UNIDIR

UNITED NATIONS
INSTITUTE FOR
DISARMAMENT
RESEARCH

The Role of Data in Algorithmic Decision-Making

A Primer

Lydia Kostopoulos

UNIDIR RESOURCES

Acknowledgements

UNIDIR's Security and Technology Programme is supported by the Governments of Germany, the Netherlands, Norway and Switzerland.

The author extends her appreciation to those who provided input and shared their thoughts at various stages of this paper's development, including Henk Sonius, Michael Klipstein, Arnel David and particularly Pablo Breuer for his assistance with tables 1 and 2.

About UNIDIR

The United Nations Institute for Disarmament Research (UNIDIR)—an autonomous institute within the United Nations—conducts research on disarmament and security. UNIDIR is based in Geneva, Switzerland, the centre for bilateral and multilateral disarmament and non-proliferation negotiations, and home of the Conference on Disarmament. The Institute explores current issues pertaining to a variety of existing and future armaments, as well as global diplomacy and local tensions and conflicts. Working with researchers, diplomats, government officials, NGOs and other institutions since 1980, UNIDIR acts as a bridge between the research community and Governments. UNIDIR activities are funded by contributions from Governments and donor foundations.

Note

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontiers or boundaries. The views expressed in the publication are the sole responsibility of the individual authors. They do not necessary reflect the views or opinions of the United Nations, UNIDIR, its staff members or sponsors.

www.unidir.org

Contents

About the author

Lydia Kostopoulos (@Lkcyber) consults on the intersection of people, strategy, technology, and national security. Formerly the Director for Strategic Engagement at the College of Information and Cyberspace at the National Defense University, a Principal Consultant for PA and a higher education professor teaching national security at several universities in the United States, Europe, the Middle East and Asia. She speaks and writes on disruptive technology convergence, innovation, tech ethics, and national security. She is a member of the IEEE's AI Policy Committee, participates in NATO’s Science for Peace and Security Programme and has spoken on AI and national security as part of a US State Department Public Diplomacy programme.

> "Just as electricity transformed almost everything 100 years ago, today I actually have a hard time thinking of an industry that I don't think AI will transform in the next several years."[1]
>
> Andrew Ng, Artificial Intelligence Scientist

1 Introduction

What Andrew Ng was referring to is the impact that artificial intelligence (AI) is going to have in the world as we know it. He expects that just as everything became *electrified,* everything will become *cognified*. Already today, large quantities of data, also referred to as 'big data', are being leveraged through AI to derive insights that could not otherwise be pulled out of this data through human analysis alone.

Algorithmic decision-making is already being routinely used in many sectors. In medicine, AI can more precisely diagnose diseases, offer treatment recommendations, free up doctors' time to focus more on a patient, and with wearables, more data can be created and analysed to more accurately understand a patient. In finance and banking, AI is being used to determine eligibility for credit, offer personalized service with virtual financial assistance, analyse spending, and improve the process for detecting fraud and money laundering. Human relationships in social media at the professional, personal, and romantic level are being curated by algorithms that respond to an individual's interaction on the platform, and those same algorithms decide what content to show in the feed, including what advertisements would be most appropriate for the particular user. This also happens with entertainment platforms, which suggest what movies, music or books a user might like based on prior histories of what was watched, read, and at what speed, and what types of music was repeatedly listened to. With the increasing volumes of data on purchases, online shopping platforms have become more sophisticated in their suggestions of items and services to pair with products in the cart.

However, as algorithmic decision-making—and the collection and processing of data upon which it depends—is increasingly used throughout society, it has also generated controversy and concern.[2]

[1] Lynch, Shana. (2017). Andrew Ng: Why AI is the new electricity. Stanford News. https://news.stanford.edu/thedish/2017/03/14/andrew-ng-why-ai-is-the-new-electricity/.

