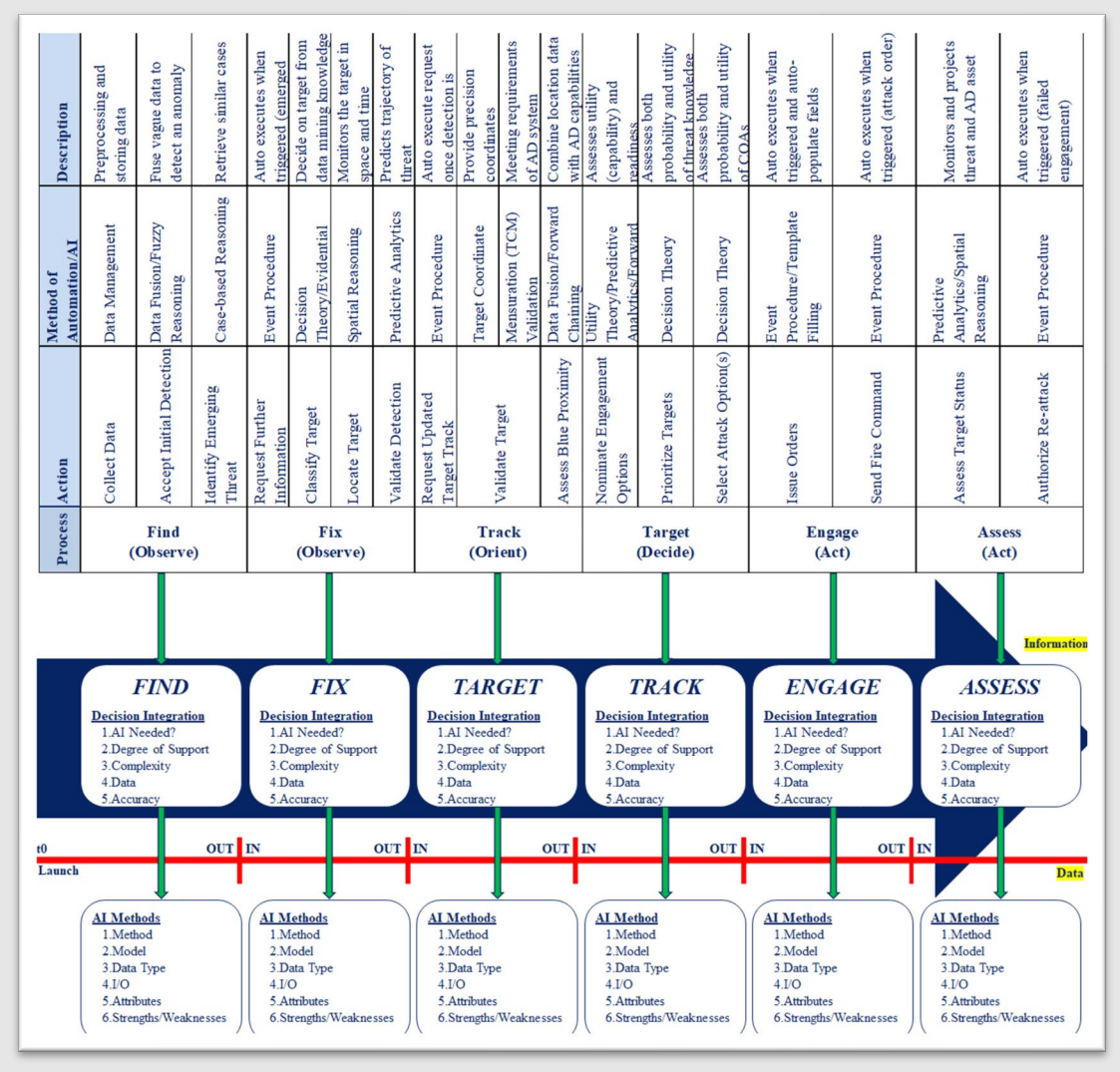



Figure 35. Functional Methodology and Analysis. Adapted from Jones et al. (2020, tab.-17).

## A. MODELING FRAMEWORK

Due to the inherent complexity of the AMD kill chain, the team's approach to establishing model framework was to break down kill chain structure into single functions for further evaluation. This process, referred throughout this report as single function analysis, included breaking down all 28 identified kill chain functions at lower levels to understand the events and communications pathways necessary to execute each function. Using the definitions cited in the appendix as the starting point, the following section will demonstrate this methodology through the analysis of function 1.2.4-*Identify*.

### 1. Single Function Analysis

The Identify function occurs during the Fix step of the kill chain and is executed in parallel to the Classify and Locate functions depicted in Figure 36.

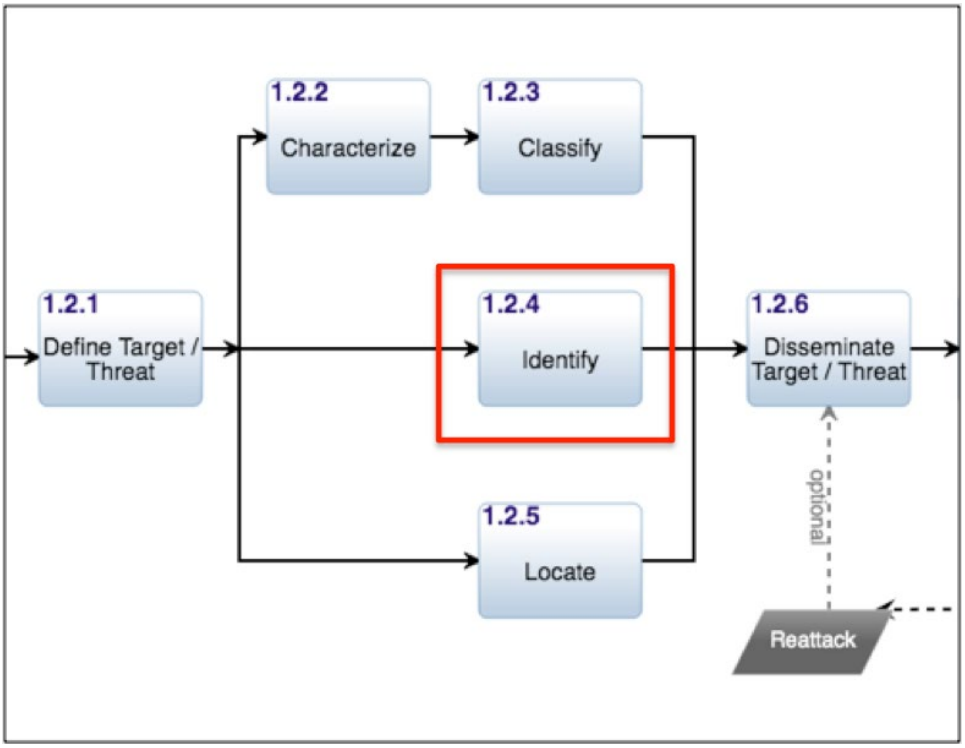



Figure 36. Kill Chain Fix Step

At this point in the kill chain, initial detection has already occurred and the C2 system executing the kill chain has begun the Define Target / Threat function which involves the processing of an "emerging target." An emerging target is defined by JP 3-60 as a detection that "meets sufficient criteria to be evaluated as a potential target" and typically requires further analysis for validation. The validation required facilitates the need for functions 1.2.2 through 1.2.5 (Joint Chiefs of Staff 2019).

The Identify function itself encompasses a significant layer of processes required to facilitate proper identification (ID). As a component of the greater Target Acquisition (TA) and ID process outlined in JP 3-09: Joint Fire Support, ID of a potential target can happen through a variety of means whether that be as simple as a visual confirmation of a

potential target, or something derived from electronic signature or sensors, such as radar detection (Joint Chiefs of Staff 2019). In most operational environments, there are several ID processes that run in parallel that all contribute to execute the Identify function. The Multi-Service Tactics, Techniques, And Procedures for Air and Missile Defense publication developed by the Air Land Sea Application (ALSA) Center outlines three methods of ID (ALSA Center 2019):

***a. Positive Identification (PID) or Procedural ID***

- PID: Derived through visual recognition, point of origin, electronic support systems, Identification Friend or Foe (IFF) systems, or other physics-based ID techniques.
- Procedural ID: Derived through compliance to established airspace coordination measures or rules.

***b. ID Criteria and Symbols Retained from ID Authorities***

- ID Criteria: Attributes and characteristics of a track enabling determination of its nature and classification (Approved by the JFC as part of the Area Air Defense Plan).
- Symbols: Seven track classifications (Pending, Unknown, Neutral, Assumed Friend, Friend, Suspect, and Hostile).

***c.*** **Auto-ID Systems:** Weapon systems with embedded Auto-ID functions (AEGIS, ***PATRIOT, etc.)***

Some of these elements and the level of complexity that they entail are demonstrated in the various tools used by watch personnel in operations. One example is the sample ID Matrix presented by ALSA shown in Figure 37 (ALSA Center 2019).

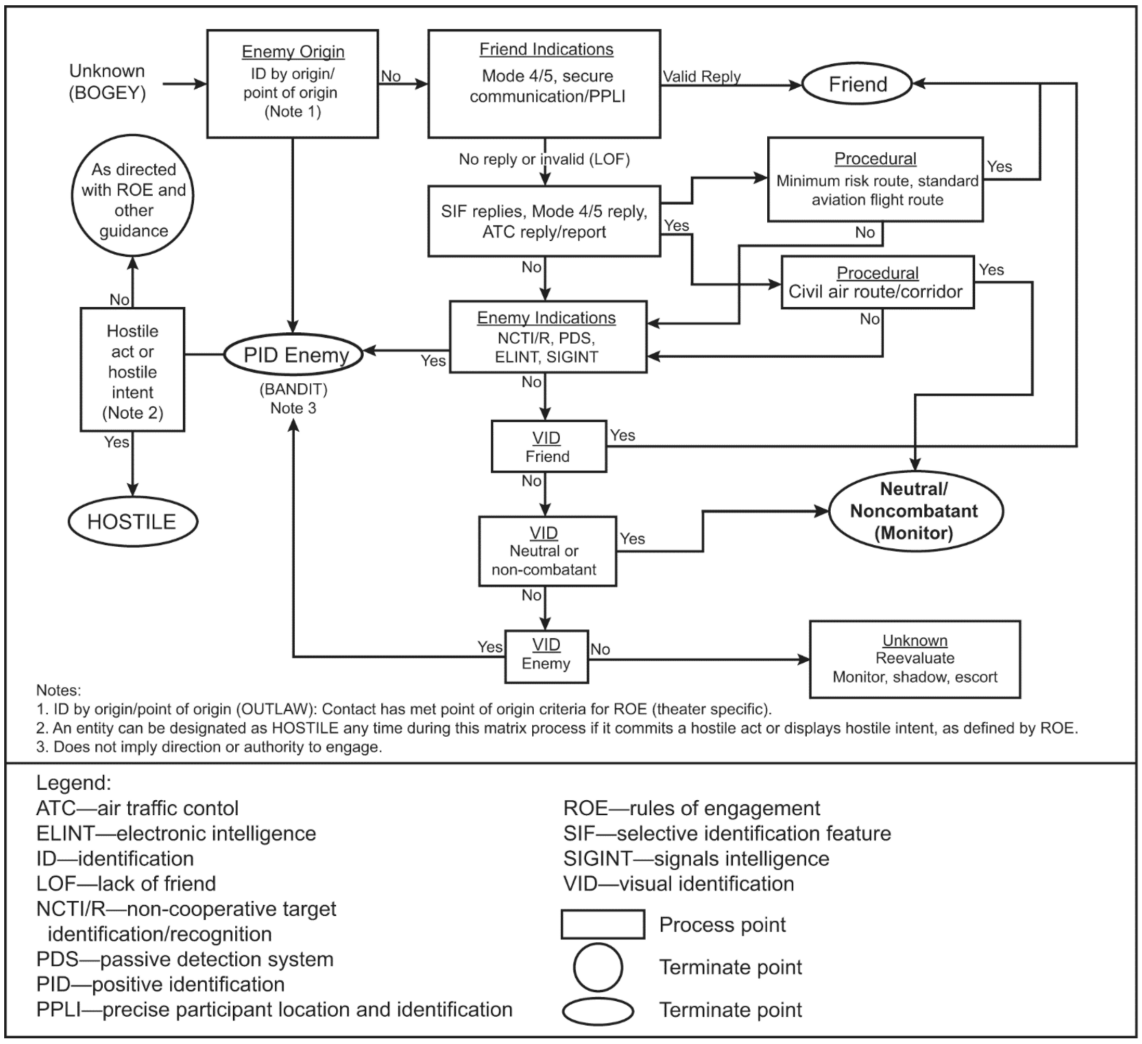



Figure 37. ID Matrix. Source: ALSA Center (2019).

This matrix is meant to be a logic tree that can be used to assist personnel in the ID of an unknown track that is picked up by a sensor. After initial detection, the unknown track is designated as a "Bogey" and watch personnel can cross examine the characteristics of the track with the criteria established in the matrix (ALSA Center 2019). This process implies multifaceted layers of decision making and criteria that are potentially required to validate a single function within the kill chain.

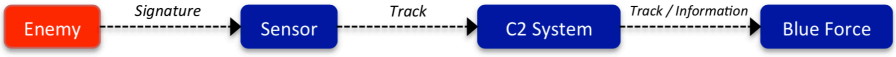
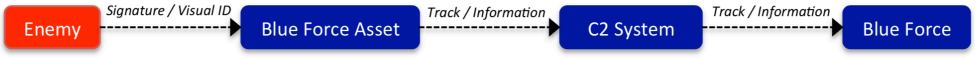
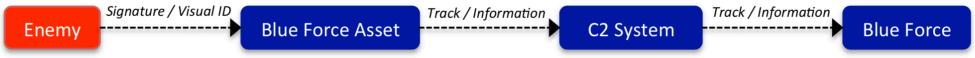
From this analysis, the team concluded that even a single function offered high levels of complexity beyond basic execution. Thus, the same Single Function Analysis process was applied to all 28 functions with the team determining the need to further

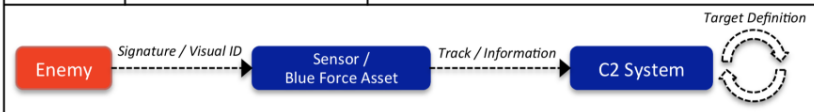
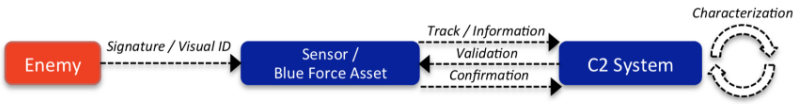
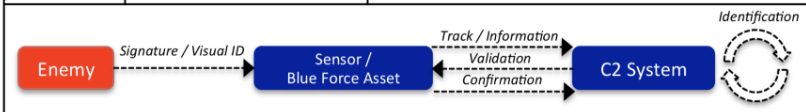
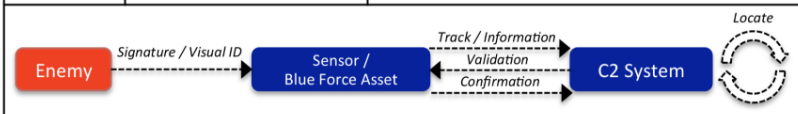
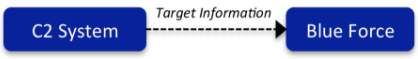
understand the data flow corresponding to each step as the kill chain is executed. As the next step of developing the modeling framework, the team conducted analysis specific to the data flow and event tracing of each function. The next section will go into detail on the model developed from this process.

## 2. Event and Data Tracing

The event and data tracing process builds on the characteristics derived from the Single Function Analysis. As each function presents a unique layer of detail, the team found it beneficial to map out the data being processed in and out as each step is being executed. The goal of this analysis was to gain a higher level of understanding of the various components involved with each kill chain function and the general flow of information as each event progresses. A sample of the team's findings are presented in Table 7 with the rest of the analysis provided in the appendix at the end of the report.

Table 7. Event and Data Tracing

| Step | Number | Function | Definition |
|---|---|---|---|
| Find | 1.1.1 | Initial Detection | Detection of enemy units by radar or other military equipment. |
| | |  | |
| | 1.1.2 | Battle Damage Assessment (BDA) Detection | Detection of enemy units as a result of BDA conducted by assets or capabilities. |
| | |  | |
| | 1.1.3 | Re-Task Detection | Detection of enemy units conducted by assets or capabilities that have been diverted. |
| | |  | |

| | | | |
|---|---|---|---|
| Fix | 1.2.1 | Define Target/Threat | Target or threat definition is initiated. |
| |  | | |
| | 1.2.2 | Characterize | Target or threat is characterized as a result of data fusion and confirmation of identity and location. |
| |  | | |
| | 1.2.3 | Classify | Sensors are focused to confirm initial track classification of target or threat. |
| |  | | |
| | 1.2.4 | Identify | Sensors are focused to confirm identity of target or threat. |
| |  | | |
| | 1.2.5 | Locate | Sensors are focused to confirm the precise location of target or threat. |
| |  | | |
| | 1.2.6 | Disseminate Target /Threat | Target or threat information is disseminated to Blue Force via command and control system. |
| |  | | |

Notice that this analysis builds on the information provided during the OV-6c. However, whereas the OV-6c provided a high-level event trace of a particular scenario, this table goes into specific detail on each function.

This analysis allowed the team to identify how information flowed through the components of the kill chain and presented intersection points that could be further assessed for mapping suitability in development of the final model. As an example, looking again at the Identify function shows that there are several opportunities for mapping on specific components and various levels of decisions that are being made. When information flows into the Blue Force C2 system it arrives as a track that must be processed for targeting. As the team learned in the Single Function Analysis, this process has several layers involved to drive decision making therefore presenting a large opportunity for potential mapping of suitable AI methodology. As the focus of the next section, the team expanded on the concept of decision making as a driving factor in determining mapping suitability.

## B. AI/ML APPLICABILITY AND DECISION INTEGRATION

When addressing AI/ML applicability as shown in Figure 38, the decision-makers need to decide if AI or ML is required and to what level of support will be required from these tools and methods. Expanding on what was introduced in the previous chapter and illustrated in Figure 69, AI/ML can either provide the decision-maker support or not, augment his decision-making process or provide (almost) full automation as shown in Figure 38. Determining what level of AI/ML can help determine which method will be mapped to the kill chain. AI/ML method mapping will be covered in a later section of this chapter.

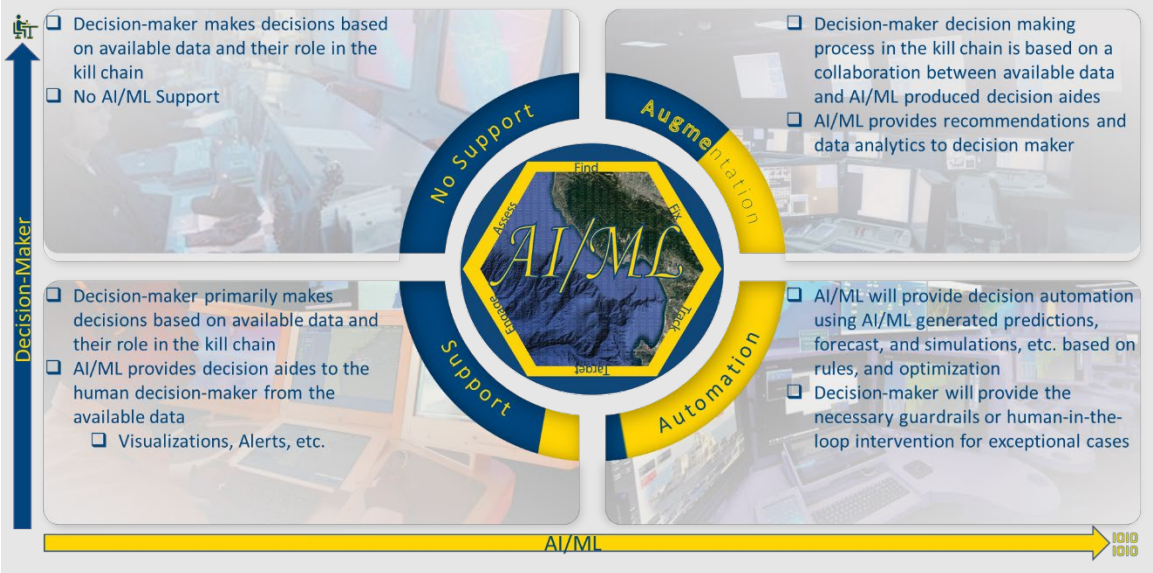



Figure 38. AI/ML Applicability to Decision Making. Adapted from Starita (2021b).

In the four quadrants shown in Figure 38, each quadrant depicts the level of the human-in-the-loop decision makers dependence on AI/ML methods. When no AI/ML support is required, the decision maker in the kill chain relies on the available data without the use of AI/ML decision aides. AI/ML can provide support to the decision maker by providing alerts and visualizations to aide in the process. The next level would be AI/ML augmentation which will provide kill chain recommendations and data analytics to the decision maker. To move "at the speed of relevance," automation will provide kill chain decisions based on AI/ML generated predictions, forecast, and simulations based on the optimal predicted outcome and rules of the kill chain. However, in all quadrants there will be a human in the loop option that "allows commanders and operators to exercise appropriate levels of human judgment over the use of force" (Sayler 2019, 15).

## C. AI/ML METHODS MAPPING CRITERIA

Common Artificial Intelligence and Machine Learning Methods were discussed in detail in Chapter III with additional detail provided in Appendix C as applicable. As outlined at the end of Chapter III, the analysis framework is on that seeks to pose a series of questions in the form of decision points to the user; the answers to these questions will

highlight the optimal AI/ML methods for a particular function. These decision points along with all answers will be discussed in the following sections. Each AI/ML method will be assigned a score at each decision, the summation of scores for each method will facilitate ranking of the possible AI/ML Methods.

The following AI/ML Methods will be scored in this section:

- linear regression
- logistic regression
- clustering
- association
- random forest
- neural networks
- GAN
- naïve bayes

Table 8. Scoring Criteria

| Score | Details |
|---|---|
| +1 | Method is well suited for the task |
| 0 | Method is suited for the task but sub optimal |
| -1 | Method is ill-suited for the task |

## 1. Decision Point #1

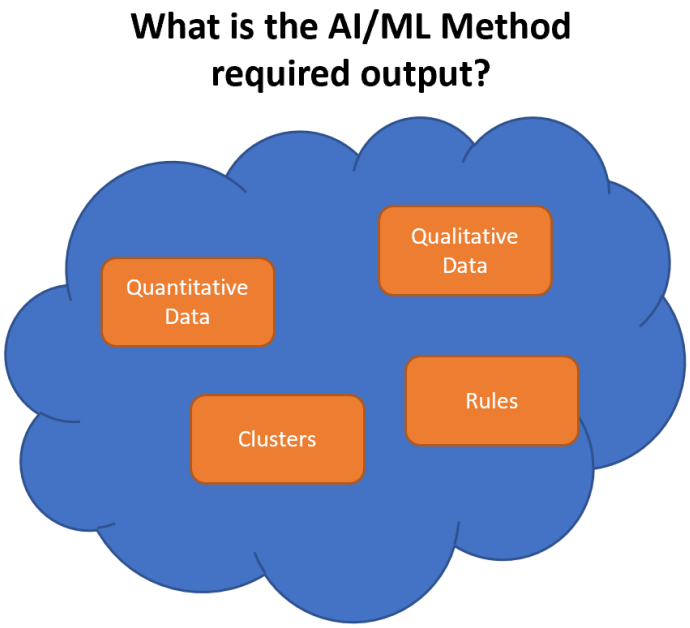



Figure 39. Decision Point #1 Options

### *a. Quantitative Data (Decision Point #1)*

The output/response contains real number values. Infinite number of values.

Table 9. Decision Point #1: Answer = Quantitative Data

| Method | Score |
|---|---|
| Linear Regression | +1 |
| Logistic Regression | 0 |
| Clustering | 0 |
| Association | 0 |
| Random Forest | +1 |
| Neural Networks | +1 |
| GAN | 0 |
| Naïve Bayes | 0 |

- Linear Regression, Random Forest and Neural networks can generate an output that is quantitative in its raw form.

- Logistic Regression and Naïve Bayes generate a categorical or probabilistic response that must be transformed into a quantitative number.
- Clustering is looking for patterns and treats the response the same as one of the predictors.
- GAN, if programmed properly, can develop a quantitative output but when compared to the other methods a great deal of additional processing is required. (Due to the adversarial networks used)

### *b. Qualitative Data (Decision Point #1)*

The output/response consists of categorical data. This is data that has been assigned to one of the predefined categorical values (numbers or strings).

Table 10. Decision Point #1: Answer = Qualitative Data

| Method | Score |
|---|---|
| Linear Regression | 0 |
| Logistic Regression | +1 |
| Clustering | +1 |
| Association | +1 |
| Random Forest | +1 |
| Neural Networks | +1 |
| GANs | 0 |
| Naïve Bayes | +1 |

- Naïve Bayes, Logistic regression and Random Forest will output a response that is categorical or probabilistic in its raw form.
- Clustering will assign the response variable based upon the assigned cluster.
- GAN if programmed properly can develop a qualitative output but when compared to the other methods a great deal of additional processing is required. (Due to the adversarial networks used)

### *c. Clusters (Decision Point #1)*

The output/response consists of clusters of data that is grouped by strongly associated qualities. This is useful when trying to find patterns in the dataset.

Table 11. Decision Point #1: Answer = Clusters

| Method | Score |
|---|---|
| Linear Regression | -1 |
| Logistic Regression | 0 |
| Clustering | +1 |
| Association | 0 |
| Random Forest | 0 |
| Neural Networks | 0 |
| GANs | 0 |
| Naïve Bayes | 0 |

- Clustering will generate a response that in its raw form assigned the data to clusters
- Logistic Regression, Association, Random Forest, Neural Networks and Naïve bayes can generate a categorical output which can then be used to cluster the data
- Linear Regression will generate a quantitative response which will be difficult to assign to categories and cluster.
- GAN if programmed properly can develop a qualitative output but when compared to the other methods a great deal of additional processing is required. (Due to the adversarial networks used)

### *d. Rules (Decision Point #1)*

The output/response consists of a series of if/then rules. One Common application of this is recommender systems.

Table 12. Decision Point #1: Answer = Rules

| Method | Score |
|---|---|
| Linear Regression | -1 |
| Logistic Regression | -1 |
| Clustering | -1 |
| Association | +1 |
| Random Forest | -1 |
| Neural Networks | -1 |
| GANs | -1 |
| Naïve Bayes | -1 |

- Except for association all the methods listed above will generate quantitative/qualitative/cluster data that is ill suited to defining a set of if/ then statements
- Association by its nature will generate a response that is a series of if/then rules.

**2. Decision Point #2**

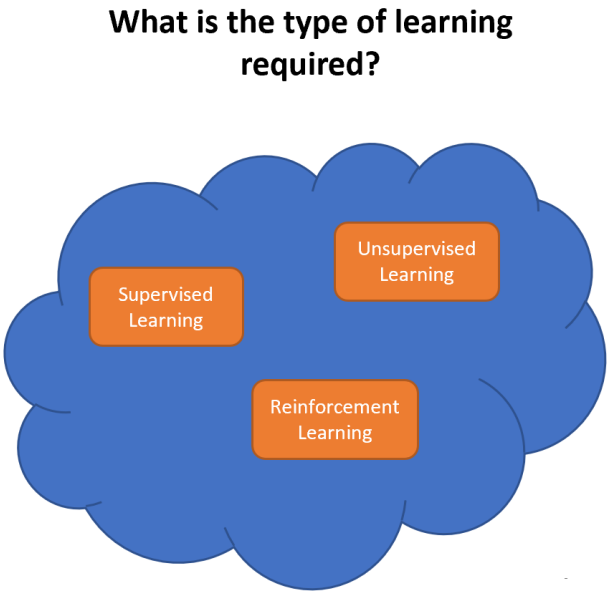



Figure 40. Decision Point #2 Options

### *a. Supervised Learning (Decision Point #2)*

A fully labeled dataset containing the predictors and response variables will be available to the method for training and model creation.

Table 13. Decision Point #2: Answer = Supervised Learning

| Method | Score |
|---|---|
| Linear Regression | +1 |
| Logistic Regression | +1 |
| Clustering | 0 |
| Association | +1 |
| Random Forest | +1 |
| Neural Networks | +1 |
| GANs | +1 |
| Naïve Bayes | +1 |

- Clustering does not require a full labeled (predictors and response) training set. It can operate and cluster with the response considered as another predictor.
- The remaining methods all perform optimally when using a labeled dataset.

### *b. Unsupervised Learning (Decision Point #2)*

No response variables, only predictors in the dataset used for training and model creation.

Table 14. Decision Point #2 Unsupervised Learning

| Method | Score |
|---|---|
| Linear Regression | -1 |
| Logistic Regression | -1 |
| Clustering | +1 |
| Association | -1 |
| Random Forest | -1 |
| Neural Networks | -1 |
| GANs | -1 |
| Naïve Bayes | -1 |

- Clustering is the only method that is well suited to working with a training set that lacks the response variable.
- The remainder of the methods rely on predicting the response thus are ill suited to work without a response in the data of the training set.

### *c. Reinforcement Learning (Decision Point #2)*

Fully labeled dataset not available (partial or non-labeled available), general rules are defined such that the method can generate feedback as it learns.

Table 15. Decision Point #2: Answer = Reinforcement Learning

| Method | Score |
|---|---|
| Linear Regression | 0 |
| Logistic Regression | 0 |
| Clustering | 0 |
| Association | 0 |
| Random Forest | 0 |
| Neural Networks | +1 |
| GANs | +1 |
| Naïve Bayes | 0 |

- GANs and Neural networks implement feedback mechanisms in their base algorithms thus are ideally suited.
- The remaining methods can be used but require external framework to implement the feedback network required.

### 3. Decision Point #3

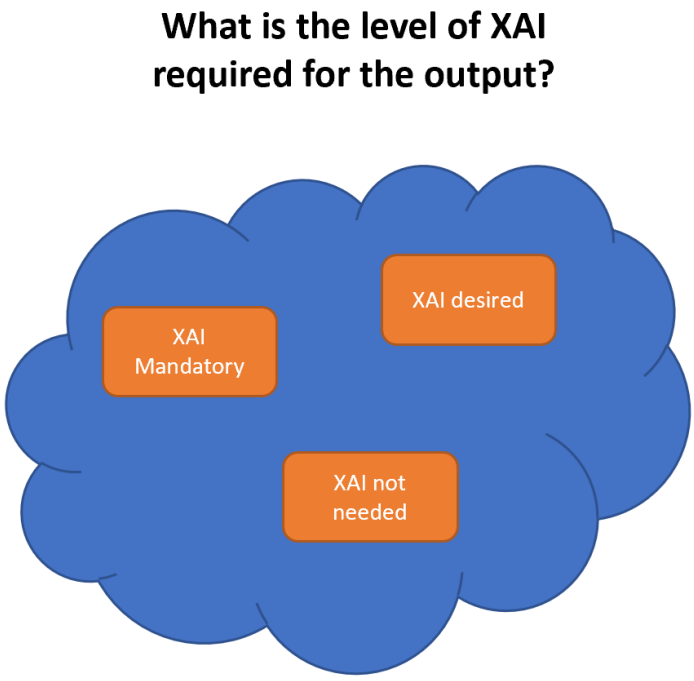



Figure 41. Decision Point #3 Options

#### *a. XAI mandatory (Decision Point #3)*

The output of the method must include or allow an easy translation to an explainable output.

Table 16. Decision Point #3: Answer = XAI Mandatory

| Method | Score |
|---|---|
| Linear Regression | +1 |
| Logistic Regression | +1 |
| Clustering | +1 |
| Association | +1 |
| Random Forest | 0 |
| Neural Networks | -1 |
| GANs | -1 |
| Naïve Bayes | 0 |

- Linear Regression, Logistic Regression, Clustering and Association operate using algorithms whose innerworkings are described in detail and can be used to create an explainable reason a response variable was chosen.
- Random Forest and Naïve bayes can generate explainable reasoning, but it may be too complicated to be of use.
- Neural Networks and GANs both use neural networks whose inner workings are nominally hidden from the method. Thus, making an explanation obtuse.

### *b. XAI desired (Decision Point #3)*

The output of the method may include or allow an easy translation to an explainable output, but it is not required.

Table 17. Decision Point #3: Answer = XAI Desired

| Method | Score |
|---|---|
| Linear Regression | +1 |
| Logistic Regression | +1 |
| Clustering | +1 |
| Association | +1 |
| Random Forest | +1 |
| Neural Networks | 0 |
| GANs | 0 |
| Naïve Bayes | +1 |

- Similar reasoning to that shown in the previous section.

### *c. XAI not required (Decision Point #3)*

The output of the method is not required to have the ability to explain its reasoning for generating the response.

