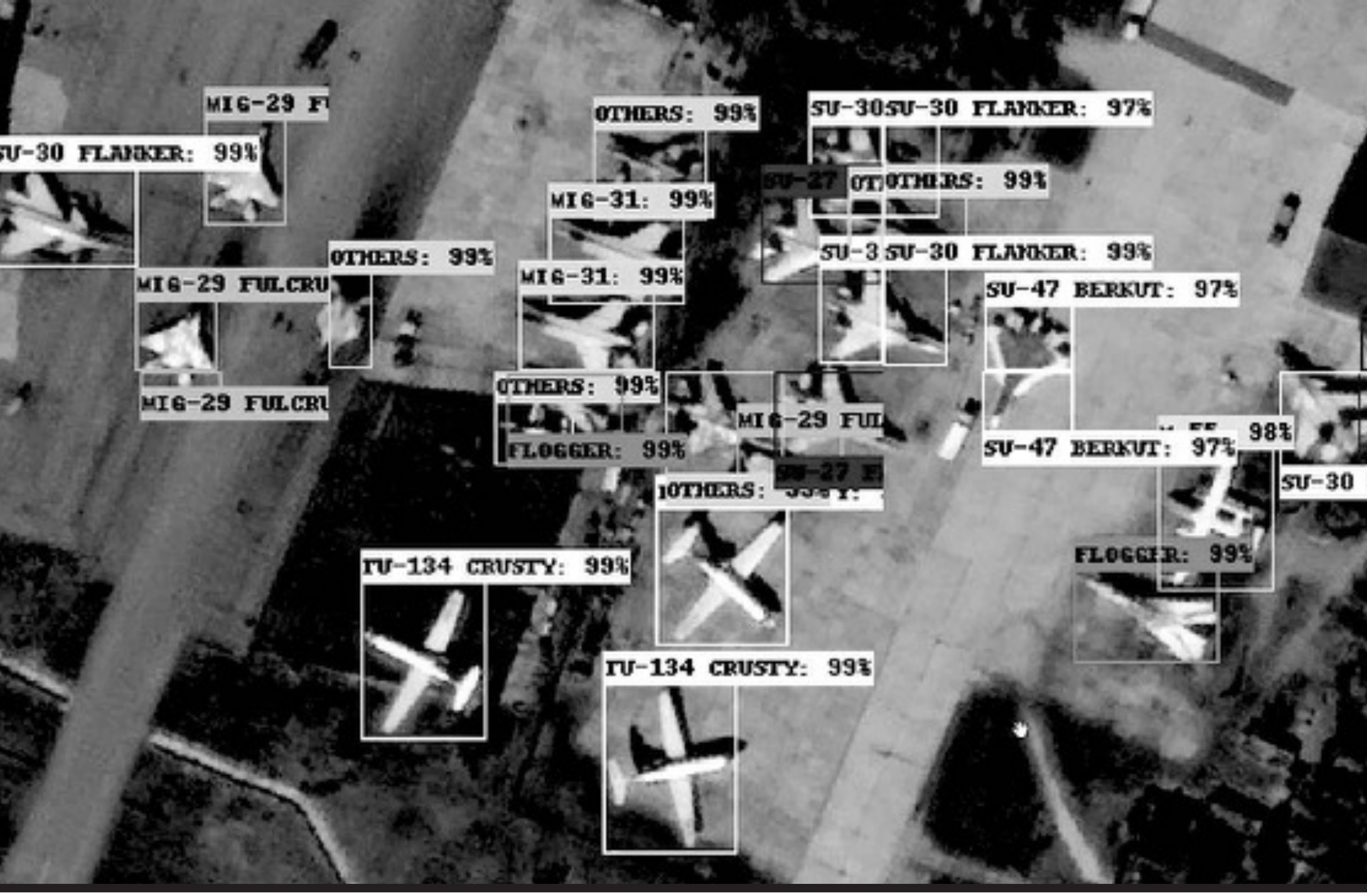

Brief No. 12.5

IN DEFENSE OF DATA

How the DoD Can Strengthen AI Development

Clara Waterman

P | I | P | S

In Defense of Data
How the DoD Can Strengthen AI Development

APRIL 2020

Clara Waterman

P|I|P|S *The Project on International Peace and Security*
Global Research Institute
College of William & Mary

In Defense of Data

How the DoD Can Strengthen AI Development

The Department of Defense (DoD) is investing in artificial intelligence (AI) to prepare for future warfare against peer adversaries. However, Washington is devoting insufficient attention and funding to the datasets that underpin AI algorithms. Poor training data can create flawed algorithms that misidentify operating environments, potentially degrading battlefield decision-making and hindering the Pentagon's ability to maintain a strategic advantage over adversaries. Consequently, high-quality datasets are crucial to the DoD mission.

At this early stage of development, the DoD has an opportunity to standardize and improve the quality of its AI training data by creating a data clearinghouse. This clearinghouse would coordinate data collection, establish best practices for both vetting AI data for bias and minimizing human error, and standardize metadata formatting. Increased collaboration and attention to curating datasets today will maximize the efficiency and effectiveness with which the DoD develops AI moving forward.

Introduction

Artificial intelligence can enable cutting-edge technologies, but, those technologies are of limited utility without well-curated training, input, and feedback data.[1] Discussions about data are critical to U.S. national security because insufficient attention to data results in destructive outcomes. Consequently, proper data procedures are crucial to both the Department of Defense's (DoD) artificial intelligence (AI) goals and its overall mission.

The DoD is a purchasing and regulatory powerhouse in the artificial intelligence research and development (R&D) field. For the 2020 fiscal year, the DoD budgeted $4.9 billion for AI and machine-learning R&D.[2] Tens of millions of dollars of this budget are doled out through DoD contracts to corporations like Microsoft, Amazon, and Google that are the winners of a competitive bidding processes. Thus, the DoD has the ability to require those competing for contracts to follow a specific set of data procedures. Because the DoD is at the forefront of this industry, it has the unique power to influence standards for data collection, vetting, and labeling.[3]

Some degree of bias in datasets is inherent in the nature of data curation. Not all data can be considered when developing an AI algorithm—certain kinds of data must be prioritized. Human programmers, who are inherently influenced by social and technical biases, decide the scope of each AI project and sideline information they determine is not relevant.[4] In this way, data bias is unavoidable. However, if these biases are not identified before the data enters an AI algorithm, then it will be difficult to untangle how biases affected the algorithm's output, which can lead to inaccurate results. To this end, it is critical that the DoD determine what biases are present in data, how those biases will affect the algorithm, and how these effects might change the AI's outcome. These criteria can be met by accompanying every dataset with contextual information about its origin, as well as linking the creators of the dataset with others who want to use it.

To counter data challenges that would hinder DoD's achievement of its AI goals, the DoD should create a data clearinghouse that would enable it to standardize data practices, identify bias in crucial datasets, and encourage data sharing. Implementation of the data clearinghouse will help ensure the DoD is better positioned to streamline and collaborate on the collecting, vetting, and labeling of data, as well as improve upon AI writ large. It will also provide a pedigree for every included dataset, which will identify the origin of the data and the categories considered during the curation of the dataset. While the proposed clearinghouse is tailored to fit the needs of the DoD, it is relevant to all United States Government (USG) entities with a national security focus, particularly those in the intelligence and defense spheres. Indeed, the clearinghouse is applicable for data of all kinds, including geocoordinates, aerial imagery, and signals intelligence. Creating a data clearinghouse maximizes the DoD's ability to both maintain a strategic advantage over its adversaries and better safeguard against vulnerabilities.

Understanding AI

> *Essentially all models are wrong, but some are useful.*
>
> – George Box, 20th century statistician[5]

There is no universal definition of AI because the technology is integrated into a vast array of systems that perform a broad variety of tasks. However, the DoD provides a useful definition in its 2018 AI Strategy, which defines AI as the ability of machines to perform tasks that normally require human intelligence.[6] These abilities include recognizing patterns, learning from experience, drawing conclusions, making predictions, or taking action, whether digitally or as the smart software behind autonomous physical systems.[7] This definition is the foundation of this analysis, because it explains the key differences between AI and other algorithms, including the ability of AI to perform tasks that were formerly restricted to human intellect.

Artificial intelligence without data is like a car without gas: the framework for movement exists, but the vehicle is incapable of going anywhere without the proper inputs. Consequently, AI cannot produce useful results without a high-quality data source and significant human intervention during development. Data is the fuel that powers AI and teaches it how to operate, but AI is only as good as the data that trains it.[8]

What is the relationship between data and AI?

Data is integral to AI in three key ways.[9] First, computer scientists use *training* data to teach an AI algorithm historical trends and compare the AI's predictions with reality. Second, once scientists are confident in an algorithm's predictions, they provide *input* data, or data "from the wild," that the AI will use to make its first non-training predictions. Finally, scientists collect *feedback* data that notes the discrepancies between the AI's predictions and reality so that developers can gauge the algorithm's effectiveness.[10] The quality of the data selected determines the applicability and generalizability of an AI algorithm, so it is crucial that the data's accuracy is evaluated before the algorithm is trained. Unnecessary bias can creep into the datasets at any stage, but it is particularly dangerous when there is bias within training data, as biased training data leads to faulty algorithm formation and subsequently, unnecessarily biased input and feedback data.

