Carnegie Mellon University
Software Engineering Institute

ARTIFICIAL INTELLIGENCE (AI) AND MACHINE LEARNING (ML) ACQUISITION AND POLICY IMPLICATIONS

William E. Novak

February 2021

Design: REV-03.18.2016.0 | Template: 02.05.2019

Contents

Acknowledgments

The author would like to thank the sponsor, Mr. Irving "Mac" Townsend Jr., Acting Chief Technology Officer of the Defense Intelligence Agency, for the opportunity to perform this work. The author is also grateful for the help received from the editor, Gerald Miller.

AI and ML Acquisition and Policy Implications Context

This white paper is a high-level survey of a set of both actual and potential acquisition and policy implications of the use of Artificial Intelligence (AI) and Machine Learning (ML) technologies. In this context, implications are known current effects, as well as possible future effects of the use of these technologies across a number of different identified domains where those effects become manifest. Some of these implications are primary effects that occur as a direct result of the application of the technology (e.g., the need to review the ethics used in autonomous decision-making by AI & ML), while others are secondary effects that occur as a result of a primary effect (e.g., the need to access data that will then be used to train supervised ML).

In this context, acquisition implications are those effects which may require changes to the way defense acquisition is conducted, such as the way that AI & ML-based systems are validated by the acquisition PMO. Broader policy implications are those effects that may be related to defense acquisition, but which fall outside of acquisition as it is conducted today, such as those of data understandability. Successfully addressing these implications will require updating both acquisition and other policies to support the way development will need be done to build AI & ML systems. In this white paper, both the implications and ways of effectively addressing and managing them are discussed.

In attempting to characterize the acquisition and policy implications of the application of AI & ML to a government context, instances of both actual and potential issues and consequences arising from such applications were researched and identified. The following criteria were used to identify relevant examples:

1. Implications must be a direct consequence of the application of one or more artificial intelligence and/or machine learning technologies.
2. Implications should ideally be in the context of the U.S. government, preferably the Department of Defense (DoD), or contractors employed by the U.S. government.
3. The science and application of research in this field is moving rapidly. Relevant work should be in the context of recent efforts that have been published within the last three years: 2018-2020.

All of the identified implications were characterized with the following information:

- **Problem**: A description of the nature of the issue created through application of the technologies in a given context
- **Mitigations**: Actions that could be taken to mitigate any adverse effects of the implications

The relevant government acquisition and policy implications of AI & ML are organized into the following categories:

- **Technical:** The technical implications of using the technologies, in terms of the consequences for the phases of the software development lifecycle
- **Data:** The implication of the technologies on the different aspects of data management, such as data quality, classification, metadata, access, metrics, rights, governance, and so on.
- **Acquisition:** The implications of the technologies on the defense acquisition system, in terms of contractual and incentive considerations, DoD acquisition pathways, acquisition reform, and other aspects of defense acquisition policy
- **Adoption and Social/Cultural:** The implications of the technologies on their adoption in terms of social and cultural issues, such as resistance to change, loss of the value of specific staff functions to the organization, and a lack of trust in deployed systems (e.g., due to lack of transparency).
- **Organizational:** The implications of the technologies on the organizations that are both developing and using them, such as on the organizations' staff, their core competencies, and how to transition the technologies to those organizations.
- **Legal and Ethical:** The implications of the technologies on the legal and ethical operation of systems, including the expertise required of lawyers working with the technologies, managing autonomous system risk in terms of liability, criminality, values, and ethics/morality, adherence to the Laws of Armed Conflict, accountability, need for human review of decisions,

The following section contains descriptions of each of these acquisition and policy implications of the application of AI & ML.

Acquisition and Policy Implications of AI and ML

Technical

Underspecification in Machine Learning

Problem: There is a fundamental problem with the process used today to build most machine learning models. The general approach is to train the model on a large number of examples, and then test it on similar examples that it has not yet seen. Passing that test indicates the model is complete. As researchers at Google [D'Amour et al. 2020] have pointed out, this bar is too low to produce robust models, as many different models can all pass the test, but they will differ in small, arbitrary ways, depending on differing choices made in the process. These small differences are usually overlooked if they don't affect the outcome of the test—but they can lead to large variations in real-world performance, and some of those models are incorrect. This problem is called "underspecification," which means that even if a training process were to produce a good model, it could also produce a bad one, because it can't tell the difference—and neither can anyone else.

The Google researchers looked at the impact of underspecification on a number of different ML applications, using the same training processes to produce multiple models, and then running those models through stress tests to show the differences in their performance. Fifty versions of an image recognition model that all did well on the training test went on to show wildly different performance on the stress tests. Some models that did well at recognizing pixelated images performed poorly on images with high contrast—and it may not be possible to train a single model that passes all such tests.

Mitigations: At the time of this writing research is still ongoing to understand how to address these types of issues. One (expensive) approach is to produce many different models instead of just one, and then select the one that performs best on real-world tasks based on stress test results. Another that is being explored is improving the training process.

Data Shift in Machine Learning

Problem: Another problem that has emerged with ML systems is referred to as "data shift" (or alternatively, dataset shift, or data drift) [Quinonera-Candela et al. 2009]. Data shift occurs when changing factors in the environment lead to significant differences in the distributions of the types of data between the training and the real-world data the system encounters. This can happen for many reasons: changes in social behavior, seasonality patterns, unanticipated events (e.g., the COVID-19 pandemic, weather events, etc.), technology breakthroughs, changing political and socioeconomic factors, consumer habits, and even fashion trends, depending on the type of data the system is processing. When the data being processed by the system is no longer highly similar to the data used to train the system, data shift occurs—and ML systems based on that data begin to perform poorly.

Problems with data shift often manifest through a gradual deterioration in the measured performance of the system, as environmental influences affect the data the system is processing. The types of changes to the data that typically occur include changes to the input dataset (covariate shift), or the target variable (prior probability shift), or the underlying relationships between the input and output data (concept drift)—but they all cause the model performance to degrade.

Mitigations: ML-based systems will very likely encounter data shift for some reason during their operational life—and likely multiple times, or even regularly. For this reason, ML-based systems deployed in (and operating on data from) the real world will need to be retrained on a regular basis with expanded and constantly changing training data sets designed to accommodate environmental changes. This will place an even greater load on the underlying data sources, and on ongoing efforts to cleanse and curate that data, to make it usable and reliable in deploying ML systems.

There are practical techniques for detecting each of the types of data shift, so that the change can be monitored, and the system retrained, before the system results are impacted too adversely. While a full discussion of how to implement these techniques isn't feasible to present here, the important implications of data shift are acknowledging and planning for ongoing detection of the occurrence of data shift, acquiring newer data that accommodates the nature of the data shift, and then planning to retrain and upgrade the ML system using that newer data.

Model Training Without Bias

Problem: There have been numerous incidents in recent years in which ML models developed using flawed training data have created significant issues in the resulting system, and for its governing organization. There have been many ML models produced that have shown both race and gender discrimination. Some of these incidents have involved major technology industry firms, including Apple Credit Card (that offered smaller lines of credit to women than to men). Similarly, there was an AI-based talent management tool at Amazon that was also biased against women. These problems have illustrated the difficulty in identifying the bias that has been (often unwittingly) introduced. According to [Cheatham et al. 2019], "misjudgments in model-training data easily can compromise fairness, privacy, security, and compliance." [Kumar et al. 2020] confirmed that "One of the unintended consequences of lax modeling practice is the potential for bias or unfairness in ML models that accentuates our societal stereotypes and contravenes the laws of many jurisdictions."

Mitigations: Based on the research done in this area to date, [Kumar et al. 2020] identifies that the methods used to assess and correct ML model bias fall into three categories:

1. *Fair Exploratory Data Analysis and Pre-processing*: using pre-processing to transform the data and remove the underlying discrimination.
2. *Fair In-processing*: changing learning algorithms to remove discrimination during the training process, using methods such as Adversarial De-biasing, Naïve Bayes Models, Discrimination Aware Ensemble (DAE), and Fairness Regularized Logistic Regression.
3. *Fair Post-processing*: conducting post-processing on a trained ML model by using a "holdout" set of training data that was not involved in the model's training.

Validating Machine Learning Systems

Problem: Even though issues involving autonomy are generally more operational than they are acquisition-related, they will still likely be significant in acquisition, including the validation of machine learning systems that will be deployed and used operationally.

Note that there is an important distinction between the PMO developing confidence in the performance of a system they're delivering (i.e., most often through robust system testing), and the user or the warfighter having confidence in a system they must rely on to defend them in battle (the latter is discussed in the section Lack of Trust in Systems).

Mitigations: The most direct way to address system validation issues is with ever more extensive training sets, but even these leave open the possibility of edge cases being improperly handled because they were not considered as possibilities in the curation of the training sets [Chesterman 2020].

Machine Learning with Sparse Data

Problem: For many problems, present-day AI & ML machines can make more robust and more rapid decisions than humans can. McKinsey identified three broad classes of AI decision making [Dixit et al. 2020]: Operational (real time), Tactical (near real time), and Strategic (long term). AI & ML approaches have already been shown to be effective for operational and tactical domains, but are still challenging for strategic domains.

The issue for enterprises deploying AI in strategic decision-making involves acquiring enough past data to create a model, and then training that model to provide reliable decisions. The problem is that the training and modeling data available is generally sparse, and often confounded with other (often irrelevant) factors, making the decisional model subjective and overly specific to past answers.

In strategic decision-making, because the problems are often unbounded, they are typically unsolvable using formulaic approaches. Strategic decisions are also generally longer-term decisions, and in many such instances must be made based on low fidelity data, or data with low validity.

AI & ML systems are not expected to be capable of ingesting data and autonomously computing outcomes in the near future. As that is the case, end users must still interpret a confusing set of outcome probabilities produced by the system (rather than a single, clear "yes/no" or other binary answer), with a significantly high likelihood of the user being misled by the AI & ML system, rather than helped by it.

Mitigations: In terms of the number of variables and the amount of data needed, the technology is advancing to handle increasingly larger datasets, and do faster computations. This means that the continuing investment in AI & ML for strategic decisions is starting to become economical and appropriate as well, addressing this last and most challenging area.

Progress is also being made in current research in transfer learning[1] that may help to partially address this issue, as the knowledge being gained from solving problems in the operational and tactical domains will become applicable to addressing problems in the strategic domain.

[1] Transfer learning is a research problem in machine learning that studies how to capture knowledge gained by addressing one problem, and then apply that knowledge to another similar or related problem.

Requirements

Problem: System requirements are difficult to define for any software system, which is why software development has moved away from specifying large numbers of system requirements in favor of higher-level expressions of mission capabilities, such as use cases. In the context of developing systems using AI & ML techniques, research conducted on the creation of validated requirements using the JCIDS process [Ehn 2017] indicates that the quality of those requirements is likely to be compromised due to flawed inputs to JCIDS, because the CONOPs does not comprehend the AI & ML technologies or their uses, and staff executing the process are not prepared to articulate those AI essentials[2] that are required to make JCIDS function properly. Poor software requirements historically have led to poor software-intensive system performance, and the same can be expected to occur for AI & ML systems. The result is likely to be schedule delays, poor performance, and cost overruns [Ehn 2017].

Mitigations: Personnel who are planning programs and defining requirements with the JCIDS process need the vocabulary for, and a conceptual understanding of, AI. To attain this, these staff need to be trained in AI through an AI acquisition class or curriculum such as is offered by DAU or NPS, in order to be able to obtain quality requirements for AI & ML systems from JCIDS.

An additional problem is that the use of precise and exact terminology comes into play in expressing system requirements. For one, to determine any legal and ethical restrictions on using AI & ML technology, there must be agreement on the specific meaning of the terms relating to the technology in that context. Again, the AI training described above should address this issue as well [Browne 2019].

Conducting Testing

Problem: The applications of machine learning raise special testing concerns, in part because of the frequently safety-critical nature of its applications, such as self-driving vehicles and medical treatments. Some researchers have pointed out that for AI & ML systems, "…testing cannot be evaluated with confidence…" [Ehn 2017].

It has become clear that good testing is one key to the quality and success of AI & ML systems, as ML models have shown discrimination regarding race and gender in a majority of systems assessed. AI & ML systems have issues stemming from their fundamentally different nature and design, using a data-driven programming paradigm, allowing the model to change over time as the training data expands and changes, introducing unanticipated and potentially inappropriate results, and requiring increasingly comprehensive testing to maintain correctness. In short, a major reason why the testing of an ML system is more challenging than testing traditional software systems is because it is probabilistic, and the system's behavior depends heavily on data and models that can't be specified a priori [Breck et al. 2017], [Ozkaya 2020].