Table 18. Decision Point #3: Answer = XAI Not Required

| Method | Score |
|---|---|
| Linear Regression | +1 |
| Logistic Regression | +1 |
| Clustering | +1 |
| Association | +1 |
| Random Forest | +1 |
| Neural Networks | +1 |
| GANs | +1 |
| Naïve Bayes | +1 |

- All methods equally weighted if XAI not required, thus this decision point contains no information.

## 4. Decision Point #4

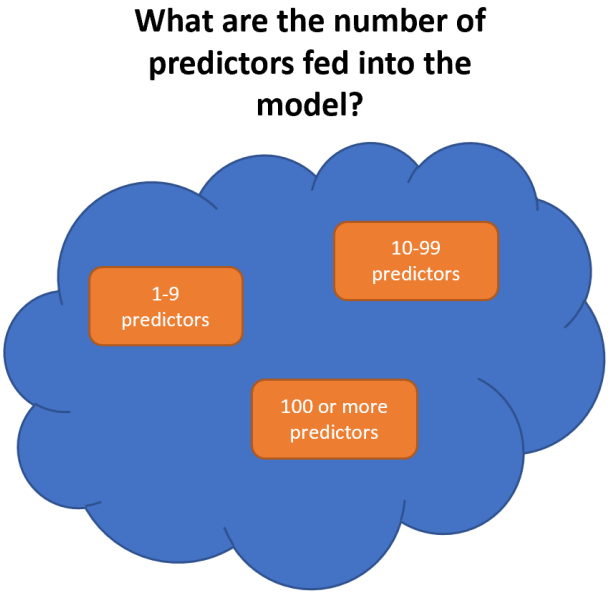



Figure 42. Decision Point #4 Options

### *a. 1-9 predictors (Decision Point #4)*

Table 19. Decision Point #4: Answer = 1–9 Predictors

| Method | Score |
|---|---|
| Linear Regression | +1 |
| Logistic Regression | +1 |
| Clustering | +1 |
| Association | +1 |
| Random Forest | +1 |
| Neural Networks | +1 |
| GANs | +1 |
| Naïve Bayes | +1 |

- All the methods are well suited to operating with 1–9 predictors. Thus, this decision point contains no information.

### *b. 10-99 predictors (Decision Point #4)*

Table 20. Decision Point #4: Answer = 10–99 Predictors

| Method | Score |
|---|---|
| Linear Regression | 0 |
| Logistic Regression | 0 |
| Clustering | 0 |
| Association | 0 |
| Random Forest | 0 |
| Neural Networks | +1 |
| GANs | +1 |
| Naïve Bayes | 0 |

- Neural Networks and GANs are well suited to operate with 10–99 predictors.
- The remaining methods can operate using 10–99 predictors, but the output/response may suffer.

### *c. 100 or more predictors (Decision Point #4)*

Table 21. Decision Point #4: Answer = 100 or More Predictors

| Method | Score |
|---|---|
| Linear Regression | -1 |
| Logistic Regression | -1 |
| Clustering | -1 |
| Association | -1 |
| Random Forest | -1 |
| Neural Networks | +1 |
| GANs | +1 |
| Naïve Bayes | -1 |

- Neural Networks and GANs are well suited to operate with 100 or more predictors.
- Due to the high number of predictors the remaining models are ill suited to handle that many predictors.

## 5. Example Score Generation

For example: If the user choses the following decision points.

- Decision Point #1 = Qualitative Data
- Decision Point #2 = Supervised Learning
- Decision Point #3 = XAI required
- Decision Point #4 = 10–99 inputs

Table 22. Example Scorecard

| AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
|---|---|---|---|---|---|
| Linear Regression | 0 | 1 | 1 | 0 | 2 |
| Logistic Regression | 1 | 1 | 1 | 0 | 3 |
| Clustering | 1 | 0 | 1 | 0 | 2 |
| Association | 1 | 1 | 1 | 0 | 3 |
| Random Forest | 1 | 1 | 0 | 0 | 2 |
| Neural Networks | 1 | 1 | -1 | 1 | 2 |
| GANs | -1 | 1 | -1 | 1 | 0 |
| Naïve Bayes | 1 | 1 | 1 | 0 | 3 |

The Methods best suited to the decision points are Logistic Regression, Association and Naïve Bayes. The Method the least suited to the above decision points are GANs. This scorecard can be used to identify the best and worst methods. If the result is a tie between methods, then the user can consult the following sections to further score the methods.

## 6. Other AI Method Considerations

Sections 1 through 5 above present the framework of a decision process geared towards identifying ML methods suitable to accomplish a specific task. ML is a subset of AI that deals with knowledge acquisition, growth, improvement, and retention. In modern applications, ML represents a substantial space within the AI domain, however, the remaining space is consumed by multidisciplinary fields in psychology, mathematics, cognitive sciences, and computer science with the goal of mimicking human reasoning and decision-making.

Table 23 presents relevant knowledge domains that do not map neatly at a task level; however, they are still extremely pertinent to successful implementation of AI. Table 23 presents the domain, along with strengths, weakness (if any), and some example relevant applications.

Table 23. Knowledge Domains

| AI Domain | Strengths | Weaknesses/ Considerations | Relevant Applications |
|---|---|---|---|
| **Decision Theory** | Facilitator of XAI – can highlight differences in decision-making between human and machine agents<br><br>Solid foundation for Decision Support Systems (relates importance of input with Utility theory) – especially useful decision analysis<br><br>Accounts for risk preference<br><br>Can highlight optimal decision based on current, or predicted state (Markov Decision Process) | Usually considers decisions of single agents (no interactions)<br><br>May be overly simplistic when used alone | Decision Support Systems<br><br>Mission Planning Systems<br><br>Decision Explanation (XAI or Decision Graphs)<br><br>Decision analysis (when paired with simulations) |
| **Fuzzy Logic** | Simple, easily modified<br><br>Allows for partial truths<br><br>Mimics human reasoning under uncertainty<br><br>Learning can be introduced when combined with modern ML methods<br><br>Inferences about unknowns<br><br>Does not require substantial resources<br><br>Genetic algorithms improve performance of fuzzy systems, especially in large solution spaces. (See Genetic Fuzzy Trees) | Inherent ambiguity<br><br>May require substantial effort to establish rule base<br><br>Accuracy can be impacted<br>Likely require substantial V&V | Intelligent Control of systems (See ALPHA description above)<br><br>Classification<br><br>Pattern recognition<br><br>Initial classification |
| **Game Theory** | Considers decision-making during multi-agent interactions<br><br>Can be used to strengthen existing AI/ML technologies<br><br>Can determine optimal force allocation, strategic defense strategies, and advantageous positional strategies | | Decision Support Systems<br><br>Mission Planning Systems<br><br>Decision analysis (when paired with simulations)<br><br>Computational/Algorithmic Game Theory |
| **Information / Data Fusion** | Combine heterogeneous data from varying sensors<br><br>Enables potential process improvements – e.g., dimensionality reduction<br><br>Reduce ambiguity/noise in training data<br><br>Enhanced situational awareness | Affected by bandwidth limitations – advancements in "edge intelligence" have worked to mitigate this issue | Decision Support Systems<br><br>Edge processing (" Edge Intelligence")<br><br>ML algorithm improvement (dimensionality reduction) |

| AI Domain | Strengths | Weaknesses/ Considerations | Relevant Applications |
|---|---|---|---|
| **Spatial-Temporal Reasoning** | Conceptualizes three-dimensional relations of object in space<br><br>Enables navigation through environment<br><br>Facilitates the observation of resulting states of object/agent interactions<br><br>Contextual awareness | | Navigation and Object Avoidance (especially in cluttered environments)<br>- Bearing/Range Clear<br><br>Behavioral Inference |
| **Evolutionary / Genetic Algorithms** | Solve optimization problems in machine learning by mimicking genetics and natural selection to provide solutions<br>- Shrinks the solution space<br><br>Effective in large solution spaces | May be computationally expensive.<br><br>Difficult implementations – improper implementation can lead to system crashes and suboptimal solution recommendations | Image/Radar Processing<br><br>Code breaking / Cyber Defense Applications<br>- Could be used for code strengthening as a result<br><br>Artificial Creativity – enabling human-like creativity in AI systems |
| **Predictive Analytics** | Facilitates prediction of potential occurrences based on historical trends. In human-machine teaming systems, this capability not only provides valuable insight but also stands to increase user confidence in AI system<br><br>When paired with ML methods, can provide insight into new and unknown strategies | Requires knowledge of past events and established trends or accurate representation of such | Decision Support Systems<br><br>Simulations (Tactical and Training)<br>- Can be used to improve unit and Strike Group level organic and distributed training scenarios<br><br>Mission Planning Systems |
| **Prescriptive Analytics** | Generate recommendations for optimal actions in response to stimuli<br><br>Increase user confidence in AI system<br><br>When paired with predictive analytics, it can facilitate large-scale, high-fidelity simulations that present the user multiple scenarios, response actions and outcomes. | | Decision Support Systems<br><br>Simulations (Tactical and Training)<br>- Can be used to improve unit and Strike Group level organic and distributed training scenarios<br><br>Mission Planning Systems |

Table 23 aims to provide insight into applications dependent on problem / solution spaces. Utilizing the table to guide further advancement along with the ML criteria presented will ensure an optimal decision-making process resulting in a more refined product for Fleet end users at significantly reduced life cycle costs.

## D. FINDINGS SUMMARY

Figure 43 is a visual illustration of all the intervening steps between Kill Chain function identification and AI method selection/Ranking. This is a high-level summary of the previous sections of Chapter IV. This process will be used to present the final model and evaluation criteria in the final section of this report.

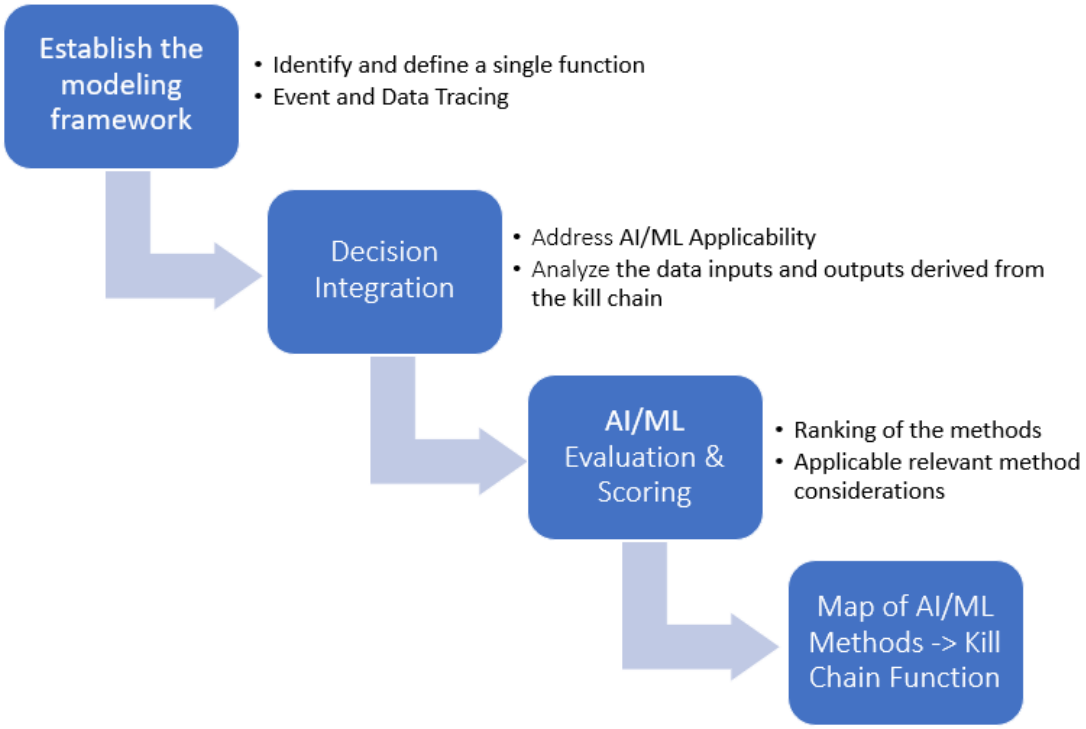



Figure 43. Analysis Roadmap

THIS PAGE INTENTIONALLY LEFT BLANK

# V. CONCLUSION AND FUTURE WORK

This chapter represents a culmination of the capstone team's work over two comprehensive literature reviews, framework analysis and development of mapping methodology. To produce the final map, the team applied the methodology discussed in Chapter IV to all 28 kill chain functions. The AI mapping methodology served two purposes; the process for mapping and overall source that the team would use for evaluation. Through the analysis of the tactical scenario related to each kill chain function, assessed decision points, and AI/ML score generation, the team produced a table of AI methods mapped to each kill chain function. The team was then able to assess the resulting score generation to evaluate each AI method for suitability (or lack of suitability) against every function in the kill chain until a recommended map was finalized.

This chapter concludes with a discussion of this capstone's potential benefits and the team's recommendations for future work and application of this research topic.

## A. FINAL MAPPING PRESENTATION AND EVALUATION CRITERIA

The final mapping process involved conducting a thorough assessment of all 28 kill chain functions and developing assumptions on the tactical scenario to build operational context. Once the team's assumptions were solidified, decision points were selected based on each kill chain function's unique context within an AMD scenario. Like the process described in Chapter IV, the decision points then drove associated scores that were totaled up to identify suitable mapping to individual AI/ML Methods. A sample of this process is shown in Figure 44.

| Step | Number | Function | Decision Points | AI/ML Score Generation |
|---|---|---|---|---|

| AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
|---|---|---|---|---|---|
| Linear Regression | -1 | -1 | +1 | 0 | -1 |
| Logistic Regression | 0 | -1 | +1 | 0 | 0 |
| Clustering | +1 | +1 | +1 | 0 | 3 |
| Association | 0 | -1 | +1 | 0 | 0 |
| Random Forrest | 0 | -1 | 0 | 0 | -1 |
| Neural Networks | 0 | -1 | -1 | +1 | -1 |
| GAN's | 0 | -1 | -1 | +1 | -1 |
| Naïve Bayes | 0 | -1 | 0 | 0 | -1 |

| Step | Number | Function | Decision Points |
|---|---|---|---|
| Find | 1.1.1 | Initial Detection | DP1: Clustering<br>DP2: Unsupervised Learning<br>DP3: XAI Mandatory<br>DP4: 10-99 Predictors |

Assumptions: Unknown signature is detected by Blue Force Asset / Sensor. Clusters of data could be used to assist system in detection and signature properties are unknown at this time. Explainable output is mandatory and number of predictors is medium to allow flexibility.

| AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
|---|---|---|---|---|---|
| Linear Regression | -1 | -1 | +1 | 0 | -1 |
| Logistic Regression | 0 | -1 | +1 | 0 | 0 |
| Clustering | +1 | +1 | +1 | 0 | 3 |
| Association | 0 | -1 | +1 | 0 | 0 |
| Random Forrest | 0 | -1 | 0 | 0 | -1 |
| Neural Networks | 0 | -1 | -1 | +1 | -1 |
| GAN's | 0 | -1 | -1 | +1 | -1 |
| Naïve Bayes | 0 | -1 | 0 | 0 | -1 |

| Step | Number | Function | Decision Points |
|---|---|---|---|
| Find | 1.1.2 | Battle Damage Assessment (BDA) Detection | DP1: Clustering<br>DP2: Unsupervised Learning<br>DP3: XAI Mandatory<br>DP4: 10-99 Predictors |

Assumptions: Unknown signature is detected by Blue Force Asset / Sensor within the battle space. Clusters of data could be used to assist system in detection and signature properties are unknown at this time. Explainable output is mandatory and number of predictors is medium to allow flexibility.

| AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
|---|---|---|---|---|---|
| Linear Regression | -1 | -1 | +1 | 0 | -1 |
| Logistic Regression | 0 | -1 | +1 | 0 | 0 |
| Clustering | +1 | +1 | +1 | 0 | 3 |
| Association | 0 | -1 | +1 | 0 | 0 |
| Random Forrest | 0 | -1 | 0 | 0 | -1 |
| Neural Networks | 0 | -1 | -1 | +1 | -1 |
| GAN's | 0 | -1 | -1 | +1 | -1 |
| Naïve Bayes | 0 | -1 | 0 | 0 | -1 |

| Step | Number | Function | Decision Points |
|---|---|---|---|
| Find | 1.1.3 | Re-Task Detection | DP1: Clustering<br>DP2: Unsupervised Learning<br>DP3: XAI Mandatory<br>DP4: 10-99 Predictors |

Assumptions: Blue Force Asset / Sensor is re-tasked to detect unknown signature within the battle space. Clusters of data could be used to assist system in detection and signature properties are unknown at this time. Explainable output is mandatory and number of predictors is given a medium range to allow flexibility.

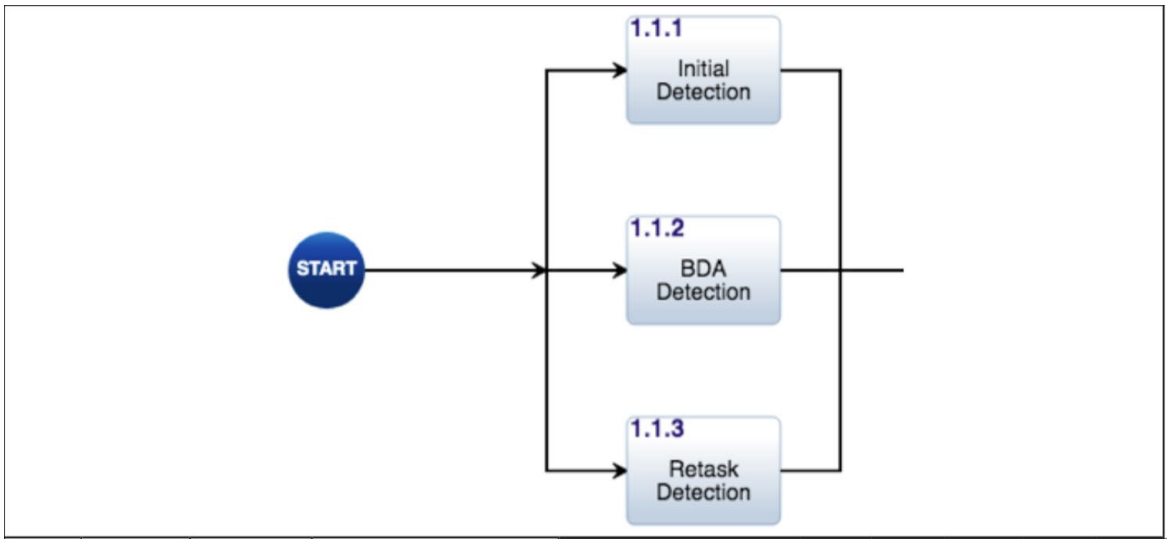



Figure 44. "Find Step – Sample Scoring Generation"

In this sample, the capstone team investigated the suitability of AL/ML Methods suitability specific to kill chain functions 1.1.1 - 1.1.3. Using 1.1.1 - *Initial Detection a*s an example, the operational context described in the assumption section is based on the team's understanding that at this point in the kill chain, detection of an unknown signal is occurring through a Blue Force sensor or asset. Knowing that the Blue Force sensor is likely processing numerous signals giving off large amounts of data with grouping potential (e.g., altitude, speed, direction, etc.), the team identified clustering as the logical choice for decision point #1. The team followed this process to determine all four decision points on each kill chain function.

After establishing each decision point, scores were generated that showed the suitability of each AI/ML method corresponding to each kill chain function. In the case of 1.1.1 - *Initial Detection,* the AI/ML method that was found most suitable was clustering. This was identified by determining the highest point total in the data set, highlighted in green. Similar to the example scorecard shown in Chapter IV, AI/ML methods that scored a value of zero or lower were highlighted red to indicate a lack of suitability for that particular kill chain function. AI/ML methods that were positive, but not the highest score was deemed to have demonstrated mapping potential and were highlighted yellow. Thus,

score generation served as both a process for mapping AI/ML methods to functions and evaluation criteria to determine the suitability of each result. The entire scoring table is included in Appendix A for reference.

After examining the scores of the entire kill chain, the team arrived at the final map, shown in Table 24.

Table 24. Final Map

| Step | Number | Function | AL/ML Method |
|---|---|---|---|
| Find | 1.1.1 | Initial Detection | Clustering |
| | 1.1.2 | Battle Damage Assessment (BDA) Detection | Clustering |
| | 1.1.3 | Re-Task Detection | Clustering |
| Fix | 1.2.1 | Define Target/Threat | Association |
| | 1.2.2 | Characterize | Clustering |
| | 1.2.3 | Classify | Logistic Regression, Association |
| | 1.2.4 | Identify | Logistic Regression, Association |
| | 1.2.5 | Locate | Clustering |
| | 1.2.6 | Disseminate Target /Threat | Association |
| Track | 1.3.1 | Generate / Update Track | Clustering |
| | 1.3.2 | Sort | Linear Regression |
| | 1.3.3 | Determine Target / Threat Urgency | Linear Regression |
| | 1.3.4 | Assess Blue Force Proximity | Association |
| | 1.3.5 | Validate Target / Threat | Association |
| Target | 1.4.1 | Nominate Engagement Option | Logistic Regression, Association |
| | 1.4.2 | Prioritize Target / Threat | Linear Regression |
| | 1.4.3 | Determine Time Available | Linear Regression |
| | 1.4.4 | Maintain Track | Clustering |
| | 1.4.5 | Select Attack Option | Logistic Regression, Association |
| | 1.4.6 | Verify Rules of Engagement (ROE) | Association |
| Engage | 1.5.1 | Issue Order | Association |
| | 1.5.2 | Attack Target / Threat | Linear Regression |
| | 1.5.3 | Track Weapon | Clustering |
| | 1.5.4 | Confirm Impact | Clustering |
| | 1.5.5 | Task Re-Attack | Linear Regression |
| Assess | 1.6.1 | Conduct Dynamic Assessment | Clustering |
| | 1.6.2 | Evaluate | Clustering |

While most evaluations resulted in a clear lead AI/ML method for suitability, there were four kill chain functions that were assessed to have more than one potential method for selection. Of the eight scored AI/ML methods, only four scored high enough to make it into the final map: Clustering, Association, Logistic Regression and Linear Regression. Table 25 offers a holistic view of all AI/ML Methods against a total score.

Table 25. AI/ML Methods Scoring Total by Kill Chain Function

| | | AI Method | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Linear Reg | Logistic Reg | Clustering | Association | Random Forrest | Neural Networks | GAN's | Naïve Bays |
| Kill Chain Function | 1.1.1 | -1 | 0 | 3 | 0 | -1 | -1 | -1 | -1 |
| | 1.1.2 | -1 | 0 | 3 | 0 | -1 | -1 | -1 | -1 |
| | 1.1.3 | -1 | 0 | 3 | 0 | -1 | -1 | -1 | -1 |
| | 1.2.1 | 1 | 1 | 0 | 3 | 0 | 0 | 0 | 0 |
| | 1.2.2 | 0 | 1 | 4 | 1 | 0 | -1 | -1 | 0 |
| | 1.2.3 | 3 | 4 | 3 | 4 | 3 | 3 | 1 | 3 |
| | 1.2.4 | 3 | 4 | 3 | 4 | 3 | 2 | 1 | 3 |
| | 1.2.5 | 0 | 1 | 4 | 1 | 0 | -1 | -1 | 0 |
| | 1.2.6 | 2 | 2 | 1 | 4 | 1 | 0 | 0 | 1 |
| | 1.3.1 | 0 | 1 | 4 | 1 | 0 | -1 | -1 | 0 |
| | 1.3.2 | 4 | 3 | 2 | 3 | 3 | 2 | 1 | 2 |
| | 1.3.3 | 4 | 3 | 2 | 3 | 3 | 2 | 1 | 2 |
| | 1.3.4 | 2 | 2 | 1 | 4 | 1 | 0 | 0 | 1 |
| | 1.3.5 | 2 | 2 | 1 | 4 | 1 | 0 | 0 | 0 |
| | 1.4.1 | 3 | 4 | 3 | 4 | 3 | 2 | 1 | 3 |
| | 1.4.2 | 4 | 3 | 2 | 3 | 3 | 2 | 1 | 2 |
| | 1.4.3 | 4 | 3 | 2 | 3 | 3 | 2 | 1 | 2 |
| | 1.4.4 | 0 | 1 | 4 | 1 | 0 | -1 | -1 | 0 |
| | 1.4.5 | 3 | 4 | 3 | 4 | 3 | 2 | 1 | 3 |
| | 1.4.6 | 2 | 2 | 1 | 4 | 1 | 0 | 0 | 1 |
| | 1.5.1 | 2 | 2 | 1 | 4 | 1 | 1 | 1 | 1 |
| | 1.5.2 | 4 | 3 | 2 | 3 | 3 | 2 | 1 | 2 |
| | 1.5.3 | 0 | 1 | 4 | 1 | 0 | -1 | -1 | 0 |
| | 1.5.4 | 0 | 1 | 4 | 1 | 0 | -1 | -1 | 0 |
| | 1.5.5 | 4 | 3 | 2 | 3 | 3 | 2 | 1 | 2 |
| | 1.6.1 | 2 | 1 | 3 | 1 | 1 | 0 | -1 | 0 |
| | 1.6.2 | 1 | 2 | 4 | 2 | 1 | 0 | -1 | 1 |
| **Total Score** | | **47** | **54** | **69** | **66** | **34** | **13** | **0** | **26** |

From these results, the capstone team determined that Clustering, Association, and Logistic Regression represented the most suitable AI/ML Methods, scoring positive over 80% cumulatively across all kill chain functions. Linear Regression, Random Forrest, and Naïve Bays represented suitability potential scoring positive over 50% overall. Finally, Neural Networks and GANs demonstrated a significant lack of suitability, scoring less than 50% across all kill chain functions. Therefore, the capstone team concluded that the final map was appropriate across both individual functions and cumulative assessment.

## B. CONCLUSION AND CONTRIBUTIONS

The final mapping provides a viable starting point for future research into validating the integration of AI/ML methods into the F2T2EA kill chain. While the majority of the capstone team's work was invested in development of a functional mapping process, the results pinpoint the most suitable AI/ML methods for integration at this stage of research while also providing insight on methods that may not be appropriate.

The AI 6 team met the objectives of this project successfully. The deliverables that were crucial were: a presentation of criteria for useful and transparent feedback of future AI implementation, evaluation criteria for each kill chain function, applicable AI methods, and a determination of which AI methods best map to which functions within the kill chain. From the outset established in Chapter I this project focused on naval kill chain functions. Also, the team developed evaluation criteria for each function to determine the efficacy of specific AI methods. The following sections restate the objectives and work performed within them.

### (1) Useful Transparent Feedback for Future AI Implementation Criteria

This objective was to determine criteria for useful and transparent feedback for future AI implementation by working in three phases: needs analysis, development of solution concepts, and an analysis of alternatives. This was accomplished by focusing on the prior NPS AI-OODA cohort, detailed literature reviews on AI and kill chain resources along with studying machine learning models. This research enabled development of evaluation criteria later in the project.

### (2) Evaluation of each Kill Chain Function

This team conducted an analysis of AI benefits to the execution of AMD kill-chain scenarios based on the F2T2EA kill-chain model. This project contends that the OODA loop to be a decision cycle that is contained within every function of the kill-chain rather than mapping specifically to one or more kill-chain functions.

(3) Applicable AI/ML Methods

Next, by looking at the works by DARPA and their waves model along with AI specialized topics this project was able to create applicable AI methods by looking at four decision points: output required, learning required, explanation required, and number of predictors. Scoring criteria for the following AI/ML methods was utilized:

- Linear Regression
- Logistic Regression
- Clustering
- Association
- Random Forest
- Neural Networks
- GAN
- Naïve Bayes

Scores ranged from +1 to –1 were applied to each method listed above to help determine a ranking for the applicable methods.

(4) Best AI Mapping Methods to Kill Chain Functions

In the end, this project presents the model for the best AI mapping method to the kill chain. This is accomplished by a large table with each kill chain function presented and scored. Each function is labeled and broken down with the winning AI/ML method highlighted in green.

## C. POTENTIAL BENEFITS

The kill chain represents the collective ability of the U.S. Department of Defense to integrate systems and effectively operate across a wide range of contingencies. The potential to incorporate AI takes current watch floor processes beyond human capacity and

unlocks a previously unachievable opportunity to maximize efficiency through ML. In addition to reducing the cognitive load and stress placed on watch personnel, refining kill chain processes with AI assisted execution better prepares the U.S. Military for future conflict with peer and near-peer threats that are emerging globally.

## D. FUTURE WORK AND APPLICATION

The importance of having accurate data available to make time critical and high-stakes decisions in the OODA loop based on uncertain situational knowledge in dynamic operational environments will require continued research. As the AI6 team used the findings (2020) provided by the AI-OODA team as a baseline for this capstone, the AI6 team would like for the findings found here within to be continued and expanded upon in the future to provide this OODA loop and kill chain data. As introduced in Chapter III, expand on the kill chain application of game theory (wargaming and algorithmic game theory) to determine its potential impact into the integration into combat systems training and tactics, techniques, and procedures development. Another area of potential interest for training and TTP is to research the use of feature engineering to determine the most relevant variables from the available date are used in the predictive model development. Additionally, the continued research of the utilization of enabling AI to provide the "combining of sensor data, or data derived from sensory data, into a common representational format" (Mitchell 2007, 1) for autonomous data fusion and threat warning functionalities to provide time critical information to the decision makers in the kill chain.

Furthermore, it has been demonstrated and discussed that there is often a substantial amount of ambiguity surrounding the term "AI." Table 24 presented readers with a list of applicable interdisciplinary topics that are often mentioned in relation to AI engineering. Utilizing this table, there is opportunity for continued research on how concepts of these knowledge domains can help better delivered AI products to warfighters. Appendix C offers examples of real-world uses of many of these knowledge domains.

As demonstrated in this capstone and as the potential basis for continued research, the AI6 team has provided decision makers with a method to determine the optimal AI/ML methods for each function in the kill chain.