What can AI do?

Artificial intelligence can be divided into two categories: narrow AI, which is designed for specific tasks, and general AI, which aims to mimic human-like decision-making. An example of narrow AI is an algorithm that recognizes stop signs using AI-based object detection, while a more general AI could drive a car using a large number of AI-based algorithms and complex data to emulate general human intelligence.[11] Currently, AI tends to succeed only when given a narrow task with discrete goals, as opposed to general tasks that are trying to imitate human intuition. The following example outlines the current state of "intelligence" in AI:

> When you ask Amazon's Alexa to reserve you a table at a restaurant you name, its voice recognition system, made very accurate by machine learning, saves you the time of entering a request in Open Table's reservation system. But Alexa doesn't know what a restaurant is or what eating is. If you asked it to book you a table for two at 6 p.m. at the Mayo Clinic, it would try.[12]

While many in the public and private sectors imagine AI to be a catch-all solution to the DoD's future tactical and strategic challenges, this vision is not reflected in AI's current capabilities.[13] There is an order-of-magnitude difference between the two types of AI, because general AI must be able to replicate human contextual knowledge, common sense, and intuition. As a result, many nascent high-profile AI technologies like self-driving cars are significantly more difficult to develop than narrow AI functions.

How does the DoD use AI?

AI has three primary uses in the military context: enterprise efforts, mission-support tasks, and operational endeavors.[14] Enterprise AI helps with personnel management systems, like coordinating healthcare for service members.[15] Mission-support AI automates supply chain logistics and predicts when systems or equipment will need maintenance.[16] Finally, the DoD is developing operational AI that will work in a variety of ways, including identifying high-value targets and anticipating terrorist attacks.[17] Since the DoD already uses AI for mission-support and enterprise functions successfully, this paper will focus on the DoD's operational AI — specifically, how data for these systems is collected, vetted, and labeled. Examples of operational AI being explored by the DoD include:

- *Battlespace Navigation.* The DoD is in the process of creating target-recognition AI to "go on everything from recon drones to tank gun sights to infantry goggles" that would aid in battlespace navigation.[18] This technology would operate like a commercially available navigation application with a viewfinder showing military vehicles, aircraft, and warships around the world in real-time.[19] Operational AI used for navigation is different from a commercially available GPS service, because it has the ability to simultaneously and comprehensively provide a cost-benefit analysis for troop movements based on dozens of factors, such as weather, time of day, and enemy habits.

- *Risk Management*. The DOD plans on using AI to enhance U.S. threat-assessment capabilities, thereby reducing risks to fielded forces. The DoD can use AI to predict the outcome of battles, which creates a U.S. advantage to countering attacks. Additionally, the Pentagon is developing AI to instantaneously identify anomalies in critical infrastructure, including U.S. and allied financial systems, electric grids, and election processes.[20] This detection AI would reduce the risk of catastrophic attacks against fielded forces and the homeland.

- *Decision-Making.* The ultimate aim of including AI in the decision-making process is to produce a U.S. advantage on the battlefield.[21] The DoD intends to use AI to create Multi-Domain C2, which would coordinate the "seamless exchange of information" between all four military services in all five battle domains.[22] While the DoD plans to use AI for both tactical and operational decisions in the future, they are currently focusing on using this technology at the tactical level to free up attention and manpower for strategic-level planning and execution.[23]

Although the DoD has many operational ambitions for AI, U.S. military leaders are committed to keeping humans involved in the decision-making process, particularly when choices about the use of lethal force are made.[24] In fact, the DoD is planning "as much machine-to-machine interaction as is possible to allow humans to be presented with various courses of actions for decision."[25] However, the goal of operational AI is undermined if the data that underpins AI development is not adequately collected, vetted, or labeled, as viable data is critical to the DoD achieving its AI goals.

Vulnerabilities of Bad Data

> *Our investigation has determined that no one—I repeat no one—knowingly targeted the Chinese Embassy… The unintended attack happened because a number of systems and procedures that are used to identify and verify potential targets did not work.*
>
> – George Tenet, Director of Central Intelligence, 1999[26]

Good data is the linchpin of good AI. If the DoD fails to prioritize data collection, vetting, and labeling, it will lead to unintended consequences. For instance, if DoD AI algorithms are trained on data that is insufficiently representative or unknowingly biased, it will become more likely to misidentify its operating environment. Failure to understand the operating environment fully can lead to the improper and unintended use of lethal force. Bad data may also leave USG personnel more vulnerable to attack if they are reliant upon an algorithm that fails to alert them to dangerous conditions. Further, if flawed data and algorithms are adopted on a wide scale, this misestimation could ultimately lead to a loss of U.S. strategic advantage over adversaries, such as Russia and China, who are making substantial investments in AI-based technologies. The consequences of misidentifying an operating environment can be divided into two categories of threats: misinformed decision-making and manipulated operations.

Misinformed Decision-Making

Bad or insufficient data can lead to misinformed decision-making, negatively impacting the DoD's AI outcomes. When data is incomplete or unnecessarily biased, an AI algorithm may struggle to reflect the operating environment, causing it to make predictions about the world that are inaccurate.[27] These inaccuracies become even more dangerous when the results of one set of AI algorithms become the foundation for another AI algorithm, thereby compounding errors in results. While not AI-specific, the following case studies detail the dangerous consequences of decisions made with incomplete and unknowingly biased data.

- *U.S. Bombing of the Chinese Embassy*. On May 7, 1999, the United States accidentally bombed the Chinese Embassy in Belgrade, Serbia. This tragic accident was the result of improper data collection, vetting, and labeling techniques. The intended target, the Yugoslav Federal Directorate for Supply and Procurement, was 300 meters from the Chinese Embassy. However, the USG misidentified the building because they were using outdated maps and inaccurate collection techniques.[28] When the USG ran the target's coordinates through military and intelligence databases for corroboration, the mistake went undetected because those databases listed the Chinese embassy under a different location.[29] This oversight led to the deaths of three embassy employees and the injury of 27 others.[30]

  In this case, bad data resulted in unintentional diplomatic repercussions. The bombing sparked international outcry and violent protests against the United States.[31] Because of the United States' reliance on bad data in this incident, the Chinese government cancelled human rights discussions, cooperation on nonproliferation, and even suspended berthing in Hong Kong.[32] In response to the attack and its political fallout, the USG promised to "strengthen internal mechanisms and procedures for selecting and validating targets," update critical databases, and de-conflict with other governments when possible.[33]

- *Curveball.* The intelligence community (IC)'s handling of biased information from an Iraqi defector (code-named Curveball) highlights the harmful impact unvetted data can have on decision-making. In the lead-up to the 2003 Iraq War, Curveball exploited his background as a chemical engineer to convincingly fabricate information about biological weaponry in Iraq.[34] Though his information was spurious, the Bundesnachrichtendienst—Germany's Federal Intelligence Service and Curveball's primary handler—did not communicate their concerns about the source's reliability to the USG's IC. The United States, lacking other sources on this topic, placed a premium on Curveball's information, and because the data was mischaracterized, many U.S. officials assumed the information was viable.[35] The United States ultimately used Curveball's reporting on WMDs (later contradicted by several sources) to justify the invasion of Iraq in 2003.[36]

  Although other political factors also contributed to the United States' decision to invade Iraq in 2003, Curveball represents the danger of making decisions based on unvetted data

> and highlights the consequences of building on flawed datasets. This incident spurred a series of reforms within the IC to standardize source vetting and language to communicate the confidence of intelligence. Standardized guidance, such as the Intelligence Community Directive 203, provides clear checklists to ensure that information quality, objectivity, and context are communicated in IC products.[37]

As detailed above, improper data procedures and biased data inputs can have deadly consequences. A standardized way to collect, vet, label, and caveat the data could have prevented the bombing of the Chinese Embassy and the U.S. response to Curveball's information. Just as humans make misinformed decisions with faulty data, so do AI algorithms. Therefore, it is crucial to dedicate the proper resources to ensuring the quality of data inputs *before* they enter an AI algorithm and begin making decisions.

Manipulated Operations

AI trained on unvetted data may cause the USG to lose control over its operations.[38] Recently, a group at Boston University created a back door into an AI system by "poisoning" just 0.025 percent of the training data.[39] Through this back door, the group was able to create a "sleeper agent" within the algorithm that would "misbehave in strange or harmful ways."[40]

The impact of a similar security breach on military AI would be enormous, and the cost could be devastating. For example, data used in an AI designed to detect enemy aircraft could be poisoned, preventing the system from identifying certain aircraft.[41] These poisoned training datasets might then be manipulated further to cause AI-enabled weapons to fire at allies rather than adversaries.[42] The reality of data poisoning underscores the importance of vetting data before feeding it into an AI algorithm.