[2] For example, there have been concerns about ownership of medical patient data and consent in the use of them. In finance, some models have been demonstrated to discriminate against low income households or marginalized communities from having a chance to acquire credit. With regards to health insurance, there are some companies that are asking customers to use wearable devices in exchange for lower premiums. While data collected can be used to help health insurance companies to identify customers who will have greater health insurance needs, there are some who are concerned that insurance companies are already creating algorithms that favour the healthy and thus reduce insurance access to those with pre-existing conditions, genetic predispositions or less healthy lifestyles. For an overview of these topics, see for example, Medical Futurist. (2019). Top Smart Algorithms in Healthcare. https://medicalfuturist.com/top-ai-algorithms-healthcare; Maskey, Sameer. (2018). How Artificial Intelligence Is Helping Financial Institutions. Forbes Technology Council. https://www.forbes.com/sites/forbestechcouncil/2018/12/05/how-artificial-intelligence-is-helping-financial-institutions/#2f9be144460a; Balasubramanian, Ramnath. Libarikian, Ari. McElhane, Doug. (2018). Insurance 2030—The Impact of AI on the Future of Insurance https://www.mckinsey.com/industries/financial-services/our-insights/insurance-2030-the-impact-of-ai-on-the-future-of-insurance; Titlow, John Paul. (2017). Instagram is Using AI-Human Hybrids to Shape What You See. Fast Company. https://www.fastcompany.com/90147890/instagram-is-using-ai-human-hybrids-to-shape-what-you-see-next; Toh, Allison. (2019). Are You Still Watching? How Netflix Uses AI to Find Your Next Binge-Worthy Show. NVIDIA. https://blogs.nvidia.com/blog/2018/06/01/how-netflix-uses-ai/; Rejoiner. (2019). The Amazon Recommendations Secret to Selling More Online. https://rejoiner.com/resources/amazon-recommendations-secret-selling-online.

This has resulted in an increasing level of social awareness and civic debate on how data is collected, who owns it, who is permitted to use it, what it is used for, under whose jurisdiction the data resides and whose interests it furthers. Concerns include questions about consent, privacy, discrimination, opacity about how decisions are made, as well as the limited avenues of recourse for those who feel that they have been harmed as a consequence of predictive decision-making.

Predictive Policing: From Data to Actionable Intelligence

Data and algorithmic decision-making is widely used in the fields of law enforcement and criminal justice, where an extremely large corpus of data permits leveraging AI to solve crimes, identify and assess suspects, predict the risks of individuals to engage or re-engage in unlawful behaviour, identify high-risk areas in order to inform where to deploy limited resources, and other applications. AI can help find patterns in large datasets of information, identify people in videos using facial recognition, improve forensics analysis and be used for "predictive policing".[3] The use of predictive analytics to flag potential future behaviour, such as recidivism, has been particularly controversial as the data used may be biased against particular groups or communities, resulting in discriminatory decisions.[4] Similarly, there are those who are voicing concern about using historical criminal data to train risk assessment tools as well as automated tools for processing prisoners through the legal system because historical data would be a reflection of the past (including its biases and discriminations) which would be replicated into present criminal justice practices.[5]

Learning from concerns raised about the transparency, reliability and legality of predictive policing, some analysts have recommended that militaries should be "transparent about how, when, why and on what legal basis the military is using predictive algorithms" in order to both "improve the quality of military decision making and enhance public support".[6]

The international discussion on military applications of machine learning has focused much more on *algorithmic decision-making* than on *data.* Yet without data, an algorithm cannot be trained, and even after training, an algorithm requires data in its processing to make decisions. Just as there is growing sensitivity about algorithmic biases,[7] there is also an element of due diligence in the choice of the collection and use of data used to train and utilize these algorithms.

Just as other sectors are leveraging data and algorithmic decision-making to improve performance, derive actionable intelligence, create greater situational awareness, produce new services, and more precisely tailor their traditional offerings, militaries and defence industries are also keen to exploit the potential of data and the value of algorithmic decision-making tools. Within the military context, there are processes, circumstances and contexts where AI will be of added value, such as

[3] Rigano, Christopher. (2018). Using Artificial Intelligence to Address Criminal Justice Needs. Office of Justice Programs. National Institute of Justice. https://www.nij.gov/journals/280/Pages/using-artificial-intelligence-to-address-criminal-justice-needs.aspx.