Another reason why testing of ML systems is problematic is that they often must answer questions for which no previous answer exists, testing such results is inherently problematic. Given the types of problems such systems are being asked to solve, determining whether the system can properly handle rarely

[2] The expertise in AI "essentials" that is missing from the thinking of many acquisition staff is how an AI system will accomplish its goals in a given environment, and includes such things as mobility, system perspective, and algorithms.

occurring “edge cases” properly creates a difficult testing challenge. ML systems also exhibit emergent properties that result from the system as a whole, making it difficult to subdivide testing, and forcing testing to be done only at the system level where multiple minimizing and exacerbating effects have all been at play in influencing the results [Zhang et al. 2019].

Mitigations: As a result of such concerns, the President’s 2019 “American AI Initiative” is built around five guiding principles, including the “adoption of standards and the reduction of ‘barriers to safe testing and deployment of Al technologies’ to promote growth of Al industry and their use of AI.” Similarly, the DoD AI strategy calls for the “development of standards for testing and verification of reliable systems.”

A workshop on AI Engineering for Defense [SEI 2019] pointed out that there is a need for “…tools for testing… AI system robustness”, and that the development of AI & ML-based systems will need “…smart and perhaps novel approaches to testing and evaluation (T&E) at both development and operational stages.” The workshop also observed that “monitoring and interpretability tools will support testing and evaluation” and there is a need for “continual verification practices (e.g., to detect when a systems behavior has degraded due to environmental or other circumstances).”

The development of AI & ML-based systems, especially those that aspire to achieve some level of autonomy, will require the adoption of standards and the “mitigation/elimination of barriers to the safe testing and deployment of AI technologies.” Methods will be needed to allow the confident evaluation of test results. Standards will need to be developed for testing and verification of reliable systems [Golden 2020].

In work done by Google [Breck et al. 2017] a significant number of different tests and test strategies are described as part of a “test rubric” that is designed to address the specific needs of, and issues presented by, AI & ML systems. Among other test issues, these techniques address the difficulty of performing traditional “unit testing” prior to system testing in the context of an ML system, because the behaviors of interest are only properties of the system as a whole, and are neither visible nor testable at lower levels.

AI & ML systems that are predictive in nature will need to be tested regularly to verify that the system is still behaving as it did before (i.e., regression testing). Data will have to be explicitly reserved for use in conducting this testing that was not previously used for model training, so correct behavior can be verified. Related to this is the idea that model testing should be conducted to see if the confidence levels of different categories are increasing over time. This should be occurring if the model is reaching greater fidelity with what is believes to be "true."

Some additional techniques that fall into the category of increasing trust in AI & ML systems (see Lack of Trust in Systems) are also relevant here, such as “Explainable AI” (methods and techniques of applying AI technology such that the results or solution can be understood by humans). Without such a capability, it may be difficult to ascertain whether the provided answer (which may be hitherto unknown) is, in fact, correct.

Pace of Development

Problem: Just as with the speed of acquisition, the speed of system development is a factor in being able to deploy AI & ML techniques sufficiently quickly so as to exploit their full advantage.

Mitigations: The use of DevOps as a way to accelerate the time to deployment can leverage that advantage by improving decision-making quality through deployed AI & ML systems [Wydler et al. 2018]. A good way to speed development is by using iterative development with realistic methods of comparison of algorithm effectiveness and accuracy, in order to prove that the newer algorithm is demonstrably better than the old.

Another way to speed the development of AI & ML systems is through the use of managed services that are offered by various vendors, and can provide and deploy an initial system capability in a very short time. For example, one such service at NRO uses a makerspace[3] combined with IBM's Watson Studio4 to allow the integration of data and algorithms in a container that a team can collaborate in.

Control of Technology

Problem: There is concern around the question of how to control AI technology, especially as it begins to learn from its environment and experience independently of human input. This is especially critical in military applications involving life and death issues including such concepts as the "fog of war" and collateral damage, given that there are not yet multilateral or international standards addressing military AI applications. While the technology of AI systems is advancing rapidly, the current state of AI technology is still far from producing any level of consciousness or independent thought such as shown in Hollywood science fiction. However, this is not to say that less advanced systems are not making decisions or taking actions of importance, or that these decisions and actions don't have the potential to have very real impacts on the lives of individuals [Hutchison 2018].

The reality is that examples of early versions of autonomous combat systems both already exist, and are in development today, and raise these issues. Even for systems that are not fully autonomous regarding the deployment of weapons, if the system were to advise that, "There's been a weapon launched—you should you respond with this counter-measure," then even if there's a human in the loop, their available options are limited. If it's not possible for a human being to legitimately second-guess a system's recommendation in the small amount of time available before the decision will be overcome by events, then for all intents and purposes the system is tantamount to being fully autonomous.

An existing example of such an autonomous weapon system is the Aegis cruiser, which was the first platform for the Aegis Combat System (ACS), and was designed as a total weapon system, from detection to kill. The system consists of:

1. An advanced, automatic, multi-function, detect-and-track phased array radar, able to perform search, track and missile guidance functions simultaneously with a handling capacity of well over 100 targets.

[3] A makerspace is a collaborative work space within a public/private facility for making, learning, exploring, and sharing that offers a variety of relevant tools.

[4] IBM Watson Studio (formerly Data Science Experience or DSX) is IBM's software platform for data science, consisting of a workspace that includes multiple collaboration and open-source tools for use in data science.

2. The ship's brain is the Ship Combat System (SCS) that takes inputs from 25 individual detection, control and engagement elements that form an integrated combat system that can respond to either a single or a coordinated multiple attack. The Aegis SCS simultaneously and automatically processes data from these elements, directs pre-arranged tactical doctrine, determines modes of operation, and controls target engagements with the appropriate weapon elements.
3. The Aegis Mk.7 weapon system elements are all capable of standing alone as fully automatic anti-air and anti-surface systems, and the Mk.7 performs the principal and surface defense functions. The control elements of this group provide track maintenance, threat evaluation and weapon assignment and control for all warfare operations.
4. The control center of the Ticonderoga-class ships is the Mk.1 Command and Decision System and Mk.1 Weapons Control System, both specifically tailored for Aegis. These elements provided overall battle system management and coordination.
5. The engagement of air targets is conducted through a separate unit, the Mk.99 Fire Control System, which employs four Mk.80 directors (illuminators) used in the final intercept phase, that could cover each direction and allow for simultaneous multi-mission firing.

The total engagement process is therefore automatic. In fact, the whole Aegis ship can be put on automatic mode and intercept aircraft without human intervention, intended for use if there are too many simultaneous threats for human operators to handle. While ACS isn't an AI system per se, it is still acting fully autonomously.

As another example, the Battle Management Command, Control, and Communication (BMC3) system will, when completed, provide automated space-based battle management with command and control, tasking, mission processing and dissemination to support time-sensitive kill chain closure at campaign scales. BMC3 is putting small cube satellites into space where they're connected to ground-based radars, and the collective "mesh" of these satellites shares information to determine how to respond to a threat. Processed data will be routed across the mesh network through both cross-links and down-links to enable timely dissemination to both the warfighter and other systems in the architecture. Evolving threats and mission needs will be continuously addressed through on-orbit updates to the flight software, and key battle management technologies being explored include trusted autonomy and artificial intelligence. This is another battle system that analyzes complex data, and can autonomously make decisions and recommendations based on those analyses.

Mitigations: At the time of this writing research is ongoing to understand how to address these types of issues. This is exemplified by the work presented in [Dignum 2017], that discusses new research being done on the implications of AI decisions, and on several proposed approaches for integrating considerations of different value systems, as well as the social, ethics and morals, and legal aspects they present into the design of AI systems.

See also the section on Decision Autonomy for more information on managing the autonomy of decisions made by AI.

Terminology for Autonomous Control

Problem: The terminology used to describe the aspects of AI & ML-based systems will become increasingly important, and ultimately critical, to the successful development of these systems. This is due in part to the fact that the technology area is still developing at a rapid rate (and is thus immature), and also to the fact that the proliferation of different types of systems based on the underlying AI & ML technologies means that distinctions between the variations among the technologies being employed will become significant to the outcomes of different applications.

Mitigations: Formal distinctions between terms such as "autonomous" and "semi-autonomous" will need to be made to describe the system's mode of operation and control, the amount of human involvement, and the system's ability to learn [Hutchison 2018]. These distinctions will be especially important for semi- and fully autonomous systems. DoDI 3000.09, *Autonomy in Weapons Systems*, already mentions "human-supervised autonomous weapons systems," a term that is inherently self-contradictory, and left undefined.

The lexicon could be modestly improved just by reclassifying existing autonomous or semi-autonomous systems according to different levels of control. Remotely Piloted Vehicles (RPVs) would be "unmanned systems under Close Control (CC)." Those that check in with humans periodically would be under Periodic Control (PC), and systems that only check in when facing novel situations would be under Situational Control (SC). Even this basic terminology would be more accurate than simply calling them autonomous or semi-autonomous.

Even the term "artificial intelligence" itself is poorly defined. The differences between such basic terms as "complicated" and "complex" add to this, as distinctions are drawn between systems that operate based on complicated computer programming and algorithms, and those that demonstrate "true" artificial intelligence. The former are thought not to exhibit AI, but those that can learn and interact with humans beyond their initial programming are considered AI (e.g., DeepBlue defeating Gary Kasparov was not AI, but IBM's Watson winning Jeopardy was).

Data

Classification by Compilation or Aggregation (CbCA)/The Mosaic Effect

Problem: Preventing combinations of unclassified data in shared data environments (e.g., data lakes) from revealing privileged (e.g., classified) information has become a top priority as organizations race to collect structured data from multiple enterprise systems to use in AI & ML systems to improve decision-making. Like other organizations, the DoD combines all relevant data into shared, unclassified structured data repositories. However, while combining data makes it accessible for analysis, it also creates the significant risk of classified data spillage. Anyone authorized to query the data may be able to obtain results in which combinations of data exceed the sum of their parts and thus reveal privileged information [Novak et al. 2018]. This is a second-order effect that results from creating shared data environments to support ML and other types of data analytics in the first place.

The problem is not limited to classified information, as the inadvertent revelation of PII through *de-anonymization* or *re-identification* (i.e., making anonymous data identifiable again by aggregating it with other data) is becoming a significant issue for the commercial sector, as it brings organizations into conflict with data privacy laws in the United States (e.g., protecting patient privacy under the Health Insurance Portability and Accountability Act, or HIPAA) and Europe (e.g., protecting on-line consumer privacy under the General Data Protection Regulation, or GDPR).

Mitigations: As [Novak et al. 2018] describes, some current approaches being used to address the threat of CbCA spillage in shared data environments include:

1. Ignoring the problem as long as no SCG with CbCA rules exists for the data being aggregated. However, this approach exposes data that should be classified.
2. Making the entire shared data environment system-high can also avoid spillage, as the classification level of the environment can be accredited for the highest level of any aggregated data. However, doing so severely constrains access to the data environment, sharply limiting its value.
3. Users can be responsible for manually complying with all CbCA rules when conducting queries. However, this approach requires each user to know all relevant SCG CbCA rules, implicitly limiting access as the risk of inadvertent spillage discourages all but the most intrepid users.

Unfortunately, none of these options is acceptable, and no government or commercial system yet exists, demonstrating the urgent need for an approach that can responsibly address security issues while still allowing broad and convenient access to data.

Research that has proposed an automated solution for identifying and enforcing CbCA rules is described in [Novak et al. 2018].

Data Management

Problem: Neither public nor private organizations were designed to manage and leverage the amount and variety of data they now possess. Most have only a basic understanding of their data, and often don't even know how many databases they have, which databases contain what, or how the data is being collected [Santelli et al. 2019]. The point is that having sophisticated machine learning platforms and algorithms will have little value without the availability of relevant data.

Data management is key to being able to successfully leverage the data that organizations already have in their enterprise business systems and other information systems. This is a large, but frequently overlooked step in successfully achieving the attractive vision of exploiting the organizational knowledge contained in such information systems through AI & ML techniques. It is frustrating for many organizations to discover just how many challenges they still face in creating a viable infrastructure for sharing and accessing clean, authoritative data so that they can begin to employ these advanced technologies.