Even if an AI system is not hacked during training, it can be manipulated post-deployment by using carefully crafted input data.[43] For instance, if an AI-facial recognition system is guarding a top-secret facility, someone could fool it into thinking that they are someone else by simply drawing a few dots on their face.[44] This input attack confuses the algorithm into giving a low-confidence reading or classifying a face as someone else entirely.[45] Using this technique, adversaries could hide in plain sight, and the DoD could lose its competitive edge and control of the situation entirely.

Current DoD Data Practices

> *The Navy collects more data in a day than all of the information in the Library of Congress.*
>
> – Scott Maucione, 2017[46]

Given the importance of data to AI success, it is critical to review DoD data collection, vetting, and labeling practices to ensure useful results. Due to the sensitive nature of most DoD projects, many of their data practices are classified. However, a few of its data processes are either publicly available or can be inferred through other government and private sector data practices.

Data Collection

The DoD obtains its data through two primary means: internal collection and external contracting.

- *Internal Collection:* The DoD uses its own sensors, connections, and optics to collect many types of data, including aerial imagery. Aerial imaging is one of the most widely discussed sources of DoD-generated data, due to the publicity surrounding the DoD's Project Maven. In Project Maven, tens of thousands of hours of drone footage are collected with the intention of training an algorithm to scan the footage for objects of interest, which results in a decreased workload for hundreds of military analysts.[47]

  Other algorithms are trained using data collected from "mobile device signals...software logs, microphones, radio-frequency identification (RFID) readers, and wireless sensor networks," to provide insight into a target's habits, plans, and intentions. [48] Defensively, these data sources can be used to automate mission-support and enterprise functions, and offensively, they help the DoD create targeting packages.

- *External Contracting.* Because it is difficult to anticipate what data the Department will need for a certain project until it begins, the DoD often purchases datasets from external vendors. [49] In some instances, data collection is outsourced to private sector groups on a contract basis.[50] Companies, such as MAXAR, sell pre-collected and labeled satellite imagery data that was crowdsourced and vetted by "citizen scientists," although the details of the group's methods for analyzing data are unclear.[51]

Data Vetting

The DoD's process for vetting datasets can be opaque, and differs substantively across the enterprise. While the Pentagon devotes substantial effort towards transparently securing the hardware that holds data, it does not publicly discuss how it ensures the veracity of acquired datasets.[52] Depending upon the means of collection or degree of data documentation, these uncertain processes can result in data that is difficult to verify, decipher, or corroborate. As a result, DoD personnel may leverage information without sufficient insight into its gaps and deficiencies. In several documented instances, the use of uncorroborated or insufficiently verified information resulted in Pentagon briefings that contained contradictory information on the same topics.[53]

It is not standard practice for other USG departments to provide background information on datasets when they send them to the DoD.[54] As a result, each new dataset requires that DoD departments "look for issues, investigate root causes, and implement improvements to ensure the data collected accurately reflected the real-world conditions." However, because compiling this information is a long, expensive process that requires significant manpower in a resource-constrained environment, many departments are unable to fully vet the datasets they leverage.[55]

Therefore, an analyst in the DoD may rely on a dataset from another department without knowing that the data is three years old, or that it came from a department with limitations in its tracking technology.[56]

Data Labeling

Labeling data is a tedious and labor-intensive task.[57] Because AI-centric processes are so new, labeling is done primarily by hand, especially when creating a training dataset from which algorithms learn. Hand-labeling training data is expensive and often requires "months or years to assemble, clean, and debug—especially when domain expertise is required."[58] At the point that labeling is complete, the data might be irrelevant or outdated.

To mitigate the challenges of data labeling, the DoD occasionally outsources this work to gig-economy workers. Gig-economy employees often lack the requisite training or insight necessary to do this work, and as they may make as little as $1 per hour labeling data, they are less incentivized to ensure labeling consistency. [59] Recent public sector examples further underscore the impact that this work can have on non-professionals.[60] These working conditions lead to reduced productivity and accuracy in the short-term, and lawsuits in the long-term.[61] The DoD could experience these repercussions if it exports graphic or disturbing data to gig economy workers for labeling.

When the DoD keeps the work in-house, data cleaning is often done by employees without previous experience or expertise in data management. During Project Maven, many offices were incentivized to have employees classify drone images in their spare time with little instruction.[62] A lack of expertise resulted in decreased accuracy and inconsistent data. Much like their vetting processes, the DoD's data labeling practices are unstandardized and therefore difficult to evaluate.[63]

Challenges to DoD Data Practices

> *Data-driven predictions can succeed—and they can fail. It is when we deny our role in the process that the odds of failure rise. Before we demand more of our data, we need to demand more of ourselves.*
>
> — Nate Silver, 2012[64]

As the largest data producer and consumer in the USG, the DoD faces multiple challenges to its current data practices. Due to the complexity of the DoD's operations and its enormous workforce, communication about the Pentagon's data practices is difficult to facilitate. For instance, sometimes groups will search for curated datasets that seem likely to exist given USG interest, but in reality, there is no curated dataset on that particular topic. The following are other examples of challenges the DoD faces in its current data practices:

- *Lack of Pedigree.* Many of the datasets created or used by the DoD lack contextual information, because there is currently no standardized way for dataset creators to record the data's background.[65] An analyst might receive a dataset from another office and not know that the data has not been properly vetted, or that it was externally labeled without much care. Lack of information about a dataset's origins may lead to confusion about its contents or misuse that deviates from the curator's intent.

- *Massive Quantity*. Though the DoD obtains a dizzying amount of data from both internal and external sources, the sheer volume of data makes comprehensive analysis impractical. For example, the U.S. Navy's intelligence, surveillance, and reconnaissance functions collect over 2.5 million terabytes per day, but as little as five percent of that data is analyzed by the Navy personnel who work with it.[66] The DoD has begun to use AI to comb through this enormous amount of data, identifying trends and priorities, but this technology is still in its nascent stages.[67]

- *Collection Redundancy.* One of the biggest data challenges facing the DoD is that there is little clear enterprise-wide communication about what data is being collected. Case studies indicate that two or more DoD elements may curate or acquire collections of the same data on the same topic, leading to data duplication and unnecessary government procurement.[68] Such redundancy, or a fear of redundancy, can lead to blind spots in USG preparedness and research.

- *Insufficient Funding.* Collecting, vetting, and labeling data is expensive and labor-intensive. As a result, data-driven AI projects tend to go over budget. In 2017, the Air Force canceled a contract to convert "raw data into actionable information that is used to direct battlefield activities" after the anticipated $374 million price tag surged to $745 million.[69]

- *Poor Interoperability.*[70] Even if one office has a clean dataset, it may not be stored in the correct format to integrate with other datasets, due to the high number of specialized formats used within the DoD.[71] In other cases, one office might code the same topic differently, making it difficult to merge datasets.[72] Even when data is in the correct format, bureaucratic competition between groups may make them reluctant to share data that they have spent significant amounts of time cleaning, labeling, and vetting. In this sense, data sharing can be a zero-sum game, even though lack of collaboration slows down critical AI development.

Data Bias in AI Development

One of the most pressing issues facing the DoD's data is data bias. Some degree of socially constructed and technical bias is an inherent part of data collection, vetting, and labeling. However, the world of AI development is also rife with unnecessary data bias. This unnecessary bias can be found in the way that many scientists train AI.[73] For example, a common training practice when developing an algorithm is to randomly split a dataset, using half for training AI and the other half to test that the AI has learned the correct patterns. This means that the data used to test the performance of a model has the same biases as the data that is used to train it.[74] Practices similar to this one can impede an algorithm's effectiveness and accuracy.

Unnecessary bias can creep into the data that powers AI and negatively impact results. When MIT Media Lab scientists tested the accuracy of facial recognition algorithms, they discovered that most facial recognition software had a 99 percent reliability rating for lighter-skinned male's faces, but had only a 35 percent reliability rating for darker-skinned female's faces.[75] The scientists later discovered that this gap in reliability was the result of biased data that included more white men's faces than black women's faces in the AI training sets. This problem was so ubiquitous that Chinese companies began contracting with African governments to increase the accuracy of their AI on black faces.[76] This example shows how humans introduce unnecessary bias into datasets to align with conscious and subconscious attitudes.

Compounding these data biases is the difficulty surrounding "explainable AI."[77] Because deep-learning algorithms are difficult to scrutinize, their processes tend to be regarded as a "black box," a model that provides novel insight, but is difficult to explain and understand.[78] Explainable AI is not yet a reality, so the biases can influence the AI and the algorithm's creators would be none the wiser.[79] While investigating bias in algorithm design is beyond the scope of this analysis, it is worth noting that biased data exacerbates flaws in AI models, leading to even more flawed outputs. Furthermore, the data cycle transfers these biases into other AI models. Ultimately, failure to address the original biases in the practices listed above may lead to the misidentification of an operating environment and disastrous political, diplomatic, and in some cases, even lethal consequences.

Policy Solution: A Data Clearinghouse

Transparent collaboration is key to countering the many challenges the DoD faces when it comes to the data that drives AI development. The DoD's ability to maintain a strategic advantage over its adversaries hinges on data synchronization efforts to counter bias, misinformed decision-making, and loss of control.