[4] Chicago Tribune. (2018). 'Predictive Policing': Big-City Departments Face Lawsuits. http://www.chicagotribune.com/news/sns-bc-us--predictive-policing-challenges-20180705-story.html.

[5] Hao, Karen. (2019). AI is Sending People to Jail—And Getting It Wrong. MIT Technology Review. https://www.technologyreview.com/s/612775/algorithms-criminal-justice-ai/.

[6] Deeks, Ashley. Predicting Enemies. (2018). Virginia Law Review, Vol. 104:1529.

[7] For an overview of algorithmic bias, see UNIDIR. Algorithmic Bias and the Weaponization of Increasingly Autonomous Technologies. http://www.unidir.ch/files/publications/pdfs/algorithmic-bias-and-the-weaponization-of-increasingly-autonomous-technologies-en-720.pdf.

logistical planning, administrative operational processes, and elements of intelligence, surveillance and reconnaissance, including targeting.

This primer[8] explores data in the context of military decision-support tools and increasingly autonomous weapons systems by briefly discussing the links of the data chain (creation, collection, organization and use), potential challenges to data integrity in adversarial environments, and concludes with a few forward-looking questions for policymakers considering military applications of increasingly autonomous systems.[9]

[8] This paper is intended to be an introductory primer for non-technical audiences. Because of the rapidly developing nature of this field, this paper can only provide a snapshot in time. However, many of the underlying concepts about data are likely to remain applicable in the short to medium term.

[9] Autonomous functions do not necessarily require machine learning. Machine learning, however, always requires big data.

2 The Data Chain: Creation, Collection, Organization and Use

Data is at the core of computational systems; it is necessary to train algorithms—without it, algorithms have no basis with which to recognize patterns, perform predictive analysis or make decisions. It is data that informs algorithms what to sense, how to adapt, and how to determine the suitable course of action. The data chain could be broken down into four stages:

- data creation,
- collection,
- organization, and
- use.

Both humans and machines may play a role in each phase of the data chain. Humans sometimes create data themselves, for example by taking a picture, and other times machines create data, for example data logs of machine activity such as sign-in activity. In the next phase humans can be involved in the collection of data, for example searching through satellite images and saving them into a folder, or AI could be scanning for certain images to store. In the data organization phase, humans could decide how it is organized, or a machine might do so based on parameters, bounds and objectives. And in the last phase, humans currently play the largest role because they determine how data is used and set the parameters for its use; however, it is technically possible for humans to delegate an increasing amount of decision-making about use to machines—which has been one of the primary points of discussion in the Group of Governmental Experts on Lethal Autonomous Weapon Systems within the framework of the Convention on Certain Conventional Weapons.

There are several technological convergences which position data as a decisive element in an increasingly data-driven military operational environment. The key technologies that are playing an important role in current military activity are:

- the proliferation of sensors,[10] which generate an ever-increasing amount of data;
- the increasing power and capacity of cloud computing. With external servers, storages, databases, analytics and intelligence this collective computing service done in the cloud allows for greater speeds and more data to be handled simultaneously;[11]
- the prevalence of connected Internet of Things (IoT), allowing for more rapid decentralized decision-making; and
- the continuous development of AI as decision-support infrastructure, as well as a tool that supports offensive operations.

This synergy of technologies enables greater situational awareness of the operational environment, the potential for real-time combined autonomous manoeuvres and integrated cognitive (or self-learning) systems, combat agility, speed and precision in force application. It also enables a wide constellation of machine to machine (M2M) communications and autonomous cooperation—one

[10] Stratfor Worldview. (2019). Sensor Proliferation Is Changing How We Wage War. Real Clear Defense. https://www.realcleardefense.com/articles/2019/04/12/sensor_proliferation_is_changing_how_we_wage_war_114328.html.

[11] Microsoft. (2019). What is Cloud Computing: A Beginner's Guide. https://azure.microsoft.com/en-us/overview/what-is-cloud-computing/.

such example is drone swarms which are “made up of cooperative, autonomous robots that react to the battlefield as one”.[12]

At the centre of this convergence is data and the data chain, which includes the creation, collection, organization and use of data, as well as its integrity throughout its life cycle. Integrity of data refers to the accuracy, completeness, consistency of data and ensuring that the data is safe from external manipulation. However, integrity of data alone is not viable without context—for example algorithmic contextualization of data is where the benefits of these technological convergences will play a role in decisions concerning discrimination and proportionality in the use of force.