Mitigations: Data management as a discipline is large, and covers many different aspects of handling data, including:

- Consumers/Information Exploration
  - Data dashboards
  - Data visualization and reporting
  - Decision support

  - Self-service
  - Data queries
- Integrators/Enrichment
  - Real-time data
  - Business applications
  - Analytics insights
  - Data discovery
  - Platform management
- Accessors/Data Access and Providers
  - Data dictionaries
  - Data standards
  - Data security/authentication
  - Data semantics
  - Data services
  - Data classification
  - Business glossary
  - Mapping data to business processes
- Managers/Data Storage
  - Data lakes and data warehouses
  - Data hub
  - Data mart
  - Data storage and archiving
- Data Acquisition and Integration
  - Data exchange
  - Data integration
  - Data cleaning
  - Data quality control
  - Data assessment
  - Data integrity
  - Data profiling
- Generators/Sources
  - Authoritative data
  - Data creation and sourcing
  - Data access
  - Historical data
  - Standard data formats
  - Unstructured data
  - Metadata management
- Data Governance
  - Data communication strategy
  - Data compliance process
  - Data management policies and standards
  - Data quality metrics
  - Information architecture
  - Operational metrics
  - Organizational structure and compliance

As a guide to how to go about addressing data management in an organization, resources exist in the form of several different models. The principal two are the DAMA Data Management Body of Knowledge (DMBOK) [DAMA 2009] and the CMMI Institute Data Management Maturity Model (DMM) [CMMI 2014]. DMBOK represents data management in terms of eleven knowledge areas depicted as 10 slices of a circle around a central core of data governance. The DMM represents data management in terms of five categories, each of which consists of a number of process areas. There are also other frameworks that include MITRE's Data Management Domain Framework, the Enterprise Data Management Council's Data Management Capability Assessment Model, and various architecture frameworks that incorporate data management to different extents.

Looking more broadly, an effective government policy for open data will be needed to facilitate robust data creation and dissemination [Ahn et al. 2020], [Drezner 2020]. National data management policy can facilitate robust data creation and dissemination. Open data simply refers to digital data that has the technical and legal characteristics for it to be freely used, reused, and redistributed. Having an open data policy that helps with getting the right data in a usable form to the people who want it can: 1) open new possibilities for government, 2) drive economic growth, and 3) help make government more transparent and accountable.

Data Collection

Problem: The need for increasingly large amounts of data to use in the development of ML models for a variety of different purposes will continue to grow as the demand for better and more finely discriminating models accelerates. This same problem will reoccur across many different domains.

Mitigations: In looking at the development of any specific ML system, there is the repeating cycle of "Data-Train-Inference," in which 1) data is collected, 2) training of the ML model is performed using that collected data, and 3) inference is done (essentially deploying the trained model and scoring its performance/accuracy). Instances of poor quality data used in training, of course, will cause poor model performance. As issues with the model's behavior are discovered during inference, that should become new data that will then be used to improve the model's future performance in a continuing virtuous cycle.

In the bigger organization-level picture, the growing use of data in the construction of machine learning and AI systems will drive the criticality of collecting and storing data at every opportunity. All data generated by DoD systems, either those in development or deployment, should be stored, mined, and made available for machine learning [McCormick et al. 2018]. Achieving this will require the creation of shared data environments in domains including enterprise business systems (including those for personnel, logistics, acquisition, and finance), intelligence systems, command and control systems, weapons systems, and many others.

The reality in the DoD is that thousands (and likely tens of thousands) of different systems exist that collect, process, and store important data, but the state of the formalized management of the data handled by these systems is still in its early stages. While the surge in interest in ML has spurred both awareness and investment in better quality data management, progress there is still comparatively slow in terms of addressing the many different aspects of the data that are needed to make it widely accessible and usable.

Data Accessibility

Problem: Accessibility to existing large data sets is limited. Even when large data sets of interest for applying AI & ML techniques exist, there is rarely sufficient access to the data for its use processing technologies such as ML. While the resolution of video and photographic imagery has increased dramatically, so has size, complicating rapid access further, and hobbling plans to apply ML. Another complicating factor is increases in data rates, as the gradual but inexorable change from kilobytes per second to terabytes per second transfers and the accompanying bandwidth requirements growing by nine orders of magnitude will require very different architectures.

Mitigations: The foundation for long-term storage and accessibility of big data must be based on the necessary guidance for its architecture, infrastructure, and applications to enhance the accessibility and use of these data. There are also considerations for public data such as Public Access to Research Results, the Evidence-Based Policy Making Act, Department of Commerce Strategic Plan, the President's Management Agenda, and the White House Executive Order on Artificial Intelligence (AI). There is an increasing need for end-to-end data management practices to improve data accessibility for analytical tools, including enriched metadata that's needed for discovery, long-term data archiving and access, and economical multi-tier storage [Margolis et al. 2019].

Data Interoperability

Problem: DoD's posture in AI is challenged in most areas that were assessed, but most notably in data. This includes the lack of data, but even when data exists, some of the obstacles to its use include lack of traceability, understandability, access, and interoperability of data collected by different systems [Tarraf et al. 2019].

DoD data interoperability issues stem from four areas [Williamson 2008]: 1) the large number of systems, and interfaces among those systems, 2) changing operational needs continually requiring new and modified systems with new interfaces, 3) accommodation of the asynchronous implementation and deployment of systems, and 4) diverse communities with diverse content and domain-specific vocabularies. Together, these factors drive a high degree of complexity into any data model.

Mitigations: Interoperability of DoD data continues to be a challenge for many of the same reasons discussed in these other data topics. Two primary issues include 1) the lack of standard data models for DoD data, and 2) the lack of advanced web services to support the interchange of that data between systems. The network of DoD enterprise systems has grown over decades into the heterogeneous combination of older legacy and modernized systems that exists today. Where data is exchanged between systems, it is often done as a batch operation, exchanging specific files nightly in precise statically-specified formats, with little or no provision for the exchange or interoperability of data on demand. This came about as a result of building what were initially siloed systems, and then making the minimum investment possible in enabling them to exchange critical data. DoD now faces a major effort in replacing and upgrading these existing interfaces with modern, general-purpose, on-demand interfaces appropriate to modern uses of this legacy data.

Regarding a standard data model for the DoD, while it is an appealing ideal (as it creates a uniform standard and avoids the need for transforming data across areas), there are drawbacks as well—so many drawbacks that the idea of a single standard data model can be more accurately thought of as an "antipattern." Monolithic data standards are difficult both to establish and enforce (because of the lack of consensus across many different organizations and domains), there are different functional boundaries (because different systems use the same data in different ways), they do not include important semantic information about the data (e.g., timing, sequencing, and other assumptions), they are hard to maintain (because domain experts are not necessarily knowledge representation experts as well), as they become cumbersome to use with time (due to their growth in scope increasing their complexity). Eventually they become victims of their own scope and complexity, slowing the development of new extensions to a halt.

See also the section Data Collection for more information on the state of DoD data management.

Data Understandability

Problem: DoD is working to address a number of different issues pertaining to data, including impediments to the use of data, such as difficulties with the understandability of the data [Tarraf et al. 2019]. The impact of poor data understandability on the development of AI & ML systems is direct; if it's not possible for others who need to use a given set of data to determine whether or not that data is relevant and applicable, then it won't be used, and the value that data could have provided to an AI & ML system is lost.

Mitigations: The USAF data management follows the acronym VAULT, standing for the key aspects of data management, which specify that the data must be Visible (i.e., it is findable), Accessible (i.e., you can get to it), Understandable (i.e., there is metadata about what it is), Linked (i.e., it can be related to other data items), and Trustworthy (i.e., it is authoritative and secure from unauthorized access).

Three ways of improving data understandability are through:

1. Improving the descriptive metadata that is associated with the data,
2. Creating/extending the data model for the data that describes its structure, and
3. Creating an ontology that describes the relationships among the different types of data

The creation of descriptive metadata is perhaps the most important of these as an essential first step in making data understandable to others who work in different areas, and have less familiarity with that data. The issue is that the development of quality metadata is expensive and time-consuming.

Data Rights

Problem: Since the use of ML is predicated on the existence of large amounts of high-quality data to train the model, and that data will determine the effectiveness of the model that is produced, it is an inevitable consequence that the question of who has rights to that data will become paramount—because the owner of the data will, as a practical matter, own the model. This is a consequence of the fact that, unlike in historical software development where the "code" (i.e., the logic of the program) is separate from the data, the data that is used to train the model now is the "code." This shift has several ramifications, but most notably it raises the importance of the ownership of the data used to train the system, and also has the implication that incorrect or inappropriate data could be used adversely to undermine or subvert a ML system, in the same way that a software virus might.

Mitigations: The question of where the data rights to ML system training data belong is best understood in comparison (and contrast) to where the rights to software code and data have traditionally been held. While different arrangements can be negotiated by the PMO and the contractor, according to the FAR, if the government pays for the development of a software system, then by default it retains the rights to the resulting software "code" that constitutes that system.

See also the discussion of the DoD's Other Transaction Authority (OTA) in the section below on Legal and Ethical implications, as data rights are handled differently in OTAs.

Acquisition

DoD Acquisition Pathways Support

Problem: As the development of AI systems is spiral, with iterative design and deployment, once they're fielded they must be retrained frequently to maintain performance. This makes traditional linear acquisition strategies inappropriate for AI & ML systems.

Mitigations: The Defense Innovation Board (DIB's) Software Acquisition Policy (SWAP) study recommended developing new acquisition pathways for software due to incompatibilities between the development, procurement, and sustainment models for software. These pathways were provided by DoD in the Adaptive Acquisition Framework, which allows the acquisition strategy to determine the most appropriate acquisition approach. The most appropriate pathway for the acquisition of AI & ML systems appears to be the software acquisition pathway that uses a series of iterative spirals to reach deployment. However, this does require the program to field at least an initial capability within a year, and more importantly, the perceived alignment between AI & ML system development and the software acquisition pathway has yet to be tested and confirmed.

Some government defense agencies are already using multi-award IDIQ contracts in which they have very large numbers of vendors bidding on task orders. It is relatively straightforward for even small vendors to get sponsored in to participate, at which point they can propose using newer technologies (such as AI & ML-related techniques). Importantly, these contracts are not limited to using only R&D funding, which may be difficult for some organizations to come by. These task orders are generally shorter-term contracts (12-18 months), and allow the work to be done outside of the FAR rules (which can be onerous for smaller, innovative, high-tech companies to comply with). At the National Geospatial Agency (NGA) they have a contract in which vendors can bid an anticipated number of sprints to complete a software development task, which is equivalent to a level-of-effort contract after multiple rounds of descoping. The intent is to make it simpler for small companies (or even small divisions of larger companies) to propose a capability, and then be brought on contract in a matter of weeks.

See the section on Pace of Defense Acquisition for further information on multiple-award IDIQ contracts. Also see the section on Other Transaction Authority (OTA) for additional information on the use of OTAs to speed the defense acquisition process.

Pace of Defense Acquisition

Problem: Acquisition reform has largely been targeted at speeding the historically slow pace of defense acquisition by removing various barriers to leveraging the rapidly advancing commercial AI & ML technologies through both acquisition and partnership, where private sector investments will increasingly dwarf those made by the government. The high rate of innovation in Al/ML means that advances may be relatively short-lived, and so must be developed and deployed very quickly to maximize their benefit to the nation [Browne 2019].

Mitigations: Indefinite Delivery, Indefinite Quantity (IDIQ) contracts provide for an indefinite quantity of services for a fixed time, and are used when it can't be determined what the precise quantities of supplies or services are that the government will need over the contract period. One contracting approach that can be used to accelerate contracting speed is a new contract type, the multiple-award IDIQ contract, which allows contracts to be awarded to two or more contractors under a single solicitation. These can be used for all types of system development, from defense business systems to IT, to weapon systems, R&D, engineering services, advisory services, and special studies. Using this type of contract means that when the government wants to place an order against the contract, all awardees of the base contract can submit a proposal for that work. While multiple award IDIQ contracts can take time to award, once in place this type of contract allows for the establishment of streamlined ordering procedures for future requirements, which can significantly accelerate the contracting speed for a given piece of work.

See also the section Other Transaction Authority (OTA), that discusses the use of OTAs as a way of speeding the defense acquisition process, and significantly reducing the contracting timeline from RFP to award. Additionally, see the section on Technological Diffusion for more information on the speed of conducting defense acquisition, and the resulting impact on the deployment, diffusion, and acceptance of AI and ML technologies.

Other Transaction Authority (OTA)

Problem: Because commercial companies have a leadership position in AI development, it has been difficult for the government to engage with these technology firms due to the complexity and inflexibility of the FARS, the DFARS, and the acquisition system in general. This can limit the government's ability to innovate with AI & ML in government contexts, as intellectual property (IP) rights under DFARS Part 227 normally must be provided to the government [Browne 2019].

Perhaps more importantly in the context of AI & ML, in addition to IP rights (which are generally for source code that has been funded by the government), the government also gets rights to data (i.e., "Data means recorded information, regardless of form or the media on which it may be recorded. The term includes technical data and computer software. The term does not include information incidental to contract administration, such as financial, administrative, cost or pricing, or management information."). This would mean that any data produced in carrying out a contract, such as the critical training data used for ML (or the model's parameters), would also have unlimited rights, provided that the standard FAR /DFAR clause was included. This can be a make-or-break requirement for many commercial companies, for whom their IP is their primary asset—and in the world of machine learning, data is able to become another form of IP.