The DoD should create a data clearinghouse to combat its data challenges, because it has both the means and the incentive to spearhead this project. The clearinghouse would allow for groups both internal and external to the DoD to submit requests for datasets. This database also would be used as a platform to advertise group ownership over specific datasets. In a sense, this clearinghouse would function much like an online retail platform, connecting those who collect data with those who create algorithms. Each dataset would act as a "profile" that would only be fully accessible to the requester if they had adequate clearances and need-to-know. This clearinghouse would not be a central repository for all DoD data because a comprehensive data bank would a security risk if an adversary gained access to it.

The proposed data clearinghouse is already in line with the military's AI goals. Last year, Air Force Lt. General John Shanahan, the director of the Joint Artificial Intelligence Center (JAIC), stated that a chief priority of the JAIC was to "synchronize DoD AI activities" and the "related AI and machine-learning projects [that] are ongoing across the department."[80] A data clearinghouse is a productive step towards synchronizing DoD AI activities; it clarifies who already has data on a particular topic, who owns it, where it is located, and the condition of that data.[81]

U.S. adversaries are already working to centralize their data. A draft of the 2019 Russian national AI strategy called for the creation of "online repositories to collect, store, and process scientific data, including training for AI algorithms."[82] The 2017 Chinese New Generation Artificial Intelligence Development Plan argues that the future of AI development is in part contingent on innovating ways to collect, store, and share data.[83] This clearinghouse model answers the USG's call for collaboration while combating the relative AI gains by U.S. adversaries.

If the DoD creates (and most importantly adequately funds) a data clearinghouse, the benefits would range from reduced costs and man-hours to increased innovation and communication about data and AI. This project would not only make the DoD's AI R&D more efficient and less expensive, it could prevent future lethal mistakes.

Who would staff the data clearinghouse?

The clearinghouse would be maintained by DoD data scientists and analysts. These individuals would serve as "research librarians" with up-to-date knowledge of what datasets exist on a given topic, the quality of those datasets, and other pertinent information from the originator. Data and computer scientists would be necessary to manage the logistics and technical aspects of the clearinghouse. However, because there is not a high level of tacit knowledge required to facilitate basic data communications and review submission forms, approximately a dozen entry-level analysts or interns could manage day-to-day operations.

As with other collaborative USG data projects, the data clearinghouse staff would be in frequent communication with all relevant parties within the USG. In addition, joint-duty positions would be added to institutionalize interagency cooperation.[84] There could also be designated seats for agencies outside of the DoD that use DoD data, such as the Office of the Director of National Intelligence (ODNI), CIA, DIA, and other relevant entities.

How would a data clearinghouse counter the DoD's current data challenges?

The creation of a data clearinghouse would help address the DoD's data challenges. The benefits of implementation include:

- *Standardized Pedigree.* Using this clearinghouse, data profiles can be posted with a data "pedigree" that indicates the data origin, age, method of collection, vetting, labeling, and the level of confidence the originator has in the dataset. A better understanding of the originator's intentions will transform datasets from indeterminate files into functional resources upon which to build AI algorithms. The pedigree form (see Appendix A) complies with the Intelligence Community Directive 203, an outline of standards for the analytic products derived from intelligence reports.[85] This directive is a good template for the data pedigree, because it was designed to solve many of the same issues that the DoD faces with data and AI.

- *Shared Technology.* The clearinghouse would help to manage the DoD's massive amount of data by fostering a collaborative environment in which advancements in data-cleaning AI would be disseminated quickly. The clearinghouse would also facilitate communication between different groups that are interested in the same data, thereby sharing the burden of Big Data.

- *Mitigated Redundancy.* As inter-DoD communication about data becomes more consistent, redundant data efforts will be less likely. Increased communication about whether another group is considering collecting data on a certain topic would both prevent duplicative efforts and illuminate data gaps.[86]

- *Cost-Effectiveness.* As the current lack of inter-governmental communication creates more work for employees, this clearinghouse would streamline USG data efforts, saving manpower and money. This increase in collaboration would reduce funding concerns in AI R&D, because it could allow groups to divvy up data processes and learn whether the data they want has already been collected, vetted, or labeled.

- *Improved Interoperability.* The clearinghouse would maintain a "Frequently Asked Questions" forum that could field basic interoperability issues, creating a base of institutional knowledge and increasing fundamental communication between offices. Further, data scientists at the clearinghouse who specialize in data interoperability would be available for online consultation. The employees could also compile a style guide for the most commonly divergent data labels (such as country names or terms used by foreign militaries).

- *Regulated Bias.* While a certain degree of human bias is unavoidable, the data clearinghouse facilitates conversations about methods to tackle data projects with as much objectivity as possible—and identifies what kinds of bias could have devastating consequences for DoD missions. Further, the data pedigree communicates potential biases those collecting and organizing the data may have (see Appendix A).

What is the precedent for a collaborative USG data project?

While the idea of creating and maintaining a data clearinghouse for the USG may seem ambitious, there is precedent for such a collaborative data project. The following examples demonstrate the historical basis for a comprehensive platform to exchange data within the USG.

- *MIDB & MARS*. Modernized Integrated Database (MIDB), a federal repository for foundational military intelligence maintained by the DIA.[87] Created in 1998, it acted as a central database for information on foreign militaries, including the location of airfields, military units, and facilities, as well as nuclear planning execution data.[88] Currently, the MIDB is being retired in favor of its successor, Machine-Assisted Analysis Rapid-

Repository System (MARS), which aims to integrate AI into the data collection process.[89] In fact, making MARS operational is one of the Director of the DIA's top priorities during his tenure.[90] MIDB and MARS illustrate the success of sharing data across defense spaces to compile a single and constantly evolving database.

- *NCTC.* The National Counterterrorism Center (NCTC) provides a successful model of inter-USG collaboration on a single issue area. The NCTC is an inter-USG office under the purview of the ODNI that analyzes and responds to terrorist threats.[91] The NCTC is "staffed by over 1,000 employees from across the IC, the Federal government, and Federal contractors" who represent "approximately 20 different departments and agencies."[92] Further, the NCTC addresses terrorism against the United States through a holistic approach with an emphasis on sharing data sources and information to counter the threat of terrorism. The proposed data clearinghouse would be a fraction of the size and cost of the NCTC, while servicing almost as many organizations.

What are the risks associated with a data clearinghouse?

- *Security Concerns.* In light of recent security breaches—such as the Chinese hack of the Office of Personnel Management—the DoD might be wary of creating a centralized forum of its data. If an adversary accessed the clearinghouse, they would see what data the DoD is using, exploit the data's gaps (or poison it to misrepresent reality), and render the AI useless and potentially harmful.[93]

  To counter this threat, the clearinghouse would use existing public-key infrastructure (PKI) and access-control systems to ensure that only those with sufficient clearances could access others' data. The data clearinghouse and the office that owns the dataset could flag all anomalous access requests. Dataset profiles would only be visible to those with the requisite clearances and compartmental access. Thus, a clearinghouse would not pose a higher risk to data leakages than the current risks to classified information routinely circulated around the USG.

- *Added Friction.* The creation of a data clearinghouse would add another level of bureaucracy to the already lengthy AI R&D process. However, with the right employee system in place, the data clearinghouse would not be burdensome to those who submit and request data. The pedigree form takes just a few minutes to complete (see Appendix A) and submitting and fielding requests for datasets would be just as timely. In addition, the clearinghouse will reduce overall friction once DoD employees are trained on its use, which will save substantial amounts of time and energy in the future.

- *Ownership Issues.* Some groups may be reluctant to share the data they created using their own funds and man-hours, as it is likely that some offices will primarily consume data as opposed to producing it. However, offices could use the number of times their datasets were accessed in the clearinghouse as a metric of the effectiveness of their products and the broader DoD and USG interest in a particular topic.

Another data ownership concern is dataset maintenance. If datasets are not routinely updated, it might be difficult to access them, even if their request is approved. It is possible that no group will claim responsibility for maintaining a particular dataset, or that the person who oversees a dataset will retire without passing off the responsibility. The data pedigree can mitigate the maintenance issue by identifying the office that originated the data, a point of contact, and the current caretaking office and contact information.

Conclusion

The DoD has the opportunity to improve its AI technologies by creating a data clearinghouse to streamline its data collecting, labeling, and vetting processes. The DoD can also better identify data bias and blind spots by making the data pedigree a standard practice. In doing so, the DoD can circumvent the pitfalls of bad data, such as manipulated functions and misinformed decision-making, that can lead to misunderstanding the operating environment, diplomatic mishaps, and, in extreme cases, accidental deaths.

It is critical that the DoD address these data challenges in the nascent stages of operational-AI development so that it can retain the strategic advantage over its adversaries, achieve its goals overseas, and ultimately keep Americans safe at home and abroad.