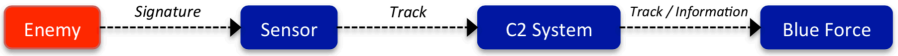
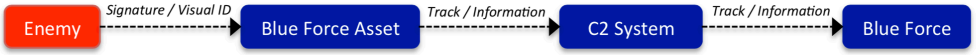
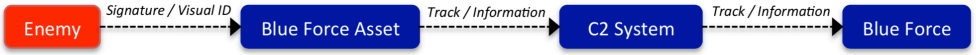
# APPENDIX A. KILL CHAIN FUNCTION DEFINITIONS AND MODELING

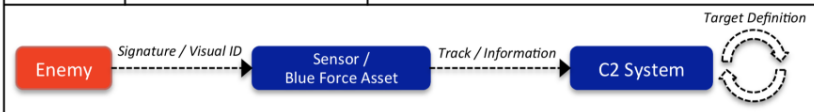
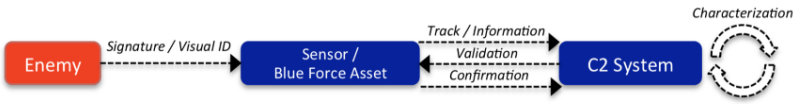
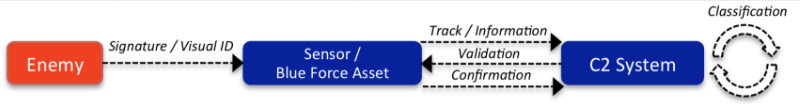
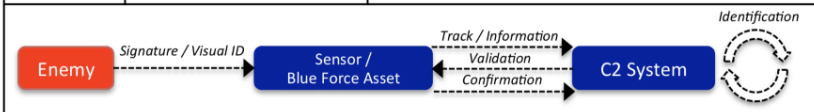
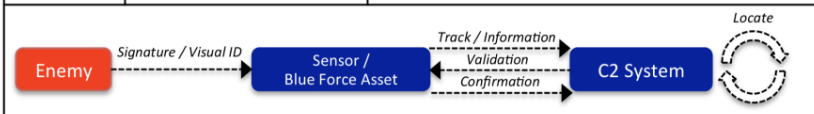
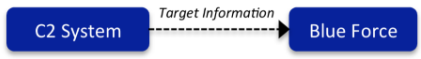
Table 26. Kill Chain Function Definitions

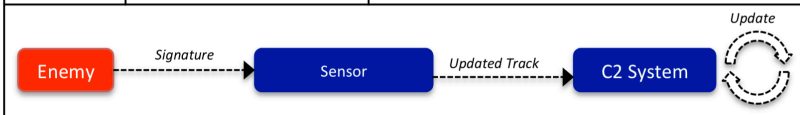
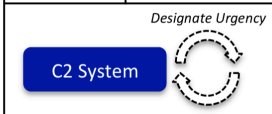
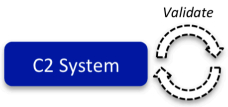
| Step | Number | Function | Definition |
|---|---|---|---|
| Find | 1.1.1 | Initial Detection | Detection of enemy units by radar or other military equipment. |
| | 1.1.2 | Battle Damage Assessment (BDA) Detection | Detection of enemy units as a result of BDA conducted by assets or capabilities. |
| | 1.1.3 | Re-Task Detection | Detection of enemy units conducted by assets or capabilities that have been diverted. |
| Fix | 1.2.1 | Define Target/Threat | Target or threat definition is initiated. |
| | 1.2.2 | Characterize | Target or threat is characterized as a result of data fusion and confirmation of identity and location. |
| | 1.2.3 | Classify | Sensors are focused to confirm initial track classification of target or threat. |
| | 1.2.4 | Identify | Sensors are focused to confirm identity of target or threat. |
| | 1.2.5 | Locate | Sensors are focused to confirm the precise location of target or threat. |
| | 1.2.6 | Disseminate Target /Threat | Target or threat information is disseminated to Blue Force command and control system. |
| Track | 1.3.1 | Generate / Update Track | Sensors produce / update target or threat track information throughout Blue Force command and control system. |
| | 1.3.2 | Sort | Target or threat is sorted by priority according to determination made during characterization. |
| | 1.3.3 | Determine Target / Threat Urgency | Target or threat level of urgency is designated according to determination made during characterization. |
| | 1.3.4 | Assess Blue Force Proximity | Blue Force command and control system assesses target or threat proximity in relation to all friendly assets or units. |
| | 1.3.5 | Validate Target / Threat | Target or threat meets objectives and criteria outlined in commander's guidance. |
| Target | 1.4.1 | Nominate Engagement Option | Initial nomination of asset(s) designated for target or threat engagement. |
| | 1.4.2 | Prioritize Target / Threat | Target or threat is given final prioritization level. |
| | 1.4.3 | Determine Time Available | Time requirements for target or threat engagement are assessed by Blue Force command and control system. |
| | 1.4.4 | Maintain Track | Sensors are focused to maintain track. |
| | 1.4.5 | Select Attack Option | Confirmation of asset designated for target or threat engagement. |
| | 1.4.6 | Verify Rules of Engagement (ROE) | Engagement of target or threat is validated in accordance with law of war or ROE, ensuring that it is not otherwise restricted. |
| Engage | 1.5.1 | Issue Order | Engagement order is issued by proper authority and transmitted to asset. |
| | 1.5.2 | Attack Target / Threat | Asset engages target or threat. |
| | 1.5.3 | Track Weapon | Sensors are focused to track asset engagement of target or threat. |
| | 1.5.4 | Confirm Impact | Sensors are focused to confirm asset engagement of target or threat. |

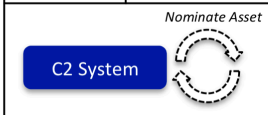
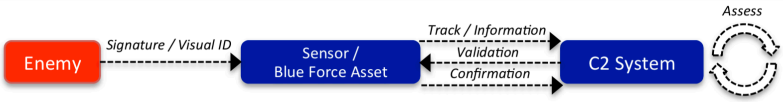
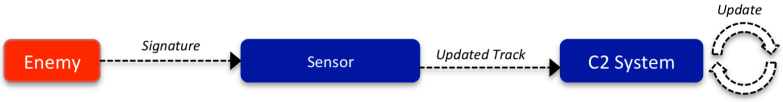
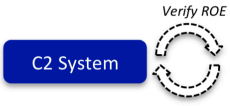
| Step | Number | Function | Definition |
|---|---|---|---|
| | 1.5.5 | Task Re-Attack | Re-attack order is issued by proper authority and transmitted to asset. |
| Assess | 1.6.1 | Conduct Dynamic Assessment | Sensors are focused to provide BDA of target or threat. |
| | 1.6.2 | Evaluate | Blue Force command and control system evaluates target or threat status. |

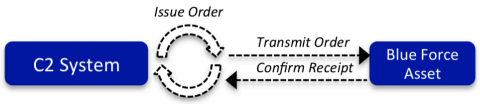
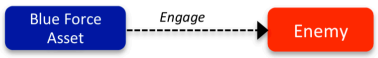
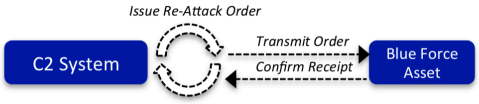
Table 27. Kill Chain Function Definitions and Models

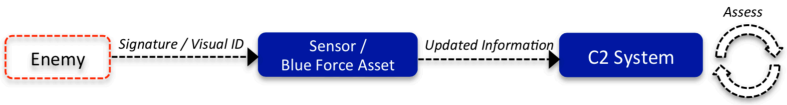
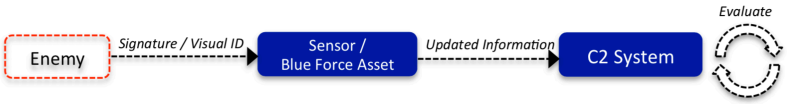
| Step | Number | Function | Definition |
|---|---|---|---|
| Find | 1.1.1 | Initial Detection | Detection of enemy units by radar or other military equipment. |
| |  | | |
| | 1.1.2 | Battle Damage Assessment (BDA) Detection | Detection of enemy units as a result of BDA conducted by assets or capabilities. |
| |  | | |
| | 1.1.3 | Re-Task Detection | Detection of enemy units conducted by assets or capabilities that have been diverted. |
| |  | | |

| | | | |
|---|---|---|---|
| Fix | 1.2.1 | Define Target/Threat | Target or threat definition is initiated. |
| |  | | |
| | 1.2.2 | Characterize | Target or threat is characterized as a result of data fusion and confirmation of identity and location. |
| |  | | |
| | 1.2.3 | Classify | Sensors are focused to confirm initial track classification of target or threat. |
| |  | | |
| | 1.2.4 | Identify | Sensors are focused to confirm identity of target or threat. |
| |  | | |
| | 1.2.5 | Locate | Sensors are focused to confirm the precise location of target or threat. |
| |  | | |
| | 1.2.6 | Disseminate Target /Threat | Target or threat information is disseminated to Blue Force via command and control system. |
| |  | | |

| | | | |
|---|---|---|---|
| Track | 1.3.1 | Generate / Update Track | Sensors produce / update target or threat track information throughout Blue Force command and control system. |
| |  | | |
| | 1.3.2 | Sort | Target or threat is sorted by priority according to determination made during characterization. |
| |  | | |
| | 1.3.3 | Determine Target / Threat Urgency | Target or threat level of urgency is designated according to determination made during characterization. |
| |  | | |
| | 1.3.4 | Assess Blue Force Proximity | Blue Force command and control system assesses target or threat proximity in relation to all friendly assets or units. |
| |  | | |
| | 1.3.5 | Validate Target / Threat | Target or threat meets objectives and criteria outlined in commander's guidance. |
| |  | | |

| | | | |
|---|---|---|---|
| Target | 1.4.1 | Nominate Engagement Option | Initial nomination of asset(s) designated for target or threat engagement. |
| |  | | |
| | 1.4.2 | Prioritize Target / Threat | Target or threat is given final prioritization level. |
| |  | | |
| | 1.4.3 | Determine Time Available | Time requirements for target or threat engagement are assessed by Blue Force command and control system. |
| |  | | |
| | 1.4.4 | Maintain Track | Sensors are focused to maintain track. |
| |  | | |
| | 1.4.5 | Select Attack Option | Confirmation of asset designated for target or threat engagement. |
| |  | | |
| | 1.4.6 | Verify Rules of Engagement (ROE) | Engagement of target or threat is validated in accordance with law of war or ROE, ensuring that it is not otherwise restricted. |
| |  | | |

| | | | |
|---|---|---|---|
| Engage | 1.5.1 | Issue Order | Engagement order is issued by proper authority and transmitted to asset. |
| |  | | |
| | 1.5.2 | Attack Target / Threat | Asset engages target or threat. |
| |  | | |
| | 1.5.3 | Track Weapon | Sensors are focused to track asset engagement of target or threat. |
| |  | | |
| | 1.5.4 | Confirm Impact | Sensors are focused to confirm asset engagement of target or threat. |
| |  | | |
| | 1.5.5 | Task Re-Attack | Re-attack order is issued by proper authority and transmitted to asset. |
| |  | | |

<table>
<tr><td rowspan="4">Assess</td><td>1.6.1</td><td>Conduct Dynamic Assessment</td><td>Sensors are focused to provide BDA of target or threat.</td></tr>
<tr><td colspan="3"> 
</td></tr>
<tr><td>1.6.2</td><td>Evaluate</td><td>Blue Force command and control system evaluates target or threat status.</td></tr>
<tr><td colspan="3"> 
</td></tr>
</table>

THIS PAGE INTENTIONALLY LEFT BLANK

# APPENDIX B. FULL MODEL TABLE

Table 28. Full Model Table

| Step | Number | Function | Decision Points | AI/ML Score Generation |
|---|---|---|---|---|
| Find | 1.1.1 | Initial Detection | DP1: Clustering<br>DP2: Unsupervised Learning<br>DP3: XAI Mandatory<br>DP4: 10–99 Predictors | |

| AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
|---|---|---|---|---|---|
| Linear Regression | -1 | -1 | +1 | 0 | -1 |
| Logistic Regression | 0 | -1 | +1 | 0 | 0 |
| Clustering | +1 | +1 | +1 | 0 | 3 |
| Association | 0 | -1 | +1 | 0 | 0 |
| Random Forrest | 0 | -1 | 0 | 0 | -1 |
| Neural Networks | 0 | -1 | -1 | +1 | -1 |
| GAN's | 0 | -1 | -1 | +1 | -1 |
| Naïve Bayes | 0 | -1 | 0 | 0 | -1 |

Assumptions: Unknown signature is detected by Blue Force Asset / Sensor. Clusters of data could be used to assist system in detection and signature properties are unknown at this time. Explainable output is mandatory, and number of predictors is medium to allow flexibility.

| Step | Number | Function | Decision Points |
|---|---|---|---|
| Find | 1.1.2 | Battle Damage Assessment (BDA) Detection | DP1: Clustering<br>DP2: Unsupervised Learning<br>DP3: XAI Mandatory<br>DP4: 10–99 Predictors |

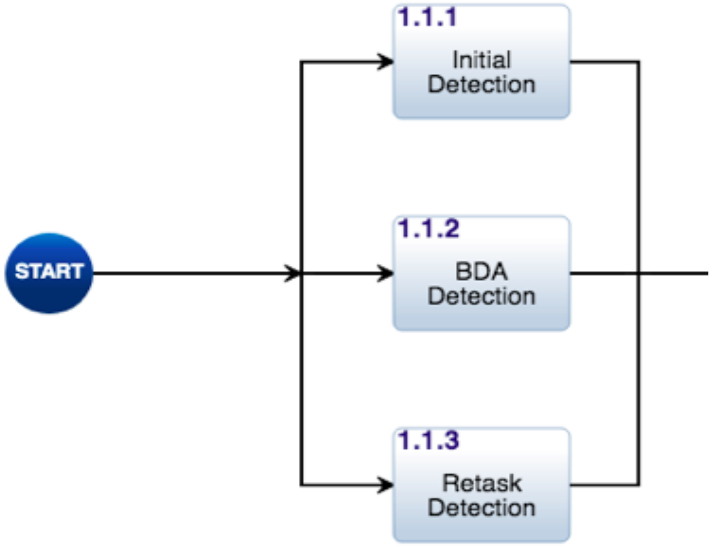
| AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
|---|---|---|---|---|---|
| Linear Regression | -1 | -1 | +1 | 0 | -1 |
| Logistic Regression | 0 | -1 | +1 | 0 | 0 |
| Clustering | +1 | +1 | +1 | 0 | 3 |
| Association | 0 | -1 | +1 | 0 | 0 |
| Random Forrest | 0 | -1 | 0 | 0 | -1 |
| Neural Networks | 0 | -1 | -1 | +1 | -1 |
| GAN's | 0 | -1 | -1 | +1 | -1 |
| Naïve Bayes | 0 | -1 | 0 | 0 | -1 |

Assumptions: Unknown signature is detected by Blue Force Asset / Sensor within the battle space. Clusters of data could be used to assist system in detection and signature properties are unknown at this time. Explainable output is mandatory, and number of predictors is medium to allow flexibility.

| Step | Number | Function | Decision Points |
|---|---|---|---|
| Find | 1.1.3 | Re-Task Detection | DP1: Clustering<br>DP2: Unsupervised Learning<br>DP3: XAI Mandatory<br>DP4: 10–99 Predictors |

| AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
|---|---|---|---|---|---|
| Linear Regression | -1 | -1 | +1 | 0 | -1 |
| Logistic Regression | 0 | -1 | +1 | 0 | 0 |
| Clustering | +1 | +1 | +1 | 0 | 3 |
| Association | 0 | -1 | +1 | 0 | 0 |
| Random Forrest | 0 | -1 | 0 | 0 | -1 |
| Neural Networks | 0 | -1 | -1 | +1 | -1 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | GAN's | 0 | -1 | -1 | +1 | -1 |
| | | | | Naïve Bayes | 0 | -1 | 0 | 0 | -1 |
| | Assumptions: Blue Force Asset / Sensor is re-tasked to detect unknown signature within the battle space. Clusters of data could be used to assist system in detection and signature properties are unknown at this time. Explainable output is mandatory, and number of predictors is given a medium range to allow flexibility. | | | | | | | | |




| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Fix | 1.2.1 | Define Target/ Threat | DP1: Rules<br>DP2: Supervised Learning<br>DP3: XAI Desired<br>DP4: 10–99 Predictors | AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
| | | | | Linear Regression | -1 | +1 | +1 | 0 | 1 |
| | | | | Logistic Regression | -1 | +1 | +1 | 0 | 1 |
| | | | | Clustering | -1 | 0 | +1 | 0 | 0 |
| | | | | Association | +1 | +1 | +1 | 0 | 3 |
| | | | | Random Forrest | -1 | +1 | 0 | 0 | 0 |
| | | | | Neural Networks | -1 | +1 | -1 | +1 | 0 |
| | | | | GAN's | -1 | +1 | -1 | +1 | 0 |
| | | | | Naïve Bayes | -1 | +1 | 0 | 0 | 0 |
| | Assumptions: Target definition process has begun based on established doctrinal procedures. If/then structure to execute target definition process. Explainable output is desired, and number of predictors is a medium range to allow flexibility. | | | | | | | | |
| | 1.2.2 | Characterize | DP1: Clustering<br>DP2: Unsupervised Learning<br>DP3: XAI Mandatory<br>DP4: 1–9 Predictors | AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
| | | | | Linear Regression | -1 | -1 | +1 | +1 | 0 |
| | | | | Logistic Regression | 0 | -1 | +1 | +1 | 1 |
| | | | | Clustering | +1 | +1 | +1 | +1 | 4 |
| | | | | Association | 0 | -1 | +1 | +1 | 1 |
| | | | | Random Forrest | 0 | -1 | 0 | +1 | 0 |
| | | | | Neural Networks | 0 | -1 | -1 | +1 | -1 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | GAN's | 0 | -1 | -1 | +1 | -1 |
| | | | Naïve Bayes | 0 | -1 | 0 | +1 | 0 |

Assumptions: Unknown signature is characterized as potential target or other target type requiring prosecution. Clusters of data could be used to assist characterization process. Explainable output is mandatory, and number of predictors is low to enable higher accuracy.

| | | | AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
|---|---|---|---|---|---|---|---|---|
| 1.2.3 | Classify | DP1: Qualitative<br>DP2: Supervised Learning<br>DP3: XAI Mandatory<br>DP4: 1–9 Predictors | Linear Regression | 0 | +1 | +1 | +1 | 3 |
| | | | Logistic Regression | +1 | +1 | +1 | +1 | 4 |
| | | | Clustering | +1 | 0 | +1 | +1 | 3 |
| | | | Association | +1 | +1 | +1 | +1 | 4 |
| | | | Random Forrest | +1 | +1 | 0 | +1 | 3 |
| | | | Neural Networks | +1 | +1 | -1 | +1 | 3 |
| | | | GAN's | 0 | +1 | -1 | +1 | 1 |
| | | | Naïve Bayes | +1 | +1 | 0 | +1 | 3 |

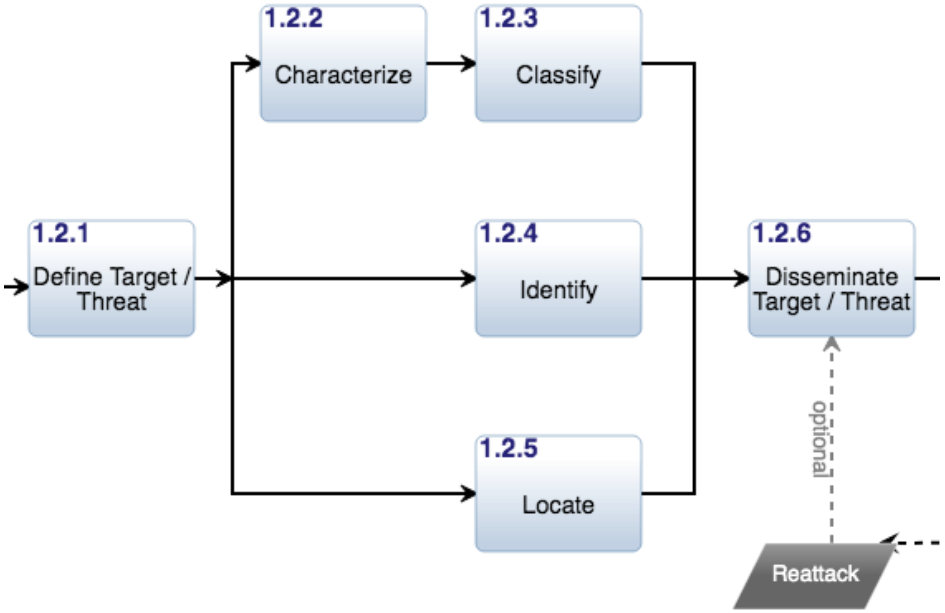
Assumptions: Initial track classification is confirmed. Data will require assignment to predefined categorical values with pre-existing datasets available based on identity. Explainable output is mandatory, and number of predictors is low to enable higher accuracy.

| | | | AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
|---|---|---|---|---|---|---|---|---|
| 1.2.4 | Identify | DP1: Qualitative<br>DP2: Supervised Learning<br>DP3: XAI Mandatory<br>DP4: 1–9 Predictors | Linear Regression | 0 | +1 | +1 | +1 | 3 |
| | | | Logistic Regression | +1 | +1 | +1 | +1 | 4 |
| | | | Clustering | +1 | 0 | +1 | +1 | 3 |
| | | | Association | +1 | +1 | +1 | +1 | 4 |
| | | | Random Forrest | +1 | +1 | 0 | +1 | 3 |
| | | | Neural Networks | +1 | +1 | -1 | +1 | 2 |
| | | | GAN's | 0 | +1 | -1 | +1 | 1 |
| | | | Naïve Bayes | +1 | +1 | 0 | +1 | 3 |

Assumptions: Identity of target or threat is confirmed. Data will require assignment to predefined categorical values with pre-existing datasets available based on identity. Explainable output is mandatory, and number of predictors is low to enable higher accuracy.

| | | | AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
|---|---|---|---|---|---|---|---|---|
| 1.2.5 | Locate | DP1: Clustering<br>DP2: Unsupervised Learning<br>DP3: XAI Mandatory<br>DP4: 1–9 Predictors | Linear Regression | -1 | -1 | +1 | +1 | 0 |
| | | | Logistic Regression | 0 | -1 | +1 | +1 | 1 |
| | | | Clustering | +1 | +1 | +1 | +1 | 4 |
| | | | Association | 0 | -1 | +1 | +1 | 1 |
| | | | Random Forrest | 0 | -1 | 0 | +1 | 0 |
| | | | Neural Networks | 0 | -1 | -1 | +1 | -1 |
| | | | GAN's | 0 | -1 | -1 | +1 | -1 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Naïve Bayes | 0 | -1 | 0 | +1 | 0 |
| | Assumptions: Location is confirmed. Clusters of data could be used to assist characterization process. Explainable output is mandatory, and number of predictors is low to enable higher accuracy. | | | | | | | | |
| | 1.2.6 | Disseminate Target /Threat | DP1: Rules<br>DP2: Supervised Learning<br>DP3: XAI Mandatory<br>DP4: 1–9 Predictors | **AI/ML Method** | **DP #1** | **DP #2** | **DP #3** | **DP #4** | **Total** |
| | | | | Linear Regression | -1 | +1 | +1 | +1 | 2 |
| | | | | Logistic Regression | -1 | +1 | +1 | +1 | 2 |
| | | | | Clustering | -1 | 0 | +1 | +1 | 1 |
| | | | | Association | +1 | +1 | +1 | +1 | 4 |
| | | | | Random Forrest | -1 | +1 | 0 | +1 | 1 |
| | | | | Neural Networks | -1 | +1 | -1 | +1 | 0 |
| | | | | GAN's | -1 | +1 | -1 | +1 | 0 |
| | | | | Naïve Bayes | -1 | +1 | 0 | +1 | 1 |
| | Assumptions: Target or threat requirements are confirmed and disseminated among Blue Force. If/then structure is required to execute dissemination process. Explainable output is desired, and number of predictors is a low range to enable higher accuracy. | | | | | | | | |




| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Track | 1.3.1 | Generate / Update Track | DP1: Clustering<br>DP2: Unsupervised Learning<br>DP3: XAI Mandatory<br>DP4: 1–9 Predictors | **AI/ML Method** | **DP #1** | **DP #2** | **DP #3** | **DP #4** | **Total** |
| | | | | Linear Regression | -1 | -1 | +1 | +1 | 0 |
| | | | | Logistic Regression | 0 | -1 | +1 | +1 | 1 |
| | | | | Clustering | +1 | +1 | +1 | +1 | 4 |
| | | | | Association | 0 | -1 | +1 | +1 | 1 |
| | | | | Random Forrest | 0 | -1 | 0 | +1 | 0 |
| | | | | Neural Networks | 0 | -1 | -1 | +1 | -1 |
| | | | | GAN's | 0 | -1 | -1 | +1 | -1 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Naïve Bayes | 0 | -1 | 0 | +1 | 0 |

Assumptions: Track is generated/updated across Blue Force C2 System. Clusters of data could be used to assist characterization process. Explainable output is mandatory, and number of predictors is low to enable higher accuracy.

| | | | AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
|---|---|---|---|---|---|---|---|---|
| 1.3.2 | Sort | DP1: Quantitative<br>DP2: Supervised Learning<br>DP3: XAI Mandatory<br>DP4: 1–9 Predictors | Linear Regression | +1 | +1 | +1 | +1 | 4 |
| | | | Logistic Regression | 0 | +1 | +1 | +1 | 3 |
| | | | Clustering | 0 | 0 | +1 | +1 | 2 |
| | | | Association | 0 | +1 | +1 | +1 | 3 |
| | | | Random Forrest | +1 | +1 | 0 | +1 | 3 |
| | | | Neural Networks | +1 | +1 | -1 | +1 | 2 |
| | | | GAN's | 0 | +1 | -1 | +1 | 1 |
| | | | Naïve Bayes | 0 | +1 | 0 | +1 | 2 |

Assumptions: Target or threat is sorted by priority. Data will require assignment to predefined categorical values with pre-existing datasets available based on priority. Explainable output is mandatory, and number of predictors is low to enable higher accuracy.

| | | | AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
|---|---|---|---|---|---|---|---|---|
| 1.3.3 | Determine Target / Threat Urgency | DP1: Quantitative<br>DP2: Supervised Learning<br>DP3: XAI Mandatory<br>DP4: 1–9 Predictors | Linear Regression | +1 | +1 | +1 | +1 | 4 |
| | | | Logistic Regression | 0 | +1 | +1 | +1 | 3 |
| | | | Clustering | 0 | 0 | +1 | +1 | 2 |
| | | | Association | 0 | +1 | +1 | +1 | 3 |
| | | | Random Forrest | +1 | +1 | 0 | +1 | 3 |
| | | | Neural Networks | +1 | +1 | -1 | +1 | 2 |
| | | | GAN's | 0 | +1 | -1 | +1 | 1 |
| | | | Naïve Bayes | 0 | +1 | 0 | +1 | 2 |

Assumptions: Prosecution urgency of target or threat is determined. Data will require assignment to predefined categorical values with pre-existing datasets available based on established criteria. Explainable output is mandatory, and number of predictors is low to enable higher accuracy.

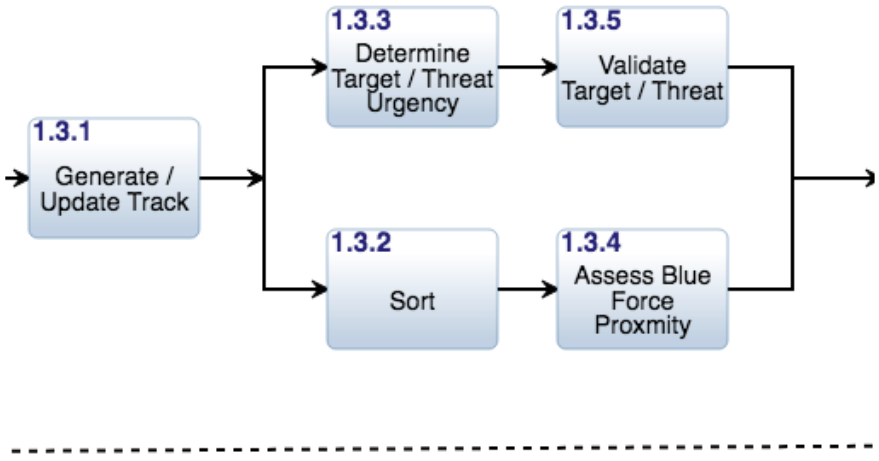
| | | | AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
|---|---|---|---|---|---|---|---|---|
| 1.3.4 | Assess Blue Force Proximity | DP1: Rules<br>DP2: Supervised Learning<br>DP3: XAI Mandatory<br>DP4: 1–9 Predictors | Linear Regression | -1 | +1 | +1 | +1 | 2 |
| | | | Logistic Regression | -1 | +1 | +1 | +1 | 2 |
| | | | Clustering | -1 | 0 | +1 | +1 | 1 |
| | | | Association | +1 | +1 | +1 | +1 | 4 |
| | | | Random Forrest | -1 | +1 | 0 | +1 | 1 |
| | | | Neural Networks | -1 | +1 | -1 | +1 | 0 |
| | | | GAN's | -1 | +1 | -1 | +1 | 0 |
| | | | Naïve Bayes | -1 | +1 | 0 | +1 | 1 |

Assumptions: Target or threat proximity to Blue Force is assessed. If/then structure is used for confirmation of friendly locations. Explainable output is desired, and number of predictors is low to enable higher accuracy.

| | | | |
|---|---|---|---|
| 1.3.5 | Validate Target / Threat | DP1: Rules<br>DP2: Supervised Learning<br>DP3: XAI Mandatory<br>DP4: 1–9 Predictors | |

| AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
|---|---|---|---|---|---|
| Linear Regression | -1 | +1 | +1 | +1 | 2 |
| Logistic Regression | -1 | +1 | +1 | +1 | 2 |
| Clustering | -1 | 0 | +1 | +1 | 1 |
| Association | +1 | +1 | +1 | +1 | 4 |
| Random Forrest | -1 | +1 | 0 | +1 | 1 |
| Neural Networks | -1 | +1 | -1 | +1 | 0 |
| GAN's | -1 | +1 | -1 | +1 | 0 |
| Naïve Bayes | -1 | +1 | 0 | +1 | 0 |

Assumptions: Target or threat is validated using pre-existing conditions. If/then structure is used in order to assess Target or threat against established criteria. Explainable output is desired, and number of predictors is low to enable higher accuracy.




Target

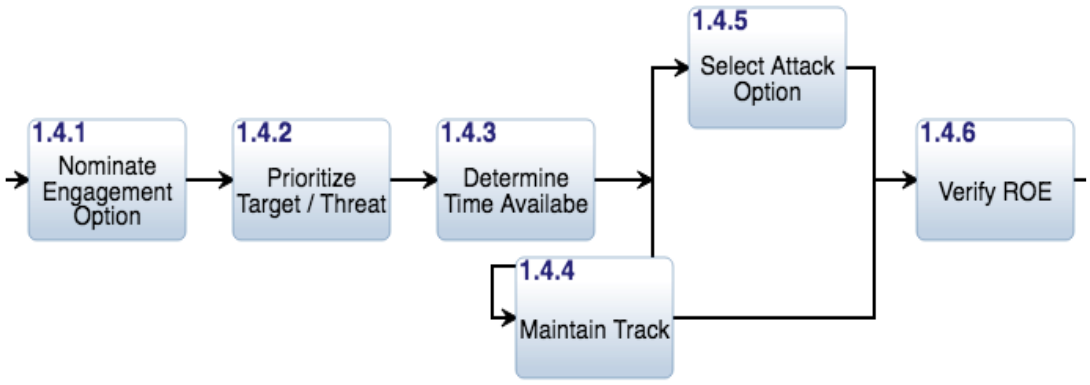
| | | | |
|---|---|---|---|
| 1.4.1 | Nominate Engagement Option | DP1: Qualitative<br>DP2: Supervised Learning<br>DP3: XAI Mandatory<br>DP4: 1–9 Predictors | |

| AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
|---|---|---|---|---|---|
| Linear Regression | 0 | +1 | +1 | +1 | 3 |
| Logistic Regression | +1 | +1 | +1 | +1 | 4 |
| Clustering | +1 | 0 | +1 | +1 | 3 |
| Association | +1 | +1 | +1 | +1 | 4 |
| Random Forrest | +1 | +1 | 0 | +1 | 3 |
| Neural Networks | +1 | +1 | -1 | +1 | 2 |
| GAN's | 0 | +1 | -1 | +1 | 1 |
| Naïve Bayes | +1 | +1 | 0 | +1 | 3 |

Assumptions: Blue Force asset is nominated for target prosecution. Data will require assignment to predefined categorical values with pre-existing datasets available based on identity. Explainable output is mandatory, and number of predictors is low to enable higher accuracy.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1.4.2 | Prioritize Target / Threat | DP1: Quantitative<br>DP2: Supervised Learning<br>DP3: XAI Mandatory<br>DP4: 1–9 Predictors | AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
| | | | Linear Regression | +1 | +1 | +1 | +1 | 4 |
| | | | Logistic Regression | 0 | +1 | +1 | +1 | 3 |
| | | | Clustering | 0 | 0 | +1 | +1 | 2 |
| | | | Association | 0 | +1 | +1 | +1 | 3 |
| | | | Random Forrest | +1 | +1 | 0 | +1 | 3 |
| | | | Neural Networks | +1 | +1 | -1 | +1 | 2 |
| | | | GAN's | 0 | +1 | -1 | +1 | 1 |
| | | | Naïve Bayes | 0 | +1 | 0 | +1 | 2 |
| Assumptions: Target is given final priority level. Data will require assignment to predefined categorical values with pre-existing datasets available based on priority. Explainable output is mandatory, and number of predictors is low to enable higher accuracy. | | | | | | | | | |
| 1.4.3 | Determine Time Available | DP1: Quantitative<br>DP2: Supervised Learning<br>DP3: XAI Mandatory<br>DP4: 1–9 Predictors | AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
| | | | Linear Regression | +1 | +1 | +1 | +1 | 4 |
| | | | Logistic Regression | 0 | +1 | +1 | +1 | 3 |
| | | | Clustering | 0 | 0 | +1 | +1 | 2 |
| | | | Association | 0 | +1 | +1 | +1 | 3 |
| | | | Random Forrest | +1 | +1 | 0 | +1 | 3 |
| | | | Neural Networks | +1 | +1 | -1 | +1 | 2 |
| | | | GAN's | 0 | +1 | -1 | +1 | 1 |
| | | | Naïve Bayes | 0 | +1 | 0 | +1 | 2 |
| Assumptions: Engagement time is assessed. Data will require assignment to predefined categorical values with pre-existing datasets available based on priority. Explainable output is mandatory, and number of predictors is low to enable higher accuracy. | | | | | | | | | |
| 1.4.4 | Maintain Track | DP1: Clustering<br>DP2: Unsupervised Learning<br>DP3: XAI Mandatory<br>DP4: 1–9 Predictors | AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
| | | | Linear Regression | -1 | -1 | +1 | +1 | 0 |
| | | | Logistic Regression | 0 | -1 | +1 | +1 | 1 |
| | | | Clustering | +1 | +1 | +1 | +1 | 4 |
| | | | Association | 0 | -1 | +1 | +1 | 1 |
| | | | Random Forrest | 0 | -1 | 0 | +1 | 0 |
| | | | Neural Networks | 0 | -1 | -1 | +1 | -1 |
| | | | GAN's | 0 | -1 | -1 | +1 | -1 |
| | | | Naïve Bayes | 0 | -1 | 0 | +1 | 0 |
| Assumptions: Track is updated/maintained across Blue Force C2 System. Clusters of data could be used to assist characterization process. Explainable output is mandatory, and number of predictors is low to enable higher accuracy. | | | | | | | | | |