Some might think that data is simply a collection of ‘facts’ and as such data is objective. However, humans create data, determine what data should be collected and how it should be processed.[13] Its use or misuse, automated or autonomous collection, and subsequent algorithmic processing may also exhibit error or bias. In an adversarial environment, data may also have been modified via a cyberattack, deceptively altered at the point of collection, or in some way intercepted before received by the intended sensor or human.

2.1 THE CREATION AND COLLECTION OF DATA IN THE MILITARY CONTEXT

An algorithmic decision is an output to some form of data input and data processing, where data input could have been from a sensor, camera or other means of data capture. In the military context, there are several types of data and means of collection of particular interest for military operations from intelligence, surveillance and reconnaissance (ISR), to decision-making support tools and weapons systems.

There are unique types of data across the five operational environments of land, air, sea, space and cyberspace, and there are certain types of data that overlap across environments. Each operational environment is different in terms of the volume of data it produces and the ease in which this data can be collected and processed. These types of data can be used in the selection of a target (identifying and selecting it) as well as in the monitoring of a target (tracking it, following or actively pursuing it). Data can be collected with lesser or greater amounts of human oversight and supervision, which can later be combined for processing by algorithmic decision-support tools.

Table 1 includes a sample of types of data from detection methods commonly used in the military context while Table 2 offers two examples of intelligence collection disciplines that will benefit from big data analysis and algorithmic decision support.[14]

[12] Scharre, Paul. (2018). How Swarming Will Change Warfare. Bulletin of the Atomic Scientists Journal. Volume 74, 2018 - Issue 6: Special issue: Existential Nexus: The Intersection of Technological Threats. https://www.tandfonline.com/doi/full/10.1080/00963402.2018.1533209.

[13] Similarly, algorithms were once widely perceived to be objective or neutral yet there is now a deeper understanding of how algorithms can reflect or amplify human biases. For an overview of algorithmic bias, see UNIDIR. Algorithmic Bias and the Weaponization of Increasingly Autonomous Technologies. http://www.unidir.ch/files/publications/pdfs/algorithmic-bias-and-the-weaponization-of-increasingly-autonomous-technologies-en-720.pdf.

[14] Other intelligence disciplines that have the potential to autonomous and connected are Communications Intelligence (COMINT), Electronic Intelligence (ELINT), Imagery Intelligence (IMINT)

Table 1: Examples of Media, Detection Methods and Use Cases

MEDIUM	EXAMPLE OF DETECTION METHOD	DESCRIPTION	EXAMPLES OF USE CASES
Radio waves	Radio detection, ranging and radio frequency	Active detection: Transmitter sends out radio wave and reads waves reflected off of an object	Detect, identify, tag and/or track aircrafts, ships, spacecrafts, telescopes, weather, terrain, missiles Used for surveillance, electronic warfare, jamming and/or spoofing
		Passive detection: Sensor detects radio wave emanation of defined frequency bands	Detect aircraft, ships, military ground units, human movement, anti-aircraft emplacements
Sound waves	Sound navigation and ranging	Active detection: Transmitter sends out sound wave and reads waves reflected off of an object	Detect, identify or track ships, submarines, vehicles, torpedoes, spacecraft and geothermal events
		Passive detection: Microphones detect sound emitted from object	Detect, identify and/or track ship, submarines, vehicles, torpedoes, spacecraft, anti-ship mines, and geothermal events
Heat	Thermal imaging	Sensors read infrared radiation (heat) emitted from object	Detect and identify vehicles, human bodies, equipment, missiles as well as activities within buildings
Wireless/wired electromagnetic or optical waves	Digital communications	Identifying a computation/transmitting entity by observing responses to non-standard queries and a failure of protocol standards to address every issue.	Identify satellites, modems, ICS/SCADA, IoT, communications TCP/Internet Protocol (for voice, video, email, Bluetooth, WiFi) Identify computing assets (such as computer or phone, including model, make, version number, operating system, version, patch level, etc.) Used for electronic warfare, eavesdropping, corrupting data and jamming
Light waves	Spectroscopy	Identification by analysing emitted electromagnetic waves (to include light)	Identify material composition, telescopes, satellites, spacecraft
Vibration	Seismology	Identifying movement through vibrations through the Earth	Detect vehicle, aircraft or personnel movement