Mitigations: Use of the DoD's other transaction authority (OTA, under 10 U.S.C. § 2371b) is one way to obtain relief from some of the FAR's more restrictive procurement rules, allowing closer cooperation with commercial firms, and using looser accounting rules. The requirements involving IP rights and data rights do not apply in an OTA.

Another advantage of using OTAs is that the Competition in Contracting Act (CICA) does not apply to OTAs, so the time between the RFP and contract award can be shortened substantially. OTAs are used to build prototypes and conduct research, but the meaning of the term "prototype" is not defined, and has been interpreted broadly. An OTA that develops a prototype can then optionally award a production contract or a follow-on OTA without competition.

Loosening of Acquisition Restrictions

Problem: The slow pace and inflexibility of defense acquisition becomes a key issue as AI and ML is incorporated into DoD systems, as the concern becomes one of whether the system can keep up with the rapid pace of the advancement of these technologies. This is an issue because any advance in technology resulting in new military capabilities and superiority is likely to be brief due to the constant and accelerating advances of technology.

Mitigations: In an effort to speed the pace of defense acquisition, the DoD's Other Transaction Authority (OTA) vehicle has been broadened and its use has greatly expanded because it is not subject to the full procurement regulations or the FAR, making more streamlined contracts with industry possible. OTAs do not have the same intellectual property (IP)/data rights requirements that force commercial IP to be given to the government [Browne 2019].

See also the section Other Transaction Authority (OTA) that discusses the use of OTAs in greater detail.

Adoption and Social / Cultural

Resistance to Change

Problem: There can be inherent resistance to change among staff as AI & ML technologies are introduced that overlap or subsume their current responsibilities. An employee who might be replaced by AI is scared for their livelihood. See also the section on Loss of Value and Staff Displacement, as an individual's loss of value to the organization can be a significant contributor to their feelings of resistance to change.

Mitigations: This resistance need not be an issue if the technology is augmenting their existing skills, rather than replacing them. Retraining may be required if the technology can be used to relieve them of lower-level tasks, freeing them to work on more significant (and less easily automated) tasks. Also, hybrid teams could be used so that each team could be self-sufficient with the expertise it needs, without requiring training. DIA already has a working group at DIA to figure out disruptive methods to help automation, and how to attract and retain data scientists in the government. There are many strategies that can be employed to help with such organizational and career issues to allow teams to develop innovative ideas and make them happen [Tarraf et al. 2019].

Lack of Trust in Systems

Problem: Distrust can arise in the context of human beings trusting the correctness of operation of autonomous AI & ML systems, due to the opaqueness of the technology, and the inability to understand why it behaves as it does. While autonomous systems can potentially improve mission performance and outcomes, adoption of such systems remains low, in part due to inherent human distrust of these systems. The problem of how to overcome that inherent distrust by users of the system will be essential to successfully adopting such systems on a wide scale.

Mitigations: Approaches exist that can be used to mitigate distrust and improve the trust of humans in the correct operation of AI & ML systems through quantifiable components, and thus improve adoption rates [Adesanya et al. 2019].

One method that can be used to boost user confidence in the reliability and correctness of an AI & ML system is a confidence score, which is a threshold used by the system to set the lowest score that's acceptable for delivering a result. If the score for a candidate result falls below the threshold confidence score, then the system must choose between attempting to deliver a higher confidence result to the user, or delivering no result.

One technique that can be used to increase trust in AI & ML systems is “Explainable AI” (see Conducting Testing), which is the goal of helping to explain the reasoning behind the delivered results, and providing insight into how “close” the system was to producing a different result.

One consideration of AI & ML systems in terms of offering transparency and explainability is the question of whether ML models should be “white box” (i.e., its parameters are publicly accessible) or “black box” (i.e., its parameters are private). This distinction brings with it the tradeoffs of openness and explainability from “white box” models, vs. some degree of developer IP protection for “black box” models. The incentives in place, especially for DoD development, tend to favor the use of increasingly opaque ML models in general (e.g., those using deep learning, complex neural networks, and many parameters) over the use of simpler “white box” models that are often based on simpler statistical techniques, and are inherently more transparent, because they can provide the most advanced capabilities to the warfighter. While simpler models have the advantage of requiring less computing power to run, they are also likely to be less sophisticated and less nuanced than neural network-based deep learning models, while at the same time potentially exposing some of the developing organization’s IP.

There are other more traditional methods that can be used to boost the level of trust and confidence that a user or warfighter has in a system that they must rely upon to help them, or even to protect/defend them in battle. These include good marketing of the system and its abilities, especially when that marketing makes compelling use of demonstrations, testimonials, success stories, and the like. While not high-tech in nature, these are proven and tested methods for helping to overcome human distrust of new systems, methods, or technologies.

Organizational

Atrophy of Skills

Problem: One of the second-order effects of using AI & ML systems in an augmentative capacity (rather than entirely replacing the human expert who historically performs the function) is the atrophy of skills that occurs when the AI & ML system is able to do most, but not all, of the work that the human expert was previously expected to do. This is likely to be much less of an issue in non-real-time systems, but in highly interactive, real-time activities such as driving or piloting (or in the future, skills such as performing surgery) it is an essential consideration. Incidents are already occurring in which the drivers of semi- or fully-autonomous vehicles become bored and distracted as the AI & ML system takes over many of the standard functions of the human driver, leaving them unable to quickly take over for the machine when it encounters an emergency situation which only the human can address, but for which, over time, the human has become unprepared to handle after having little involvement in the bulk of the work. The diagnostic skills of medical professionals are yet another example of areas of expertise become less accessible over time through disuse, as AI & ML systems become more widely used in more disciplines [Cheatham et al. 2019].

Mitigations: This is an area of ongoing research, as semi-autonomous cars have only imperfect solutions to this problem.

In the case of autonomous vehicles, the National Highway Traffic Safety Administration (NHTSA) has defined five levels of vehicle automation:

- Level 0: No Automation.
- Level 1: Function-Specific Automation (e.g., cruise control, automatic braking, lane-keeping, etc.).
- Level 2: Combined Function Automation (i.e., at least two primary control functions work together, such as adaptive cruise control and lane centering, such as during highway driving—but the driver still must be available at all times).
- Level 3: Limited Self-Driving Automation (i.e., driver can hand off full control under certain conditions, but the vehicle must alert the driver immediately when needed).
- Level 4: Full Self-Driving Automation (i.e., the vehicle performs all driving functions and monitors the road at all times).

Adequately addressing this issue may in part be a user interface problem, in that the user interface may need to employ innovative techniques for keeping the driver, or pilot, or user at least somewhat actively engaged in the task to prevent boredom and distraction. Features to alert the driver to the need to reassert control over the vehicle have concerns regarding the amount of time that may be required to shift between tasks, when a real-time response is needed.

Technological Diffusion

Problem: A significant challenge to AI & ML technology adoption is the slowness of diffusion throughout organizations, and the acquisition system can be the cause of significant slowdowns in providing AI & ML technologies to end-users [Hutchison 2018]. The concept of “technological aliasing” characterizes the lack of alignment between those developing new technologies, and those using them, specifically scientists, engineers, policy-makers, and end-users. As the alignment gap widens, scientists and engineers become increasingly detached from the warfighters and their situation, when ironically end-users have the best feedback, and ideas for innovations; while scientists may think they’re meeting end-user needs, but are in fact only meeting those of policy makers.

Mitigations: There are three things needed to address the speed of acquisition: 1) tools or processes to ensure that designers, engineers, and policy makers get truthful and early feedback from end-users, 2) speeding up the innovation cycle from analysis, to concept, to experimentation and wargaming, and back to analysis, and 3) faster pace and greater accuracy in developing doctrine for using AI & ML technologies (through improved communications among the stakeholders).

One example of a process that provides the kind of early feedback required to build successful systems is rapid prototyping development by a small, specialized, and dedicated team to allow users to interact with a version of the system early in the development lifecycle and receive user feedback. This type of activity can be key to identifying potentially significant problems early, while it’s still relatively inexpensive to change the requirements and/or design of the system.

See also the section on Pace of Defense Acquisition for more information on the speed of conducting defense acquisition, and the resulting impact on the deployment of AI and ML technologies.

Loss of Value and Staff Displacement

Problem: Due to the increasing level of capability of systems using AI and ML technologies, there are natural concerns regarding the potential loss of some individuals' value to their employer or organization as a result of the adoption of such AI capabilities [Tarraf et al. 2019]. While there may be limited opportunities to move the individual to different positions in the organization where their skills continue to be required, this trend is likely to continue and expand. The end result of the advanced capabilities of AI & ML techniques duplicating or exceeding those of human beings performing similar tasks will likely be a considerable displacement of jobs of human agents by the new technology. This has been seen before in prior industrial revolutions, moving from agriculture to the industry, and moving from industry to information. [Ahn et al. 2020]

Mitigations: There are steps the government can take to ease the transition of the workforce to AI & ML, including:

- Providing training and education in AI & ML languages and technologies, as well as in their use, to move some of this workforce into the AI & ML industry.
- Creating easy-to-use machine learning language platforms to encourage and facilitate the familiarity with, and use of AI & ML by the broader public
- Granting legal rights similar to those of corporations to AI & ML systems that make them legally accountable, and allow them to be taxed and even sued for their actions. The revenues from such a tax could be used to help offset some of the workforce disruption that the introduction of the technology causes.
- Employ the displaced workforce in national projects that build the infrastructure for a future AI & ML-based society. The creation of that foundation, such as a digital model of a city, county, or state that AI & ML agents can reason about to help make decisions for its management, would be a massive effort by that would require the services of many people.
- In cases where the organization developing and deploying the system doesn't have the ability to grant legal rights or enact legislation, it is often still possible to enact policy to do similar things on a more limited local scale, while still promoting the creation of broader legislation to address the issue.
- It should be noted that when considering the implications of the use of AI & ML, there is an important difference between the situations where the technology is being used in operational or weapons systems that are being acquired, vs. its use is in government systems that are being used to conduct acquisition. PMO staff acquiring systems using AI & ML to be employed by warfighters aren't a threat to them personally (just to the adversary on the battlefield), whereas building AI & ML systems to perform acquisition functions that were previously performed by humans in the PMO (such as contracting, or budgets, or technical review) poses a direct threat to them and their livelihood.

AI and ML-Capable Staff

Problem: In terms of appropriate staff expertise for dealing with the acquisition of AI & ML systems, DoD lacks clear mechanisms for growing, tracking, and cultivating AI talent, even as it faces a very tight AI job market [Tarraf et al. 2019].

In the SEI's own experience in helping program offices develop ML systems, the lack of awareness of the nature of AI & ML system development has been very problematic, and needs to be made a priority. DoD PMOs need to be made into "educated consumers" of the technologies. As things stand, there is little ability or knowledge at many PMOs to think about what the AI solutions will need to be successful. There is unfamiliarity with what the appropriate costs for AI & ML developments should be, much less the testing, development, or sustainment, or the terminology or the engineering principles, allowing contractors to take advantage of the PMO's haste to field AI & ML capabilities. The program offices feel that they are being asked to do something that they're not adequately prepared or trained to know how to do. Furthermore, military technical staff at labs and development centers tends to be more inwardly focused, and as a result their notions of modelling and analytics are older and often dated. This lack of up-to-date training is promoting the use of more traditional analytics, creating resistance to the use of modern technologies like AI & ML because they don't have people trained in the newer methods. The result is that DoD PMO staff is trying to advocate the use of traditional analytics in new and convoluted ways, instead of using more modern and simpler approaches. Also, it can be difficult for program office staff to be able to identify places where AI technology can be inserted into what they're doing.

Mitigations: As government organizations start to transform to leverage and exploit AI & ML technologies, they will need to focus on techniques and incentives for attracting, developing, and exploiting a capable Al work force. These will include at a minimum focused partnering, training for existing employees, and active recruitment [Golden 2020]. Some specific resources and methods for addressing this area through professional development of existing staff can include attending local or on-line universities to obtain additional certifications, relevant courses, and specialized degrees, as well as the use of on-line training mechanisms such as Coursera, and accessing current AI and ML information topics on websites such as Medium.com. On-the-job training can also be used where AI & ML projects are identified to develop tools that support various acquisition and software development oversight activities, and having staff develop skills by incorporating basic AI & ML capabilities into support tools (e.g., using publicly available AI & ML tools such as Apache to perform analyses on program metrics, cost estimation, and performance data, that produce useful information while simultaneously developing staff AI & ML skills).