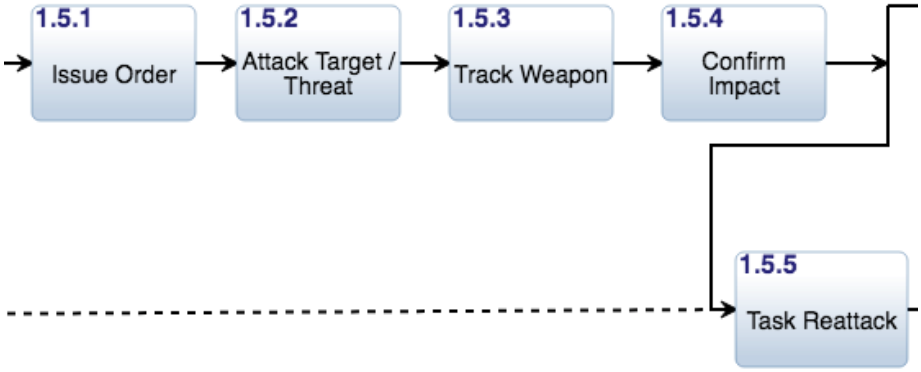
| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1.4.5 | Select Attack Option | DP1: Qualitative<br>DP2: Supervised Learning<br>DP3: XAI Mandatory<br>DP4: 1–9 Predictors | AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
| | | | | Linear Regression | 0 | +1 | +1 | +1 | 3 |
| | | | | Logistic Regression | +1 | +1 | +1 | +1 | 4 |
| | | | | Clustering | +1 | 0 | +1 | +1 | 3 |
| | | | | Association | +1 | +1 | +1 | +1 | 4 |
| | | | | Random Forrest | +1 | +1 | 0 | +1 | 3 |
| | | | | Neural Networks | +1 | +1 | -1 | +1 | 2 |
| | | | | GAN's | 0 | +1 | -1 | +1 | 1 |
| | | | | Naïve Bayes | +1 | +1 | 0 | +1 | 3 |
| | Assumptions: Blue Force asset is selected for target prosecution. If/then structure is used in order to assess Target or threat against established criteria. Explainable output is desired, and number of predictors is low to enable higher accuracy. | | | | | | | | |
| | 1.4.6 | Verify Rules of Engagement (ROE) | DP1: Rules<br>DP2: Supervised Learning<br>DP3: XAI Mandatory<br>DP4: 1–9 Predictors | AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
| | | | | Linear Regression | -1 | +1 | +1 | +1 | 2 |
| | | | | Logistic Regression | -1 | +1 | +1 | +1 | 2 |
| | | | | Clustering | -1 | 0 | +1 | +1 | 1 |
| | | | | Association | +1 | +1 | +1 | +1 | 4 |
| | | | | Random Forrest | -1 | +1 | 0 | +1 | 1 |
| | | | | Neural Networks | -1 | +1 | -1 | +1 | 0 |
| | | | | GAN's | -1 | +1 | -1 | +1 | 0 |
| | | | | Naïve Bayes | -1 | +1 | 0 | +1 | 1 |
| | Assumptions: Blue Force C2 System verifies fulfillment of ROE. If/then structure is used in order to assess Target or threat against established criteria. Explainable output is desired, and number of predictors is low to enable higher accuracy. | | | | | | | | |




| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Engage | 1.5.1 | Issue Order | DP1: Rules | AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | DP2: Supervised Learning<br><br>DP3: XAI Mandatory<br><br>DP4: 1–9 Predictors | Linear Regression | -1 | +1 | +1 | +1 | 2 |
| | | | | Logistic Regression | -1 | +1 | +1 | +1 | 2 |
| | | | | Clustering | -1 | 0 | +1 | +1 | 1 |
| | | | | Association | +1 | +1 | +1 | +1 | 4 |
| | | | | Random Forrest | -1 | +1 | 0 | +1 | 1 |
| | | | | Neural Networks | -1 | +1 | -1 | +1 | 1 |
| | | | | GAN's | -1 | +1 | -1 | +1 | 1 |
| | | | | Naïve Bayes | -1 | +1 | 0 | +1 | 1 |
| | Assumptions: Engagement authority issues order. If/then structure is used in order to assess Target or threat against established criteria. Explainable output is desired, and number of predictors is low to enable higher accuracy. | | | | | | | | |
| | 1.5.2 | Attack Target / Threat | DP1: Quantitative<br><br>DP2: Supervised Learning<br><br>DP3: XAI Mandatory<br><br>DP4: 1–9 Predictors | AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
| | | | | Linear Regression | +1 | +1 | +1 | +1 | 4 |
| | | | | Logistic Regression | 0 | +1 | +1 | +1 | 3 |
| | | | | Clustering | 0 | 0 | +1 | +1 | 2 |
| | | | | Association | 0 | +1 | +1 | +1 | 3 |
| | | | | Random Forrest | +1 | +1 | 0 | +1 | 3 |
| | | | | Neural Networks | +1 | +1 | -1 | +1 | 2 |
| | | | | GAN's | 0 | +1 | -1 | +1 | 1 |
| | | | | Naïve Bayes | 0 | +1 | 0 | +1 | 2 |
| | Assumptions: Blue Force asset engages target. Data will require assignment to predefined categorical values with pre-existing datasets available based on priority. Explainable output is mandatory, and number of predictors is low to enable higher accuracy. | | | | | | | | |
| | 1.5.3 | Track Weapon | DP1: Clustering<br><br>DP2: Unsupervised Learning<br><br>DP3: XAI Mandatory<br><br>DP4: 1–9 Predictors | AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
| | | | | Linear Regression | -1 | -1 | +1 | +1 | 0 |
| | | | | Logistic Regression | 0 | -1 | +1 | +1 | 1 |
| | | | | Clustering | +1 | +1 | +1 | +1 | 4 |
| | | | | Association | 0 | -1 | +1 | +1 | 1 |
| | | | | Random Forrest | 0 | -1 | 0 | +1 | 0 |
| | | | | Neural Networks | 0 | -1 | -1 | +1 | -1 |
| | | | | GAN's | 0 | -1 | -1 | +1 | -1 |
| | | | | Naïve Bayes | 0 | -1 | 0 | +1 | 0 |
| | Assumptions: Blue Force asset/sensor tracks weapon delivery to target. Clusters of data could be used to assist characterization process. Explainable output is mandatory, and number of predictors is low to enable higher accuracy. | | | | | | | | |
| | 1.5.4 | Confirm Impact | DP1: Clustering<br><br>DP2: Unsupervised Learning | AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
| | | | | Linear Regression | -1 | -1 | +1 | +1 | 0 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | DP3: XAI Mandatory<br>DP4: 1–9 Predictors | Logistic Regression | 0 | -1 | +1 | +1 | 1 |
| | | | | Clustering | +1 | +1 | +1 | +1 | 4 |
| | | | | Association | 0 | -1 | +1 | +1 | 1 |
| | | | | Random Forrest | 0 | -1 | 0 | +1 | 0 |
| | | | | Neural Networks | 0 | -1 | -1 | +1 | -1 |
| | | | | GAN's | 0 | -1 | -1 | +1 | -1 |
| | | | | Naïve Bayes | 0 | -1 | 0 | +1 | 0 |
| | Assumptions: Blue Force asset/sensor tracks confirm asset engagement on target. Clusters of data could be used to assist characterization process. Explainable output is mandatory, and number of predictors is low to enable higher accuracy. | | | | | | | | |
| | 1.5.5 | Task Re-Attack | DP1: Quantitative<br>DP2: Supervised Learning<br>DP3: XAI Mandatory<br>DP4: 1–9 Predictors | AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
| | | | | Linear Regression | +1 | +1 | +1 | +1 | 4 |
| | | | | Logistic Regression | 0 | +1 | +1 | +1 | 3 |
| | | | | Clustering | 0 | 0 | +1 | +1 | 2 |
| | | | | Association | 0 | +1 | +1 | +1 | 3 |
| | | | | Random Forrest | +1 | +1 | 0 | +1 | 3 |
| | | | | Neural Networks | +1 | +1 | -1 | +1 | 2 |
| | | | | GAN's | 0 | +1 | -1 | +1 | 1 |
| | | | | Naïve Bayes | 0 | +1 | 0 | +1 | 2 |
| | Assumptions: Engagement authority issues order for re-attack. Data will require assignment to predefined categorical values with pre-existing datasets available based on priority. Explainable output is mandatory, and number of predictors is low to enable higher accuracy. | | | | | | | | |



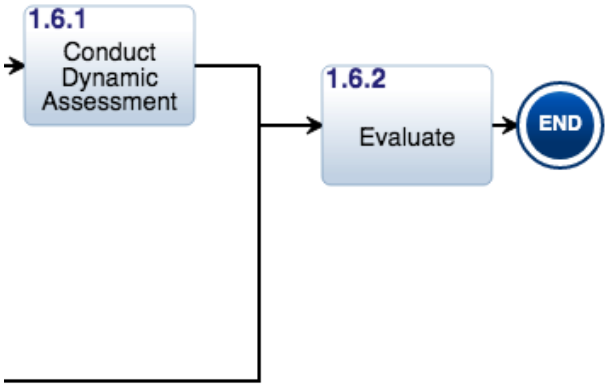

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Assess | 1.6.1 | Conduct Dynamic Assessment | DP1: Quantitative<br>DP2: Unsupervised Learning<br>DP3: XAI Mandatory<br>DP4: 1–9 Predictors | AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
| | | | | Linear Regression | +1 | -1 | +1 | +1 | 2 |
| | | | | Logistic Regression | 0 | -1 | +1 | +1 | 1 |
| | | | | Clustering | 0 | +1 | +1 | +1 | 3 |
| | | | | Association | 0 | -1 | +1 | +1 | 1 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Random Forrest | +1 | -1 | 0 | +1 | 1 |
| | | | | Neural Networks | +1 | -1 | -1 | +1 | 0 |
| | | | | GAN's | 0 | -1 | -1 | +1 | -1 |
| | | | | Naïve Bayes | 0 | -1 | 0 | +1 | 0 |

Assumptions: Blue Force asset/sensor conducts assessment on target. Data will require assignment to predefined categorical values with pre-existing datasets available based on priority. Explainable output is mandatory, and number of predictors is low to enable higher accuracy.

| | | | AI/ML Method | DP #1 | DP #2 | DP #3 | DP #4 | Total |
|---|---|---|---|---|---|---|---|---|
| 1.6.2 | Evaluate | DP1: Qualitative<br>DP2: Unsupervised Learning<br>DP3: XAI Mandatory<br>DP4: 1–9 Predictors | Linear Regression | 0 | -1 | +1 | +1 | 1 |
| | | | Logistic Regression | +1 | -1 | +1 | +1 | 2 |
| | | | Clustering | +1 | +1 | +1 | +1 | 4 |
| | | | Association | +1 | -1 | +1 | +1 | 2 |
| | | | Random Forrest | +1 | -1 | 0 | +1 | 1 |
| | | | Neural Networks | +1 | -1 | -1 | +1 | 0 |
| | | | GAN's | 0 | -1 | -1 | +1 | -1 |
| | | | Naïve Bayes | +1 | -1 | 0 | +1 | 1 |

Assumptions: Blue Force C2 System evaluates target or threat status. If/then structure is used in order to assess Target or threat against established criteria. Explainable output is desired, and number of predictors is low to enable higher accuracy.



THIS PAGE INTENTIONALLY LEFT BLANK

# APPENDIX C. MACHINE LEARNING TOPICS DETAIL

This appendix provides greater detail to methods and topics presented in Chapter III of this report and aims to ensure that the report provides value to readers with varying levels of AI/ML understanding. As such, the information that follows can be viewed as supplementary to ideas presented in Chapter III with level of detail greater than likely required to understand the analysis conducted by the study.

## 1. Statistical Learning Methods

Machine learning uses algorithms developed from these statistical methods and applies them in ways that automatically enhance learning techniques through a combination of increasing data and experience. Modern statistical learning techniques are focused on developing models to predict new observations. Methods are chosen based on the type of response values: numerical or categorical, or the structure of the provided dataset. Learning implies developing a function to characterize the behavior of the dataset's predictor variables (X) and their response values (Y) or identify patterns and structure in the dataset where there is no explicit response variable. Additional techniques are used to evaluate and identify the model with best performance for the given dataset. These techniques include validation, and breaking up the source data into training, validation, and test sets. Overcoming bias and variance, and overfitting are also problems that require addressing in statistical learning models. Techniques for adjusting for overfitting include fine tuning performance-based attributes (hyperparameters) for complexity control. Performance based metrics are used to assess model quality.

### *a. Supervised Learning*

Supervised learning is a form of statistical learning. "In supervised learning, the goal is to predict the value of an outcome measure based on a number of input measures" (Hastie, Tibshirani, and Friedman 2017, xi).

> For each observation of the predictor measurement(s) $x_i$, $i = 1, \ldots, n$ there is an associated response measurement $y_i$. We wish to fit a model that relates the response to the predictors, with the aim of accurately predicting the

> response for future observations (prediction) or better understanding the relationship between the response and the predictors (inference). (James et al. 2017, 26)

Supervised learning methods can be divided into three main technique areas:

(1) Regression

- Linear Regression – Linear relationships are identified between input predictors (measures) (X) and output variables (Y) to develop functions that create predictive models. Response values are numerical. Discussed in greater detail in section V.D.1.a(4)

(2) Classification

- Logistic Regression – relationships and predictive models are developed based on a categorical response. Discussed in greater detail in section V.D.1.a(5)

- KNN - The K-Nearest Neighbors method uses a prediction algorithm to predict the response of new observations by finding a "K" number of observations in the training set that are closest or most similar to a new observation, and then take the average of the responses.

- Naïve Bayes – Probabilistic classifier based on Bayes' theorem.

(3) Trees

- Trees (combo of regression and classification) "involve stratifying or segmenting the predictor space into a number of simple regions. To make a prediction for a given observation, we typically use the mean or the mode of the training observations in the region to which it belongs. Since the set of splitting rules used to segment the predictor space can be summarized in a tree, these types of approaches are known as decision tree methods" (James et al. 2017, 303).

- CART - The Classification and Regression Tree model uses a predictive "decision-tree" algorithm to partition the training set predictor space into multidimensional boxes. Trees are used to express predictability of model responses.

- Random Forest - The Random Forest model builds on the CART model by generating many trees via bootstrapping. Bias is minimized by growing large trees and variance is reduced by averaging the results over the trees (bagging).

- Boosting – Similar to random forest but with many trees generated in a sequential manner. Representative of best "off-the-shelf" algorithms.

Supervised learning emphasizes learning based on training data. After first determining the response data type and selecting the appropriate method, the dataset is split into training, validation, and test sets. For overly complex models, and to avoid overfitting, regularization methods can be incorporated to tune hyperparameters and control complexity. Cross-validation can be used to identify the best performing model when choosing between multiple supervised learning methods. For categorical response variable models such as logistic regression, the optimal binary response value can be determined and applied to the model. The model is then evaluated on the test set and accuracy analyzed based on the appropriate performance metric for the chosen supervised learning method.

(4) Regression

Regression is a form of supervised learning, and thus the principal objective of regression methods is estimating relationships between variables with a numerical response, as in the dataset has a dependent response variable. Regression is one of the most common machine learning models. Regression is predictive and models are reliant on historical data. This data consists of continuous dependent and independent variables; however, regression models can also include categorical predictor variables.

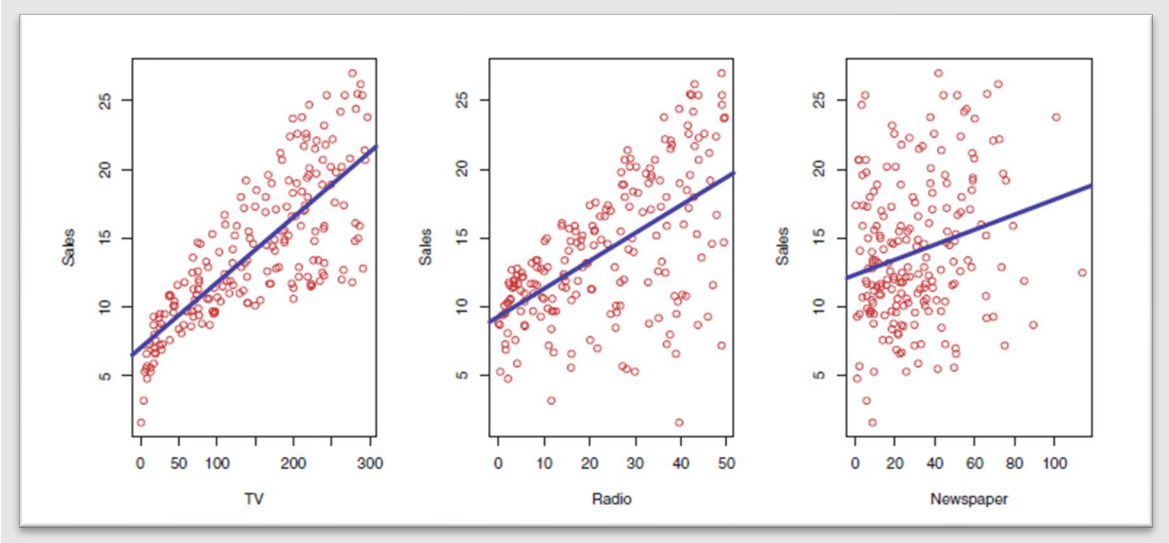



Figure 45. Sales as a Function of Budget. Source: James et al. (2017, 16).

Figure 45 "displays sales, in thousands of units, as a function of TV, radio, and newspaper budgets, in thousands of dollars, for 200 different markets...each blue line represents a simple model that can be used to predict sales using TV, radio, and newspaper, respectively" (James et al. 2017, 16).

The term regression can be considered a misnomer as for example, logistics regression is typically a classification method. "Least squares linear regression is used with a quantitative response, whereas logistic regression is typically used with a qualitative (two-class, or binary) response. As such it is often used as a classification method. But since it estimates class probabilities, it can be thought of as a regression method as well" (James et al. 2017, 28).

Regression is a form of supervised learning; thus, the model's dataset must be broken up into training, validation, and test sets. Models cannot be evaluated with the same data used to build them. Training data is used to teach the model to identify relationships in the data and define a functional model. Validation is then performed to evaluate model selection by estimating the test prediction error. "The model is fit on the training set, and the fitted model is used to predict the responses for the observations in the validation set. The resulting validation set error rate—typically assessed using MSE (mean squared error) in the case of a quantitative response—provides an estimate of the test error rate" (James

et al. 2017, 176). One form of validation is cross-validation in which many training and validation set splits are created. Overly complex models can lead to overfitting. As complexity increases the training error decreases. Validation allows for determining the optimal test error that balances prediction error and model complexity, bias, variance. Once validation has been performed, the model can finally be applied to the test set for assessment of final model prediction performance.

To address overly complex models and overfitting, additional techniques can be applied to regression methods. These techniques are referred to as regularization, also known as shrinkage. "This approach involves fitting a model involving all p predictors. However, the estimated coefficients are shrunken towards zero relative to the least squares estimates" (James et al. 2017, 204). Regularization techniques include Lasso, Ridge, and ElasticNet. These regularization techniques add varying degrees of penalties for model complexity to reduce overfitting. These penalties are based on the result of tuning model hyperparameters that specify the degree of complexity penalization.

Regression methods should be chosen when training data is available, and the model's predictive output response is continuous and numerical.

(5) Classification

Classification is a form of supervised learning, and thus also requires training data, but unlike regression methods, classification is used for predicting an output response data type that is discrete and categorical (qualitative). Apart from linear regression, the regression methods outline above can also be used for classification problems.

> Predicting a qualitative response for an observation can be referred to as classifying that observation, since it involves assigning the observation to a category, or class. On the other hand, often the methods used for classification first predict the probability of each of the categories of a qualitative variable, as the basis for making the classification. In this sense they also behave like regression methods. (James et al. 2017, 127)

Logistic regression, like linear regression, is a supervised statistical learning technique.

> Logistic regression models are used mostly as a data analysis and inference tool, where the goal is to understand the role of the input variables in explaining the outcome. Typically, many models are fit in a search for a parsimonious model involving a subset of the variables, possibly with some interactions terms. (Hastie, Tibshirani, and Friedman 2017, 121)

Logistic regression is also predictive, but unlike linear regression predicts a categorical output response as opposed to a numerical value.

> Logistic regression models the probability that Y belongs to a particular category. Logistic regression is typically used with a qualitative (two-class, or binary) response. As such it is often used as a classification method. But since it estimates class probabilities, it can be thought of as a regression method as well. (James et al. 2017, 130)

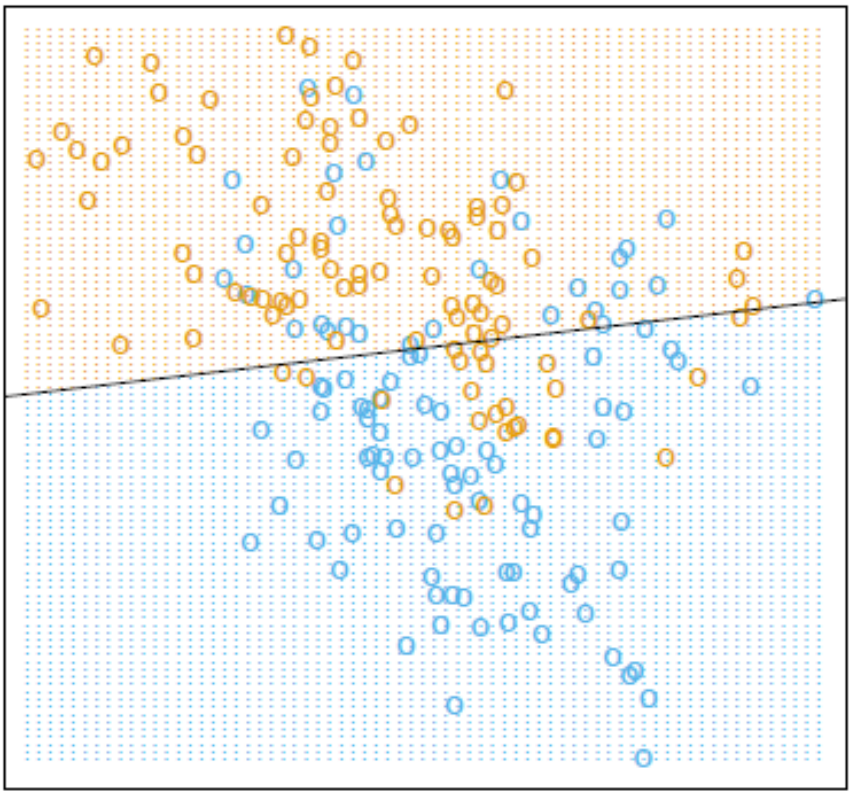


Figure 46. Classification Example. Source: Hastie, Tibshirani, and Friedman (2017, 13).

Figure 46 "illustrates a classification example in two dimensions. The classes are coded as a binary variable (BLUE = 0, ORANGE = 1), and then fit by linear regression.

The orange shaded region denotes that part of input space classified as ORANGE, while the blue region is classified as BLUE." (Hastie, Tibshirani, and Friedman 2017, 13).

Models that predict a quantitative value can be compared using the Mean Squared Error (MSE) between the predicted response and actual response. Models that predict a qualitative value can be compared using Recall, Precision and F-score. Recall and Precision are typically presented in a confusion matrix which is a table that the predicted vs. actual response variables for a set of test data for which the true values are known.

For example, we want to evaluate whether a piece of equipment (missile, vehicle, ship, etc.) is a foe. Here is an example in a simple table with a binary classifier of 259 data points (predictions) where a positive (yes) is a foe:

Table 29. Confusion Matrix (# of foes)

| n=259 (total number of predictions) | Predicted: No | Predicted: Yes |
|---|---|---|
| Actual: No | 165 | 6 |
| Actual: Yes | 8 | 80 |

As one can see the model correctly predicted 165 times when there was no foe and 80 times when there was a foe. Now, for some terminology and definitions from based on the website dataschool,io ("Simple Guide to Confusion Matrix Terminology" 2014):

- **true positives (TP):** These are cases in which we predicted yes (they are a foe), and they are a foe.
- **true negatives (TN):** We predicted no, and they are not a foe.
- **false positives (FP):** We predicted yes, but they are not a foe. (Also known as a "Type I error.")
- **false negatives (FN):** We predicted no, but they are a foe. (Also known as a "Type II error.")

Listed below are rates computed from Table 30's confusion matrix and closely based on the steps and the common formulas as provided from the website dataschool.io ("Simple Guide to Confusion Matrix Terminology" 2014):

- **Accuracy**: The classifier is correct by this proportion
- $$\frac{(TP+TN)}{total}=\frac{(165+80)}{259}=0.95$$
- **Misclassification Rate**: This shows what proportion it is wrong
- $$\frac{(FP+FN)}{total}=\frac{(8+6)}{259}=0.05$$
- equivalent to 1 minus Accuracy
- also known as "Error Rate"
- **False Positive Rate**: How often does it predict yes, when no
- $$\frac{FP}{actual\_no}=\frac{6}{171}=0.04$$
- **True Negative Rate**: It predicts no when it is no
- $$\frac{TN}{actual\_no}=\frac{165}{171}=0.96$$
- equivalent to 1 minus **False Positive Rate**
- also known as "**Specificity**"
- **Prevalence**: The yes condition happens in our sample by this proportion
- $$\frac{actual\_yes}{total}=\frac{88}{259}=0.34$$
- **Precision**: How often is it correct when it predicts yes?

- $\frac{TP}{predicted\_yes} = \frac{80}{86} = 0.93$

- **Recall or Sensitivity or True Positive Rate**: When it is yes, how often does it predict yes

- $\frac{TP}{actual\_yes} = \frac{80}{88} = 0.91$

- **F-score**: Weighted average of **Recall** and **Precision:**

- $F = \frac{2}{\left[\frac{1}{\mathrm{Re}call} + \frac{1}{\mathrm{Pr}ecision}\right]}$ or,

- $F = \frac{TP}{\left[TP + 0.5 \times (FP + FN)\right]}$

- In our example: $F = \frac{80}{\left[80 + 0.5 \times (6 + 8)\right]} = \frac{80}{87} = 0.92$

The F-score is one way to blend Precision and Recall into a single number. However, there may be different weights for each measure. The Matthews correlation coefficient (MCC) is considered a better means of evaluating a binary evaluation classification due to this difference in importance.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

This formula may be a little involved, but it is regarded as a much better measure because it considers all four members of the confusion matrix (TN, TP, FN, FP). The above formula with our values, solved:

$$MCC = \frac{[(80 \times 165) - (6 \times 8)]}{\sqrt{[(80 + 6) \times (80 + 8) \times (165 + 80) \times (165 + 8)]}} = \frac{13{,}152}{17{,}919} = 0.734$$

Another popular classification algorithm is the Naïve Bayes classifier which is based on Bayes' rule which underlies most modern AI systems for probabilistic inference. The general case for Bayes rule is as follows:

$$P(Y \mid X) = \frac{P(X \mid Y) \times P(Y)}{P(X)}$$

When one initially views Bayes rule it seems like we substitute one probability for three thus making the calculation more complicated. When we are calculating the unknown probability P(Y|X) with known probabilities P(X), P(Y) and P(X|Y). If x=cause and y=effect, then Bayes rule is transformed to:

$$P(effect \mid cause) = \frac{P(cause \mid effect) \times P(effect)}{P(cause)}$$

For a labeled training set the effect is the variable you want to predict, and the cause is the features that lead to the effect. If we have multiple causes (i.e., multiple input features) then bayes rule becomes:

$$P(Y \mid (X1 \& X2 \& \cdots Xn)) = \frac{P((X1 \& X2 \& \cdots Xn) \mid Y) \times P(Y)}{P(X1 \& X2 \& \cdots Xn)}$$

To simplify the rule, we can assume independence of the input features, this is called Naïve Bayes.

$$P(Y \mid (X1 \& X2 \& \cdots Xn)) = \frac{P(X1 \mid Y) \times P(X2 \mid Y) \times \cdots P(Xn \mid Y) \times P(Y)}{P(X1 \& X2 \& \cdots Xn)}$$

With Naïve Bayes we can calculate the cause-and-effect probabilities from a labeled training set. When presented with a set of causes we can find the effect by applying the above theorem. Naïve Bayes can only be applied when a categorical response is needed. Some examples of Naïve Bayes are target identification and text classification.

(6) Trees

Random Forest methods generate many decision trees and average the results over the entire population of trees. Numerous trees must be generated, and those trees must have low correlation to reduce variance. Bootstrapping is a method for generating this new data and it does so by sampling from the original training set. “The bootstrap method provides a direct computational way of assessing uncertainty, by sampling from the training data” (Hastie, Tibshirani, and Friedman 2017, 261). “Bootstrap aggregation, or bagging, is a general-purpose procedure for reducing the variance of a statistical learning method,” (James et al. 2017, 316) and is a tool that creates many bootstrap datasets, analyses them and averages the result over all the results of the bootstrap sets. One detractor to using a Random Forest is the low explain ability of the output model.

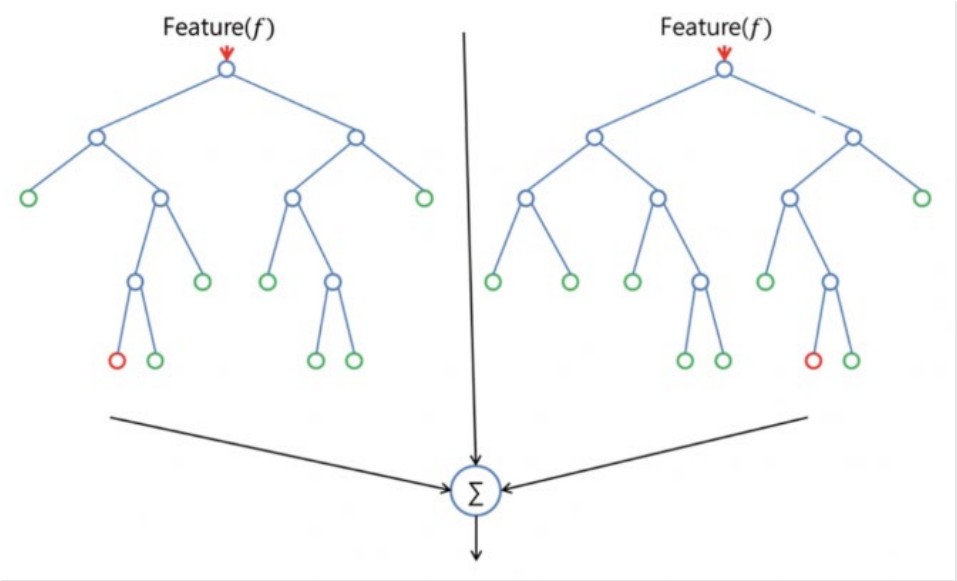



Figure 47. Random Forest Example (2 of Many Trees Shown) Unsupervised Learning. Source: Donges (2021).

### *b. Unsupervised Learning*

Unsupervised learning is a type of machine learning that uses algorithms to discover data patterns or groupings in unlabeled/untagged data sets. Unsupervised learning is beneficial when analyzing unknown data to reveal patterns in the data. The main

differences between supervised vs. unsupervised learning are demonstrated in Figure 48 is that supervised learning provides labels to find patterns in existing structure where no labels are provided to find structure in unsupervised learning (Jones, Kruger, and Johnston 2020). Supervised learning algorithm will be able to classify the images as cats and dogs where the unsupervised learning will be able to cluster images by similarities or differences.