Table 2: Examples of Intelligence Disciplines, Media and Use Cases

INTELLIGENCE COLLECTION DISCIPLINE	DESCRIPTION	MEDIUM	EXAMPLES OF USE CASES
SIGINT *Signals Intelligence*	Passive sensors receive and characterize emissions from objects (i.e. radar, communications, etc.)	Radio waves	Obtain intelligence from radio-frequency signals. This includes communications radios, mobile phones, or any other electromagnetic signals used for communications.
MASINT *Measures and Signals Intelligence*	Similar to SIGNIT, but combines multiple spectrums and characterization with analysis	Heat, light, radio waves, sound waves	Obtain intelligence on terrain to identify how it has been altered (indicating, for example, tank tracks, mass graves, etc.) Analyse telemetry, for example to track the engine power and flight profile of rockets and missiles

2.2 DATA ORGANIZATION

While we live in increasingly technology-enabled environments, it isn't just machines and software making decisions all by themselves. It is the human engineers, user interface designers, platform architects and designers who together make our technologies and shape our digital environments. The same applies for algorithms—humans play a fundamental role not just in the design of the algorithm but in the data chain. They do this through their intentional and unintentional choices about what data to collect, how to collect it, decisions about what data is valuable to use in processing, and how to organize it. The collected data alone has no value if it has not been organized to meet its intended objective.

In the military context, these choices are based on human understandings of the operational environment, preferences and styles of manoeuvre as well as doctrinal teachings in operational planning. Choices concerning collection and organization of data may be influenced by heuristics, which are mental shortcuts used in human judgement and decision-making to make sense of information and situations. Heuristics help humans make decisions and understand information in an environment by centring on the most relevant aspects of a problem.[15] Analogous to the human bias in algorithms, human bias is present in the organization, representation and value ascribed to data and the datasets created for use in algorithmic decision-making tools and increasingly autonomous systems.

The following are some of the most common ways that human choices, intentional or otherwise, might influence the data chain:

[15] Lewis, Alan. (2008). The Cambridge Handbook of Psychology and Economic Behavior. Cambridge University Press. p. 43. ISBN 978-0-521-85665-2.

- **Anchoring bias:** when people make judgements that rely heavily on an initial piece of information which colours their frame of reference and perspective of the situation.[16] For example, data scientists who were first introduced to data present in the context of electronic warfare (EW) and then introduced to how EW supports and enables other air, land and sea capabilities, may weigh EW related data higher than others. This thinking would produce a data validity problem.
- **Confirmation bias:** when people interpret information that aligns with their existing beliefs, values or perception of the situation.[17] For example, it might be explained to a data analyst that certain types of air manoeuvres constitute a form of deception and the data that would be used to detect this type of manoeuvre. The data scientist may then code this data into an algorithm to label such manoeuvres as deception, when in reality such a manoeuvre may not always be a form of deception.
- **Availability heuristic:** when people make judgements about a situation or the likelihood of something happening based on information that they know and to which they can quickly remember or refer. [18] For example, a human analyst may judge the intention of a military escalation based on information that she recently received in a briefing.
- **Bandwagon effect:** when people chose to believe in something or take part in something because other people are doing it. This can sometimes occur even when these beliefs and actions are contrary to an individual's previous perspectives.[19] For example, data collected from social media to flag conflict-related trends and provide predictive analysis may interpret a geopolitical situation based on data that went viral on reputable news channels because those channels have been assigned higher trust values. Similarly, data from likes or comments may make the information appear more trustworthy.