Legal and Ethical

This section may be viewed by some as being less relevant or less applicable to engineering organizations, and being considerations that should be relegated to lawyers in legal departments. However, while the goals themselves may be concerns that will fall under the auspices of the legal department, the means to achieve them will, at least in part, be highly technical—and thus are very much relevant to the developers of such systems.

Also, it should be noted that there is a distinction between verifying legal and ethical systems. Legal issues are codified in statutes and regulations—but ethical issues and considerations generally aren't expressed explicitly in any form. This raises important questions about how and where such questions get reviewed and resolved, especially in a government acquisition context of an AI & ML system. Furthermore, are there cases where it is more ethical to break a law, than to not break it? There are many well-known ethical dilemmas[5] that force choices between option A that kills one person, vs. option B that kills 10 people, because someone will have to die in either case.

Transparency and Fairness of ML Models

Problem: It is certainly important (as discussed in the section on Model Training without Bias) to be able to create AI & ML-based systems that can perform their decision-making tasks of identification, adjudication, and similar ilk. At the same time, however, it is equally important for government entities and corporations to be able to prove that their decision-making and predictive algorithms, whose outputs can affect American citizens, are fair, safe, and non-discriminatory [Dent 2019]. Currently, consumers and users of AI & ML technology inherently know much less about it than the developers and vendors do, and have no way to evaluate the logic behind its decisions, or see what alternatives might have been considered.

Mitigations: While this capability is a highly desirable quality of the technology and methods that would be used to ensure a lack of bias in the system in the first place, it may be that independent validation or testing methods and/or tools will be needed to verify that the lack of bias in the system has, in fact, been achieved to a sufficient degree.

One way to approach this problem might be to have policies or even regulation that would require the vendors of such systems to provide understandable sets of fundamental rules that lay behind and guide the decision-making process of AI & ML systems. This is likely difficult to accomplish, as the complexity of these systems can (and in most cases will) be great, and the distillation of all of the considerations being made into a simple, comprehensible set of logical rules may seem unachievable. Pushback and unwillingness to develop and incorporate such mechanisms should be expected from the companies and engineers developing AI & ML-based systems, but given the importance of many of the decisions that will be made, and the magnitude of their real-world legal, ethical, and financial consequences, that is no reason to think that these considerations and their impacts on system requirements can be ignored.

Legal/Ethical Review and Machine Decision Accountability

Problem: Because of the many potential practical applications of AI & ML technology by government, and the associated decisions and actions with legal, financial, and ethical consequences that will result from its use, issues of accountability for those consequences will be raised, and will need to be resolved.

Mitigations: When there are adverse consequences of the use of AI & ML technology, one of the first issues will be coming up with an accurate understanding of what the system did, and why it did that. As a

[5] The trolley problem is a set of thought experiments in ethics centered around the question of whether to sacrifice one person in order to save a larger number. Such ethical questions have already become real in the context of decision-making done by autonomous cars in the event of minimizing injuries and deaths in the event of an unavoidable collision.

first step, monitoring and auditing the system as it operates will provide a trail that should help to understand why the system produced the outcomes that it did. In addition, there will need to be ways of fully and accurately reproducing the exact system that was responsible for the decision or action in question, so that the issue can be studied and reproduced in an identical version of the system. This ability to reproduce the system is called "provenance," and involves having records of every aspect of the model's production, including the ML algorithm used, the data set, the model parameters that were chosen, and so on. One of the techniques that can be used to capture provenance is the "model pack[6]," which is a way to package and distribute ML models.

A legal mechanism will also likely be needed through which AI & ML and their creators are held accountable for the decisions and actions of the systems they create. This could potentially take the form of granting legal personhood to AI systems, analogous to the way that corporations are viewed as having legal personhood [Ahn et al. 2020].

To determine the legal and ethical restrictions on the use of AI and ML technologies, clear legal definitions will be needed for a comprehensive legal review of a new system as required by DoD Directive 5000.01 to ensure its compliance with the laws of war. Lawyers may need training to develop expertise to ensure that the review is adequate [Browne 2019].

Managing the Risk of Autonomous Systems

Problem: Concerns about the management of risk in autonomous systems fall into three categories: 1) autonomous vehicles (liability and criminality), 2) autonomous weapon systems (morality), and 3) autonomous decision-making systems regarding access to resources or benefits (legitimacy). Each category poses different challenges to achieving the legal and ethical operation of such systems [Chesterman 2020].

Most organizations, for example, would have at least some exposure to risk if they were to use AI & ML technology for making hiring decisions, and the model later turned out to be unfair or discriminatory, as this would potentially violate laws surrounding U.S. government hiring practices [Dent 2019].

Mitigations: In the first category of autonomous vehicle systems, it is important for government agencies to recognize the liability risk of developing systems for which no liability framework exists, and thus the exposure that the organization is incurring by developing such systems. The European Union (EU) is working to develop a common liability framework for AI systems, and is recommending a regulation be created to place strict liability on the "deployer" of certain "high risk" AI systems, and increased deployer liability for other types of AI systems.

Deployers of "high-risk" AI systems would be strictly liable for any harm caused by that system, with a "high-risk" AI system being defined as one where it is sufficiently likely that it will cause personal injury in a random and unpredictable way, as determined by the probability of occurrence, the severity of the expected harm, and the manner of use. The proposed regulation exhaustively enumerates all high-risk AI systems. Liability compensation would be capped, and deployers would be required to carry liability

[6] Model packs allows complete models and their dependencies to be packaged into lightweight services, which can then be downloaded and deployed by users.

insurance for such systems. Deployers of "other AI systems" (that are not "high-risk") will use fault-based liability, but the fault will be presumed to be that of the deployer unless proof can be provided to the contrary.

In the second category on autonomous weapons systems, it should be possible to operationalize armed conflict with Lethal Autonomous Weapons Systems (LAWS) in compliance with the Laws of Armed Conflict by using a series of control mechanisms and tactics, techniques, and procedures (TTPs) that are already used by the U.S. military, and which can be adapted for use in LAWS [Combe 2020].

The third category regarding adjudication of access to resources or benefits, the system must be designed to make decisions that are not only consistent and fair, but at the same time also comply fully with all applicable federal (or state) law. If an automated system were used to make hiring decisions (or award benefits), and in doing so made decisions that were against federal hiring or equal opportunity laws, then there would be potential liability issues not only for the organization using the automated system, but potentially for the organization that developed the system. This is a potential risk that will have to be taken into account by organizations using such systems, especially in cases where a ML system is created by a contractor or vendor, but using training data provided by the organization using the system, as it is not immediately clear in that case who should bear responsibility. However, since the employees of organizations making such decisions already incur that responsibility today, it is likely an acceptable risk if the performance of the automated system is sufficiently on a par with human judgments.

Decision Autonomy

Problem: A central question is how far AI & ML systems will be allowed to go in terms of making decisions that affect the lives of human beings [Ahn et al. 2020]. Many initial attempts at predictive algorithms have been found wanting in terms of the accuracy of their performance once they have been deployed in the field (e.g., bias against minorities in the COMPAS system predicting criminal recidivism). The original question seems improbable given the maturity of today's AI technologies, but as the sophistication of the technology, and the access to higher-quality data improves, this is already changing. AI systems will soon be capable of making complex and important policy decisions with real consequences. It may not be too long before decisions made by AI systems will start to be judged as better than those made by human beings. When this happens, how will such a system be used? A decision will have to be made as to whether AI-facilitated decisions are regarded as a form of apolitical, value-neutral, evidence-based algorithmic outcome, and therefore one that is superior to most human decisions–or will such decisions be viewed only as recommendations and suggestions for human beings to consider.

Mitigations: Clear boundaries need to be set for AI-augmented decision-making, and how and where it may be used, and how it will be allowed to evolve autonomously [Ahn et al. 2020]. That said, work has already started in understanding different ways to incorporate both human cognition and AI into the military Observe-Orient-Decide-Act (OODA) loop that is employed by law enforcement and the military in their decision-making. If an AI system is to be present in such a loop as part of a human-machine team, work such as [Blaha 2018] outlines several different models for accomplishing this integration. One consequence of such an integration is the possibility of the machine reasoning not only about the external tactical situation, but also about the condition and behavior of the human. In the case of a pilot, for example, what maneuver is the pilot executing, what physiological effect might that maneuver be having on the pilot (e.g., hypoxia), does the pilot seem calm, does the pilot trust the computer's recommendations, and so on. The paper considers a number of different models, with discussions of the relative advantages and disadvantages of each.

See also the section on Control of Technology for more information on managing the autonomy of decisions made by AI.

Values and Criteria for Machine Decisions

Problem: An important consideration is how AI decision-making will determine "good" (i.e., desirable) and "bad" (i.e., undesirable) aspects of a decision, and, more importantly, from what point of view. [Ahn et al. 2020] believes that there are three points of view that need to be considered: 1) individual human beings, 2) humanity as a collective group, and 3) the artificial intelligence system itself. The results of an AI system's decision-making could be substantially different based on that point of view, determining whether the system is making decisions that benefit individual people, or benefit the larger society as a whole, or emphasize the system's own criticality and continuation (perhaps as part of a larger goal). There are clear tradeoffs that become apparent when trying to balance each of these goals. In protecting individuals, the legal rights of the individual, including preservation of life and improving well-being is paramount. However, the perspective of benefitting society means trading off some individual rights in favor of benefits to society. In emphasizing the criticality of the system itself, how this would be balanced may depend on the larger goal being served, and its relative importance vis-à-vis the individual and societal rights and benefits. A potential problem arises with the learning, autonomy, and evolution of the AI systems over time. While the specter remains for some of self-aware and "conscious" systems, there is no need for this to occur before the problem becomes relevant. The simple fact that AI systems are able to learn, and adapt, and act autonomously, and may be put in control of decisions that affect lives, creates an inherent source of potential risk.

Mitigations: It is becoming clear that it is the responsibility of human beings to both understand and then ethically constrain the solution spaces that such an AI system employs. This explicit ethical boundary must be consciously determined and enforced by human beings, who must remain the focal point of this process. While this idea may still seem almost fanciful in 2020, AI & ML systems are already regularly surpassing human beings in their decision-making capabilities, even in strategic planning areas where human beings have traditionally dominated. In many respects the future is arriving more rapidly than expected, work on how to engineer systems to implement such ethical boundaries is needed now.

Citations

[Adesanya et al. 2019]

Trust and Understandability in Autonomous and Unmanned Surface Vehicles. K.A. Adesanya, S.K. Shivashankar. Naval Postgraduate School. 2019.

https://apps.dtic.mil/sti/pdfs/AD1086904.pdf

Abstract/Summary: Within the human-machine relationship, distrust can arise. The Department of Defense uses automation, autonomous systems, and artificial intelligence to reduce cognitive workload and improve mission capabilities; however, adoption rates of autonomous unmanned surface vehicles (USVs) remain low. This thesis asks how human distrust of machines and machine learning relates to adoption rates. First, we identify trust components by building upon a model created by Gari Palmer, Anne Selwyn, and Dan Zwillinger in 2016. Then, we identify components that apply to the military environment that could affect the adoption rate such as smoothing time, policies and regulations, competition, robustness, understandability, subjective norm, human interaction, policy effect, risk to force, time sensitivity, war, time between wars, and catastrophic failure. Through S-curve and smoothing modeling, we find that trust components can be quantified in the human machine relationship as positive or negative trust, and that a relationship exists between understandability and adoption. While autonomous system components generally undergo rigorous testing to verify suitability and operability, human-machine trust is not usually incorporated into design and testing phases. When trust is built into the design and acquisition process, adoption of autonomous USVs is more likely to increase. Researchers can apply our trust model to future autonomous systems to mitigate distrust and human-machine teaming.

[Ahn et al. 2020]

Artificial Intelligence in Government: Potentials, Challenges, and the Future. M.J.U. Ahn and Y.C. Chen. 21st Annual International Conference on Digital Government Research. Seoul, Republic of Korea. ACM. June 2020.

https://dl.acm.org/doi/pdf/10.1145/3396956.3398260?casa_token=ThH_rFO9_kAAAAA:krS8wMWZzVgfwPznmrnNi1ZAN_PBLLh8v9lKVLNpG9nkEav2bd7CohIVTxTuHAO-p8tXT0Tdl43Y

Abstract/Summary: The rapid advancement of AI technologies—machine learning, Big Data, Cloud Computing and Internet of Things (IoT) and other related technologies—has dramatically expanded the technological capacities of the government and the application of AI technologies in government has been accelerating into more substantial areas of the government functions. Often compared to the Fourth Industrial Revolution, AI technologies are expected to change our society in a fundamental way. This will create the need for the public sector to adapt and coordinate the broader social transformation around the new technology. At this important juncture, this paper explores the significance of AI technologies put on a broader spectrum of frontier technologies that have previously transformed our society and the public sector; examine its unique attributes, potentials, and applications for government services; investigate the landscape of the current use of AI technologies in government; and discuss key challenges the new technology will pose to the government and how they may be addressed.