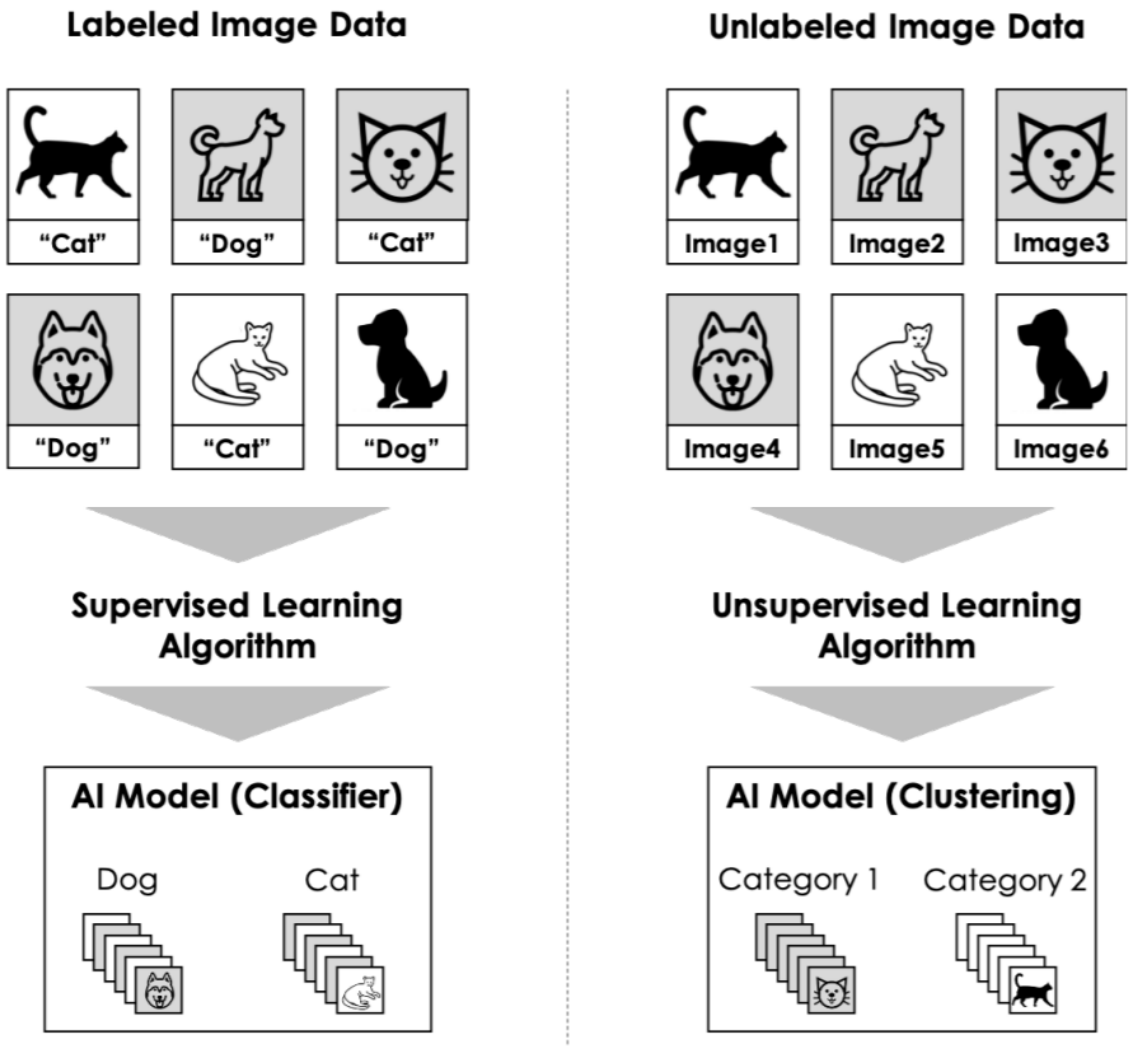



Figure 48. Supervised vs. Unsupervised Algorithms. Source: Jones, Kruger, and Johnston (2020).

This section will focus primarily on two of the most common unsupervised learning methods:

4. Clustering

- K-means – Algorithm that attempts to partition datasets into a pre-defined number (k) of clusters.
- Hierarchical – Clustering method that does not require a pre-defined number of clusters. Can be top-down (divisive) or bottom-up (agglomerative).

5. Association

Figure 49 highlights the general difference between the two unsupervised learning methods.


Figure 49. Clustering vs. Association. Source: Diaz (2021).

(1) Clustering

Clustering is a form of unsupervised learning which separates datasets into groups based on similarities in properties and/or features which potentially uncover meaningful relationships and patterns amongst samples in the dataset. Clustering can serve multiple purposes depending on the need of the user. These include:

- Data reduction – finding a representative set of data for a corresponding group of data
- Natural clustering – discovering natural clusters or useful data types and gaining insight into unknown properties
- Outlier detection – can enable the identification of unusual data

Two popular methods of clustering are k-means and hierarchical clustering. K-means is an algorithm that begins with a predefined number of clusters (k). Analysis of data, oftentimes by visualization or by hyperparameter tuning, is usually required to determine a suitable number of clusters. Initially, center points for the k clusters are initialized at random. The algorithm will cycle through each of the N data points, determine the closest cluster center, and assign the data point to the associated cluster. Once all N data points have been assigned to clusters, the cluster centers are recalculated and the process repeats. This process will continue until clusters cease to change. Usually, the clustering process is repeated multiple times to ensure the best clustering solution is found typically determined by comparing total variation amongst all clustering models.

Hierarchical methods are also popular for clustering. There are two strategies in hierarchical clustering algorithms: agglomerative and divisive. In Agglomerative Hierarchical Clustering, each data point is defined as its own cluster. The algorithm then calculates the distances of each data point from all other data points, resulting in a distance matrix. The closest two points are grouped together as a cluster and then distances are recalculated. This process repeats until either some predefined threshold for termination or until there is one large cluster containing all data points. Following the completion of the clustering a dendrogram can be generated which represents the hierarchical tree generated by the algorithm. This graphical representation provides the user with the capability to visualize how clusters are represented at varying tree heights. Divisive Hierarchical Clustering is similar, except rather than starting with N clusters for N data points, it begins with one large cluster containing all N data points (hence the top-down description). Like the agglomerative method, this process will continue until either a defined termination threshold is met or until all N data points are in their own clusters.

(2) Association

Association, commonly referred to as Association Rule Mining, is another form of unsupervised learning that examines large datasets to find relationships between variables. Association rules are often represented as if/then statements and are commonly seen in recommender system implementations. Consider shopping on Amazon and seeing suggested products – these suggestions are selected based on a perceived similarity between you and other customers with similar purchase histories. A popular association rule mining algorithm is the a-priori algorithm. This algorithm begins by calculating the support, or the frequency, of each itemset with the goal of reducing the amount of itemsets that must be considered. Itemsets not meeting a pre-defined support threshold are pruned from consideration. The process continues by generating every combination of remaining itemsets— it then performs the support calculation and pruning step again. This continues until a point at which there are no longer any candidate itemsets for pruning. The resulting output is several itemsets that represent discovered association rules. To determine usefulness of the discovered rules, the following metrics are calculated:

- Confidence: indication of how often the rule has been found to be true

$$\text{Confidence}(\{X\}\text{-}\{Y\}) = \frac{\text{Transactions containing both X and Y}}{\text{Transactions containing X}}$$

- Lift: ratio of the observed support to that expected if X and Y were independent

$$\text{LIFT}(\{X\}\text{-}\{Y\}) = \frac{(\text{Transactions containing both X and Y})(\text{Transactions containing X})}{\text{Fraction of transactions containing Y}}$$

Confidence can be inflated based on the popularity of the consequent – if the popularity of the consequent is high, there will be a higher chance that an itemset contains both X and Y. Thus, Lift is calculated and compares confidence with expected confidence while controlling for the popularity of Y. A lift value greater than 1 signifies that the consequent is likely to occur with the antecedent; lift < 1 signifies that the occurrence of the consequent has a negative effect on the presence of the antecedent (and vice versa); lift

= 1 signifies that the probabilities of the antecedent and of the consequent are independent of each other and have no meaningful rule relationship.

### *c. Reinforcement Learning*

Reinforcement learning is a form of Machine Learning that utilizes an agent that learns from the feedback (rewards and agent state) from an action completed in its intended environment. In Figure 50 below, Pac-Man is the agent that completes an action (moves) in the maze (environment). The feedback is the state of the agent state (avoid ghost – alive or dead) and did he gain a reward (food). Pac-Man in a reinforcement learning environment would use a mathematical model for decision making called the Markov decision processes to analyze and learn from the cause and of effects (rewards and agent state) of actions that result in his environment (maze). In this example, Pac-Man would automatically learn to maximize rewards while avoiding ghosts by just the direct maze interaction without external intervention or supervision (Sutton and Barto 2018). Supervised learning would have Pac-Man respond to new mazes based on the training data alone to predict the correct action where reinforcement learning would learn from each new maze interaction to maximize reward. “In interactive problems it is often impractical to obtain examples of desired behavior that are both correct and representative of all the situations in which the agent has to act” (Sutton and Barto 2018, 2). Unlabeled data or unsupervised learning would help Pac-Man learn the hidden structure of the new mazes (potentially) but would not teach Pac-Man how to avoid ghosts and eat the maximum amount of food (maximize award).

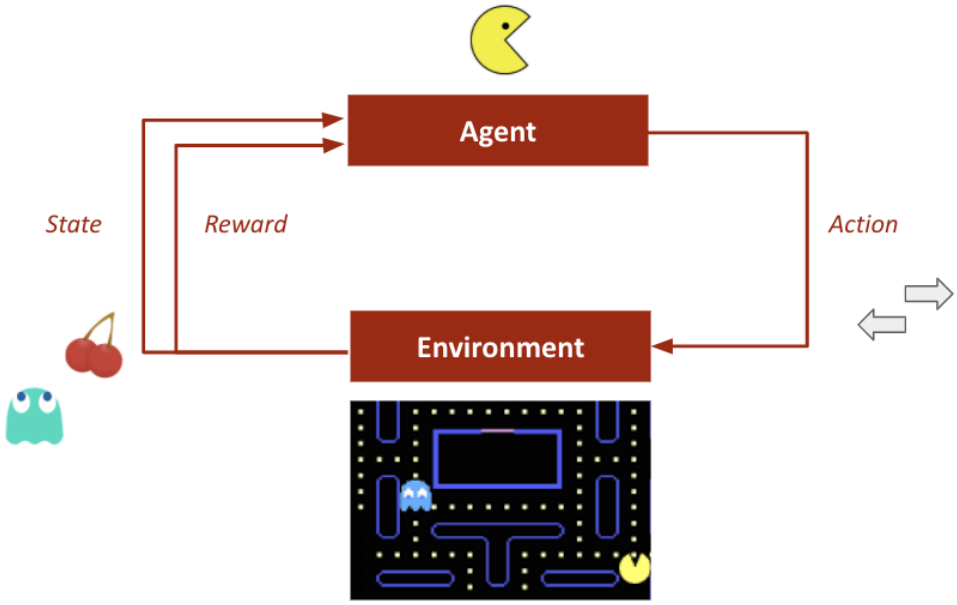



Figure 50. Reinforcement Learning Environment Components. Source: Faik (2021).

To utilize reinforcement learning, the agent must be able to receive feedback from the environment (rewards, state) and understand the basic goal (avoid ghost – eat food). Additionally, the agent must be able to take actions in the intended environment that affect achieving this basic goal. To use reinforcement learning, there is "trade-off between exploration and exploitation" (Sutton and Barto 2018, 3). The reward seeking actions of the agent when presented with a new environment are based on positive cause-and-affect actions from the past; therefore, there is a potential for this new action may result in a negative outcome. Iteratively, the agent learns from these actions to make the best award seeking action in the future but not without risk. Additionally, the agent takes learned action that may receive an immediate reward but also takes actions that also considers the immediate actions that may affect the future or a collective reward strategy. Pac-Man may avoid an immediate food reward to avoid a ghost in the future or to set himself up to gain more rewards over the course of the game.

(a) Generational Adversarial Network (GAN)

A Generational Adversarial Model (GAN) is a mix of unsupervised learning and reinforcement learning. The GAN methodology utilized two unique types of models:

- The Generator Model, this is the model generation step, and it can take real work input or statistical inputs from a user defined distribution (typically gaussian). The goal is to generate models that the discriminator model classifies as real, thus an errant classification.
- The Discriminator Model takes the output from the generator model and predicts whether it thinks its fake or real (binary output). This output is then fed back into the generator model for updates, this allows the generator model to create better representations of the data. After the training process the discriminator model is discarded as we are concerned with the generator model.

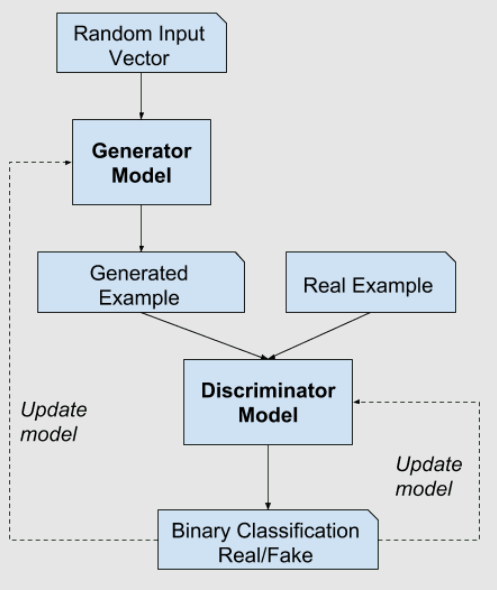



Figure 51. Generative Adversarial Model (GAN). Source: Brownlee (2014).

### 2. Contextual Reasoning/Adaptation

#### *a. Neural Networks*

Neural networks tend to shine when the number of inputs becomes computationally cumbersome to Wave 2 machine learning algorithms. Computer vision, natural language processing and simulations are disciplines taking advantage of the neural networks and their properties. The name and structure of neural networks are inspired by the human brain, mimicking the way that biological neurons signal to one another. At an elevated level, a neural network is a means to map a set of inputs to an output and allow learning along the way.

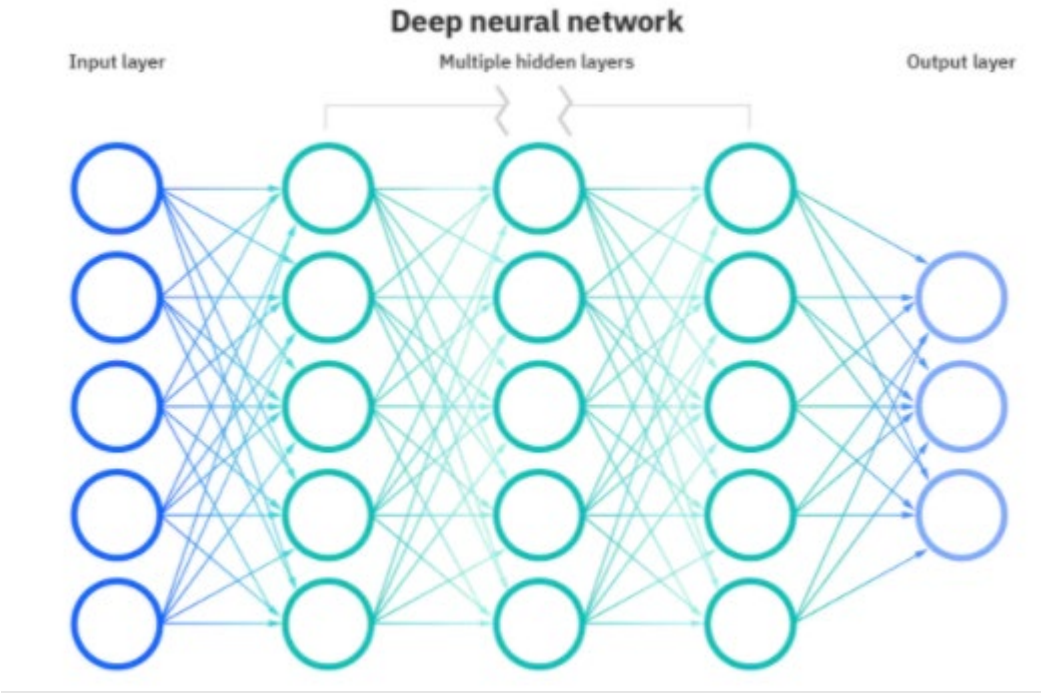



Figure 52. Neural Network Diagram. Source: IBM Cloud Education (2020).

Each layer can be thought of as a data transformation step. Neural Network Terminology:

- **Input Layer:** input features

- **Output Layer:** transforms its input from the hidden layers into an output using a defined activation function
- **Forward Propagation:** cyclical process that allows the weights to be updated based upon the error of the output.
- **Backward Propagation:** Once the output has been scored how to update the weights to improve the score of the data at the output layer. This is how the neural network learns.
- **Loss Function:** Score of the values at the output layer (compared against a training dataset) (Mean Squared Error, log-loss, etc.….)
- **Activation functions:** Mathematical function that operates on the input data to produce a bounded output. The exact function is selectable by the user.

Note that a neural network consisting of more than three layers is considered Deep Learning (Wave 3). A neural network consisting of 1 or 2 layers is considered part of Machine Learning (Wave 2).

(1) Alpha Go example

As an example of Artificial Intelligence implementation, the Google AlphaGo algorithm will be explored. AlphaGo is an AI computer program that plays the board game "Go." In March of 2016 AlphaGo beat the 18-time human world champion in a tournament winning 4 of 5 games.

Figure 53. Go Board Game. Source: Getty Images.

The board game Go can be describes as follows:

> The rules of the several-thousand-year-old game of Go are extremely simple. The board consists of 19 horizontal and 19 vertical black lines. Players take turns placing either black or white stones on vacant intersections of the grid with the goal of surrounding the largest area and capturing their opponent's stones. Once placed, stones cannot be moved again. Despite the simplicity of its rules, Go is a mind-bogglingly complex game—far more complex than chess. A game of 150 moves (approximately average for a game of Go) can involve 10^360 possible configurations, more than there are atoms in the Universe. (Hölldobler, Möhle, and Tigunova 2017, 92)

Chess is less complex in terms of total move combinations when compared to Go, thus the AI for chess can simulate all potential moves to determine the correct next move, this is referred to as a brute force approach. A brute force approach is computationally difficult with Go given the current state of the art processors. The AlphaGo Algorithm utilizes two distinct types of machine learning: decision tree and neural networks. The decision tree is used to keep track of the state of the board and possible moves. The Neural Network is used to simulate a game to completion.

The moves in Go can be represented by a decision tree. Descending through the tree one can represent the sequential moves of the game. The optimal move is found by searching through the complete game tree. These are the moves that the algorithm determines have the highest probability of victory at the end of the game. As mentioned

previously a complete game tree from start to every end is possible with chess but due to complexity, this is unfeasible with Go.

To overcome the inherent complexity of Go one can modify the decision tree algorithm. This modification is a Monte Carlo Tree Search (MCTS).

> Briefly, in MCTS during the descent each move is chosen according to its value, which is accumulated by making random simulations, each one representing a complete game. The value for the move reflects the information of the number and the outcome of the simulations that have run through it. This approximation is justified by a Central Limit Theorem, which says that the Monte Carlo values (mean of all outcomes) converge to the normal distribution. If the tree is explored to a fair extent, the strategy using MCTS converges to the optimal strategy. (Hölldobler, Möhle, and Tigunova 2017, 94)

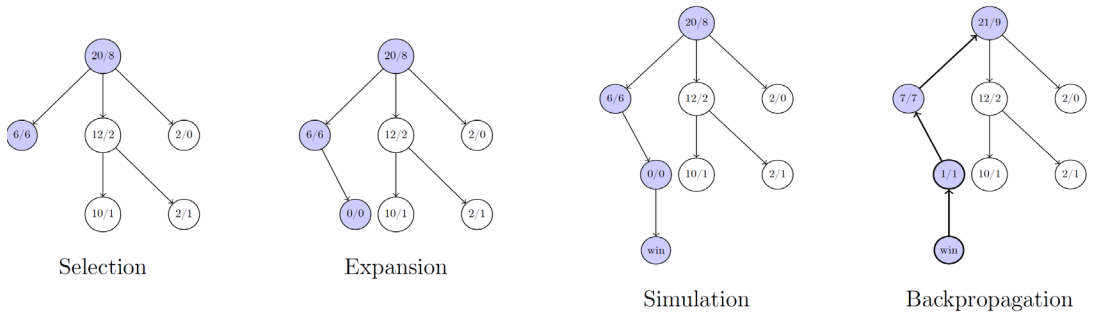



Figure 54. MCTS (All Visits/Winning Visits is the Value in Each Node). Source: Hölldobler, Möhle, and Tigunova (2017, fig. 1).

Sequential steps of the MCTS (Hölldobler, Möhle, and Tigunova 2017):

1. **Selection:** The algorithm determines the next node by seeking a balance between exploitation (selecting the nodes with the highest win rate) and exploration (seeking the least explored moves)
2. **Expansion:** At a node we have yet to encounter, this node is added, and a random value is assigned for all visits/winning visits)
3. **Simulation:** After adding a node you need to adjust the assigned value using a random simulation of the game to the terminal point.

4. **Backpropagation:** When we reach the end of the game, we determine win/lose and update the values of each parent node with this data. On the way down the tree increment the visit count and, on the way, up add the reward for win/lose.

MCTS does not require any inherent game knowledge, it will determine the path based upon the simulation (which does require game knowledge). This broadens the reach of MCTS because it can be thought of as a general algorithm applicable to any game that can be simulated. With any Monte Carlo method to converge to a true solution one must have infinite simulations, to shortcut this requirement we can use Neural Networks to mimic the playing style of a human. This reduces the number of modes because randomization is no longer used.

AlphaGo uses neural networks to predict human moves and simulate a Go game to completion. Neural Networks at an elevated level are a computer program that is meant to simulate a human's thought process by means of a neural pathway. Neural Networks can lean in two ways via supervised learning or via unsupervised (reinforcement) learning. For the Supervised Learning (SL) the algorithm analyzed recorded Go games. For Reinforcement Learning (RL) the algorithm plays itself, note this usually follows supervised learning (see below). The output of the learning is a value network which given a position in the game outputs a single value denoting win or loss.

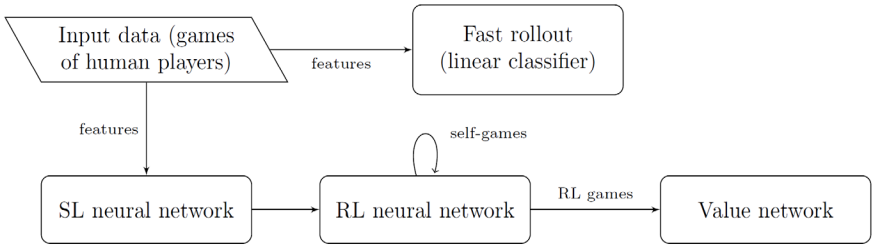



Figure 55. Learning Pipeline of AlphaGo (SL=Supervised Learning, RL=Reinforcement Learning). Source: Hölldobler, Möhle, and Tigunova (2017, fig. 5).

What the lessons to be learned from AlphaGo?

> The power of AlphaGo is undoubtedly vast, it managed to make a revolution in the state-of-the-art AI. The number of games AlphaGo played to train itself was more than all the people have ever played [WZZ+16]. However, AlphaGo is not a breakthrough technology, all the methods that it uses have been known and developed for a long while. Therefore, we can claim that AlphaGo is a consequence of the recent research in computer Go. (Hölldobler, Möhle, and Tigunova 2017, 100)

During the match against the world champion the AlphaGo algorithm made numerous moves that the experts deemed unconventional. These unconventional moves were honed through Reinforcement Learning. One takeaway from the AlphaGo implementation is that Reinforcement learning can explore the solution space to determine novel moves that maximize the value network.

### 3. Specialized Topics in AI

#### *a. Explainable AI*

The three main components of XAI are Explainable Models, Explanation Interface, and Psychology of Explanation. Inclusion of these components into the design process early in the life cycle will save redevelopment costs and time. Given these components one can see how XAI implementation is a key component in a Root Cause Analysis (RCA). When developing the three main components one must keep the target audience in mind to appropriately tailor the outputs. To help with this, Figure 56 illustrates the XAI target audience and their concerns/questions.

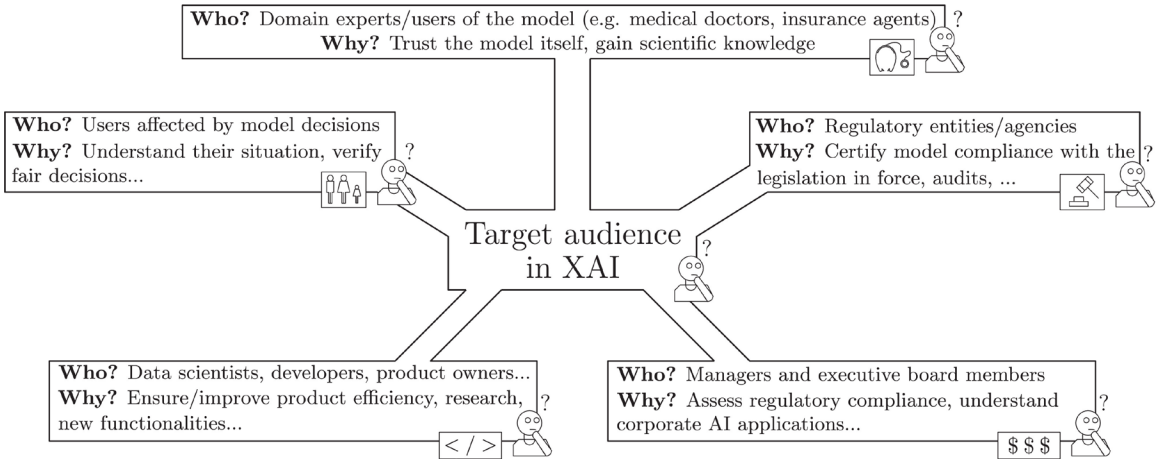



Figure 56. XAI Target Audience. Source: Gunning and Aha (2019).

"The reality of practical applications of AI and ML in sensitive areas (such as the medical domain) reveals an inability of deep learned systems to communicate effectively with their users. So emerges the urgent need to make results and machine decisions transparent, understandable and explainable" (Goebel et al. 2018, sec. Introduction). XAI and its benefits can be described visually using a dataset from an image classification algorithm seen in Figure 57.

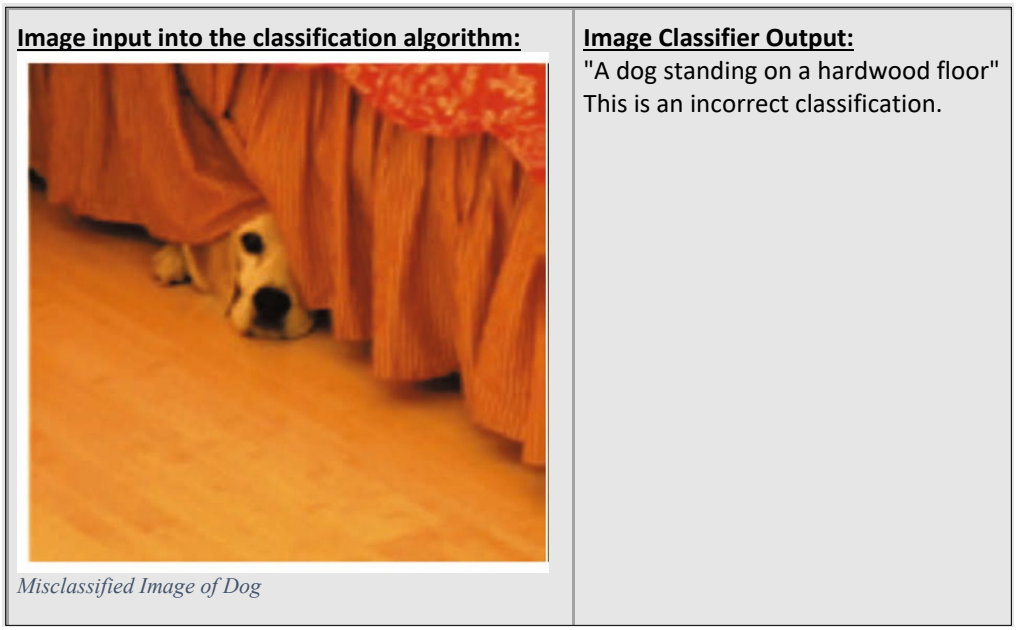



Figure 57. Image Classifier. Source: Holzinger et al. (2018).

When the image on the left is fed into an image classification algorithm the output was “A dog standing on a hardwood floor.” As you can see this image has been incorrectly identified by the AI/ML algorithm, the dog is lying down and not standing. Beyond the Boolean of did the algorithm correctly identify the image, the user has no insight into the reasoning behind that choice. During the RCA, the following questions are relevant:

- Will retraining with more standing dog pictures help the classification?
- Will training with more dogs laying down pictures help the classification?

Blackbox algorithms can leave the users grasping at straws when trying to perform an RCA. If during the machine learning algorithm development explainable interfaces are incorporated the algorithm output will contain an explanation for the classification. These XAI outputs will allow the user further insight into the reasoning behind the image classification output.

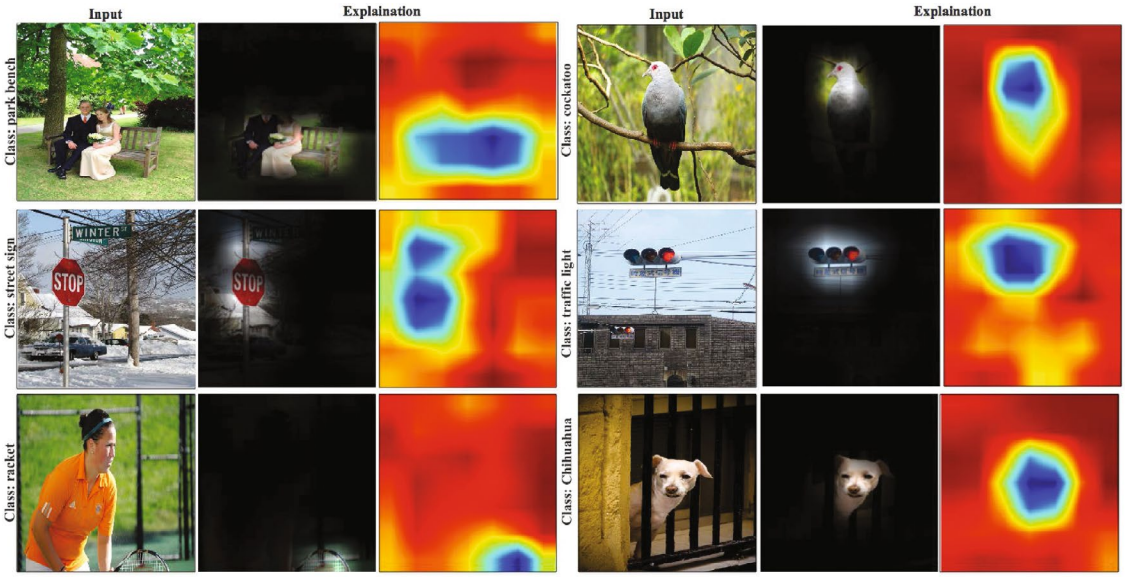



Figure 58. XAI Example. Source: Holzinger et al. (2018).

The explainable output can take the form of a text document. Or pictorial description seen in Figure 58. The first and third columns are the picture inputs, and the remaining columns are the algorithms explanation for its choice.

### *b. Big Data*

A term often associated with, and sometimes confused with AI, is Big Data. Big Data refers to large, complex, and potentially unstructured data sets. Big Data is traditionally characterized by three Vs:

1. ***Volume*** – Large sets of data being generated from a range of sources. Examples of this are source data from the multitudes of military radars and sensors that are used to establish common operational pictures from the strike group level to the global level.
2. ***Velocity*** – Influx of data has increased rapidly over time and, as a result, correlating, fusing, and acting on available data has become impossibly complex and burdensome for current manual processes.
3. ***Variety*** – Varying types of data characterizing different situations. Different radars and sensors have varying mission objectives and capabilities and thus capture varying types of data.

As time has progressed, additional vs. have been used to expand the definition of Big Data. Two of the more common ones are veracity and value. Veracity describes accuracy of the data along any inconsistencies and uncertainties in data. Value simply refers to the usefulness of data that is being captured. There are many challenges related to Big Data – how to capture the data, where to store the data, how to share data, information privacy, and data source reliability amongst others. When exploring Big Data, it is prudent to consider computer technologies necessary to leverage the large data sets. Cloud computing, parallel computing, and data storage are important considerations. While these technologies have rapidly advanced in recent years, the DOD faces challenges related to the uniqueness of military operations – various theaters of operation, mobility of units, intelligence organizations, bandwidth limitations, substantial amounts of differing sensors, etc.

Big Data serves as a foundation to advancing machine learning and AI technologies as the data serves as input to many of the methods detailed above. Data mining (this term is a misnomer—extraction of patterns is the goal) is a process of obtaining and determining

patterns in huge data sets. Data mining considers machine learning, statistics, and database systems. It enables knowledge discovery in these massive data sets that directly improve machine learning which, in turn, relates to potentially improved decisions output by AI algorithms. Additionally, there is an inverse relationship between AI and Big Data whereas AI is used to process large data sets to identify potentially "dirty" information. As mentioned above, veracity is a critical characteristic of Big Data and was explored further in the previous section.

Data mining also involves:

- data management (maintenance, common formatting...)
- data pre-processing (missing data, erroneous data...)
- differing systems have different information to process
- model and inference thinking
- "Interesting" metrics
- complexity considerations
- post-processing of discovered structures
- visualization
- updating (online, HDD, distribution...)