These 'mental shortcuts' are natural and useful both to avoid information overload and decision paralysis in situations where there is a surfeit of information, as well as in situations when it is necessary to 'fill in the gaps' due to incomplete information. However, as algorithmic technologies are poised to become an important tool for decision support and are expected to change the way humans interact with each other and machines, it is important to consider the assumptions made and methodologies employed to handle data that are utilized in the context of intelligence, surveillance and reconnaissance; increasingly autonomous weapon systems;[20] and situational awareness of the operating environment. As military applications are increasingly data dependent, much greater attention must be paid to developing concrete ways to have human bias mitigation strategies to reduce the potential harm or consequences of these heuristics.

The data, recommendations, predictive analysis and decisions resulting from algorithms have the potential to enhance human judgement. The role humans play in data organization provides spaces for human accountability and responsibility. The is particularly crucial in order to avoid what has been called "distributed responsibility", where responsibility is distributed so widely that in the end

[16] Tversky, A. Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. Science (New Series), 185, 1124-1131.

[17] Plous, Scott. (1993). The Psychology of Judgment and Decision Making, p. 233.

[18] Esgate, Anthony. Groome, David. (2005). An Introduction to Applied Cognitive Psychology. Psychology Press. p. 201. ISBN 978-1-84169-318-7.

[19] Colman, Andrew. (2003). Oxford Dictionary of Psychology. New York: Oxford University Press. p. 77. ISBN 0-19-280632-7.

[20] For a more detailed discussion, see UNIDIR. (2016). Safety, Unintentional Risk and Accidents in the Weaponization of Increasingly Autonomous Technologies.

no one person can be held accountable for negative outcomes.[21] In the context of military applications, such a circumstance would create problems for identifying who is responsible. Increased transparency across the various elements in the data chain may provide greater insight to different ways in which data was accessed, and this could improve data chain oversight.

2.3 USE: DATA AT THE HEART OF THE CHANGING CHARACTER OF WAR

Currently the military is using data and data analytics to gain greater situational awareness of operational environments. As more data is being generated and collected, intelligence collection disciplines are exploring how "new technologies for collecting, moving, storing, and organizing data could give all-source analysts access to vastly more information with more automation and productivity, thereby allowing them to concentrate their finite cognitive capacity on the hardest, highest-priority problems".[22] The synergy between data and AI is set to play an important role in combat readiness. One example is where predictive analytics will be able to "anticipate component failures and reduce the amount of unplanned maintenance" in tanks and other military vehicles.[23] In military training, data is being used to create realistic synthetic environments for training in virtual battlefields.[24] These are some examples of how data is currently being used and it is going to become increasingly more important.

In just the first two decades of the twenty-first century, the Internet and connected devices have expanded the way in which people communicate across the world, share information, sell products and generate profits. This dependency on Internet communications and increasingly connected things has resulted in a new form of critical infrastructure that didn't exist in previous centuries. As of 2019, there are over 10 billion things connected to the Internet, each creating and sharing data, and while the projected numbers for 2025 vary depending on the source, the lowest figure is double that of today.[25] The civilian corporate market is preparing itself for a world of ubiquitous and faster connectivity. All this will be powered by data produced and collected through various Internet-connected devices, machines and sensors. For national security this means is a larger attack surface area. Just as some military operations have conducted strategic bombing on civilian infrastructure (such as key roads, bridges, train stations, etc.) the Internet of Things and smart cities will be considered high-value strategic targets.

The advancements made in the civilian market with connected sensors, big data and predictive analysis will not be in isolation to defence applications. Militaries tend to develop in tandem with civilian markets in leveraging data for actionable intelligence, predictive forecasting, management of infrastructure, energy cost reduction, greater situational awareness and much more.[26] This type of decision support as well as decision-making capability derived from AI has the potential to drive

[21] Mariarosaria, Taddeo. Floridi, Luciano. How AI Can Be a Force for Good, *Science*, 24 August 2018, Vol 361, issue 6404.

[22] Symon, Paul. Tarapore, Arzan. (2015). Defense Intelligence Analysis in the Age of Big Data. Joint Force Quarterly 79. National Defense University Press. https://ndupress.ndu.edu/Media/News/Article/621113/defense-intelligence-analysis-in-the-age-of-big-data/.