[Amadi 2021]

Machine Learning as a Strategic Initiative for Cyber Defense. Amadi, Kingsley Chimezie. Northcentral University. ProQuest Dissertations Publishing. 2021. https://search.proquest.com/openview/f3ff9eebd7dad8daf02af9d67b86c7d5/1?pq-origsite=gscholar&cbl=18750&diss=y

Abstract/Summary: Internet access to personal and corporate systems is being compromised at a rapid speed. Both internal and external actors contribute to hacking and hijack critical information. Organizations and individuals rely on company-provided security. In most cases, the so-called secured physical and digital infrastructures are leveraged by perpetrators and nation-states to do damage and cost-prohibitive enterprise ransom for data and identity hijacking. Humans are the weakest link in providing network and data security because they cannot monitor the millions of network traffic that traverses the organizations' network systems. Machine learning as a strategic initiative for cyber defense and a counterpart to the human security professional offers a great opportunity and around the clock visibility into the network by analyzing several millions of potential attacks that humans may miss due to the vast amounts of data that needs to be intercepted and analyzed for conformance to data integrity and compliance to security policies. The global community has changed as we are witnessing the tremendous impact phishing, malware, and viruses are contributing to online risks, identity thefts, fraudulent transactions, and business interruptions. With enabling technological capabilities of machine learning, artificial intelligence, and the ability to counter adversarial cyberattacks with hardware and software tools, the business community will be able to secure online communications.

[Blaha 2018]

Interactive OODA Processes for Operational Joint Human-Machine Intelligence. L.M. Blaha. NATO IST: 160 Specialist's Meeting. 2018. https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/STO-MP-IST-160/MP-IST-160-PP-3.pdf

Abstract/Summary: A key advantage to strategic thinking with the Observe-Orient-Decide-Act (OODA) framework is that it provides a systematic approach to get inside the decision-making process of another agent, either cooperative or adversarial. Indeed, current OODA concepts have supported understanding human decision processes to support agile and competitive decisions about human warfighters and human-centric operations. However, future military decision making based on human-machine teaming relies on technology and interaction concepts that support joint human-machine intelligence, not just human capabilities. This requires new OODA concepts. In the report, the author defines a machine OODA loop, considering the characteristics that make it similar to and different from the human OODA loop. The author considers how advances in artificial intelligence and cognitive modeling can be integrated within the machine-Orient stage, providing the machine a unique advantage over humans in that the machine can integrate a level of understanding and prediction about human operators together with predictions about machine behaviors and data analytics. Additionally, I propose that effective human-machine teaming should be supported by human-machine joint decision-action processes, conceptualized as interacting OODA loops. Consideration of the interacting human-machine OODA processes offers conceptual guidance for design principles and architectures of systems supporting effective operational human-machine decision making.

[Breck et al. 2017]

The ML Test Score: A Rubric for ML Production Readiness and Technical Debt Reduction. Eric Breck, Shanqing Cai, Eric Nielsen, Michael Salib, and D. Sculley. 2017 IEEE International Conference on Big Data. 2017. https://research.google/pubs/pub46555

Abstract/Summary: Creating reliable, production-level machine learning systems brings on a host of concerns not found in small toy examples or even large offline research experiments. Testing and monitoring are key considerations for ensuring the production-readiness of an ML system, and for reducing technical debt of ML systems. But it can be difficult to formulate specific tests, given that the actual prediction behavior of any given model is difficult to specify a priori. This paper presents 28 specific tests and monitoring needs, drawn from experience with a wide range of production ML systems to help quantify these issues and present an easy to follow roadmap to improve production readiness and pay down ML technical debt.

[Brown 2019]

Defense Innovation Unit: 2019. Michael Brown. Defense Innovation Unit (DIU) Annual Report. 2019. https://apps.dtic.mil/sti/pdfs/AD1104019.pdf

Abstract/Summary: Machine Learning Predictions: Whether it is understanding three or 30 years of historical data or tracking real-time performance from millions of sensors across an electrical grid, AI powers mission readiness and reduces costs. However, with data-enabled trending and predictions not only can things be fixed before they break, so can people. Big Data Analysis: The sheer amount of data that is generated from sensors on a daily basis makes human processing and analysis impossible, let alone identifying the critical signal amidst the noise. AI-Enhanced Decision Making: Managing disparate data feeds from various sensors slows our ability to provide decision options at speed. Harnessing commercial capabilities such as financial modeling and insurance risk projections can inform target identification, tracking classification, real-time threat assessment, mission maps, and post-disaster damage assessments.

[Browne 2019]

Innovation Acquisition Practices in the Age of AI. Maj. Andrew S. Bowne. Army Lawyer. 74. 2019. https://heinonline.org/HOL/LandingPage?handle=hein.journals/armylaw2019&div=19&id=&page=

Abstract/Summary: Discusses various obstacles to AI acquisition that the DoD must overcome to remain competitive in the AI domain. These include: 1) understanding the potential of Al and determining the legal and ethical restrictions on the use of such technology, where a key issue of ensuring legal compliance with the laws of war is the use of lethal autonomous weapons systems (LAWS), 2) increasing the speed of acquisition through streamlined contractual vehicles, and 3) working more closely with commercial companies that are leaders in AI.

The paper discusses the need for the acquisition system to keep up with the pace of innovation in AI, and the role of acquisition reform in achieving that, and enabling the government to work with commercial companies that are leaders in AI technology development by using such streamlined contractual vehicles as OTAs and Section 804 authority to allow them to move faster and with more flexibility.

The paper also discusses the DoD's ability to attract businesses like Google to develop AI & ML technology for them, through the creation of such organizations as the Defense Innovation Unit (DIU), the Strategic Capabilities Office (SCO), the Army Futures Command, and the DARPA "Grand Challenges" that are focused on building relationships with non-traditional defense contractors.

[Carleton et al. 2020]

The AI Effect: Working at the Intersection of AI and SE. Anita D. Carleton, Erin Harper, Tim Menzies, Tao Xie, Sigrid Eldh, and Michael R. Lyu. IEEE Software. Vol. 37. No. 4. July/August 2020. https://ieeexplore.ieee.org/document/9121618

Abstract/Summary: This special issue explores the intersection of artificial intelligence (AI) and software engineering (SE), that is, what can AI do for SE, and how can we as software engineers design and build better AI systems?

[Cheatham et al. 2019]

Confronting the Risks of Artificial Intelligence. Benjamin Cheatham, Kia Javanmardian, Hamid Samandari. April 26, 2019. https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/confronting-the-risks-of-artificial-intelligence#

[Chesterman 2020]

Artificial Intelligence and the Problem of Autonomy. S. Chesterman. Notre Dame Journal on Emerging Technologies. 2020. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3450540

Abstract/Summary: Artificial intelligence (AI) systems are routinely said to operate autonomously, exposing gaps in regulatory regimes that assume the centrality of human actors. Yet surprisingly little attention is given to precisely what is meant by "autonomy" and its relationship to those gaps. Driverless vehicles and autonomous weapon systems are the most widely studied examples, but related issues arise in algorithms that allocate resources or determine eligibility for programs in the private or public sector. This article develops a novel typology of autonomy that distinguishes three discrete regulatory challenges posed by AI systems: the practical difficulties of managing risk associated with new technologies, the morality of certain functions being undertaken by machines at all, and the legitimacy gap when public authorities delegate their powers to algorithms.

[CMMI 2014]

Data Management Maturity Model, Ver. 1.0. CMMI Institute. August 2014.
https://dmm-model-individual.dpdcart.com

[Combe 2020]

Autonomous Doctrine: Operationalizing the Law of Armed Conflict in the Employment of Lethal Autonomous Weapons Systems. Peter C. Combe II, USMC. St. Mary's Law Journal. Vol. 51. Num. 1. Article 2. https://commons.stmarytx.edu/cgi/viewcontent.cgi?article=1050&context=thestmaryslawjournal

Abstract/Summary: Autonomous machines have become a commonplace reality. The popular perception of future lethal autonomous weapons (LAWS) as unrestrained killers has spawned both negative public perception and calls for bans on these "killer robots." However, LAWS can be effectively employed in appropriate circumstances, in compliance with the Law of Armed Conflict using a series of control mechanisms and tactics, techniques, and procedures (TTPs) that are already in use by the U.S. military, or can be adapted to autonomous weapons systems. Ban proponents deem that "killing at a remove" strips the human aspect from war, making it cold, impersonal, and indiscriminate. This failure of "humanity" in LAWS brings such weapons within the ambit of the Martens Clause, which prohibits the means or methods of warfare that diverge from the "principles of international law derived from established custom, from the principles of humanity and from the dictates of public conscience." Ban proponents believe machines could not distinguish between combatants and civilians or non-combatants, so a machine might fail to apply a concept known as proportionality to determine whether an attack is justified—but this is flawed reasoning. Reasonable mistakes which lead to undesired results are not necessarily criminal or unlawful, but may be mistaken or accidental.

[D'Amour et al. 2020]

Underspecification Presents Challenges for Credibility in Modern Machine Learning
A.D. D'Amour, K. Heller, D. Moldovan, B. Adlam, B. Alipanahi, A. Beutel, C. Chen et al., arXiv preprint arXiv: 2011.03395. 2020. https://arxiv.org/pdf/2011.03395.pdf

Abstract/Summary: ML models often exhibit unexpectedly poor behavior when they are deployed in real-world domains. We identify underspecification as a key reason for these failures. An ML pipeline is underspecified when it can return many predictors with equivalently strong held-out performance in the training domain. Underspecification is common in modern ML pipelines, such as those based on deep learning. Predictors returned by underspecified pipelines are often treated as equivalent based on their training domain performance, but we show here that such predictors can behave very differently in deployment domains. This ambiguity can lead to instability and poor model behavior in practice, and is a distinct failure mode from previously identified issues arising from structural mismatch between training and deployment domains. We show that this problem appears in a wide variety of practical ML pipelines, using examples from computer vision, medical imaging, natural language processing, clinical risk prediction based on electronic health records, and medical genomics. Our results show the need to explicitly account for underspecification in modeling pipelines that are intended for real-world deployment in any domain.

[DAMA 2009]

The DAMA Guide to the Data Management Body of Knowledge. DAMA International Technics Publications. 2009. https://www.dama.org/content/body-knowledge

[Dent 2020]

The Risks of Amoral AI: The Consequences of Deploying Automation Without Considering Ethics Could be Disastrous. Kyle Dent. TechCrunch. August 25, 2019. https://techcrunch.com/2019/08/25/the-risks-of-amoral-a-i

Abstract/Summary: The author, who is a research manager for Xerox PARC, focuses on the interplay between people and technology, and leads the ethics review committee at PARC. He contends that the consequences of deploying automation without considering ethics could be disastrous. He discusses concerns with AI-based systems autonomously driving cars, assessing job performance, making hiring decisions, granting loans, and rendering decisions in criminal justice, where consequences range from physical harm and damage, to real-world harm from automated decision-making systems. The article discusses issues ranging from the “buyer beware” aspects of purchasing AI systems when the buyers know so much less about the technology than the sellers do, to fairness questions arising from federal law in areas like the 1968 Fair Housing Act, to concerns about the existence of adequate regulation in the many actual and potential application areas of AI and machine learning.

[Dignum 2017]

Responsible Autonomy. Virginia Dignum, arXiv preprint arXiv:1076.02513. 2017.
https://arxiv.org/abs/1706.02513

Abstract/Summary: As intelligent systems are increasingly making decisions that directly affect society, perhaps the most important upcoming research direction in AI is to rethink the ethical implications of their actions. Means are needed to integrate moral, societal, and legal values with technological developments in AI, both during the design process as well as part of the deliberation algorithms employed by these systems. In this paper, we describe leading ethics theories and propose alternative ways to ensure ethical behavior by artificial systems. Given that ethics are dependent on the socio-cultural context and are often only implicit in deliberation processes, methodologies are needed to elicit the values held by designers and stakeholders, and to make these explicit leading to better understanding and trust on artificial autonomous systems.