### *c. Feature engineering*

For a given object at a fixed moment in time multiple features can be defined resulting in a feature vector for the object. Features can be clustered into several feature types:

- **Categorical Feature** (finite possibilities) (eye color = (black, blue, brown, green, red, white, yellow))
- **Ordinal** (Categorical but with hierarchy Bachelors < Masters < PHD)

- **Numerical** (quantitative or continuous)
- **Ratio-scaled numerical feature** (weighted value w.r.t. some meaning)

A feature may not be useful in the ML method. The feature can be static throughout the dataset thus providing no information. The feature can be identical to a neighboring feature again providing no additional information. The ML method specific algorithm will dictate the usefulness of the feature: "The usefulness of a feature is measured ultimately in terms of the improvement the feature adds to the data analytic task at hand" (Dong and Liu 2018, 2). Feature Engineering includes a subset of topics such as:

- **Feature Transformation**- creating new features from existing features, normally using simple mathematical mapping
- **Feature Extraction-** creating new features from existing features, normally using complex mathematical mapping or pattern recognition.
- **Feature Selection and Analysis** - selecting a small set of features from a larger set, may lead to improved algorithm speed and reduced complexity. This selection is the result of the analysis of the usefulness of features.

The two main branches of feature selection are search-based and correlation-based. For the search-based feature selection the algorithm is bounded by a user defined stopping criterion.

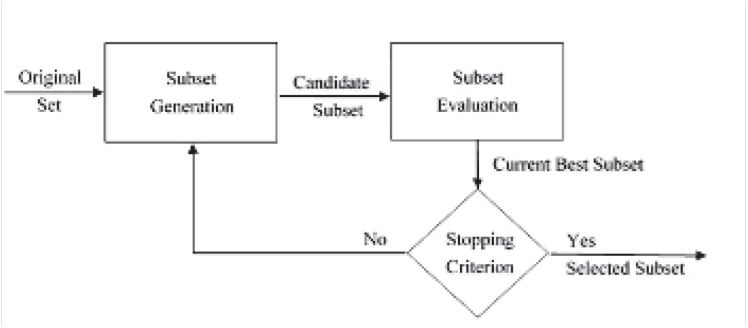



Figure 59. Search-Based Selection. Source: Dong and Liu (2018).

> Each candidate subset is evaluated and compared with the previous best one according to a certain evaluation criterion. If the newly generated subset is better than the previous one, it will be the latest best subset. The first two steps of search-based feature selection are repeated until a given stopping criterion is satisfied. (Dong and Liu 2018, 193)

Correlation based feature selection attempts to weed out the features that are highly correlated. . Specifically feature-feature correlation and feature-class correlation. (Dong and Liu 2018, 193) Feature to Feature correlation indicates redundancy and feature to class correlation indicates feature relevance. A Class can be thought of as an end state for the algorithm.

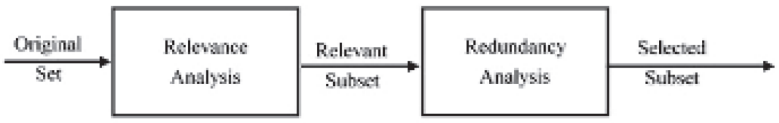



Figure 60. Correlation-Based Selection. Source: Dong and Liu (2018).

Proper application of feature engineering allows us to organize/rank the possible features to isolate the relevant features. Application of only relevant features to an AI/ML algorithm should improve performance and decrease the complexity of the algorithm and data.

### *d. AI Data Security*

Further detail regarding AI data security follows:

> Adversary's Goal: The goal of the attacker is to corrupt the learning model generated in the training phase, so that predictions on new data will be modified in the testing phase. The attack is considered a poisoning availability attack, if its goal is to affect prediction results indiscriminately, i.e., to cause a denial of service. It is instead referred to as a poisoning integrity attack, if the goal is to cause specific mis-predictions at test time, while preserving the predictions on the other test samples. This is a similar setting to that of backdoor poisoning attacks recently reported in classification settings. (Jagielski et al. 2021, 21)

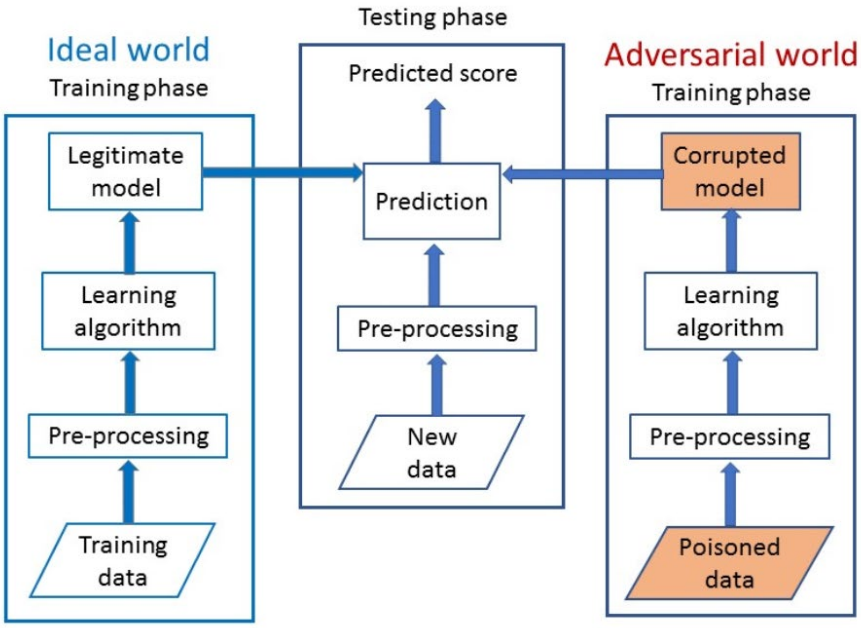



Figure 61. Data Poisoning System Architecture. Source: Jagielski et al. (2021, 21).

Figure 61 represents the general system architecture for training the algorithm in the presence of a data poisoning attack. For the adversarial world there are two main types of attacks:

- **White Box Attacks:** The attackers know the training data, the features, the learning algorithm, and the training parameters; an insider threat/ attack.
- **Black Box Attacks:** The attackers have no knowledge of the training set but can collect a representative dataset without the labeled data. The features and learning algorithm are known while the training parameters are not. However, the training parameters can be estimated using the learning algorithm and representative training dataset.

### *e. Game Theory*

Expanding on the definition presented in Chapter III, *Game Theory*, a game consists of a set of players with associated strategies and actions along with a final objective or payoff. A fundamental principle to game theory is the identification of the Nash equilibrium. This equilibrium assumes that each agent in the game has an awareness of the other agents' equilibrium strategies. Thus, achieving Nash equilibrium represents a scenario where there is no benefit for any agent to switch strategies. To illustrate, a well-known example is The Prisoners' Dilemma is presented below.

Consider the game consisting of two prisoners, prisoner A and prisoner B, who are suspected of committing a robbery together. They have each been detained and placed into different interrogation rooms and are being persuaded to confess to the robbery. Each prisoner is presented with two choices:

1. Stay silent
2. Confess

Additionally, they are informed of the varying payoffs related to each outcome depicted in Figure 62.

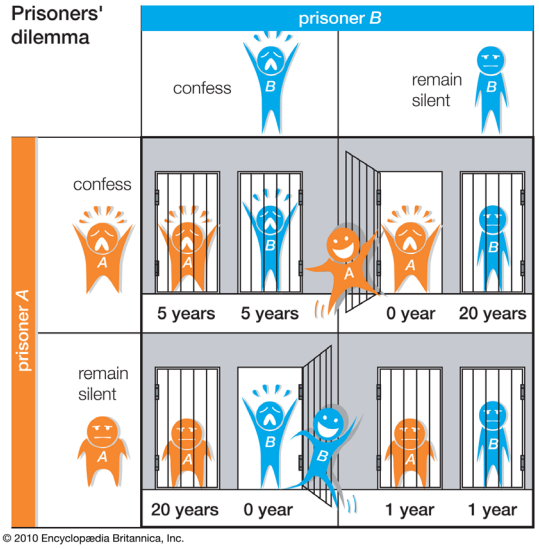



Figure 62. The Prisoners' Dilemma. Source: Encyclopedia Britannica, Inc.

- Upon initial investigation, it is intuitive to think that both prisoners would choose to remain silent as that would result in a sentence of 1 year for each of them. However, consider from prisoner A's perspective. Without knowing, for certain, what prisoner B will do, prisoner A understands that there is an evident self-interest of all game participants in minimizing the amount of imprisonment they each receive. So, as prisoner A considers their options, the rationale would look like this:

- If prisoner B confesses, it is best for me to confess because 5 years is better than 20 years

- If prisoner B remains silent, it is still best for me to confess since I would walk away with zero jail time.

- Prisoner B, being equally rational, would consider the same sequence of events. Thus, below is a summary in Table 30.

Table 30. Summary Table

| prisoner A \ prisoner B | confess | remain silent |
|---|---|---|
| confess | A: -5, B: -5 | A: 0, B: -20 |
| remain silent | A: -20, B: 0 | A: -1, B: -1 |

- The circles represent each prisoner's best option given the other prisoner's corresponding choice. The intersect of best options signifies that the best solution is {confess, confess} – this is a Nash Equilibrium strategy. This solution also addresses regret minimization as both prisoners can consider

this a “no regret” solution, even though it is not necessary the most optimal one. The possibilities exist that there are multiple Nash equilibria or none.

Table 31. Type of Games

| | |
|---|---|
| Symmetric vs. Asymmetric | Symmetric: Strategy and payoffs are the same for all agents in the interaction.<br><br>Asymmetric: Strategy and payoffs are different for different agents. |
| Zero Sum vs. Non-zero Sum | Zero Sum: Available resources of the game are constant and total benefits go to all agents in the game. Agents benefit at equal expenses to other agents.<br><br>Non-zero Sum: Gains by one agent do not necessarily correspond to a loss by another. |
| Simultaneous vs. Sequential (Extensive) | In simultaneous games, agents make moves simultaneously and have no awareness of other agents’ actions in advance.<br><br>In a sequential game, commonly referred to as an extensive game, the opposite is true. Each agent is aware of what actions the other agent is doing and can then choose actions that correspond to the other agent’s actions/states. |
| Cooperative vs. Non-Cooperative | Cooperative: Games in which agents can adopt strategies through negotiations and agreements with other agents.<br><br>Non-Cooperative: Agents decide on their own strategies to maximize their rewards, such as discussed above in the Nash Equilibrium section (self-interest) |
| Perfect vs. Imperfect vs. Incomplete Information | Perfect: Every agent has knowledge of all the possible actions other agents can take, what actions are currently being executed and the associated payouts for the actions.<br><br>Imperfect: Agents are aware of the nature and motive of the other agents and the payoffs associated with all outcomes, but they do not know what actions are currently being executed.<br>Incomplete: Agents do not have full understanding about opposing agents. They may have knowledge of current actions, but they do not know about the motivations of other agents, or the rewards associated with actions. |

**Imperfect and incomplete information games best represent real-world situations.**

Game theoretic applications have already been commissioned in active AI projects funded by government agencies. Examples of these are ARMOR which viewed LA Airport security as a game between defenders and terrorists, IRIS which is a Federal Air Marshall project that viewed the assignment of Marshalls to flights as a game of maximizing expected utility versus multi-agent adversaries and PROTECT which is a U.S. Coast Guard program that improved efficiency of U.S. port patrols. A link to a presentation of these applications is provided in the reference and give a firm view at how game theory can play a critical role in the improvement of AI in future efforts.

Further, the use of game theory has been utilized to improve AI algorithms. The most prominent example of this application is seen in the implementation of Generative Adversarial Networks (GANs) which are an approach to using deep learning methods. GANs are represented as two neural networks – a Generator Neural Network and a Discriminator Neural Network. The Generator is trained through supervised learning with a goal of establishing enough awareness of regularities and patterns in input data to generate new examples. Following the learning process, the Generator feeds examples, both real and generated, to the Discriminator whose job is to classify each instance as real or fake. These models compete against each other in a zero-sum game in which the Generator aims to fool the Discriminator at a predefined threshold frequency, usually 50% of the time. This state represents the Nash equilibrium of the game and signifies that the Generator is generating representative examples.

Another example of algorithmic game theory application can be found in EigenGame – a competitive multi-agent game used to improve principal component analysis (PCA). PCA is method used in machine learning with a goal to "identify a reduced set of features that represent the original data in a lower-dimensional subspace with minimal loss of information" (Kherif and Latypova 2020, 209). It does this by computing a set of vectors corresponding to the dimensionality of the data where each vector captures the most variance and least error while being orthogonal to the vector preceding it. Original

data points are projected onto these vectors, known as the principal components, and the result is a reduction in the number of features to consider. EigenGame reformulated the process from an optimization, or single-agent, problem to one that is a multi-agent game where each successive principal component is rewarded and penalized based on the relationship between the amount of variance it captures and its alignment with other "players," or principal components. The resulting PCA solution, when all "players" play optimally, represents the achievement of the game's Nash Equilibrium.

These are just a few examples of how Game Theory directly impacts and improves AI in domains relevant to those of interest to the DOD and its military branches, both in wargaming scenarios and in algorithmic design. The potential benefits and relevance to kill-chain functions will be explored heavily during the development of our evaluation criteria.

### *f. Utility*

Utility curves associated with the types of risk preference are depicted in Figure 63.

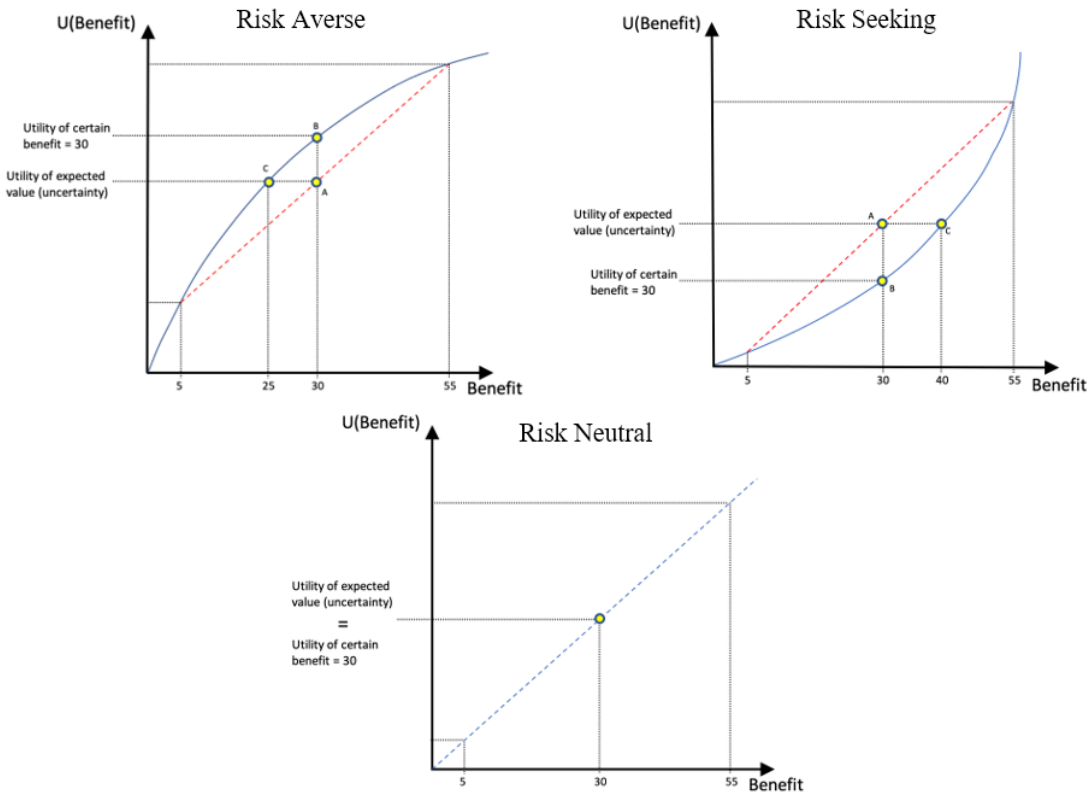



Figure 63. Utility Curves Associated with Types of Risk Preferences. Adapted from Koller (n.d.).

The utility curves above represent single attribute examples of depicting the differences between decision-making under certainty and decision-making under uncertainty. They are based on a set of lotteries {benefit = 5, p = 0.5; benefit = 55, p = 0.5}. In the risk averse graphic, the concave curve represents utility of certainty while the dashed red line represents the expected value of the lottery (benefit = 30). Observe point A, which corresponds to the utility value of the expected value of the lottery. Point B demonstrates this agent's preference for a benefit with certainty over uncertainty - the utility of B is higher than the utility of A. Point C represents the certainty equivalent of point A – the utility of a value equal to 25 with certainty is equivalent in utility to the value of 30 with uncertainty.

Contrastingly, the risk seeking graphic depicts the same three points (A, B, and C) as the risk averse graphic, though the utility curve is now convex. It should be evident that

point A, still corresponding to the utility of the expected value of 30, now has a higher utility value than point B, the point representing a value of 30 with certainty. Additionally, the certainty equivalent of point A, point C, is now a value of 40.

As seen in the risk neutral graphic, there is a single line which represents both the utility curve for certain and uncertain outcomes. The utility of a value of 30 with certainty is equal to the utility of 30 under uncertainty which is equal to the certainty equivalent. An agent of this type does not care about risk and is only concerned with the outcome.

Utility theory is closely intertwined with both decision theory and game theory. As already discussed, AI aims to model rational agents which relates significantly to how agents interact and make decisions to achieve specified objectives.

### *g. Fuzzy Logic*

Figure 64 highlights the basic architecture of a fuzzy logic system which contains four main parts – Fuzzifier, Rule base, Intelligence (more commonly Inference Engine), and the Defuzzifier. The Rule base represents the repository of if-then rules that must be supplied to the system, typically by domain experts and can also be thought of as the encoding of experience-based knowledge.

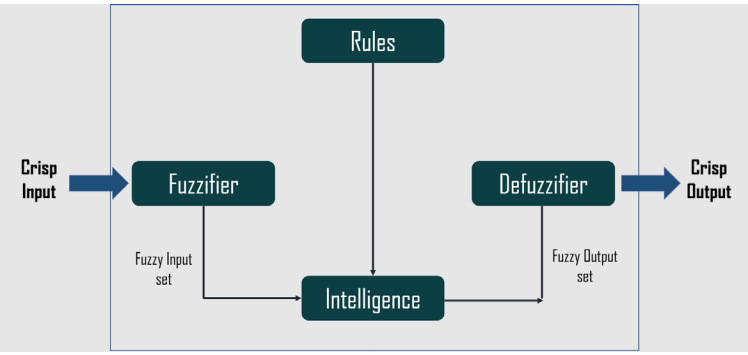



Figure 64. Fuzzy Logic Architecture. Source: Sayantini (2019).

The process begins with crisp input that is fed into the Fuzzifier. The Fuzzifier utilizes one or more membership functions and converts the crisp input into fuzzy sets with varying degrees of membership. This *fuzzy set* is passed to the inference engine where the

set variables are assessed against the rules in the rule base. Another *fuzzy set* is generated by the inference engine that is then passed to the Defuzzifier. The Defuzifier utilizes one or more membership functions and converts the fuzzy variables into crisp output.

The following example is provided to reinforce the concepts discussed thus far. This example was presented by the MATLAB team (2021) and is an oversimplified version solely intended to highlight the overarching concepts of fuzzy inference systems. The example represents a banking system that decides the risk associated with loaning money to a person based on a set of personal and financial information.

Consider the simple rule base below – italicized words represent vague concepts. What is "good credit"? Or "elevated risk"? These are terms that can potentially carry different meanings for different people.


Figure 65. Example Fuzzy Logic Rule Base

Before looking at the fuzzy inference example, imagine we distinguished between "good" and "neutral" credit by the following rule:

***If credit score >= 750, the credit is good, else credit is neutral.***

The associated graph would look something like this:

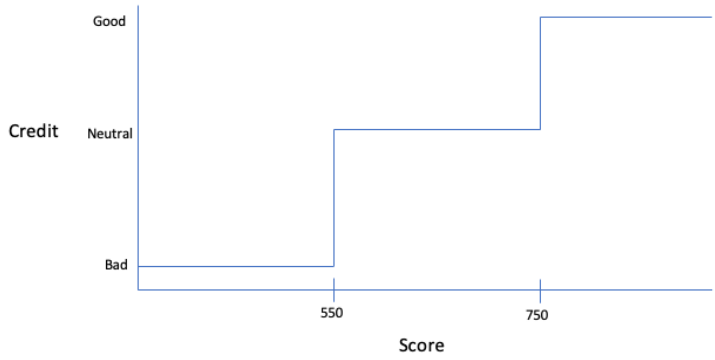



Figure 66. Binary Logical Graph

**A couple of items to note:**

- This binary rule indicates that improving your credit score by 1 point (749 to 750) makes your risk factor jump from neutral to good – drastic considering the 1-point increase.
- It also implies that there is some agreed upon, concrete definition of what "good," "neutral," and "bad" mean – which is not true.

Now, let us look at a potential fuzzy inference system implementation:

To begin, we poll 100 bankers and ask them to each provide their definitions of what is bad, neutral, and good credit. The results are detailed in Table 32:

Table 32. Credit Scores

| **Good credit membership** | |
|---|---|
| % of bankers | Credit Score |
| 100 | 750 |
| 50 | 700 |
| 0 | 650 |

| **Neutral credit membership** | |
|---|---|
| % of bankers | Credit Score |
| 100 | 650 |
| 0 | 750 |
| 0 | 550 |

| **Bad credit membership** | |
|---|---|
| % of bankers | Credit Score |
| 100 | 550 |
| 0 | 650 |

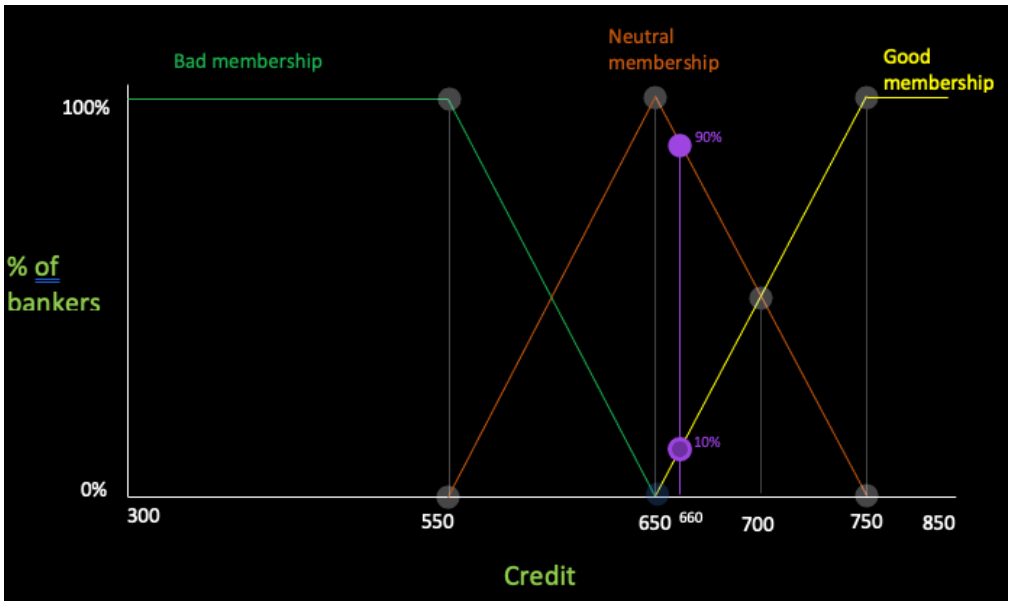



Figure 67. Membership Functions Graph (Fuzzification). Adapted from MATLAB (2021).

From the chart and associated graph (Figure 67), it can now be seen that there exist credit scores that are members of multiple sets with differing degrees of membership (e.g., a score of 700 is 50% good and 50% neutral). This represents the membership functions utilized inside the Fuzzifier. Consider an applicant is applying for a loan and has a credit score of 660. Referring to the chart above, the following fuzzy set is generated: [0, 0.9, 0.1] ([bad, neutral, good]). This fuzzy set is fed into the inference engine and assessed against the rule base in Figure 68 and the resulting new fuzzy set is [0, 0.9, 0.1] ([high, medium, low]). This is fed into the Defuzzifier to generate a crisp output from the remaining fuzzy set. Recall that the Defuzzifier also utilizes membership functions to convert the fuzzy set back to a crisp output, so again, we polled the same 100 bankers on how they would convert high, medium, or low into a crisp percentage value.

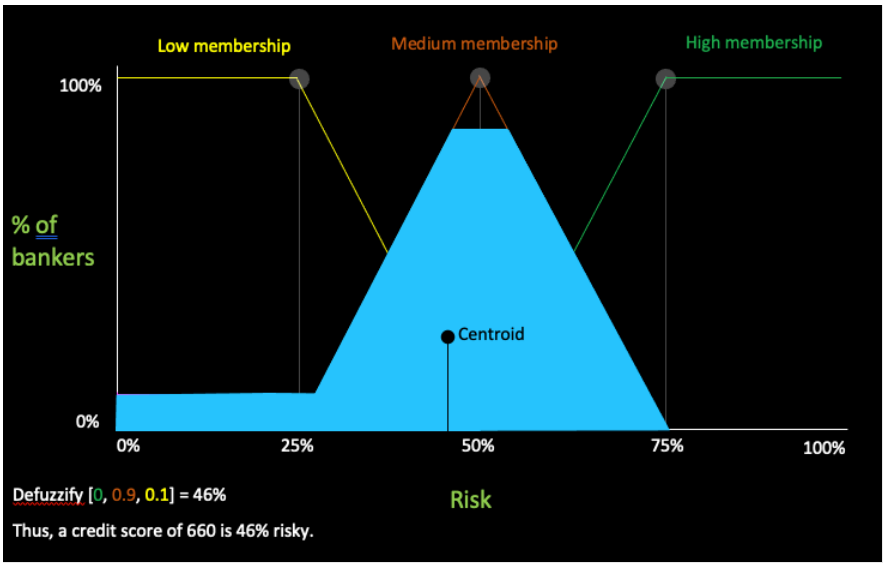



Figure 68. Membership Functions Graph (Defuzzification) with Representation of Crisp Output Value. Adapted from MATLAB (2021).

There are many ways to defuzzify, but for simplicity's sake, we will reference the fuzzy set representing the percentages of high, medium, low-risk membership. We indicated a 10% membership in low risk, so we can draw a horizontal line from the y-axis value of 10% to the opposite border of the low membership line. Do the same for the medium risk value – draw a horizontal line from border to border of the medium membership line at the y-axis value of 90%. We are left with two trapezoids; combine them to create one large geometric shape. The crisp output value will be represented by the x-coordinate of the calculated centroid of the shape. In this example, we determine that a credit score of 660 equates to a 46% risk.

We were able to approximate the risk function with the interpolation of two sets of vague terms. Fuzzy logic is a powerful tool in the decision-making domain of AI. Fuzzy inference tools are fielded in vehicle control applications, expert systems, and various decision support systems to name a few. A prime example within the DOD domain is the 2015 AI project developed at the University of Cincinnati by Nicholas Ernest, ALPHA. ALPHA is an "Artificial Intelligence that controls flights of Unmanned Combat Aerial Vehicles in aerial combat missions within an extreme-fidelity simulation environment" (Ernest and Carroll 2016, 1). ALPHA was lauded as a "breakthrough in the application of

what's called genetic fuzzy systems" (Reilly 2016) after, in October 2015, it defeated one of the Air Force's top tactical experts, Colonel Gene Lee, numerous times. ALPHA was handicapped with less capable munitions and demonstrated the ability to successfully control multiple UAVs in the tactical war fight with Lee who noted his surprise at "how aware and reactive it was" (Reilly 2016, para. 8). A final highlight of the ALPHA AI – UC researchers indicated that "the system was trained and tested on a $500 consumer-grade PC but could run on something as simple as a $35 Raspberry Pi machine" (Dalton 2016). Recent studies have explored utilizing fuzzy systems in combination with machine learning methods. A commonly mentioned limitation of fuzzy systems includes an incapability to learn and the requirement for a-priori knowledge. These limitations can be mitigated by combining with modern ML methods as has been proven by the implementation of cooperative fuzzy neural networks.

#### *h. AI/ML to Human Decision Integration*

A detailed description of the AI/ML to Human Decision Integration framework is included in this subsection.

##### (1) Efficacy Characteristics

For modern decision making to be effective it must be connected, contextual, and continuous. Connected in the sense that decisions are not independent and have wide ranging effects across an organization. "Decision making needs to become much more connected, on all levels — not only hierarchically (strategic > tactical > operational), but also in a networked sense. Sharing of data and insights across organizational boundaries is critical" (Rollings 2021, para. 10). Large amounts of data utilized to support decision making must be given context. Decisions must also be made in a continuous manner without interruption. This requires organizations to remain flexible. "Organizations must be as responsive as possible to opportunities and disruptions" (Rollings 2021, para. 10).

##### (2) Decision Factors

Several decision factors impact the identification of causes and effects in any given event. These factors include:

- quality and quantity of available data and information
- risk level
- available solution options
- human impacts
- measure of success

The quality and quantity of supporting data is paramount for AI based decision aids, and necessary for predictive models using supervised learning methods. The appropriate amount of risk in each decision must be factored and measured. What options are available because of the decision must also be factored. For example, the choice to use lethal or non-lethal weapons. Human impacts must also be considered including those of human operators and those externally impacted by the decision. Measures of success must be identified to determine the appropriate metrics that quantify accuracy and precision.

(3) Degree of AI in Decision Making

> There is a growing awareness that intelligent decision support should not be presented to the human decision maker as a distinct system, or as a working tool. Rather than being separate entities, humans and machines should collaborate. Decision making should be considered as a joint activity of humans and intelligent technology, working together in a collaborative, and coordinated fashion. (van den Bosch and Bronkhorst 2018, 6)

AI methods integrated into organizational decision-making aids must have degrees of control and influence shared between the human operator and the machine. These levels of control and influence are parameterized by level of involvement in the decision-making process:

1. Decision Support: Humans make decisions based on information from machines.
2. Decision Augmentation: Both humans and machines make decisions based on machine recommendations and repeatable tasks.
3. Decision Automation: Machines make autonomous decisions.

Scenarios where AI provides decision support allow for majority human control with machines providing insight and supporting visualizations supported by predictive analytics. Decision augmentation presents a split share of involvement in the decision-making process. The machine will provide recommendations and alternatives based on predictive analytics.

> Augmentation is ideal where actions and work are repeatable, but data can add intelligence. But in general, machines and humans each have a role in effective decision making. Human decision makers certainly shouldn't be replaced everywhere; rather, they should be complemented by the power of data, analytics and AI. (Rollings 2021, para. 9)

Full decision automation is fully reliant on the machine making decision autonomously with minimal human intervention. Decisions are made based on prescriptive and predictive analytics with the benefit of increased speed and consistency at the detriment of increased risk.

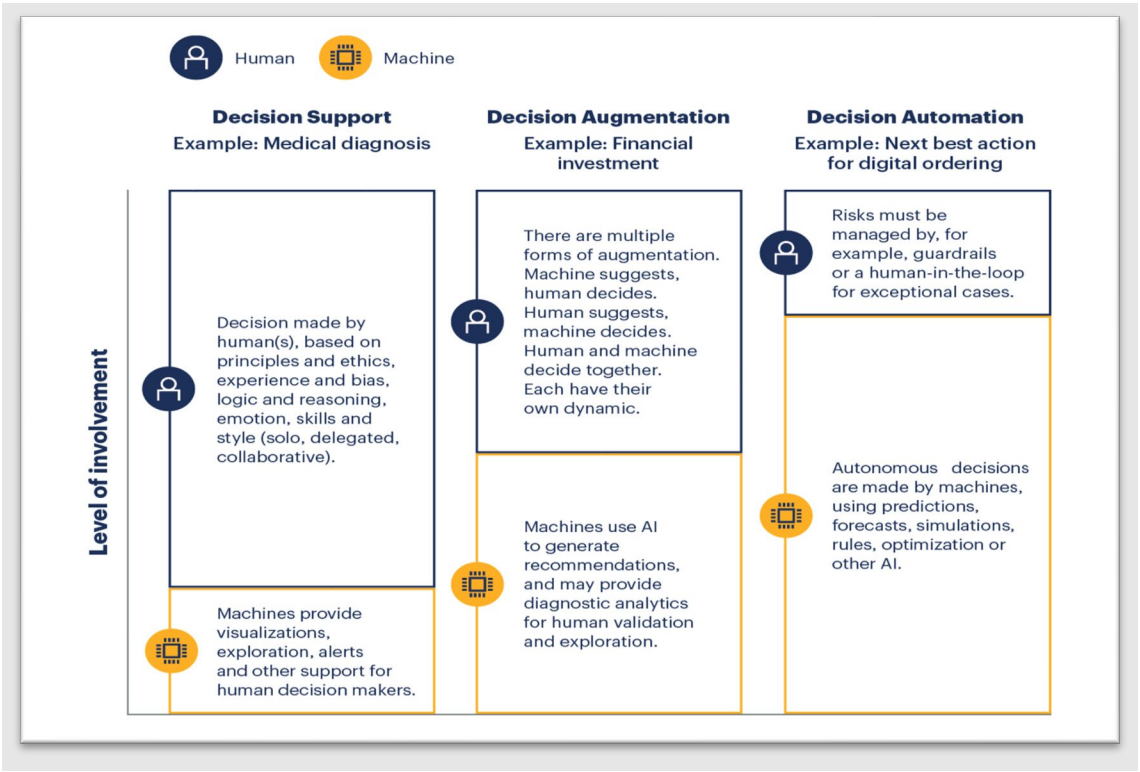



Figure 69. Human-Machine Decision Level Involvement. Source: Starita (2021b, para. 9).