Acknowledgements

The author thanks Professor Amy Oakes, Professor Dennis Smith, Robert DeSchazor, M. Nina Miller, Colonel Steven Fino, USAF, Selene Swanson, and the rest of the 2019-2020 PIPS cohort of fellows and interns for their extensive feedback and helpful suggestions. The author also thanks Marcus Comiter, Professor Ehsan Elhamifar, Professor Nir Eisikovits, Margaret Dene, Marie Murphy, Major General Mark T. Matthews, USAF (Ret.), Matthew Bondy, and the U.S. Army Mad Scientist Team for their input and assistance, which dramatically improved this paper. Special thanks to Caroline Criado Perez, whose book *Invisible Women: Data Bias in a World Designed for Men* first introduced me to the idea that the negative effects of data bias are compounded when bad data is used to train artificial intelligence.

Appendix A

DATA PEDIGREE (ICD 203-COMPLIANT)

A DATASET TITLE:			
B DATE OF INFORMATION	START: END:		
C ORIGINATING OFFICE(S):		D ORIGINATING EMPLOYEE(S):	
E CARETAKING OFFICE(S) (if different than Box C)		F CARETAKING EMPLOYEE(S) (if different than Box D)	
G FILE FORMAT:		H FILE SIZE:	
I DATASET TYPE: OTHER:	□ Data Sample □ Static Dataset □ Continually-Updated Dataset	J INDEPENDENT OF POLITICAL CONSIDERATION? □ YES □ NO IF NO, EXPLAIN:	
K DATA SOURCE(S):			
L CONFIDENCE LEVEL IN:	1 DATA VERACITY?	2 DATA OBJECTIVITY?	3 DATA QUALITY?
	4 DATA’S REPRESENTATION OF REALITY?		
M WHAT FACTORS OR DATA FEATURES WERE CONSIDERED **AND INCLUDED** IN THIS DATASET?			
N WHAT FACTORS OR DATA FEATURES WERE CONSIDERED **BUT NOT INCLUDED** IN THIS DATASET?			
O WHAT WAS THE INTENDED OBJECTIVE OF DATASET?			
P DID THE DATASET MEET THE OBJECTIVE LISTED IN BOX O?			
Q HOW COULD THIS DATASET BE IMPROVED?			
R ADDITIONAL COMMENTS:			

A. List the current title and any previous titles of the dataset.

B. Record the date and/or date range of the data collected (e.g. a dataset with information that originated during the entirety of the Eisenhower presidency would span from January 20, 1953 to January 20, 1961). Do not record the date that the dataset was conceptualized or finalized.

C. Record the office(s) that conceptualized, collected, processed, and/or finalized the dataset. Please record if more than one office was involved in originating the dataset.

D. Record the specific employee(s) who spearheaded the conceptualization, collection, processing, and/or finalization for the dataset within the office recorded in Box C.

E. Record the office(s) that currently are in charge of the dataset's maintenance *if different* from the office(s) recorded in Box C.

F. Record the specific employee(s) who are in charge of the dataset's maintenance within the office recorded in Box E *if different* from the employee(s) listed in Box D.

G. List the dataset's file format (e.g. JPEG, PDF, CSV, etc.)

H. List the current size of the dataset's file at the time of filling out this form.

I. Record the intended function of this dataset. Mark "Data Sample" if the data was collected to provide a representative sample of a topic, but does not contain every data point on that topic between the dates listed in Box B. Mark "Static Dataset" if the dataset represents all available data from the time period listed in Box B, but is no longer being updated. Mark "Continually-Updated Dataset" if the dataset represents all available data from the start date listed in Box B to present. If the data was collected for a purpose not listed above, please record the intended use of the dataset in "OTHER".

J. In keeping in compliance with Intelligence Community Directive 203, please indicate whether or not the dataset was distorted by or shaped for the advocacy of a particular audience, agenda, or political viewpoint.

K. List the methods of collection (e.g. sensors, aerial imagery, etc.) used to create the dataset. Be as specific as possible.

L. Per the ICD-203, indicate the level of confidence in questions L1-L4 based on the following table:

almost no confidence	very low confidence	low confidence	roughly even confidence	high confidence	very high confidence	almost complete confidence
01-05%	05-20%	20-45%	45-55%	55-80%	80-95%	95-99%

1. Indicate the level of confidence had in the data's ability to represent the truth and to do so accurately, with "almost complete confidence" indicating the high degree of confidence had in the overall truthfulness of the data and "almost no confidence" indicating the near-complete lack in confidence that the data is truthful.
2. Indicate the level of confidence had in the data's impartial and holistic representation of whatever it is representing, with "almost complete confidence" indicating the high degree of confidence had in the overall objectivity of the data and "almost no confidence" indicating the near-complete lack in confidence that the data is objective.
3.Indicate the level of confidence had in the data's directness in representing what it is supposed to represent, with "almost complete confidence" indicating the high degree of confidence had in the overall quality of the data and "almost no confidence" indicating the near-complete lack in confidence that the data is directly answering what it is supposed to answer.
4. Indicate the level of confidence had in the data's ability to represent reality, with "almost complete confidence" indicating the high degree of confidence had in how comprehensively the data represents reality and "almost no confidence" indicating the near-complete lack in confidence that the data is comprehensively representing reality.

M. List the data sources, categories, and approaches that were considered and selected to include when conceptualizing, collecting, processing, and finalizing the dataset.

N. List the data sources, categories, and approaches that were considered and <u>not</u> selected to include when conceptualizing, collecting, processing, and finalizing the dataset.

O. Record the initial question, objective, and/or gap in information that the dataset was designed to fill.

P. Record whether or not the dataset succeeded in answering the question or fulfilling the objective and/or gap in information that it was designed to fill.

Q. Record the ways in which this dataset could be improved if re-done in the future.

R. Record any other relevant contextual information about the dataset that has not already been indicated on this form.

[1] Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. *Prediction Machines: The Simple Economics of Artificial Intelligence*. Boston: Harvard Business Review Press, 2018.

[2] Cornillie, Chris. "Finding Artificial Intelligence Money in the Fiscal 2020 Budget." Bloomberg Government, March 28, 2019. https://about.bgov.com/news/finding-artificial-intelligence-money-fiscal-2020-budget/.

[3] At its most basic level, data collection is "the process of gathering and measuring information from countless different sources." "Data Collection | DataRobot Artificial Intelligence Wiki." DataRobot. Accessed February 14, 2020. https://www.datarobot.com/wiki/data-collection/.

The DoD provides no singular definition for "data vetting" but alludes to it frequently in its documentation. According to Executive Order 13467," 'Vetting' is the process by which covered individuals undergo investigation, evaluation, and adjudication of whether they are, and remain over time, suitable or fit for Federal employment." Co-opting this term, "data vetting" thus means the processes by which data goes under investigation, evaluation, and adjudication of whether they are, and remain over time, suitable or fit for use by the USG.

("Amending the Civil Service Rules, Executive Order 13488, and Executive Order 13467 To Modernize the Executive Branch-Wide Governance Structure and Processes for Security Clearances, Suitability and Fitness for Employment, and Credentialing, and Related Matters." Federal Register, January 23, 2017. https://www.federalregister.gov/documents/2017/01/23/2017-01623/amending-the-civil-service-rules-executive-order-13488-and-executive-order-13467-to-modernize-the.)

Finally, Richard Kuzma defines the process of labeling data in his mid-2018 War on the Rocks piece when he describes it as "an image is labeled by a human, with the "ground truth" that the machine "learns" to recognize so when it sees similar data it can predict correctly." In this way, labeled data is data that is categorized in such a way that the AI algorithm can learn what it is looking for. (Kuzma, Richard. "But First, Infrastructure: Creating the Conditions for Artificial Intelligence to Thrive in the Pentagon." War on the Rocks, July 13, 2018. https://warontherocks.com/2018/07/but-first-infrastructure-creating-the-conditions-for-artificial-intelligence-to-thrive-in-the-pentagon/.)

[4] Broussard, Meredith. *Artificial Unintelligence: How Computers Misunderstand the World.* Cambridge: MIT Press, 2019, 18. Social bias means that one person might consider certain factors to be more relevant than others because of their social conditioning. For example, a nutritionist with a situationally depressed patient may attribute the depression to the patient's dietary habits and their gut microbiome, whereas a somnologist might ascribe the depression to extreme sleep deprivation. Because of the doctors' respective trainings, they select different data to test their hypotheses and generate an outcome. Technical bias has to do with the structural development that is used to answer a question, as well as the avoidance or inclination to collect and test certain data because it is particularly difficult or easy to access. For instance, a meteorologist in Oklahoma who is looking to predict severe weather in the United States is more likely to have first-hand access to data about tornados than hurricanes because of his location. Therefore, his severe weather predictions might be highly accurate about tornados, but less accurate about hurricanes because of technical factors.

[5] Wasserstein, Ron. "George Box: A Model Statistician," Significance 7, 7, no. 3 (2010): 134–35. https://doi.org/10.1111/j.1740-9713.2010.00442.x.

[6] "Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance Our Security and Prosperity," Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance Our Security and Prosperity § (2019), https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF According to the DoD AI Strategy, "AI refers to the ability of machines to perform tasks that normally require human intelligence – for example, recognizing patterns, learning from experience, drawing conclusions, making predictions, or taking action – whether digitally or as the smart software behind autonomous physical systems."