[23] Jordan, Sonja. (2018). Army Investing in Predictive Maintenance for Bradleys. National Defense Magazine. https://www.nationaldefensemagazine.org/articles/2018/9/26/army-investing-in-predictive-maintenance-for-bradleys.

[24] Miller, Susan. (2018). Army Looks to Improve Training Simulations with Intelligent Automation. Defense Systems. https://defensesystems.com/articles/2018/12/19/army-automation-tools-ste.aspx.

[25] Insights on the Internet of Things from McKinsey. https://www.mckinsey.com/featured-insights/internet-of-things/our-insights.

[26] Stanley-Lockman, Zoe. (2018). Three A's of Military Logistics: Modernizing the Armed Forces' Tail – Analysis. https://www.eurasiareview.com/13072018-three-as-of-military-logistics-modernizing-the-armed-forces-tail-analysis/.

forward new ways of engaging in armed conflict and at greater speeds which will have will have implications on the use of force.

With the proliferation of sensors, improvements in quantum computing (which will exponentially speed up data processing)[27] and more systems and devices contributing to the data flow, there will be greater potential to understand complex relationship patterns in the operating environment and gain greater situational awareness.[28] Data is ultimately the engine for increasingly autonomous systems. The volume, variety and velocity of data combined with AI is what will enable systems (for example swarms) "to adapt in real time during the mission, based on what [they are] observing and how well [the system] is performing".[29]

[27] Biercuk, Michael. Fontaine, Richard. (2017). The Leap into Quantum Technology: A Primer for National Security Professionals. War on the Rocks. https://warontherocks.com/2017/11/leap-quantum-technology-primer-national-security-professionals/.

[28] Abadicio, Millicent. (2019). Predictive Analytics in the Military—Current Applications. Emerj. https://emerj.com/ai-sector-overviews/predictive-analytics-in-the-military-current-applications.

[29] Wilson, J.R. (2016). Today's Battle for the Electromagnetic Spectrum. Military & Aerospace Electronics. https://www.militaryaerospace.com/communications/article/16709112/todays-battle-for-the-electromagnetic-spectrum.

3 Attacks on Data

Military deception and espionage are as old as war itself and it can be expected that new technologies will be utilized to confuse adversary forces, limit their decision-making capability and deny them the ability to target friendly forces. Data itself is a high-value target in attacks on algorithmic systems. There are several ways in which data could be targeted by an adversary to gain advantage in the operational environment. Three are briefly described here: data poisoning, denial of data, and espionage.

3.1 DATA POISONING

Data poisoning refers to the manipulation of data—whether by omitting data, replacing data or adding it with the intention to corrupt or alter a learning model or algorithm.[30] There are several spaces where this can happen:

- **Data collection**: As described above, depending on the data type, collection can be done by a human and by a machine. In both instances, it is possible for adversaries to intentionally try to deceive the collector to collect false or misleading data that would be to their advantage. For example, in one well known study to test the robustness of a neural network for autonomous vehicles, the researchers presented its visual sensor with a stop sign that had stickers placed on it. Their research found that the algorithm was not able to identify it as a stop sign with this simple addition, while a human would identify the stop sign immediately, regardless of whether it had been defaced.[31] This sort of physical adversarial data input can be expected in military operational environments. This type of data alteration for the purpose of misleading a sensor need not only be done on physical objects. It can also be in a radio wave or any of the data detection methods described in table 1.
- **Training**: During the training phase of an algorithm, the training environment could be compromised through a cyberattack on the stored data and datasets. This could involve deleting, modifying or adding data with the intention to compromise the algorithm's integrity and degrade the capability or performance of the algorithm during military operations. Military entities are not the only targets; organizations that are part of the wider military ecosystem will also be targets. For example, defence contractors and military research organizations responsible for developing military-related algorithmic tools would be high-value institutional targets for data-related attacks.
- **Data classification**: Another form of attack (which could be conducted using conventional cyberattack methods) could be conducted on the data classification or labelling of data by changing the values and specifications of the data. For example, image classifier data could be altered to recognize protected objects, such as ambulances and hospitals, as valid military targets, or a cyberattack could alter the data input for the threshold of heat signatures that trigger a kinetic response.