Additionally, the levels of AI integration into the decision-making process can also be expressed in degrees of human-AI interaction and human-AI collaboration (0). These three degrees are one-directional, bi-directional, and collaborative. The one-directional application implies a one-way flow of information from AI to the decision-maker. Black box implementations are representative of this type of human-AI relationship. In the bi-directional interaction implementation communication flows in both directions between the AI and human. In this case the human must have a better understanding of AI and "request explanations on demand."

> The initiative for clarification then lies on the part of the human and requires from the AI the capability to determine the purpose of the human's request, and to select and generate a (set of) explanations that fit the purpose (query-based explanations). A more elaborated functionality is when 'explanations' can be initiated by either party (mixed-initiative). In this stage it is not only the human that can express a need for information, but also the AI that can voluntarily provide information, for example when it detects misunderstandings, possible errors of judgment (e.g., bias), or unjust exclusions of COAs during planning. (van den Bosch and Bronkhorst 2018, 8)

In a fully collaborative application "humans and AI will form a truly collaborative unity in decision making" and "humans and AIs alike strengthen their understanding of each other by harvesting the feedback and information released during their interaction" (van den Bosch and Bronkhorst 2018, 8).

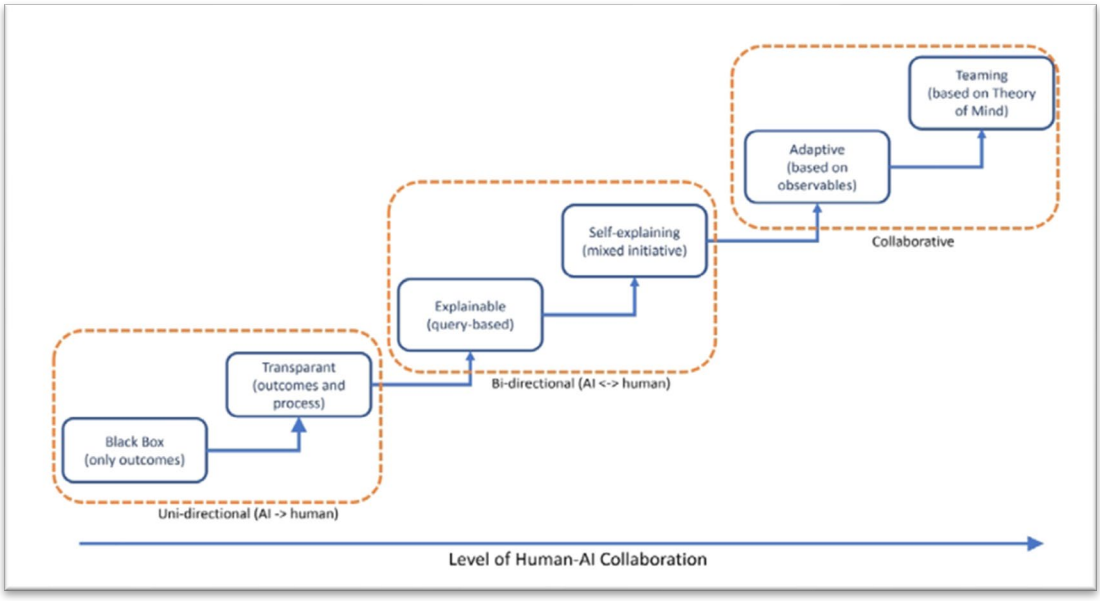



Figure 70. Levels of Human-AI Collaboration. Source: van den Bosch and Bronkhorst (2018, 8).

(4) Human/Machine Decision and Solution Complexity

The complexity of the decision "operates on a continuum" (Starita 2021a, para. 10) and can be bounded by the dimensions of time and complexity using the Cynefin framework. In this case the dimension of time can represent the period from detection to engagement on a scale from seconds to days.

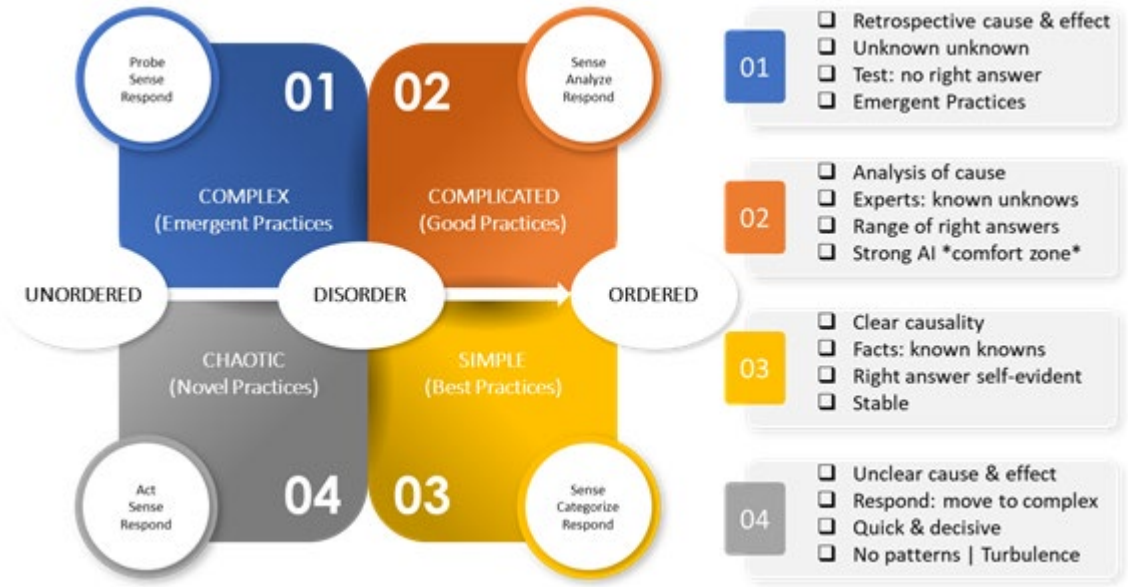



Figure 71. Cynefin Framework for Decision Complexity. Adapted from Spitz (2021); Snowden and Boone (2007).

A complex system has many interacting dynamic systems with nonlinear actions that may affect the system disproportionately. In the moment, there are no apparent cause-and-effect and as the systems evolve the retrospective determined cause-and-effect will be affected; therefore, the future is unpredictable with no immediate "stable" right answer. A decision maker should probe, sense, then respond in a complex context. A complicated environment will have at least one correct answer with the potential for multiple right answers which lends itself well to AI/ML methods. The cause-and-effect will be clear to domain experts. Leaders should sense, analyze, and respond in complicated environments. A simple context has an obvious correct solution to most and the cause-and-effect will be obvious to all. A leader should sense, categorize, then respond in a simple environment but avoid being complacent because of a stable/simple environment. Finally, the last quadrant is the chaotic contexts or the triage environment. As the term chaotic suggests there are no right answers and cause-and-effect relationships cannot be determined. "In the chaotic domain, a leader's immediate job is not to discover patterns but to stanch the bleeding" and to act, sense, then respond (Snowden and Boone 2007, para. 28). The more ordered the environment, the more that AI/ML methods can be utilized and relied upon. To further

explain the characteristics, the decision makers' job, risk, and risk response, we have provided Table 33 below.

Table 33. Decisions in Multiple Contexts: A Decision Maker's Guide. Source: Snowden and Boone (2007, sec. "Decisions in Multiple Contexts").

| | THE CONTEXT'S CHARACTERISTICS | THE LEADER'S JOB | DANGER SIGNALS | RESPONSE TO DANGER SIGNALS |
|---|---|---|---|---|
| SIMPLE | Repeating patterns and consistent events<br>Clear cause-and-effect relationships evident to everyone; right answer exists<br>Known knowns<br>Fact-based management | Sense, categorize, respond<br>Ensure that proper processes are in place<br>Delegate<br>Use best practices<br>Communicate in clear, direct ways<br>Understand that extensive interactive communication may not be necessary | Complacency and comfort<br>Desire to make complex problems simple<br>Entrained thinking<br>No challenge of received wisdom<br>Overreliance on best practice if context shifts | Create communication channels to challenge orthodoxy<br>Stay connected without micromanaging<br>Don't assume things are simple<br>Recognize both the value and the limitations of best practice |
| COMPLICATED | Expert diagnosis required<br>Cause-and-effect relationships discoverable but not immediately apparent to everyone; more than one right answer possible<br>Known unknowns<br>Fact-based management | Sense, analyze, respond<br>Create panels of experts<br>Listen to conflicting advice | Experts overconfident in their own solutions or in the efficacy of past solutions<br>Analysis paralysis<br>Expert panels<br>Viewpoints of nonexperts excluded | Encourage external and internal stakeholders to challenge expert opinions to combat entrained thinking<br>Use experiments and games to force people to think outside the familiar |
| COMPLEX | Flux and unpredictability<br>No right answers; emergent instructive patterns<br>Unknown unknowns<br>Many competing ideas<br>A need for creative and innovative approaches<br>Pattern-based leadership | Probe, sense, respond<br>Create environments and experiments that allow patterns to emerge<br>Increase levels of interaction and communication<br>Use methods that can help generate ideas: Open up discussion (as through large group methods); set barriers; stimulate attractors; encourage dissent and diversity; and manage starting conditions and monitor for emergence | Temptation to fall back into habitual, command-and-control mode<br>Temptation to look for facts rather than allowing patterns to emerge<br>Desire for accelerated resolution of problems or exploitation of opportunities | Be patient and allow time for reflection<br>Use approaches that encourage interaction so patterns can emerge |
| CHAOTIC | High turbulence<br>No clear cause-and-effect relationships, so no point in looking for right answers<br>Unknowables<br>Many decisions to make and no time to think<br>High tension<br>Pattern-based leadership | Act, sense, respond<br>Look for what works instead of seeking right answers<br>Take immediate action to reestablish order (command and control)<br>Provide clear, direct communication | Applying a command-and-control approach longer than needed<br>"Cult of the leader"<br>Missed opportunity for innovation<br>Chaos unabated | Set up mechanisms (such as parallel teams) to take advantage of opportunities afforded by a chaotic environment<br>Encourage advisers to challenge your point of view once the crisis has abated<br>Work to shift the context from chaotic to complex |

Figure 72 illustrates this framework mapped to the degree of AI decision level involvement. “Applying the dimensions of time and complexity together can enable leaders to assess individual decisions and determine the value and feasibility of automating, augmenting or supporting them” (Starita 2021a, para. 11).


Figure 72. Decision Assessment Model. Source: Starita (2021a, para. 11).

“Over time, as technology advances, leaders can expect the bounds of what can be feasibly automated to move further along the axis of complexity” (Starita 2021a, para. 11). According to van den Bosch and Bronkhorst (2018, S3-1–1):

> Thanks to the recent growth in sensor technology and analysis software, the military generally has systems available that provide large streams of information related to a decision situation. However, information becomes

> burdensome if supplied in massive quantities and in an uncontrolled manner. It may increase the decision maker's workload due to the need to process all this information. The resulting 'information clutter' thus endangers situation awareness and the quality of human decision making.

The complexity of the solution to the problem at each decision point, iteratively throughout the kill chain evolution, must be determined and must be characterized at both the human and machine level. The dimensions of human and machine solution complexity drive the bounds for scaling decision solution complexity. At the **human level**, solution complexity can be characterized by the following:

- time: seconds to days
- resource availability: few too many to too many
- communications: disrupted or flowing
- environment: the diversity and complexity of the situation
- relationships, dependencies, and interactions
- quality, quantity, and type of information. Referencing Herbert Simon's concept of "bounded rationality," Colonel Cunningham cites (1997, 323):

> [He] is limited by his unconscious skills, habits, and reflexes; he is limited by his values and conceptions of purpose, which may diverge from the organization goals; he is limited by the extent of his knowledge and information. The individual can be rational in terms of the organization's goals only to the extent that he is able to pursue a particular course of action, he has correct conception of the goal of the action, and he is correctly informed about the conditions surrounding his actions. Within the boundaries laid down by these 14 factors his choices are rational-goal-oriented.

- degrees of uncertainty
- bias
- experience, training, competencies.

- Onboard a naval warship, Sailors possess varying levels of experience and competencies because of the amount of time served, past experiences, system familiarity, and domain knowledge. The Navy places immeasurable amounts of responsibility on personnel of all ages and experience and, as such, oftentimes Sailors are typically expected to possess knowledge of both their primary jobs (rate specific – maintenance, operations) and numerous secondary jobs (watch stations, damage control, general shipboard expertise).
- fatigue.
- As mentioned above, Sailors possess numerous jobs underway on a warship. They are left to balance requirements of day-to-day work such as maintenance and division level tasking with continued training requirements and rotation watch schedules. As can be seen in each of the incidents highlighted in this paper's Introduction, watch stander fatigue was a common contributing factor to all the mishaps.

Solution complexity during combat operations underway are heavily dependent on the successful interactions between watch standers with varying roles and degrees of responsibility. Accountability for decisions can often rest with a single person (e.g., the Commanding Officer), but decisions are rarely made in totality by one person. In fact, onboard a naval combatant, problem solving during any evolution is a distributed decision-making process. The effects of this dynamic are often hard to quantify, but it should be reasonably obvious to envision the above listed characterizations affecting each individual decision maker, during each decision point, throughout the whole decision-making process which adds additional layers of complexity to the problem-solving process.

At the **machine level**, several tenets of computational complexity theory can be used to describe the scale of solution complexity at the machine level.

- time: microseconds to days.
- Can the system be faster? Is the system fast enough to handle the operations requested by the user in extreme decision scenarios?
- quantity, quality, and type of data

> The main ingredient in any ML-application is data from which the machines can learn and, ultimately, provide insight into. Military organizations are often good at collecting data for debriefing or reconstruction purposes. However, there is no guarantee that the same data can be used successfully for ML. (Svenmarck et al. 2018, 2)

- data storage and memory: local limited storage to cloud based unlimited storage. Are these functions local and limited to a tactical node or is distributed processing in a larger network available?
- system and network performance: slow to fast.
- What are the number of processors, number of gates, and computational limits and bounds of the system? How many operations per second can it execute? How much information can be transferred per second? How complex is the algorithm?

A linear relationship between the two dimensions at decision-making levels can be seen in the example heatmap in Figure 73.

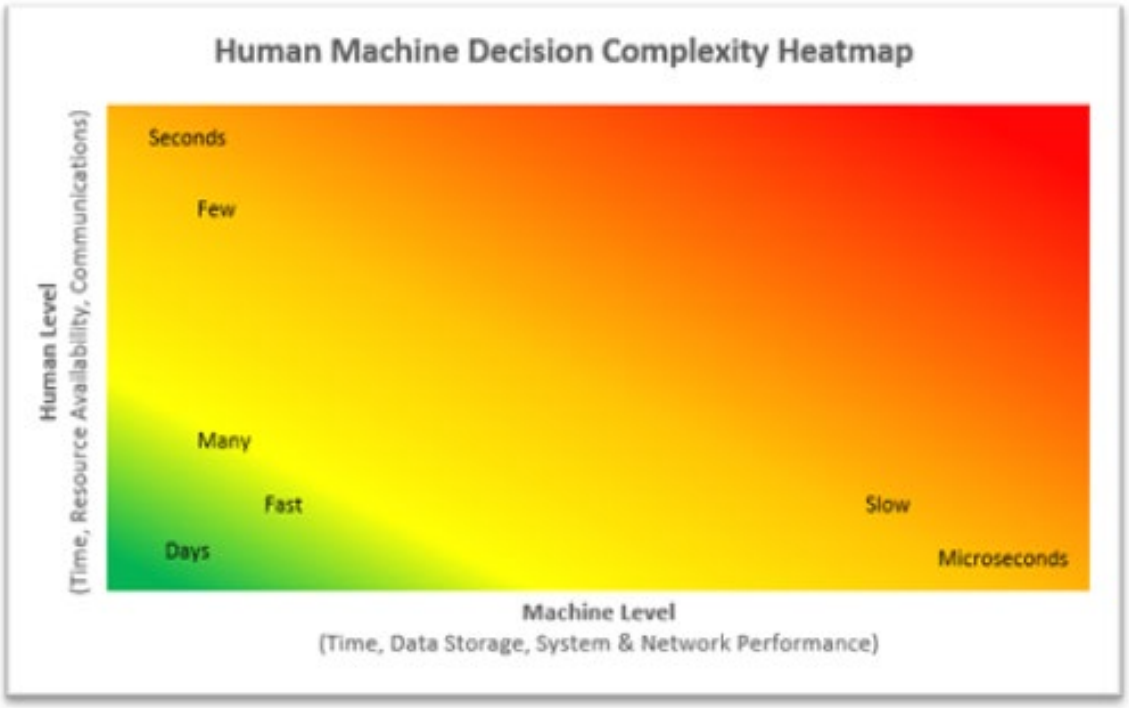



Figure 73. Human Machine Decision Complexity Map

A final aspect of decision-making under stress at sea that demands mention is risk acceptance. As is well illustrated in the AI OODA Team's capstone depiction of a moderate-stress scenario, risk acceptance can be the differentiator between success and failure scenarios. Obviously, this is not a black and white concept and is coupled closely to the discussion earlier in this paper in the Decision Theory and Utility Functions sections.

Understandably, it can seem counterintuitive to believe that an optimal solution in combat scenarios does not necessarily equate to a "perfect" solution. Fuzzy Logic and its notions of partial truths become relevant in these scenarios as well. Modern inbound missiles can achieve time on target (TOT) within minutes and, even in perfect scenarios, may only present one or two realistic opportunities to gain meaningful sensor data during weapon flight. Threat and weapons postures, theatre rules of engagement, and regional instabilities are examples of factors external to system and sensor data that must be considered when generating potential solutions. Thus, in the development of decision support systems and/or autonomous systems, engineers, developers, and military experts must collaborate to establish effective rules for handling these gray areas.

(5) Accuracy and Interpretability

For an AI/ML method to be effective in providing a usable model that is applicable to the problem, it must be accurate and the decision-making process at each step of the kill chain must be made using the most accurate data available. The decision-making processes occurring throughout the kill chain evolution, and the integration of AI/ML methods at each of those steps, can iteratively be characterized and categorized via Table 34.

Table 34. AI/ML Kill Chain Decision Accuracy Drivers

| Problem Complexity | Solution Complexity - Human Level | Solution Compexity - Machine Level | Degree of Support | Data |
|---|---|---|---|---|
| Simple<br>Complicated<br>Complex<br>Chaotic | High<br>Medium<br>Low | High<br>Medium<br>Low | Decision Support / Uni-directional<br>Augmentation / Bi-directional<br>Full Autonomy / Collaborative | Type<br>Size<br>Condition<br>Accuracy |

The goal is to reduce uncertainty and make the most accurate decision possible. Decision accuracy in the kill chain framework is comprised of both human and machine level contributions. The determination of the appropriate AI/ML method to map to each specific phase of the kill chain must account for the attributes outlined in Figure 35 and strive for maximum accuracy. Based on decision factors and constraints, such as the attributes outlined in Table 34, accuracy metrics can vary for differing scenarios. Obviously, high accuracy in decision making will always be desired, but applications of decision aids range from automation of repetitive machine learning tasks to weapon systems acting fully autonomously in their decision-making process. For example, situations that are extraordinarily complex or chaotic, with solutions necessitating a high degree of complexity at the human and machine level, quick fully autonomous responses may be required. In these types of situations involving fully autonomous responses, decision-making and AI/ML model accuracy must be the highest possible. Unfortunately, high degrees of model accuracy can come at the expense of interpretability and explain-ability.

With increasing algorithmic complexity, we are paying the cost of decreasing interpretability and trust. It means that the more complex a model is, the less likely we understand how it works. ...in high-stake environments where even a single mistake can have a dramatic impact or can cost a lot of money, you should always start with interpretable models. (Ahmad 2020, para. 1)

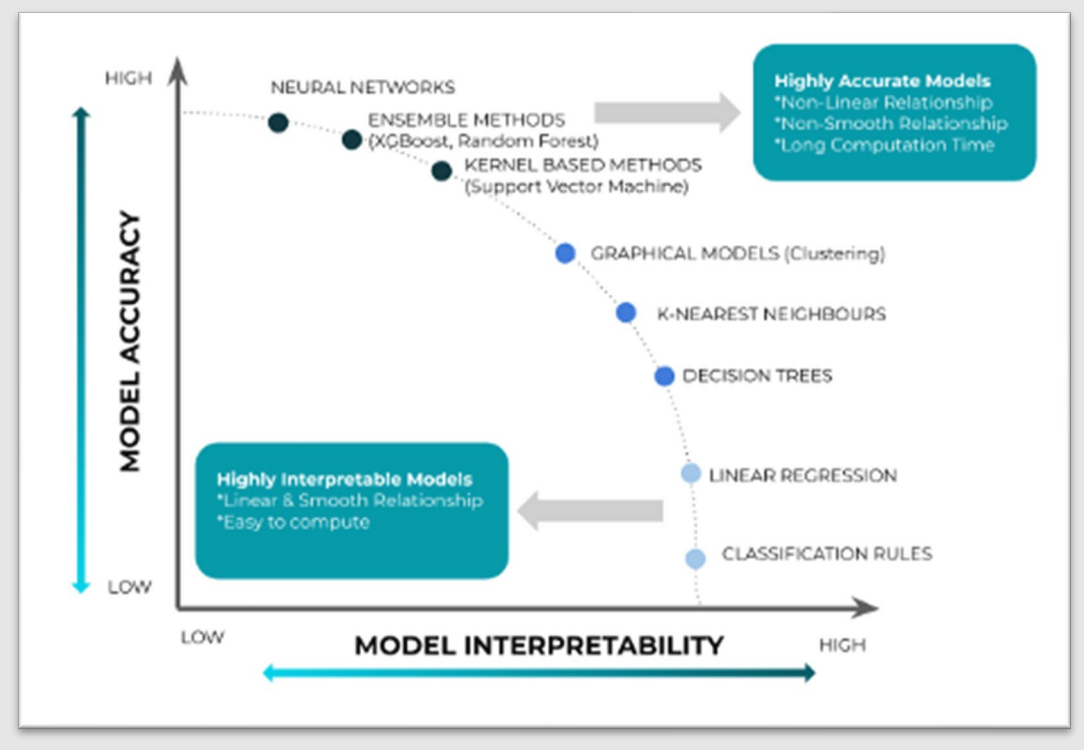



Figure 74. Model Interpretability vs. Accuracy. Source: Ahmad (2020).

Overall accuracy in the decision-making process can be improved by models that build on the principles of transparency and scalability, leading to better interpretability (explain-ability) of the chosen AI model.

> We all are very familiar with the hype which Artificial Intelligence (AI) Techniques (especially deep learning) has created globally-attributed to mostly one goal, get a higher accuracy and beat existing benchmarks. This is very prominent in almost every domain where deep learning techniques continue to be applied for instance, wherein, although the models can achieve high accuracy (in some cases, even garbage data might give you > 90% accuracy!), they suffer from the key problem of transparency, scalability and interpretability. (Chatterjee 2020, para. 1)

In many instances it may be more beneficial to select simpler models that increase interpretability and performance at the expense of minimal decreases in accuracy that ultimately come at the cost of increased computational complexity. "Transparent AI would allow you to judge why (and how) your AI model is making a decision (or not making a decision) for your data" (Chatterjee 2020, para. 4).

The factors characterizing the decision-making processes within the kill chain framework are complex and involve multiple instances of human-machine teaming. These instances provide several optimal points for integration of AI/ML decision aids. High accuracy is desired but not at the cost of interpretability and explain-ability. Simpler models can be just as effective. In addition to factoring in problem and solution complexity at the human and machine level, degrees of AI support, and data attributes, AI/ML methods adapted to kill chain decision scenarios should allow for transparency and scalability.

(6) Engagement Level

The level of engagement for the decision can be broken down into four categories:

1. strategic
2. operational
3. tactical
4. technical

> For example, on the strategic level decisions are made as to if and/or when a military mission is started within a specific operational area. On the operational level a Joint Forces Commander decides what military elements are assigned to a certain operation and specifies the desired effects that will be sought in specific operations. On the tactical level, e.g., a maritime task group Anti-Air Warfare Commander determines what frigate should engage in an incoming threat. Lastly, on the technical level it is decided what weapon is employed at what range to neutralize an adversary. (Kerbusch, Keijser, and Smit 2018, 1–2)

For AI methods to provide real value added they must be embedded in the decision-making process at all decision levels and should only be used when they are of use to the decision maker. "And each decision-making process is different, with different time constraints,

different actors, in different operational environments. This will pose different functional requirements on the solutions, including AI technologies, that are developed to be used" (Kerbusch, Keijser, and Smit 2018, 4).

Figure 75 provides an example of AI decision aid integration into the NATO Joint Targeting Cycle based on five key ideas:

1. AI-based all sources analysis for target system analysis
2. Algorithmic identification of prioritized targets from a target system analysis
3. Automated mapping of capabilities and prioritized targets
4. Computer-assisted robust and adaptive force planning and assignment
5. Automated assessment of military operation performance measures (Kerbusch, Keijser, and Smit 2018, 4)

At each step in the Joint Targeting Cycle example a decision-maker is identified, "and/or product that is strengthened, how AI provides support, and what is the added value of using this form of support" (Kerbusch, Keijser, and Smit 2018, 4). "During the final phase of Joint Targeting, data and information is gathered and analyzed to determine to what extent planned actions are executed (measure of performance) and the intended effects are being reached (measure of effect)" (Kerbusch, Keijser, and Smit 2018, 6).

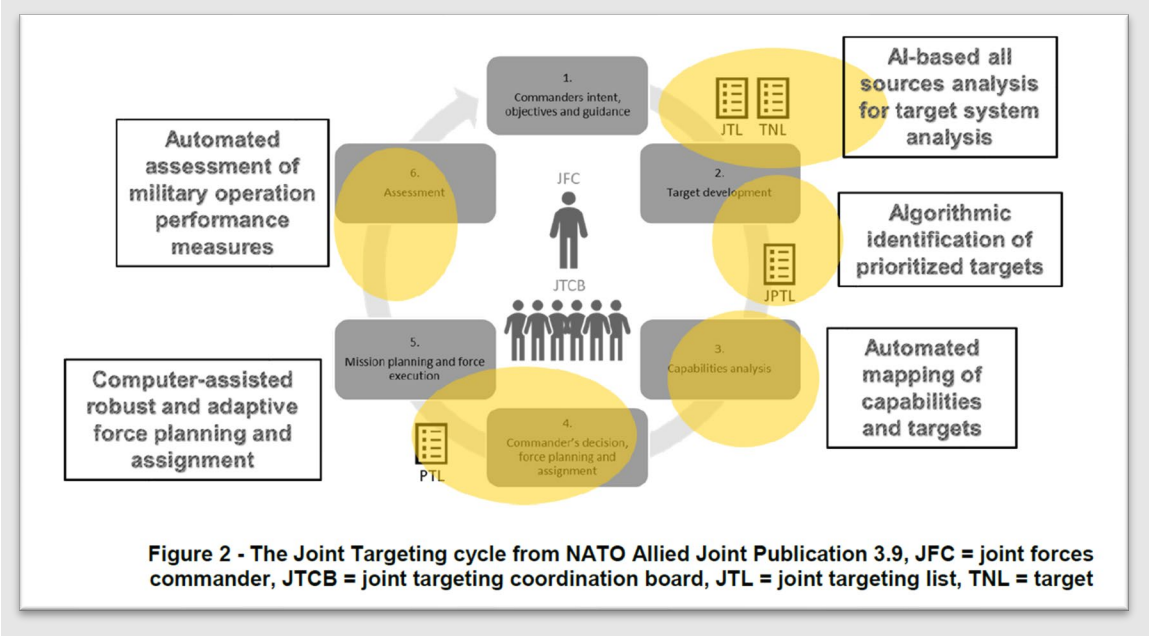



Figure 75. NATO Joint Targeting Cycle with AI Decision Aids. Source: Kerbusch, Keijser, and Smit (2018).

(7) Functional Roles

AI technologies can be related to functional roles in a military decision-making system. These roles can be mapped, for example, to phases of the OODA loop as shown in Figure 76 and are differentiated by two levels: the process level and the individual level.

- process level: the process level encompasses the entirety of the top level of the decision-making process and consists of four roles for AI technology integration and decision support:
- sensing: Sensors that identify and recognize patterns, process enormous amounts of data, provide detection warnings.
- situation understanding: technology employed in this role must make sense of the operational environment and generate suggestions and predictions.
- plan generation: AI technologies in this role create courses of action in response to detection and prediction on the operational environment.

- learning: throughout the decision loop the system must continually learn from the knowledge that it gains.
- individual level: the individual level is representative of the specific step when a decision is made. “AI technologies can be employed in different *collaborative* roles to support the human(s).” (Kerbusch, Keijser, and Smit 2018, 7) These roles are analogous to the three levels identified as the degree of AI in decision making.
- expert system support: analogous to decision support with an emphasis on XAI to support the human decision maker.
- virtual team member: analogous to augmentation.
- autonomous decision making: analogous to automation

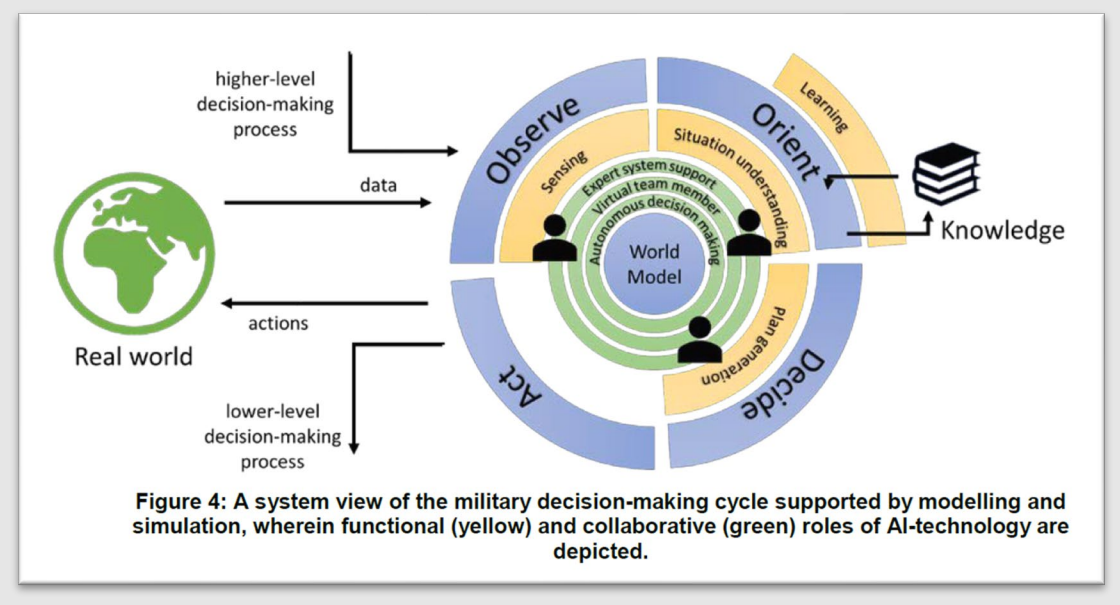



**Figure 4: A system view of the military decision-making cycle supported by modelling and simulation, wherein functional (yellow) and collaborative (green) roles of AI-technology are depicted.**

Figure 76. Functional Roles Mapped to the OODA Loop. Source: Kerbusch, Keijser, and Smit (2018, 8).