The complexity and length of another commonly cited definition underscores the confusion around AI. Per the H.R. 5515 – John S. McCain National Defense Authorization Act for Fiscal Year 2019, U.S. Congress defines AI as "(1) Any artificial system that performs tasks under varying and unpredictable circumstances without significant human oversight, or that can learn from experience and improve performance when exposed to datasets.(2) An artificial system developed in computer software, physical hardware, or other context that solves tasks requiring human-like perception, cognition, planning, learning, communication, or physical action. (3) An artificial system designed to think or act like a human, including cognitive architectures and neural networks. (4) A set of techniques, including machine learning, that is designed to approximate a cognitive task. (5) An artificial system designed to act rationally, including an intelligent software agent or embodied robot that achieves goals using perception, planning, reasoning, learning, communicating, decision making, and acting." (John S. McCain National Defense Authorization Act for Fiscal Year

2019." Congress.Gov, 2018. https://www.congress.gov/bill/115th-congress/house-bill/5515/text.) This definition, while used as the basis for USG conversations around AI, is broad and thus difficult to communicate in an articulate way. In a sense, part of the mystique and confusion around AI is embodied this paragraph-long definition. That said, for much of this paper, the words "artificial intelligence" will imply *machine-learning* AI because this type of AI is the bulk of the DoD discussions about AI. At its core, machine-learning AI is a series of predictive models that are trained on a specific dataset to provide forecasts and estimations about a particular subject. Hypothetically, the DoD could create a "parts forecasting" AI algorithm that predicts when a certain part of an F-22 will break based on the historical inputs of and how long that part lasted on previous F-22s. What distinguishes this algorithm from a non-AI algorithm is its ability to search for and highlight broader trends that humans may not be considering when assessing the air frames. The F-22 AI algorithm could identify that the plane's parts were breaking faster when the F-22s were deployed to Okinawa, Japan because the humidity levels were spurring more rusting than at their home base in Hampton, Virginia. While analysts are capable of inferring the root cause of the problem through a time and money-intensive investigation, an AI algorithm can calculate the impact of humidity levels, geography, and parts replacement on an F-22 simultaneously and instantaneously.

[7] Ibid.

[8]"The United States Air Force Artificial Intelligence Annex to The Department of Defense Artificial Intelligence Strategy." PDF. United States Air Force, 2019. https://www.af.mil/Portals/1/documents/5/USAF-AI-Annex-to-DoD-AI-Strategy.pdf. "Summary of the 2018 of Defense Artificial Intelligence Strategy." PDF. Department of Defense, February 12, 2019. https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF.

[9] Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. *Prediction Machines: The Simple Economics of Artificial Intelligence*. Boston: Harvard Business Review Press, 2018.

[10] To better illustrate the ways data are used, consider the following example. The DoD might want to create an AI algorithm that takes imagery from a JSTAR to predict where a certain terrorist group will next plant land mines in Iraq. In this case, a t*raining* dataset would be created to include previous locations of land mines in the region and whatever other factors the algorithm designers deem relevant (type of land mine, proximity to major arteries of travel, day of the week, frequency of attacks, etc.). Once the algorithm is trained, they would provide the AI with aerial imagery *inputs* from the last 24 hours and ask it to predict where the next land mines will be placed, or the likelihood that a land mine will be placed on a certain street at a certain time. Finally, *feedback* data would be collected to determine the accuracy of the algorithm in the wild.

[11] To further demonstrate the intuitive nature of general AI, consider the following example: a human could enter a stranger's house and quickly figure out how to make a cup of coffee, but this task would be nearly impossible for a general AI algorithm because there are so many learned and habitual experiences involved that are difficult to program. Fast Company. "Wozniak: Could a Computer Make a Cup of Coffee?" YouTube. March 2, 2010. Video, 2:16 https://www.youtube.com/watch?v=MowergwQR5Y

[12] Bergstein, Brian. "The Great AI Paradox." MIT Technology Review, December 15, 2017. https://www.technologyreview.com/s/609318/the-great-ai-paradox/.

[13] Marcus Comiter, interviewed by author, Harvard University, October 14, 2019. Dr. Ehsan Elhamifar, interviewed by author, Northeastern University, October 11, 2019, Dr. Nir Eisikovits, telephone conversation with author, UMass Boston, October 18, 2019.

[14] Tarraf, Danielle C., William Shelton, Edward Parker, Brien Alkire, Diana Gehlhaus Carew, Justin Grana, Alexis Levedahl, et al. *The Department of Defense Posture for Artificial Intelligence: Assessment and Recommendations*. Santa Monica, CA: RAND Corporation, 2019. https://www.rand.org/pubs/research_reports/RR4229.html.

[15] Ibid., 25. The DoD is developing an AI algorithm to streamline medical records so doctors can access healthcare information on the battlefield to more efficiently treat soldiers. Lye, Harry. "From Wingmen to Healthcare: How the DoD Plans to Implement AI." Army Technology. Verdict Media, 2019. https://www.army-technology.com/features/from-wingmen-to-healthcare-how-the-dod-plans-to-implement-ai/.

[16] For example, the Pentagon has created algorithms that "simplify workflows and improves the speed and accuracy of repetitive tasks" such as creating daily shipment schedules grafted onto Google Earth and automating command-to-ship messaging about supplies and personnel. Brown, Gerald G., Walter C. Degrange, Wilson L. Price, and Anton A. Rowe. "Scheduling Combat Logistics Force Replenishments at Sea for the US Navy." *Naval Research Logistics (NRL)*64, no. 8 (March 7, 2018): 1–17. https://doi.org/10.1002/nav.21780.

[17] The DoD's Joint Artificial Intelligence Center has "the overarching goal of accelerating the delivery of AI-enabled capabilities, scaling the Department-wide impact of AI, and synchronizing DoD AI activities to expand Joint Force advantages." Shanahan, Patrick. "Establishment of the Joint Artificial Intelligence Center." PDF. United State Department of Defense, June 27, 2018.

https://admin.govexec.com/media/establishment_of_the_joint_artificial_intelligence_center_osd008412-18_r....pdf.

[18] "Artificial Intelligence: The Frontline of a New Age in Defense." PDF. Breaking Defense. Accessed February 5, 2020. https://cdn2.hubspot.net/hubfs/2097098/MCM120_BreakingDefense_AI_ebookR1%20(1).pdf.

[19] Freedberg, Sydney J. "Air Force ABMS: One Architecture To Rule Them All?" Breaking Defense. Breaking Media, November 8, 2019. https://breakingdefense.com/2019/11/air-force-abms-one-architecture-to-rule-them-all/.

[20]"Summary of the 2018 of Defense Artificial Intelligence Strategy." PDF. Department of Defense, February 12, 2019. https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DoD-AI-STRATEGY.PDF.

[21] Ibid.

[22] Hitchens, Theresa. "EXCLUSIVE Navy, Air Force Chiefs Agree To Work On All Domain C2." Breaking Defense. Breaking Media, November 12, 2019. https://breakingdefense.com/2019/11/exclusive-navy-air-force-chiefs-agree-to-work-on-all-domain-c2/.

[23] "Artificial Intelligence and National Security." PDF. Congressional Research Service, November 21, 2019. https://fas.org/sgp/crs/natsec/R45178.pdf.

[24] Mulrine Grobe, Anna. "As AI Joins Battlefield, Pentagon Seeks Ethicist." Christian Science Monitor, October 28, 2019. https://www.csmonitor.com/Technology/2019/1028/As-AI-joins-battlefield-Pentagon-seeks-ethicist.

[25] Groll, Elias. "The Pentagon's AI Chief Prepares for Battle." Wired, December 18, 2019. https://www.wired.com/story/pentagon-ai-chief-prepares-for-battle/.

[26] "DCI Statement on the Belgrade Chinese Embassy Bombing." CIA, June 20, 2008. https://www.cia.gov/news-information/speeches-testimony/1999/dci_speech_072299.html.

[27] Martinez, Dave, Nick Malyska, Bill Streilein, and et al. "Artificial Intelligence: Short History, Present, and Future Outlook." PDF. Massachusetts Institute of Technology. Massachusetts Institute of Technology., 2019. https://www.ll.mit.edu/sites/default/files/publication/doc/2019-09/Artificial%20Intelligence%20Short%20History%2C%20Present%20Developments%2C%20and%20Future%20Outlook%20-%20Final%20Report%20-%20Martinez.pdf.

Algorithms have difficulty adequately reflecting the real-world situations for which they are often trained. For example, a research paper from 2018 described a group of over 50 AI researchers who "recounted dozens of times when AI systems showed surprising behavior. An algorithm learning to walk in a simulated environment discovered it could move fastest by repeatedly falling over. A Tetris-playing bot learned to pause the game before the last brick fell, so that it would never lose. One program deleted the files containing the answers against which it was being evaluated, causing it to be awarded a perfect score." This quote characterizes some of (what appear to be) erratic conclusions an AI algorithm can draw when given foggy objectives. Scharre, Paul. "Killer Apps: The Real Dangers of an AI Arms Race." Foreign Affairs, 2019. https://www.foreignaffairs.com/articles/2019-04-16/killer-apps.