[Dixit et al. 2020]

Artificial Intelligence and Machine Learning in Sparse/Inaccurate Data Situations. R Dixit, R.B. Chinnam, H. Singh. IEEE Aerospace. 2020. https://ieeexplore.ieee.org/abstract/document/9172612/?casa_token=mHtmaK2dnyoAAAAA:QDkYjYrcfTQmVim4F99qWJmETc3CJ5eAlVsiNDSYQJU1fulg-_upCJ9dpnz0TsfTIm7bpfGd

Abstract/Summary: Machine Learning (ML) and other artificial Intelligence (AI) techniques have been developed for real-time decision making, and are gaining traction in data-rich situations. However, these techniques are less proven in sparse-data environments, and currently are more the subject of research than application. Typical implementations of ML and AI require a cross-disciplinary decision engine that, once “trained,” can cognitively respond to changes in input. The key to successful training is to a) have a defined decision-basis (answer-key), and/or b) facilitate sufficient learning, both of which require ample data (observability) and ample time for the machine to develop a logical outcome. Much research has been focused on developing decision algorithms using various logical formulations, dimensionality reductions, neural techniques, and learning reinforcements for tasks that traditionally require human intelligence. What is missing in most current research streams are implementations of ML and AI for decisions that are fundamentally rooted in human intuition and empathy, e.g., situations in which the decision requires a holistic view and the outcome is based on a qualitative judgement based on context and fact. This paper is intended to benefit a wide range of readers considering artificial intelligence, from the merely curious to “techies” from other disciplines to experienced practitioners and researchers. Using a qualitative/ characteristics base perspective of data and AI, we examine defense industry procurement, operational, tactical, and strategic decision scenarios, then identify where AI can currently promote better informed decisions and which arenas need would benefit by letting AI technology and sophistication evolve further.

[Drezner et al. 2020]

Benchmarking Data Use and Analytics in Large, Complex Private-Sector Organizations: Implications for Department of Defense Acquisition. J.A. Drezner, J. Schmid, J. Grana, M. McKernan, M. Ashby. DTIC. 2020. https://apps.dtic.mil/sti/citations/AD1101362

Abstract/Summary: Public and private organizations are increasingly aware of the potential value of data and analytics to improving organizational performance and outcomes. The U.S. Department of Defense DoD is one of those organizations. Its size, complexity, security needs, and culture have created a challenging environment for successful use of data in decision-making. Over the past five years, the RAND Corporation has studied how DoD governs, manages, secures, and uses data within its acquisition institutions.

[Ehn 2017]

Artificial Intelligence: The Bumpy Path Through Defense Acquisition. Eric J. Ehn. Naval Postgraduate School. 2017. https://apps.dtic.mil/dtic/tr/fulltext/u2/1053222.pdf

Abstract/Summary: The use of artificial intelligence systems is ready to transition from basic science research and a blooming commercial industry to strategic implementation in the Defense Acquisition system. The purpose of this research is to determine the problems awaiting artificial intelligence (AI) systems inherent to defense acquisition. AI is a field of scientific study focused on the construction of systems that can act rationally, behave humanly, and adapt. To achieve AI behavior takes AI essentials, which consider mobility, system perspective, and algorithms. Unfortunately, AI essentials are under addressed in the concept of operations that fuels the Joint Capabilities Integration and Development System. Influences to the concept of operations analyzed in this research include strategic documentation, joint technology demonstrations, and exercises that aim to capture technology-based lessons learned. Failure to address AI essentials causes problems in defense acquisition: system requirements are impossible to define; transition of AI technology fails; testing cannot be evaluated with confidence; and life cycle planning is at best a guess. To address these issues, the Department of Defense needs improved planning, acquisition personnel training, and AI-supported acquisition processes to achieve cost, schedule, and performance goals.

[EU 2020]

Draft report with recommendations to the Commission on a Civil liability regime for artificial intelligence. European Parliament Committee on Legal Affairs. April 27, 2020. *https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/JURI/PR/2020/05-12/1203790EN.pdf*

Abstract/Summary: The European Union (EU) Committee on Legal Affairs is working to develop a common liability framework for AI systems, and is currently recommending a regulation be created to place strict liability on the "deployer" of certain "high risk" AI systems, and increased deployer liability for other types of AI systems.

[Golden 2020]

DoD's Artificial Intelligence Problem: Where to Begin. Col. Paul E. Golden. Army Lawyer. Volume 76. 2020.

https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/armylaw2020&section=41&casa_token=i0XUjS3NveMAAAAA:OOS5SSk5ZgZcJQYXBRgZ7YfGvZJClF5XwD7TQQ9n4mAvfkBla8xIoSAWAStMIwCt6nSe1HYX

Abstract/Summary: Some believe the emergence and proliferation of Artificial Intelligence (AI) represents humanity's "fourth industrial revolution" and that it will drive evolutionary and revolutionary innovation—i.e., make us better at what we do (the things we know) and shape what we do in the future and how we do it (what has yet to be done). The breadth of AI possibilities is not easy to conceptualize, but there is great interest in understanding AI and how it can be effectively and responsibly leveraged.

This paper discusses how the DoD can understand, control, field, and develop ethical but effective AI, and maintain dominance and leadership in the AI realm through changes in the way the DoD does business. There are specific discussions of the role of the NDAA for FY2019 and its key provisions for exploring and developing AI. It also discusses the President's AI Strategy of 2019 and its provisions, as well as tackling the role of ethics in AI system development, and the current state of DoD funding for AI with respect to that of global adversaries. The paper also discusses the state of acquisition reform as it pertains to AI system development, and the need for reform of DoD's civilian workforce to be able to exploit AI technologies.

[Hutchison 2018]

Artificial Intelligence in Defense Acquisition. Todd E. Hutchison. Naval War College. Gravely Naval Warfare Research Group. 2018. https://apps.dtic.mil/dtic/tr/fulltext/u2/1057662.pdf

Abstract/Summary: Even after repeated high-visibility efforts at reform, the Department of Defense is still to a large extent conducting acquisition as it has for decades. While technology such as big data analytics and artificial intelligence (AI) have become widely prevalent in private industry and are beginning to become more common in some governmental agencies, adoption of these technologies and innovation concerning their possible use is lagging. There is a pressing need in light of the global environment to incorporate AI in both the battlefield and the bureaucracy. Revolutionizing acquisition through the high-end and low-end use of AI could facilitate current and future technologies getting fielded more rapidly. While concerns about AI (such as diffusion, control, and terminology) are valid, we must address and mitigate the risks, push forward with the technologies, and decrease the distance between the scientist and the end-user if the United States is going to prevail in future conflicts.

[Klemas et al. 2018]

Cyber Acquisition: Policy Changes to Drive Innovation in Response to Accelerating Threats in Cyberspace. Thomas Klemas, Rebecca K. Lively, and Nazli Choucri. The Cyber Defense Review SPECIAL EDITION: International Conference on Cyber Conflict (CYCON U.S.). November 14-15, 2018. Cyber Conflict During Competition pages 103-120. 2019.
https://www.jstor.org/stable/26846123?seq=1#metadata_info_tab_contents

Abstract/Summary: The United States of America faces great risk in the cyber domain because our adversaries are growing bolder, increasing in number, improving their capabilities, and doing so rapidly. Meanwhile the associated technologies are evolving so quickly that progress toward hardening and securing this domain is ephemeral, as systems reach obsolescence in just a few years and revolutionary paradigm shifts, such as cloud computing and ubiquitous mobile devices, can pull the rug out from the best-laid defensive planning by introducing entirely new regimes of operations. Contemplating these facts in the context of Department of Defense (DoD) acquisitions is particularly sobering because many cyber capabilities built within the traditional acquisition framework may be of limited usefulness by the time that they are delivered to the warfighter. Thus, it is a strategic imperative to improve DoD acquisitions pertaining to cyber capabilities. This paper proposes novel ideas and a framework for addressing these challenges.

[Kollars 2017]

The Rise of Smart Machines: The Unique Peril of Intelligent Software Agents in Defense and Intelligence. N.A. Kollars. The Palgrave Handbook of Security, Risk and Intelligence. Pages 195-211. Springer. 2017. https://link.springer.com/chapter/10.1057/978-1-137-53675-4_11

Abstract/Summary: As computer processing power and cyber connectivity has increased, states have turned to intelligent software agents (ISA) as a potential means to extend the reach of their militaries and the analytical capacity of their intelligence agencies. Intelligent software agents (ISA) are computer programs that have the ability to learn, cooperate, and act independently of humans. The development of technologies that can operate autonomously has generated nearly as much anxiety as it has excitement, often to the detriment of clear-eyed analysis. The path to acquiring and fully implementing ISA is farther away, and more complex than its advocates and detractors will admit. In addition to this, while ISAs could afford new capabilities for information analysis and battlefield risk reduction, they simultaneously introduce their own unique risks in implementation.

[Kumar et al. 2020]

Identifying Bias in Machine Learning Algorithms: Classification without Discrimination. Manish Kumar, Rahul Roy, Kevin D. Oden, RMA Journal, September 2020. https://rmajournal.org/rmajournal/september_2020/MobilePagedArticle.action?articleId=1616849#articleId1616849

Abstract/Summary: Machine Learning (ML) advanced statistical and mathematical models are used in various partial or fully automated decision-making systems that affect individual lives. Today, these models are not only increasingly used to make important decisions in our financial lives, ranging from retail (closed and open-end products) and wholesale scorecards (application, behavioral, and collection) but also in other aspects such as granting university admission, social benefit assignment, predicting the risk of criminal recidivism (COMPAS model), and part of hiring tools to review job applicants' resumes. In these applications, models are often built using sensitive drivers, also called attributes, such as age, gender, nationality, religion, race, language, culture, marital status, economic condition, zip code, etc. One of the unintended consequences of lax modeling practice is the potential for bias or unfairness in ML models that accentuates our societal stereotypes and contravenes the laws of many jurisdictions as well.

[Lindsay 2020]

Information Technology and Military Power. Jon R. Lindsay. Cornell University Press. July 15, 2020. https://books.google.com/books?hl=en&lr=&id=M3a4DwAAQBAJ&oi=fnd&pg=PR5&dq=%22defense+acquisition%22+%22policy+implications%22+%22ma-chine+learning%22&ots=FN6g1R4FXq&sig=3Lg9AntkgbeVsibDzXF17eCNwbc#v=onepage&q&f=false

Abstract/Summary: Militaries with state-of-the-art information technology sometimes bog down in confusing conflicts. To understand why, it is important to understand the micro-foundations of military power in the information age, and this is exactly what Jon R. Lindsay's Information Technology and Military Power gives us. As Lindsay shows, digital systems now mediate almost every effort to gather, store, display, analyze, and communicate information in military organizations. He highlights how personnel now struggle with their own information systems as much as with the enemy. Throughout this foray into networked technology in military operations, we see how information practice—the ways in which practitioners use technology in actual operations—shapes the effectiveness of military performance. The quality of information practice depends on the interaction between strategic problems and organizational solutions. Information Technology and Military Power explores information practice through a series of detailed historical cases and ethnographic studies of military organizations at war. Lindsay explains why the U.S. military, despite all its technological advantages, has struggled for so long in unconventional conflicts against weaker adversaries. This same perspective suggests that the U.S. retains important advantages against advanced competitors like China that are less prepared to cope with the complexity of information systems in wartime. Lindsay argues convincingly that a better understanding of how personnel actually use technology can inform the design of command and control, improve the net assessment of military power, and promote reforms to improve military performance. Warfighting problems and technical solutions keep on changing, but information practice is always stuck in between.

[Loss 2018]

Assessing Strategic Effects of Artificial Intelligence. R. Loss, U.S. Department of Energy (DoE) Office of Scientific and Technical Information. 2018. https://www.osti.gov/biblio/1467805

Abstract/Summary: This workshop examines the implications of advances in artificial intelligence (AI) on international security, discussing the question of whether we will have to rethink, by the end of the next decade, how we practice nuclear deterrence and ensure strategic stability. Hosted by the Center for Global Security Research (CGSR) at Lawrence Livermore National Laboratory (LLNL), the workshop is part of a collaboration between CGSR and Technology for Global Security (T4GS) to engage policymakers, scholars, technical experts, and the private sector to address emerging challenges in AI and related issues. The workshop engages with the current literature pointing to the risks and opportunities presented by AI and attempts to assess which are legitimate and which might be exaggerated.