> At the process level of for example Joint Targeting, Phase 2 includes both Orient (target system analysis) and Decide (what to target for desired effect). Phase 3 also includes both Orient (own capabilities) and Decide (how to bring the desired effect). Theses phases share the same world model, and it is not unthinkable that the introduction of AI support to this process will lead to a merger of these steps. (Kerbusch, Keijser, and Smit 2018, 7)

(8) Event Descriptors

Event descriptors can be used as tags to not only describe the problem that is occurring at specific instances within the kill chain evolution but also create an event record. These descriptors can be grouped into categories with extensible vocabulary. An example event descriptor framework is the Hierarchical Event Descriptor (HED) tags schema. The HED tag schema was developed at the University of California San Diego for use in brain imaging experiments and data structure projects. The HED tag schema is an example of a community-maintained standard that can be adopted for appropriately tagging and managing enormous amounts of data needed for machine learning models. "HED tags are comma-separated path strings assigned from a tree-structured vocabulary called a **HED schema**" (hedtags.org, n.d.). An event descriptor framework similar to HED can be used to "ensure the continued usability and reuse of data, and to provide compatibility among recorded event descriptions" (hedtags.org n.d.). In addition to ease of maintaining data records, event descriptor tags standardize how events are described, and provide for the consistency in terms needed for building machine learning models and more quickly creating classifiers. A similar schema can be adopted for quickly characterizing and categorizing decision events throughout the kill chain evolution. This schema would serve as the mechanism for translating the real-world events bounded by the kill chain framework via decision-making processes to the machine world for potential application of AI/ML methods and technologies. This event descriptor tag schema is shown in Figure 77- Figure 79. In the schema the top-level attributes are designated as follows:

Table 35. Kill Chain Event Descriptor Schema

| Top-Level Attribute | Description |
|---|---|
| Event | Something that happens at a given time and place |
| Agent | Someone of something that takes an active role or produces a specified effect |
| Action | Primary action in the kill chain framework |
| Environment | Kill chain decision-making environment |
| Data | Attributes of data being exchanged |
| Degree of Support | Degree of AI/ML support |
| Support Relationship | Human-machine relationship type |
| Solution Complexity (Machine) | Machine computation complexity |
| Solution Complexity (Human) | Human solution complexity in given decision-making scenario |
| Problem Complexity | Complexity of problem |
| Object Classification | Target classification |
| Object Attributes | Target attributes |

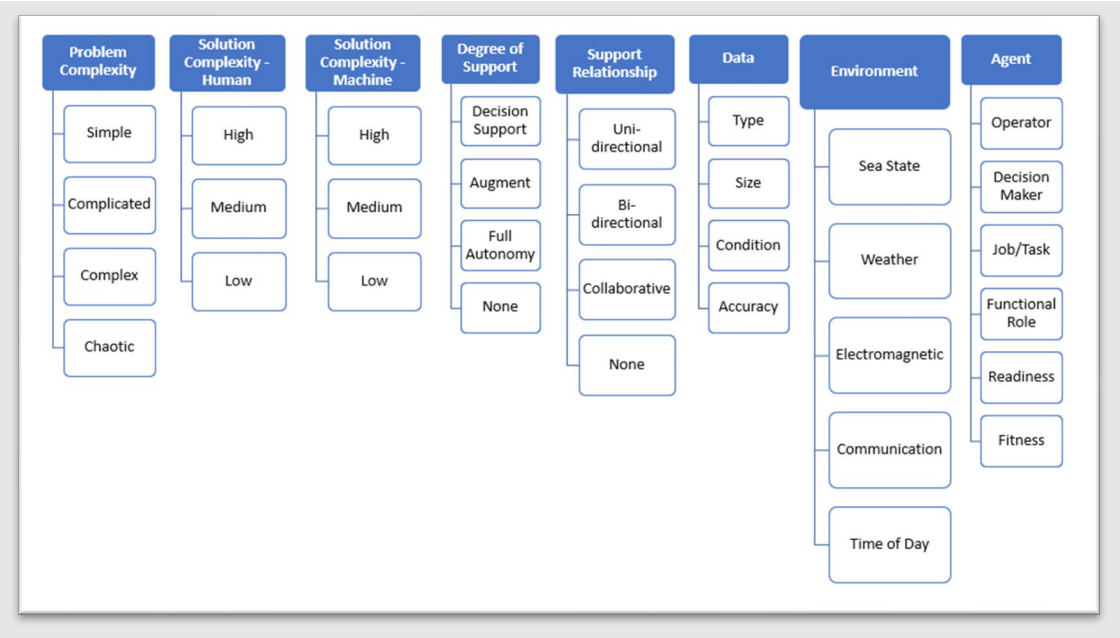



Figure 77. Kill Chain Event Descriptors

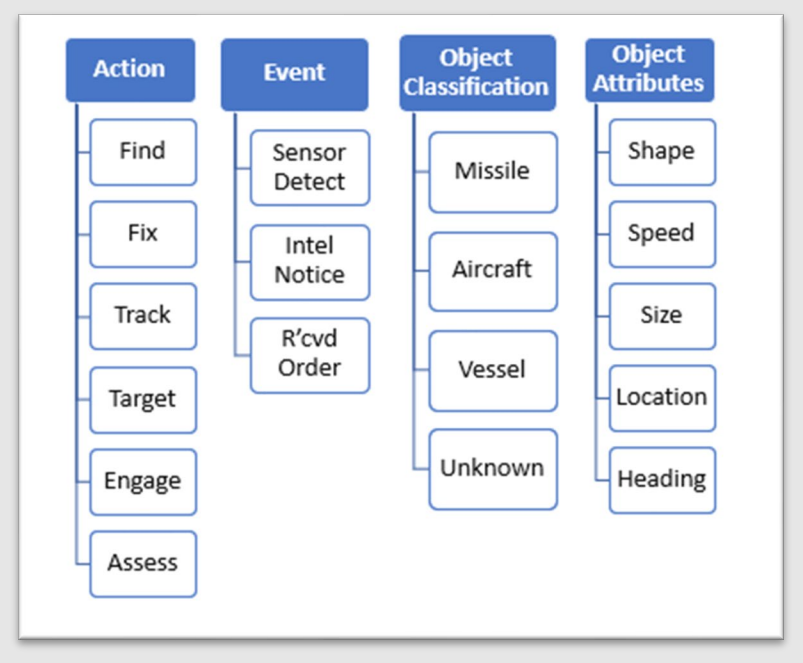



Figure 78. Kill Chain Event Descriptors

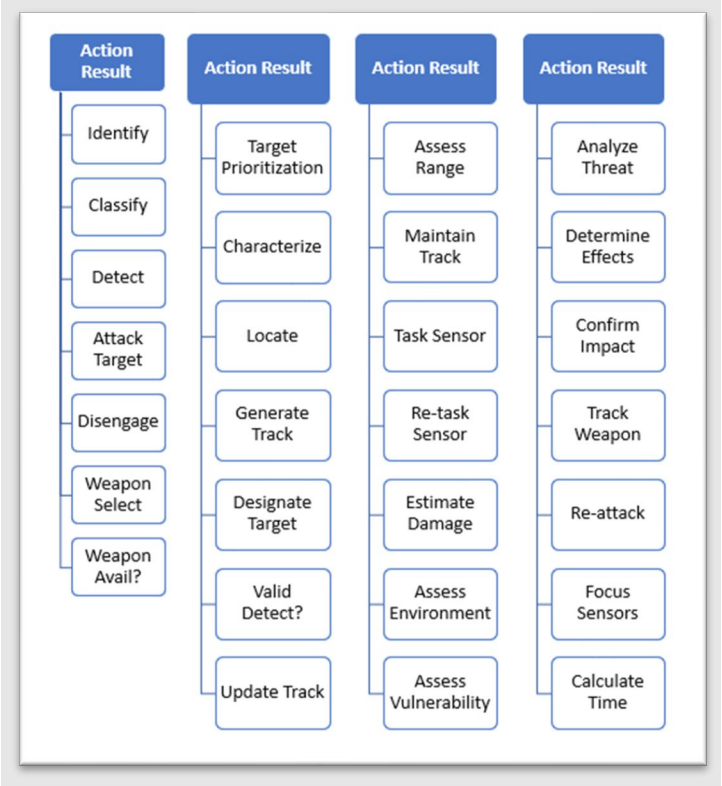



Figure 79. Kill Chain Event Descriptors

(9) Data Characteristics

As we consider the type, size, condition, and accuracy of the data as derived from the inputs and outputs from the kill chain evolution, we must also consider the data interoperability in particular the semantic and syntactic interoperability. “Semantic interoperability is the ability of different computer information systems to use and share data in a meaningful way” (Potgieter 2018, para. 1). As the data moves from input/output from one step to the kill chain to the next step in the kill, the data should not only have consistent formatting, protocols, and standards but a consistent meaning (Potgieter 2018). Syntactic interoperability allows two more disparate systems to exchange data and refers to the “ formats, schema and protocols” of the data and data transfer (Naveed et al., n.d.). As data is utilized from one step of the kill chain to the next it is important that the input and output data is not only in a consistent format but that the data has the same meaning for all users or syntactic and semantic interoperability, respectfully. This will be especially important as data from new sources and/or systems are added. Finally, in addition to these interoperability considerations, we must consider how the data is exchanged between the steps in the kill chain. Therefore, utilizing the data framework as proposed by Naveed et al., we should consider a layered data framework as shown in Figure 80.

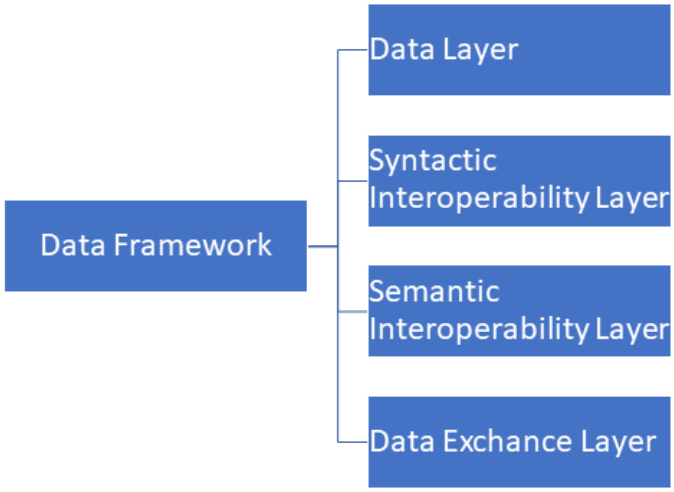



Figure 80. Layered Data Framework. Adapted from Naveed et al. (n.d.)

Borrowing from research from the healthcare and Internet of Things (IoT) community, we propose utilizing a layered data approach for the kill chain data framework as shown in Figure 80. The **Data Layer**, as already introduced, will define items such as the type, size, condition, location, data repository type, and accuracy of the data. The **Syntactic Interoperability Layer** will focus on the formats, representation and proprieties of the kill chain data allowing different systems or models to easily utilize available kill chain data. This layer could further divide the data into meaningful subcategories such as geolocation, timing, etc. Similarly, the **Semantic Interoperability Layer** will define the common protocols, structures and standardization of the kill chain data along with providing a common and consistent meaning to all users no matter where in the chain. The importance of having the similar syntactic and semantic layers is to allow multiple systems and/or models in all staged of the kill chain to accurately communicate without requiring a human-in-the-loop to define the meaning of this critical standardized kill chain data. To further distinguish the difference between syntactic and semantic interoperability the following quotes will provide clarity.

> Semantic interoperability is a mechanism to interpret information whereas syntactic (and structural) interoperability describes data in a uniform way for allowing automatic processing of shared information with ease. The relationship between two is inclusive: a pattern semantically valid will always be syntactically valid, but not the other way around. (Bhartiya, Mehrotra, and Girdhar 2016, 193)

Finally, the Data Exchange Layer will define how the kill chain data will be transferred between the different chain stage.

Additionally, the kill chain data framework shall include something analogous to the Department of Defense Architecture Framework (DODAF) AV-2: Integrated Dictionary to ensure consistency and understanding. The AV-2 is "an architectural data repository with definitions of all terms used throughout the architectural data and presentations" (DLA 2012, 18). The Chairman of the Joint Chiefs of Staff (CJCS) is the lead DOD agency that manages the common vocabulary across federal agencies and our international partners such as NATO (North Atlantic Treaty Organization) "through standardization of military and associated terminology" (Chairman of the Joint Chiefs of

Staff n.d., sec. DOD Terminology Program). Therefore, this integrated dictionary shall be based off the most current DOD Dictionary of Military and Associated Terms and policies as per the CJCS's DOD terminology program and provided guidance.

# LIST OF REFERENCES


Ahmad, Raheed. 2020. "How to Think about Explainability in Your Machine Learning Models?" *TowardsDataScience* (blog). October 9, 2020. https://towardsdatascience.com/how-can-we-build-explainable-ai-f79f4a134406.

Allen, Greg. 2020. "Understanding AI Technology," April, 20. https://www.ai.mil/docs/Understanding%20AI%20Technology.pdf.

Alpaydin, Ethem. 2010. *Introduction to Machine Learning*. 2nd ed. Adaptive Computation and Machine Learning. Cambridge, Mass: MIT Press.

ALSA Center. 2019. *Multi-Service Tactics, Techniques, and Procedures for Air and Missile Defense*. Vol. NTTP 3–01.8. Hampton, VA 23665: Air Land Sea Application (ALSA) Center. https://www.alsa.mil/mttps/iads/.

Barnett, Jackson. 2020. "AI Needs Humans 'on the Loop' Not 'in the Loop' for Nuke Detection, General Says." February 14, 2020. https://www.fedscoop.com/ai-should-have-human-on-the-loop-not-in-the-loop-when-it-comes-to-nuke-detection-general-says/.

Bhartiya, Shalini, Deepti Mehrotra, and Anup Girdhar. 2016. "Issues in Achieving Complete Interoperability While Sharing Electronic Health Records." *Procedia Computer Science* 78: 192–98.

Bosch, Karel van den, and Adelbert Bronkhorst. 2018. "Human-AI Cooperation to Benefit Military Decision Making." STO-MP-IST-160. NATO S&T. https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/Forms/All%20MPs.aspx?RootFolder=%2Fpublications%2FSTO%20Meeting%20Proceedings%2FSTO%2DMP%2DIST%2D160&FolderCTID=0x0120D5200078F9E87043356C409A0D30823AFA16F602008CF184CAB7588E468F5E9FA364E05BA5&View=%7B72ED425F-C31F-451C-A545-41122BBA61A7%7D.

Brose, Christian. 2020. *The Kill Chain: Defending America in the Future of High-Tech Warfare*. First edition. New York: Hachette Books.

Brownlee, Jason. 2014. "Discover Feature Engineering, How to Engineer Features and How to Get Good at It." A Gentle Introduction to Generative Adversarial Networks (GANs). September 26, 2014. https://machinelearningmastery.com/discover-feature-engineering-how-to-engineer-features-and-how-to-get-good-at-it/.

Campitelli, Guillermo, and Fernand Gobet. 2010. "Herbert Simon's Decision-Making Approach: Investigation of Cognitive Processes in Experts." *Review of General Psychology* 14 (4): 354–64. https://doi.org/10.1037/a0021256.

CBS News. n.d. “Past Commercial Airliners Shot down by Military, Rebels.” Accessed June 26, 2021. https://www.cbsnews.com/news/past-commercial-airliners-shot-down-by-military-rebels/.

Chairman of the Joint Chiefs of Staff. n.d. “DOD Terminology Program.” Accessed October 25, 2021. https://www.jcs.mil/Doctrine/DOD-Terminology-Program/.

Chatterjee, Joyjit. 2020. “AI beyond Accuracy: Transparency and Scalability.” *TowardsDataScience* (blog). May 13, 2020. https://towardsdatascience.com/ai-beyond-accuracy-transparency-and-scalability-d44b9f70f7d8.

Clawson, David, Sara Wallace, Gregory Little, and Keith Wheeler. 2015. “Naval Sea Systems Command > Home > Warfare Centers > NSWC Dahlgren > Resources > Leading Edge > I&I Leading Edge > Clawson.” 2015. https://www.navsea.navy.mil/Home/Warfare-Centers/NSWC-Dahlgren/Resources/Leading-Edge/I-I-Leading-Edge/Clawson/.

Cole, William. 2009. “Hawaii-Based Ship’s Grounding Detailed.” *Honolulu Advertiser*, July 7, 2009. http://the.honoluluadvertiser.com/article/2009/Jul/07/ln/hawaii907070350.h tml.

Copeland, B.J. 2020. “Artificial Intelligence - Alan Turing and the Beginning of AI | Britannica.” August 11, 2020. https://www.britannica.com/technology/artificial-intelligence/Alan-Turing-and-the-beginning-of-AI.

Cox. 2018. “Cracking Stuff: How Turing Beat the Enigma.” *Science and Engineering* (blog). November 28, 2018. https://www.mub.eps.manchester.ac.uk/science-engineering/2018/11/28/cracking-stuff-how-turing-beat-the-enigma/.

Dalton, A. 2016. “Combat AI Beats the Air Force’s Top Tactical Experts.” Engadget. June 28, 2016. https://www.engadget.com/2016-06-28-combat-ai-beats-air-force-experts.html.

Das, Somen. 2020. “Artificial Intelligence.” *Different Domains of Artificial Intelligence (AI)* (blog). November 19, 2020. https://somenplus.blogspot.com/2020/11/different-domains-of-artificial.html.

Diaz, Samuel. 2021. “Supervised vs. Unsupervised Machine Learning.” Medium. June 13, 2021. https://medium.datadriveninvestor.com/supervised-vs-unsupervised-machine-learning-732d49413986.

DLA. 2012. “Defense Logistis Agencey Enterprise Architecture,” November, 25. http://www.dla.mil/Portals/104/Documents/J5StrategicPlansPolicy/PublicIssuances/i6604.pdf.

DOD. 1988. “Formal Investigation into the Circumstances Surrounding the Downing of Iran Air Flight 655 on 3 July 1988.” Investigation Report 93-FOI-0184. Washington, DC: Department of Defense. https://www.jag.navy.mil/library/investigations/VINCENNES%20INV.pdf.

———. 2012. “Autonomy in Weapon Systems.” Department of Defense Directive DoDD 3000.09. Washington, DC: Department of Defense. https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf.

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

———. 2020. “DOD Adopts Ethical Principles for Artificial Intelligence.” U.S. Department of Defense. February 24, 2020. https://www.defense.gov/News/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/.

DoDCIO. 2021. “Joint Artificial Intelligence Center.” 2021. https://dodcio.defense.gov/About-DOD-CIO/Organization/jaic/.

Dong, Guozhu, and Huan Liu, eds. 2018. *Feature Engineering for Machine Learning and Data Analytics*. First edition. Chapman & Hall/CRC Data Mining & Knowledge Discovery Series, no. 44. Boca Raton: CRC Press/Taylor & Francis Group.

Donges, Niklas. 2021. “Random Forest Algorithms: A Complete Guide | Built In.” June 22, 2021. https://builtin.com/data-science/random-forest-algorithm.

Ernest, Nicholas, and David Carroll. 2016. “Genetic Fuzzy Based Artificial Intelligence for Unmanned Combat Aerial Vehicle Control in Simulated Air Combat Missions.” *Journal of Defense Management* 06 (01). https://doi.org/10.4172/2167-0374.1000144.

Faik, Lina. 2021. “Deep Q Network: Combining Deep & Reinforcement Learning.” Medium. August 21, 2021. https://towardsdatascience.com/deep-q-network-combining-deep-reinforcement-learning-a5616bcfc207.

Feickert, Andrew, Lawrence Kapp, Jennifer K Elsea, and Laurie A Harris. n.d. “U.S. Ground Forces Robotics and Autonomous Systems (RAS) and Artificial Intelligence (AI): Considerations for Congress,” 47.

Goebel, Randy, Ajay Chander, Katharina Holzinger, Freddy Lecue, Zeynep Akata, Simone Stumpf, Peter Kieseberg, and Andreas Holzinger. 2018. “Explainable AI: The New 42?” In *Machine Learning and Knowledge Extraction*, 295–303. Springer, Cham. https://doi.org/10.1007/978-3-319-99740-7_21.

Gunning, David, and David Aha. 2019. “DARPA’s Explainable Artificial Intelligence (XAI) Program.” *AI Magazine* 40 (2): 44–58.

Hammerer, John. 2021. “Email Correspondence with Advisor, Dr. Bonnie Johnson, 6 Jan 2021.” January 6, 2021.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2017. *The Elements of Statistical Learning Data Mining, Inference, and Prediction*. Second. Springer Series in Statistics. Springer.

hedtags.org. n.d. “Hierarchical Event Descriptor (HED) Tags.” What Is HED? https://www.hedtags.org/.

Hippold, Sarah. 2020. “6 AI Myths Debunked.” *Gartner* Smarter with Gartner (Data&Analytics). https://www.gartner.com/smarterwithgartner/ai-myths-debunked/.

Hölldobler, Steffen, Sibylle Möhle, and Anna Tigunova. 2017. “Lessons Learned from AlphaGo,” in *Proceedings of the Second Young Scientist's International Workshop on Trends in Information Processing*, 10.

Holzinger, Andreas, Peter Kieseberg, Edgar Weippl, and A Min Tjoa. 2018. “Current Advances, Trends and Challenges of Machine Learning and Knowledge Extraction: From Machine Learning to Explainable AI: Second IFIP TC 5, TC 8/ WG 8.4, 8.9, TC 12/WG 12.9 International Cross-Domain Conference, CD-MAKE 2018, Hamburg, Germany, August 27–30, 2018, Proceedings.” In , 1–8. https://doi.org/10.1007/978-3-319-99740-7_1.

Horvitz, Eric, John Breese, and Max Henrion. 1988. “Decision Theory in Expert Systems and Artificial Intelligence.” *International Journal of Approximate Reasoning* 2 (3): 247–302.

IBM Cloud Education. 2020. “What Are Neural Networks? | IBM.” August 17, 2020. https://www.ibm.com/cloud/learn/neural-networks.

Iversen, Mark, and Joseph DiVita. 2019. “Modeling of Cognitive Loading Aboard Naval Surface Combatants.”

Jagielski, Matthew, Alina Oprea, Battista Biggio, Chang Liu, Cristina Nita-Rotaru, and Bo Li. 2018. “Manipulating Machine Learning: Poisoning Attacks and Countermeasures for Regression Learning.” In *2018 IEEE Symposium on Security and Privacy (SP)*, 19–35. IEEE.

———. 2021. “Manipulating Machine Learning: Poisoning Attacks and Countermeasures for Regression Learning.” *ArXiv:1804.00308 [Cs]*, September. http://arxiv.org/abs/1804.00308.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2017. *An Introduction to Statistical Learning*. Springer Texts in Statistics. Springer.

Joint Chiefs of Staff. 2013. *Joint Targeting (JP 3-60)*. https://www.justsecurity.org/wp-content/uploads/2015/06/Joint_Chiefs-Joint_Targeting_20130131.pdf.

———. 2019. *Joint Fire Support (JP 3-09)*. https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_09.pdf.

Jones, Aaron, Christopher Kruger, and Benjamin Johnston. 2020. *The Unsupervised Learning Workshop*. Birmingham B3 2PB, UK: Packt Publishing.

Jones, Julian, Russell Kress, William Newmeyer, and Adam Rahman. 2020. "Leveraging Artificial Intelligence (AI) for Air and Missile Defense (AMD): An Outcome-Oriented Decision Aid." Systems engineering capstone report, Naval Postgraduate School. http://hdl.handle.net/10945/66088.

Kerbusch, Philip, Bas Keijser, and Selmar Smit. 2018. "Roles of AI and Simulation for Military Decision Making." STO-MP-IST-160. NATO. https://www.semanticscholar.org/paper/Roles-of-AI-and-Simulation-for-Military-Decision-Kerbusch/885b182170db541d48ca7f0380bc0447ce56c9ae.

Kherif, Ferath, and Adeliya Latypova. 2020. "Principal Component Analysis." In *Machine Learning: Methods and Applications to Brain Disorders*, 209–25. Elsevier. https://doi.org/10.1016/B978-0-12-815739-8.00012-2.

Koller, Daphne. n.d. "Utility Functions." Video presented at the Probabilistic Graphical Models 1: Representation, Stanford, CA. https://www.coursera.org/lecture/probabilistic-graphical-models/utility-functions-aLRvk.

Konrad, JOhn. 2009. "USS Port Royal Grounding – Incident Updates, Links and Photos!" GCaptain. September 29, 2009. https://gcaptain.com/uss-port-royal-grounding-incident-photo-of-the-week/.

Launchbury, John. 2017. "A DARPA Persepctive on Artificial Intelligence." PDF PPT. https://www.darpa.mil/about-us/darpa-perspective-on-ai.

MATLAB. 2021. *What Is Fuzzy Logic | Fuzzy Logic Part 1*. https://www.youtube.com/watch?v=__0nZuG4sTw.

Mitchell, Harvey B. 2007. *Multi-Sensor Data Fusion: An Introduction*. Springer Science & Business Media.

Mizokami, Kyle. 2020. "Three Years After a Fatal Collision, the USS Fitzgerald Returns to Sea." *Popular Mechanics*. February 5, 2020. https://www.popularmechanics.com/military/navy-ships/a30770027/uss-fitzgerald-return/.

Myerson, Roger B. 1997. *Game Theory: Analysis of Conflict*. Cambridge, MA: Harvard University Press.

Naveed, Arjmand, Fun Hu, Tshiamo Sigwele, Ghulam Mohi-Ud-Din, and Mumtaz Kamala. n.d. “Similarity Analyzer For Semantic Interoperability Of Electronic Health Records Using Artificial Intelligence Techniques: Initial Experiments.”

Nilsson, Nils J. 2010. *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. Cambridge ; New York: Cambridge University Press.

NTSB. 2019. “Collision between U.S. Navy Destroyer John S McCain and Tanker Alnic MC Singapore Strait, 5 Miles Northeast of Horsburgh Lighthouse August 21, 2017.” Marine Accident Report NTSB/MAR-19/01 PB2019-100970. Washington, DC, 20594: National Transportation Safety Board (NTSB). https://www.ntsb.gov/investigations/AccidentReports/Reports/MAR1901.pdf.

———. 2020. “Collision between U.S. Navy Destroyer Fitzgerald and Philippine-Flag Container Ship ACX Crystal Sagami Nada Bay off Izu Peninsula, Honshu Island, Japan July 17, 2017.” Marine Accident Report NTSB/MAR-20/02 PB2020-101007. Washington, DC, 20594: National Transportation Safety Board (NTSB). https://www.ntsb.gov/investigations/AccidentReports/Reports/MAR2002.pdf.

Office of the Federal Register, National Archives and Records Administration. 2018. “Public Law 115 - 232 - John S. McCain National Defense Authorization Act for Fiscal Year 2019.” Government. Govinfo.Gov. U.S. Government Publishing Office. August 13, 2018. https://www.govinfo.gov/app/details/PLAW-115publ232/https%3A%2F%2Fwww.govinfo.gov%2Fapp%2Fdetails%2FPLAW-115publ232.

Oppermann, Artem. 2019. “Artificial Intelligence vs. Machine Learning vs. Deep Learning | Towards Data Science.” October 29, 2019. https://towardsdatascience.com/artificial-intelligence-vs-machine-learning-vs-deep-learning-2210ba8cc4ac.

Parasuraman, Raja, Thomas Sheridan, and Christopher Wickens. 2000. “A Model for Types and Levels of Human Interaction with Automation. IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. 30(3), 286–297.” *IEEE Transactions on Systems, Man, and Cybernetics. Part A, Systems and Humans : A Publication of the IEEE Systems, Man, and Cybernetics Society* 30 (June): 286–97. https://doi.org/10.1109/3468.844354.

Pasley, James. 2020. “Tragic Photos Show How the U.S. Shot Down Iran Air Flight 655 in 1988.” Business Insider. January 9, 2020. https://www.businessinsider.com/iran-air-flight-655-us-navy-shot-down-1988-photos-2020-1.

“(PDF) Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI.” n.d. *ResearchGate*. Accessed October 26, 2021. https://doi.org/10.1016/j.inffus.2019.12.012.

Polit, Kate. 2021. “DOD Releases Roadmap for ‘Responsible AI’ Implementation.” June 7, 2021. https://www.meritalk.com/articles/dod-releases-roadmap-for-responsible-ai-implementation/.

Potgieter, Luke. 2018. “Semantic Interoperability: Are You Training Your AI by Mixing Data Sources That Look the Same but Aren’t?” *KDnuggets* (blog). October 2018. https://www.kdnuggets.com/semantic-interoperability-are-you-training-your-ai-by-mixing-data-sources-that-look-the-same-but-arent.html/.

Reilly, M.B. 2016. “Beyond Video Games: New Artificial Intelligence Beats Tactical Experts in Combat Simulation.” University of Cincinnati. June 27, 2016. https://magazine.uc.edu:8443https://magazine.uc.edu/editors_picks/recent_features/alpha.

*Reuters*. 2017. “USS John S. McCain Suffered Flooding after Collision: U.S. Navy,” August 21, 2017, sec. U.S. News. https://www.reuters.com/article/us-usa-navy-crash-damage-idUSKCN1B10LT.

Roland, Alex, and Philip Shiman. 2002. *Strategic Computing: DARPA and the Quest for Machine Intelligence, 1983–1993*. History of Computing. Cambridge, Mass: MIT Press.

Rollings, Mike. 2021. “How to Make Better Business Decisions.” *Gartner* Smarter with Gartner (Data&Analytics). https://www.gartner.com/smarterwithgartner/how-to-make-better-business-decisions/.

Russell, Stuart J., and Peter Norvig. 2021. *Artificial Intelligence: A Modern Approach*. Fourth edition. Pearson Series in Artificial Intelligence. Hoboken: Pearson.

Sayantini. 2019. “What Is Fuzzy Logic in AI and What Are Its Applications?” *Edureka* (blog). December 10, 2019. https://www.edureka.co/blog/fuzzy-logic-ai/.

Sayler, Kelley. 2019. “Artificial Intelligence and National Security.” https://www.everycrsreport.com/files/20191121_R45178_ddbcce24a6fbf02ad9e81387b5623295ac60f017.pdf.

Simon, Herbert. 1997. *Administrative Behavior.* Fourth edition. Fourth. New York, NY: The Free Press. https://books.google.com/books/about/Administrative_Behavior_4th_Edition.html?id=_obn42iD3mYC.

“Simple Guide to Confusion Matrix Terminology.” 2014. Data School. March 26, 2014. https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/.

Singh, Seema. 2018. “Cousins of Artificial Intelligence.” Medium. June 1, 2018. https://towardsdatascience.com/cousins-of-artificial-intelligence-dda4edc27b55.

Snowden, David J., and Mary E. Boone. 2007. “A Leader’s Framework for Decision Making.” *Harvard Business Review* 85 (11): 68.

Starita, Laura. 2021a. “Using AI in Decision Making: When and Why.” *Gartner* Smarter with Gartner (Data&Analytics). https://www.gartner.com/smarterwithgartner/would-you-let-artificial-intelligence-make-your-pay-decisions.

———. 2021b. “Would You Let Artificial Intelligence Make Your Pay Decisions?” *Gartner* Smarter with Gartner (Data&Analytics). https://www.gartner.com/smarterwithgartner/would-you-let-artificial-intelligence-make-your-pay-decisions/.

Sutton, Richard S., and Andrew G. Barto. 2018. *Reinforcement Learning, Second Edition: An Introduction*. MIT Press.

Svenmarck, Peter, Linus Luotsinen, Mattias Nilsson, and Johan Schubert. 2018. “Possibilities and Challenges for Artificial Intelligence in Military Applications.” STO-MP-IST-160. NATO. https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/Forms/All%20MPs.aspx?RootFolder=%2Fpublications%2FSTO%20Meeting%20Proceedings%2FSTO%2DMP%2DIST%2D160&FolderCTID=0x0120D5200078F9E87043356C409A0D30823AFA16F602008CF184CAB7588E468F5E9FA364E05BA5&View=%7B72ED425F-C31F-451C-A545-41122BBA61A7%7D.

Tarraf, Danielle C, William L Shelton, Edward Parker, Brien Alkire, Diana Carew, Justin Grana, Alexis Levedahl et al. 2019. *The Department of Defense Posture for Artificial Intelligence: Assessment and Recommendations*.

Theodoridis, Sergios. 2015. *Machine Learning: A Bayesian and Optimization Perspective*. Amsterdam Heidelberg: Elsevier.

“Tradewind | An Acquisition Business Model for AI at the DOD.” n.d. Accessed August 29, 2021. https://tradewindfaq.org/.

USNI. 2017. “7 Sailors Missing, CO Injured After Destroyer USS Fitzgerald Collided with Philippine Merchant Ship.” *USNI News* (blog). June 16, 2017. https://news.usni.org/2017/06/16/destroyer-uss-fitzgerald-collides-japanese-merchant-ship.

———. n.d. “USS Port Royal Archives.” *USNI News* (blog). Accessed June 26, 2021. https://news.usni.org/tag/uss-port-royal.

Werner, Ben. 2017. “USS John S. McCain Now in Japan for Repairs Following Deadly August Collision.” USNI News. December 13, 2017. https://news.usni.org/2017/12/13/uss-john-s-mccain-now-japan-repairs-following-deadly-august-collision.

# INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California