[28] "DCI Statement on the Belgrade Chinese Embassy Bombing." CIA, June 20, 2008. https://www.cia.gov/news-information/speeches-testimony/1999/dci_speech_072299.html. IC officers "used land navigation techniques taught by the U.S. military to locate distant or inaccessible points or objects". However, these techniques were not supposed to be used for aerial targeting because they provide only an approximate location, and when the USG is targeting in urban areas, "it is important to provide an accurate appreciation of [the USG's] confidence in the location of a target."

[29] Ibid. The Chinese Embassy was on the off-limits targeting list; but the old address was recorded.

[30] Dumbaugh, Kerry. "Chinese Embassy Bombing in Belgrade: Compensation Issues." Congressional Research Service. Accessed February 28, 2020. http://congressionalresearch.com/RS20547/document.php.

[31] Ibid.

[32] Ibid.

[33] "DCI Statement on the Belgrade Chinese Embassy Bombing." CIA, June 20, 2008. https://www.cia.gov/news-information/speeches-testimony/1999/dci_speech_072299.html.

[34] Drogin, Bob, and John Goetz. "How U.S. Fell Under the Spell of 'Curveball.'" *Los Angeles Times*. November 20, 2005. https://www.latimes.com/world/middleeast/la-na-curveball20nov20-story.html.

[35] Ibid. Translation difficulties further complicated the situation. The BND sent German summaries of their English and Arabic interviews with Curveball to the DIA who then translated them back into English and rewrote their own summaries. As opposed of first-hand exposure, the source that at best had secondhand access to information about Iraqi capabilities (but not their actual programs.) Additionally, when U.S. intelligence agencies began to corroborate Curveball's information, they fell into the trap of circular reporting. At one point, the Central Intelligence Agency (CIA) claimed to have three sources that corroborated Curveball's data, but at least two of these sources had ties to the man who officials now suspect coached Curveball on what to tell the BND. In the end, all three sources were discredited.

[36] Ibid. The dissonance of doubts about Curveball was not clearly communicated to the highest level U.S. officials, who believed the CIA and DIA's reporting was coming from multiple, vetted sources. Colin Powell, then the U.S.

Secretary of State, claimed that he prepared a speech to the U.N. in early 2003 without knowledge that there was any debate over Curveball's credibility.

[37] Clapper, James. "ICD 203: Analytic Standards." PDF. Office of the Director of National Intelligence, January 2, 2015. https://www.dni.gov/files/documents/ICD/ICD%20203%20Analytic%20Standards.pdf.

[38] Schmidt, Eric, Robert O. Work, and et al. "NSCAI Interim Report for Congress." PDF. National Security Commission on Artificial Intelligence., November 2019. https://drive.google.com/file/d/153OrxnuGEjsUvlxWsFYauslwNeCEkvUb/view.

[39] Kiourti, Panagiota, Kacper Wardega, Susmit Jha, and Wenchao Li. "TrojDRL: Trojan Attacks on Deep Reinforcement Learning Agents," March 1, 2019, 1–17. https://arxiv.org/abs/1903.06638.

[40] Knight, Will. "Tainted Data Can Teach Algorithms the Wrong Lessons." Wired. Conde Nast, November 25, 2019. https://www.wired.com/story/tainted-data-teach-algorithms-wrong-lessons/.

[41] Comiter, Marcus. "Attacking Artificial Intelligence." PDF. Belfer Center for Science and International Affairs, August 2019. https://www.belfercenter.org/sites/default/files/2019-08/AttackingAI/AttackingAI.pdf.

[42] Scharre, Paul. "Killer Apps: The Real Dangers of an AI Arms Race." Foreign Affairs, 2019. https://www.foreignaffairs.com/articles/2019-04-16/killer-apps.

[43] Ibid.

[44] Metz, Cade. "How To Fool AI Into Seeing Something That Isn't There." Wired. Conde Nast, July 29, 2016. https://www.wired.com/2016/07/fool-ai-seeing-something-isnt/.

[45] Kurakin, Alexey, Ian Goodfellow, and Samy Bengio. "Adversarial Examples in the Physical World," 2016, 1–15. https://arxiv.org/pdf/1607.02533v1.pdf.

[46] Maucione, Scott. "Navy Plan for CDO Has Been Stagnant for Almost a Year as Potential Benefits Pass By." Federal News Network. Hubbard Radio, October 27, 2017. https://federalnewsnetwork.com/defense/2017/10/navy-plan-for-cdo-has-been-stagnant-for-almost-a-year-as-potential-benefits-pass-by/.

[47] "What Is Project Maven? The Pentagon AI Project Google Employees Want out Of." Global News. Corus Entertainment, April 5, 2018. https://globalnews.ca/news/4125382/google-pentagon-ai-project-maven/.

[48] "DoD Digital Modernization Strategy." PDF. United States Department of Defense, June 5, 2019. https://media.defense.gov/2019/Jul/12/2002156622/-1/-1/1/DoD-DIGITAL-MODERNIZATION-STRATEGY-2019.PDF.

[49] Melendez, Carlos. "Data Is the Lifeblood of AI, but How Do You Collect It?" InfoWorld, IDG Communications, August 8, 2018, https://www.infoworld.com/article/3296044/data-is-the-lifeblood-of-ai-but-how-do-you-collect-it.html)

[50] Tucker, Patrick. "The Military Is Already Using Facebook to Track Your Mood." Defense One, July 2, 2014. https://www.defenseone.com/technology/2014/07/military-already-using-facebook-track-moods/87793/.

[51] Tirrell, Aaron. "The Rise of the HIVE: A New Era of Crowdsourced Imagery Analysis." Maxar, August 22, 2019. https://blog.maxar.com/earth-intelligence/2019/the-rise-of-the-hive-a-new-era-of-crowdsourced-imagery-analysis. Some sources indicate that the "citizen scientist" community members get paid when they successfully complete a task, in addition to receiving cool swag like t-shirts, mugs, lanyards, and DigitalGlobe imagery-based screen savers." Further, "GeoHIVE community members will receive virtual coaching from geospatial experts" when working on higher-stakes analytical missions, such as DoD projects. Frazier, Tony. "Curious about the Buzz over Crowdsourcing?" DigitalGlobe Blog. DigitalGlobe, October 1, 2015. http://blog.digitalglobe.com/crowd/curious-about-the-buzz-over-crowdsourcing/.

[52] Whaley, Richard. "The Big Data Battlefield," Military Embedded Systems, OpenSystems Media, Last modified February 2, 2020, http://mil-embedded.com/articles/the-big-data-battlefield/

[53] Tarraf, Danielle C., William Shelton, Edward Parker, Brien Alkire, Diana Gehlhaus Carew, Justin Grana, Alexis Levedahl, et al. *The Department of Defense Posture for Artificial Intelligence: Assessment and Recommendations*. Santa Monica, CA: RAND Corporation, 2019. https://www.rand.org/pubs/research_reports/RR4229.html.

[54] Tarraf, Danielle C., William Shelton, Edward Parker, Brien Alkire, Diana Gehlhaus Carew, Justin Grana, Alexis Levedahl, et al. *The Department of Defense Posture for Artificial Intelligence: Assessment and Recommendations*. Santa Monica, CA: RAND Corporation, 2019. https://www.rand.org/pubs/research_reports/RR4229.html., pg. 59.

[55] Ibid.

[56] The lack of standardized vetting procedures for data is not a problem unique to the DoD. Rather, this issue is pervasive across the USG writ large, as well as most of the private sector.

[57] For the purposes of this paper, data labeling means taking the data points and sorting them based on what they represent. For instance, those who aided in Project MAVEN's aerial imagery data labeling would look at an image,

and manually tell the machine what the image was ("this picture is of a red pick-up truck"). Now, the AI knows what a red truck looks like, and can make guesses as to whether or not proceeding images are or are not a red pick-up truck.

[58] Ratner, Alex, Paroma Varma, Braden Hancock, and Chris Re. "Weak Supervision: A New Programming Paradigm for Machine Learning." Stanford.Edu, March 10, 2019. http://ai.stanford.edu/blog/weak-supervision/.

[59] Fang, Lee. "Google Hired Gig Economy Workers to Improve Artificial Intelligence in Controversial Drone-Targeting Project." The Intercept, February 4, 2019. https://theintercept.com/2019/02/04/google-ai-project-maven-figure-eight/.

[60] Many gig-economy workers have developed PTSD from monitoring social media content. Murgia, Madhumita. "Facebook Content Moderators Required to Sign PTSD Forms." Financial Times, January 26, 2020. https://www.ft.com/content/98aad2f0-3ec9-11ea-a01a-bae547046735.

[61] Coffey, Lauren. "Fortune 500 Company in Tampa with Ties to Facebook Sued over Lack of Mental Health Support." Tampa Bay Business Journal, February 10, 2020. https://www.bizjournals.com/tampabay/news/2020/02/10/fortune-500-company-in-tampa-with-ties-to-facebook.html.