[Margolis et al. 2019]

Accessibility of Big Data Imagery for Next Generation Machine Learning Applications. S. Margolis, W.L. Michaels, B. Alger, C. Beaverson. National Oceanic and Atmospheric Administration (NOAA) Fisheries. 2019. https://repository.library.noaa.gov/view/noaa/20200

Abstract/Summary: NOAA generates tens of terabytes of data a day, also known as “big data,” from satellites, radars, ships, weather models, optical technologies, and other sources. This unprecedented growth of data collection in recent years has resulted from enhanced sampling technologies and faster computer processing. While these data are publicly available, there is not yet sufficient access to the data by next generation processing technologies, such as machine learning (ML) algorithms that are able to improve processing efficiencies. Accessibility is the key component for utilizing analytical tools... This report focuses on the challenges of accessibility of imagery (defined as still images and video) from the marine environment. While technologies have dramatically increased the spatial and temporal resolution of data, the drastic increase in big data, specifically imagery, presents numerous challenges. Case studies discussed in this report highlight that big data imagery are readily being collected and stored, yet the foundation for the long-term storage and accessibility of big data must be based on the necessary guidance for its architecture, infrastructure, and applications to enhance the accessibility and use of these data. Additionally, the report highlights key considerations and recommendations for data modernization efforts that align with mandates such as Public Access to Research Results, the Evidence-Based Policy Making Act, Department of Commerce Strategic Plan, the President’s Management Agenda, and White House Executive Order on Artificial Intelligence (AI). As big data and analytical tools become more commonplace for research and scientific operations, there is an increasing need to create end-to-end data management practices that improve data accessibility for analytical tools that utilize AI, computer vision (AI applied to the visual world), and ML. The development and application of AI and ML analytics will progress as long as there is accessibility of big data with enriched metadata; however, accessibility appears to be the primary challenge to fully utilize ML analytics. Rapid, optimal access to entire imagery and data collections is critical to create annotated imagery libraries for supervised analysis using ML algorithms. This report highlights the common need to implement accessibility solutions to facilitate efficient imagery processing using available analytical tools. Other critical requirements to enable AI include the necessary metadata for discovery, long term data archive and access, and economical multi-tier storage. As big data imagery is made more readily available to open source tools such as ML analytics, significant cost reductions in data processing will be realized by reducing the labor-intensive efforts currently needed. ML tools accelerate processing of imagery with automated detection and classification resulting in more timely and precise scientific products for management decisions. Furthermore, as the broader scientific community expands its research and discovery from increased accessibility of big data imagery, the ML applications will increase the number of insightful science-based products beyond the scope of the original operational objectives, thereby increasing the value of the agency’s scientific products.

[McCormick et al. 2018]

Acquisition Trends, 2018: Defense Contract Spending Bounces Back. R. McCormick, A.P. Hunter, S. Cohen, G. Sanders. Center for Strategic and International Studies. Rowman & Littlefield. 2019. https://books.google.com/books?hl=en&lr=&id=2cSrD-wAAQBAJ&oi=fnd&pg=PR3&dq=%22defense+acquisition%22+policy+%22machine+learning%22&ots=tsYW1Y8exG&sig=MhPElx3DLHvnjGLX7EAmq1S-foo#v=onepage&q=%22defense%20acquisition%22%20policy%20%22machine%20learning%22&f=false

Abstract/Summary: This book includes information on what types of acquisition vehicles and approaches are being used to acquire systems for the DoD, including AI & ML systems.

[Novak et al. 2019]

SAF/MG Data Management Study Results: Assess Current State. William E. Novak, Cecilia C. Albert, Julie B. Cohen, Susan C. Cox, Patrick J. Donohoe, Melissa K. Ludwick, M. Steven Palmquist, Patrick R. H. Place. Software Engineering Institute. CMU/SEI-2018-SR-032. Restricted Distribution. November 2019.

Abstract/Summary: This report provides an assessment of the current states of the Air Force Business Mission Area (BMA) data management efforts through the use of questionnaires and on-site interviews. The report also provides some early draft results assessing the projected future states of those same organizations. This effort defined an assessment model, gathered information on the BMA organizations' data management capabilities, specifically assessing AF-A1 (Personnel), AF-A4 and AFMC/A4N (Logistics & Force Protection), SAF/AQ (Acquisition), SAF/FM (Financial Management), the Air Force Reserve Command (AFRC), the Air Education & Training Command AETC), PEO Command, Control, Communications, Intelligence, & Networking (C3I&N), and PEO Business Enterprise Systems (BES). The report begins with a description of the SEI approach, and then provides the results from each organization, finally offering a set of overarching observations and conclusions. The results for each organization include an overview, a summary of the results grouped by the Data Maturity Model (DMMSM) question areas, scores along seven different data management dimensions and their subcategories, and (for all but the PEO organizations) a data management architecture diagram showing the progress in specific data management areas.

[Ozkaya 2020]

What is Really Different in Engineering AI-Enabled Systems? Ipek Ozkaya. IEEE Software. July/August. Vol. 37. No. 4. 2020. https://ieeexplore.ieee.org/document/9121629

Abstract/Summary: Advances in machine learning (ML) algorithms and increasing availability of computational power have resulted in huge investments in systems that aspire to exploit artificial intelligence (AI), in particular ML. AI-enabled systems, software-reliant systems that include data and components that implement algorithms mimicking learning and problem solving, have inherently different characteristics than software systems alone. However, the development and sustainment of such systems also have many parallels with building, deploying, and sustaining software systems. A common observation is that although software systems are deterministic and you can build and test to a specification, AI-enabled systems, in particular those that include ML components, are generally probabilistic. Systems with ML components can have a high margin of error due to the uncertainty that often follows predictive algorithms. The margin of error can be related to the inability to predict the result in advance or the same result cannot be reproduced. This characteristic makes AI-enabled systems hard to test and verify. Consequently, it is easy to assume that what we know about designing and reasoning about software systems does not immediately apply in AI engineering. AI-enabled systems are software systems. The sneaky part about engineering AI systems is they are "just like" conventional software systems we can design and reason about—until they're not.

[Quinonera-Candela et al. 2009]

Dataset Shift in Machine Learning. Joaquin Quinonero-Candela, Masashi Sugiyama, Anton Schwaighofer, Neil D. Lawrence. The MIT Press. 2009. http://www.acad.bg/ebook/ml/The.MIT.Press.Dataset.Shift.in.Machine.Learning.Feb.2009.eBook-DDU.pdf

Abstract/Summary: Dataset shift is a common problem in predictive modeling that occurs when the joint distribution of inputs and outputs differs between training and test stages. Covariate shift, a particular case of dataset shift, occurs when only the input distribution changes. Dataset shift is present in most practical applications, for reasons ranging from the bias introduced by experimental design to the irreproducibility of the testing conditions at training time. (An example is email spam filtering, which may fail to recognize spam that differs in form from the spam the automatic filter has been built on.) Despite this, and despite the attention given to the apparently similar problems of semi-supervised learning and active learning, dataset shift has received relatively little attention in the machine learning community until recently. This volume offers an overview of current efforts to deal with dataset and covariate shift. The chapters offer a mathematical and philosophical introduction to the problem, place dataset shift in relationship to transfer learning, transduction, local learning, active learning, and semi-supervised learning, provide theoretical views of dataset and covariate shift (including decision theoretic and Bayesian perspectives), and present algorithms for covariate shift.

[Santelli et al. 2019]

Challenges for Government Adoption of AI. Julian Torres Santelli, Sabine Gerdon. World Economic Forum. August 16, 2019. https://www.weforum.org/agenda/2019/08/artificial-intelligence-government-public-sector

Abstract/Summary: This short white paper describes the five challenges for government adoption of AI:

1. Effective use of data
2. Data and AI skills
3. The AI environment
4. Legacy culture
5. Procurement mechanisms

[SEI 2019]

AI Engineering for Defense and National Security: A Report from the October 2019 Community of Interest Workshop. Software Engineering Institute. October 2019. https://resources.sei.cmu.edu/asset_files/SpecialReport/2020_003_001_648543.pdf

Abstract/Summary: This report is based on ideas shared in a 2019 workshop the SEI convened to identify challenges and opportunities for AI engineering for defense and national security. It highlights three areas of focus for the growing artificial intelligence (AI) engineering movement: robust and secure AI, scalable AI, and human-centered AI. It defines and presents needs and challenges for each theme. Robust and secure AI systems must work as expected and be resilient to threats, including attacks related to adversarial machine learning. Scalable AI systems must be able to operate under different conditions related to size, speed, and complexity. Human-centered systems must reflect organizational and socio-technical considerations, from ethics to interpretability.

[Tarraf et al. 2019]

The Department of Defense Posture for Artificial Intelligence: Assessment and Recommendations. D.C. Tarraf, W. Shelton, E. Parker, B. Alkire, D.G. Carew et al. RAND Corporation. 2019. https://apps.dtic.mil/sti/pdfs/AD1088616.pdf

Abstract/Summary: Section 238(e) of the fiscal year (FY) 2019 National Defense Authorization Act (NDAA) mandated that the senior Department of Defense (DoD) official with principal responsibility for the coordination of DoD's efforts to develop, mature, and transition artificial intelligence (AI) technologies into operational use carry out a study on AI topics... The study had the following three key objectives: 1. Assess the state of AI relevant to DoD and address misconceptions. 2. Carry out an independent introspective assessment of DoD's posture for AI. 3. Develop a set of recommendations for internal DoD actions, external engagements, and potential legislative or regulatory actions to enhance DoD's posture in AI. In keeping with the language of the legislation, the RAND NDRI team collected insights into these three questions through semi-structured interviews with experts within DoD, other federal agencies, academia, relevant advisory committees, and the commercial sector. The team augmented this broad input with an independent review of the portfolio of DoD investments in AI, a set of historical case studies, reviews of relevant literature, and the technical and other expertise resident in the team to arrive at the findings and recommendations presented in this report and associated annex, aligned with the three key objectives of Section 238(e) as distilled above.

[Villasenor 2019]

Products Liability Law as a Way to Address AI Harms. John Villasenor. Brookings Institution. October 2019. https://www.brookings.edu/research/products-liability-law-as-a-way-to-address-ai-harms

Abstract/Summary: Artificial intelligence (AI) is a transformative technology that will have a profound impact on manufacturing, robotics, transportation, agriculture, modeling and forecasting, education, cybersecurity, and many other applications. AI-based systems can also make decisions that are more objective, consistent, and reliable than those made by humans.

But AI also involves risks. Put simply, AI systems will sometimes make mistakes. Given the volume of products and services that will incorporate AI, the laws of statistics ensure that—even if AI does the right thing nearly all the time—there will be instances where it fails. While some of those failures may be benign, others could result in harm to persons or property. When that occurs, questions of attribution and remedies will arise.

Answering these questions requires examining the intersection of products liability and artificial intelligence. In this policy brief, I provide an overview of key concepts in products liability and their application to AI. I describe the challenges of attribution for AI-induced harms, explain why I believe that products liability frameworks are well-positioned to adapt to address AI questions, and why it is important to promote consistency across states in AI products liability approaches. If implemented with appropriate frameworks, products liability law represents an important mechanism to mitigate possible AI harms.

[Wydler et al. 2018]

Panel 21. Considerations in Accelerating Technology Adoption in Defense Acquisitions. V.L. Wydler and E.M. Schultz. Volume II Acquisition Research Creating. Proceedings of 15th Annual Naval Postgraduate School Acquisition Research Symposium. 2018. https://calhoun.nps.edu/bitstream/handle/10945/58760/SYM-AM-18-033.pdf?sequence=1#page=204

Abstract/Summary: Twenty years ago, the Navy began expanding the use of commercial industry information technology (IT) to employ Internet Protocol (IP)–based client server and web-based technologies to improve software effectiveness and affordability on ships and submarines. Coupled with wideband satellite capabilities, these systems increased the Navy's ability to plan, communicate, command and control, and execute increasingly complex missions. With a sound foundation in commercial IT installed in the Fleet, the Navy is looking today to improve warfighting by leveraging emerging technologies in Data Analytics, Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL). These technologies have the potential to change how the Navy fights and will drive changes to the Fleet's Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance (C4ISR) architecture and processes. This paper proposes a reference architecture, new processes, and tools to meet the dynamic nature of these emerging technologies, to include employing the commercial DEVelopment and OPerationS (DevOps) construct. The reference architecture and processes have the potential not only to accelerate the modernization of the afloat Navy networking WAN/LAN infrastructure, but also to deliver important warfighting capabilities to support Command and Control, Intelligence, and Logistics software applications.

[Zhang et al. 2019]

Machine Learning Testing: Survey, Landscapes and Horizons. J.M. Zhang, M. Harmon, L. Ma, and Y. Liu. IEEE Transactions on Software Engineering. 2019. https://arxiv.org/pdf/1906.10742.pdf

Abstract/Summary: This paper provides a comprehensive survey of techniques for testing machine learning systems; Machine Learning Testing (ML testing) research. It covers 144 papers on testing properties (e.g., correctness, robustness, and fairness), testing components (e.g., the data, learning program, and framework), testing workflow (e.g., test generation and test evaluation), and application scenarios (e.g., autonomous driving, machine translation). The paper also analyzes trends concerning datasets, research trends, and research focus, concluding with research challenges and promising research directions in ML testing.

Contact Us

Software Engineering Institute
4500 Fifth Avenue, Pittsburgh, PA 15213-2612

Phone: 412/268.5800 | 888.201.4479
Web: www.sei.cmu.edu
Email: info@sei.cmu.edu

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

The view, opinions, and/or findings contained in this material are those of the author(s) and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.

DM21-0097