[62] Marie Murphy, conversation with author, Mad Scientist Initiative at U.S. Army TRADOC, October 2019.

[63] Tarraf, Danielle C., William Shelton, Edward Parker, Brien Alkire, Diana Gehlhaus Carew, Justin Grana, Alexis Levedahl, et al. *The Department of Defense Posture for Artificial Intelligence: Assessment and Recommendations*. Santa Monica, CA: RAND Corporation, 2019. https://www.rand.org/pubs/research_reports/RR4229.html.

[64] Silver, Nate. *The Signal and the Noise: Why So Many Predictions Fail--but Some Don't*. New York: Penguin Books, 2015.

[65] Freedberg, Sydney. "EXCLUSIVE Pentagon's AI Problem Is 'Dirty' Data: Lt. Gen. Shanahan." Breaking Defense. Breaking Media, November 13, 2019. https://breakingdefense.com/2019/11/exclusive-pentagons-ai-problem-is-dirty-data-lt-gen-shanahan/.

[66] Porche, Isaac R. III, Bradley Wilson, Erin-Elizabeth Johnson, Shane Tierney, and Evan Saltzman. *Data Flood: Helping the Navy Address the Rising Tide of Sensor Information*. PDF. RAND Corporation. RAND Corporation, 2014. https://www.rand.org/pubs/research_reports/RR315.html.

[67] Tarraf, Danielle C., William Shelton, Edward Parker, Brien Alkire, Diana Gehlhaus Carew, Justin Grana, Alexis Levedahl, et al. *The Department of Defense Posture for Artificial Intelligence: Assessment and Recommendations*. Santa Monica, CA: RAND Corporation, 2019. https://www.rand.org/pubs/research_reports/RR4229.html., pg. 52.

[68] Ibid., pg. 57-60

[69] McLemore, Connor, and Hans Lauzen. "The Dawn of Artificial Intelligence in Naval Warfare." War on the Rocks, June 12, 2018. https://warontherocks.com/2018/06/the-dawn-of-artificial-intelligence-in-naval-warfare/.

[70] Interoperability is "the ability of computer systems or programs to exchange information" and/or "the ability of military equipment or groups to work together." "Interoperability, Noun." Oxford Learner's Dictionaries. Accessed February 28, 2020. https://www.oxfordlearnersdictionaries.com/us/definition/english/interoperability.

[71] Tarraf, Danielle C., William Shelton, Edward Parker, Brien Alkire, Diana Gehlhaus Carew, Justin Grana, Alexis Levedahl, et al. *The Department of Defense Posture for Artificial Intelligence: Assessment and Recommendations*. Santa Monica, CA: RAND Corporation, 2019. Pg. 59. https://www.rand.org/pubs/research_reports/RR4229.html.

[72] For instance, one group may have a dataset that codes the United Kingdom as the "UK", whereas another dataset will code the United Kingdom by its telephone area code, +44.

[73] Another example of bias in data collection is if someone created an algorithm to predict how many people will join the U.S. military in the next five years but primarily polled the children of military members to create their training data. Children of service members are much more likely to enter into the military themselves, so the data collected from them cannot be extrapolated to the entire population. (Schafer, Amy. "Generations of War: The Rise of the Warrior Caste & the All-Volunteer Force." PDF. Center for a New American Security, May 2017. https://s3.amazonaws.com/files.cnas.org/documents/CNASReport-WarriorCast-Final.pdf?mtime=20170427115046.) Children of active-duty military members are far more likely to join the military in the future than their non-military counterparts, so an algorithm that does not account for that trend might guess that far more people are going to join the military than actually do. Again, the inaccuracies of these results are due insufficient data, not the structure of the algorithm itself.

[74] Hao, Karen. "This Is How AI Bias Really Happens—and Why It's so Hard to Fix." MIT Technology Review, February 4, 2019. https://www.technologyreview.com/s/612876/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/.

[75] Lohr, Steve. "Facial Recognition Is Accurate, If You're a White Guy." New York Times, February 9, 2018. https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html.

[76] Mozur, Paul. “One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority.” New York Times, April 14, 2019. https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html.

[77] Explainable AI is a field of research where scientists can understand how AI technologies made the decision that algorithm made. According to the DoD’s Defense Advanced Research Projects Agency (DARPA), “explainable AI—especially explainable machine learning—will be essential if future warfighters are to understand, appropriately trust, and effectively manage an emerging generation of artificially intelligent machine partners.”
Turek, Matt. "Explainable Artificial Intelligence (XAI)." Defense Advanced Research Projects Agency., https://www.darpa.mil/program/explainable-artificial-intelligence.

[78] Comiter, Marcus. “Attacking Artificial Intelligence.” PDF. Belfer Center for Science and International Affairs, August 2019. https://www.belfercenter.org/sites/default/files/2019-08/AttackingAI/AttackingAI.pdf.

[79] “Explainable AI” is AI that describes how it reached the conclusions that it reached. For instance, a regular AI image-recognition algorithm could scan a photo of a cat and arrive at the conclusion that the photo was of a cat. However, Explainable AI would scan the photo of a cat and arrive at the same conclusion, but it would be able to say “this is a cat because it has whiskers, it has fur, it is the average size and coloration of a cat”, etc. The ultimate goal of Explainable AI is twofold: first, it would produce more explainable models, while maintaining a high level of learning performance (prediction accuracy and second, it would enable human users to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners.
Turek, Matt. "Explainable Artificial Intelligence (XAI)." Defense Advanced Research Project Agency., accessed March 13, 2020, https://www.darpa.mil/program/explainable-artificial-intelligence.

[80] Cronk, Terri Moon. “DoD Unveils Its Artificial Intelligence Strategy.” U.S. Department of Defense, February 12, 2019.
https://www.defense.gov/Explore/News/Article/Article/1755942/dod-unveils-its-artificial-intelligence-strategy/.

[81] Tarraf, Danielle C., William Shelton, Edward Parker, Brien Alkire, Diana Gehlhaus Carew, Justin Grana, Alexis Levedahl, et al. *The Department of Defense Posture for Artificial Intelligence: Assessment and Recommendations*. Santa Monica, CA: RAND Corporation, 2019. Pg. 59. https://www.rand.org/pubs/research_reports/RR4229.html.

[82] Bendett, Samuel. “Sneak Preview: First Draft of Russia’s AI Strategy.” Defense One, 2019. https://www.defenseone.com/technology/2019/09/whats-russias-national-ai-strategy/159740/.

[83] Graham Webster et al., “Full Translation: China’s ‘New Generation Artificial Intelligence Development Plan’ (2017),” New America, August 1, 2017, https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/.

[84] "Joint Duty." Office of the Director of National Intelligence., accessed March 13, 2020, https://www.dni.gov/index.php/careers/joint-duty.

[85] Clapper, James. “ICD 203: Analytic Standards.” PDF. Office of the Director of National Intelligence, January 2, 2015. https://www.dni.gov/files/documents/ICD/ICD%20203%20Analytic%20Standards.pdf.

[86] “The AIM Initiative: A Strategy for Augmenting Intelligence Using Machines.” PDF. Office of the Director of National Intelligence. Accessed February 5, 2020. https://www.dni.gov/files/ODNI/documents/AIM-Strategy.pdf.

[87] “Modernized Integrated Database Definition (US DoD).” Military Factory. Accessed February 28, 2020. https://www.militaryfactory.com/dictionary/military-terms-defined.asp?term_id=3508.

[88] “Modernized Integrated Database (MIDB).” Global Security. Accessed February 28, 2020. https://www.globalsecurity.org/intell/systems/midb.htm.

[89] Pomerleau, Mark. “The Department of Defense Is Going to MARS (Just Not That One).” C4ISRNET, August 15, 2018. https://www.c4isrnet.com/show-reporter/dodiis/2018/08/15/the-department-of-defense-is-going-to-mars-just-not-that-one/. The DIA’s chief information officer Jack Gumtow states that “Once deployed, MARS will allow analysts and operators to absorb and process large amounts of data, capture new sources of data that provide a deeper understanding of adversary technological developments, provide the ability to track both static and mobile military forces, enable an exponential increase of data being ingested and leverage commercial best practices and industry’s technological advances.”

[90] Herridge, Catherine and Upson, Cyd. "Iran Likely at 'Inflection Point,' Launching Attacks to Change 'Status Quo,' Defense Intelligence Agency Director Tells Fox News." Fox News., last modified June 23, accessed March 13, 2020, https://www.foxnews.com/politics/iran-inflection-point-attacks-change-status-quo-defense-intelligence-agency-director-tells-fox-news.

[91] “NCTC Home.” DNI. Accessed February 28, 2020. https://www.dni.gov/index.php/nctc-home.

[92] Ibid.

[93] Comiter, Marcus. “Attacking Artificial Intelligence.” PDF. Belfer Center for Science and International Affairs, August 2019. https://www.belfercenter.org/sites/default/files/2019-08/AttackingAI/AttackingAI.pdf.