[30] Steinhardt, Jacob. Wei Koh, Pang. Liang, Percy. (2017). Certified Defenses for Data Poisoning Attacks. 31st Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA. https://papers.nips.cc/paper/6943-certified-defenses-for-data-poisoning-attacks.pdf.

[31] Eykholt, Kevin. Evtimov, Ivan. Fernandes, Earlence. Li, Bo., Rahmati, Amir. Xiao, Chaowei. Prakash, Atul. Kohno, Tadayoshi. Song, Dawn. (2018) Robust Physical-World Attacks on Deep Learning Visual Classification. CVPR 2018. https://arxiv.org/pdf/1707.08945.pdf.

3.2 DENIAL OF DATA

Analogous to the concept of denial of service (DoS) attacks where "legitimate users are unable to access information systems, devices, or other network resources due to the actions of a malicious cyber threat actor",[32] a denial of data would be the blocking of access by legitimate users to data.

- **Jamming**: In the operational environment sensors and other data-collection mechanisms could be denied information through jamming techniques. This would create a lack of access to pertinent data for situational awareness and data for algorithmic processing, thereby degrading or rendering unusable particular systems.
- **Noise**: Intentional spoofing could be used across various environments to create data noise that sensors and other data collection mechanisms would pick up. This data would in essence be meaningless and intentionally cause confusion for algorithmic processing, which would be a way of disrupting, denying and degrading the opponent's decision-making capability.

3.3 ESPIONAGE

The data used in the training environments as well as the way data has been classified is equally important to the algorithms that make the decisions. Understanding this proprietary information could be of strategical, operational and tactical use in the military context. In this sense, military data used in algorithmic decision-making, whether it be for operational, intelligence or weapons purposes, will be a high-value target for espionage.

- **Data exfiltration**: Cyberattacks resulting in the exfiltration of the data used for training algorithms and the classifiers will be very useful for the adversary to understand how the algorithm processes data and how the bounds of the data have been classified. This can be valuable for anticipating how a military algorithm (particularly in a weapons system) will react to given data inputs. Additionally, having this data would help the adversary better identify what aspects of the data chain to target during operational planning.

[32] US-CERT. (2009). Security Tip (ST04-015) Understanding Denial-of-Service Attacks. CISA Cyber Infrastructure. Department of Homeland Security. https://www.us-cert.gov/ncas/tips/ST04-015.

4 Conclusions and Key Questions

Those considering the military applications of AI, including in the ongoing discussions on Lethal Autonomous Weapon Systems within the framework of the CCW, may wish to consider the following questions about the creation, collection, organization and use of data:

- If governments decide to regulate increasingly autonomous weapon systems, which national or international organizations or instruments would be best placed to offer guidance or assistance on standards with regards to data meant for algorithmic processing in military contexts?
- As the software and algorithms that are embedded in weapons systems and other military systems are subject to high levels of classifications and intellectual property protection, what can designers of these algorithms do to demonstrate that they have appropriately mitigated data-related risks?
- How could militaries be more transparent about how, when, why and on what legal basis they use data and predictive analytics? Is such transparency desirable?
- What sort of indicators would assist operators and commanders to understand the quality and integrity of data used in decision-support tools in order for them to maintain trust in the system?
- As data is a high-value target, it should be protected and secured with the expectation that adversaries will attempt to target it. What concrete actions are necessary for governments to strengthen information security at all points of the data chain to prevent data from being manipulated or compromised?

When decisions are delegated to algorithms, it is important to understand how the algorithms are designed and what objectives have been embedded into them to fulfil their intended task. As algorithms become an integral component in military decision-making and operations, it will be equally important to understand the source, treatment and integrity of the data on which the algorithms based their decision. In particular, the topic of data is of fundamental importance to discussions on responsibility and accountability in the military context. Thus, deeper consideration of the issue of data has the potential to add value to the existing international discussion on increasing autonomy in weapon systems, as well as consideration of applications of AI in non-weaponized military applications, such as decision-support tools.

UNIDIR
UNITED NATIONS
INSTITUTE FOR
DISARMAMENT
RESEARCH

The Role of Data in Algorithmic Decision Making

A Primer

The international discussion on military applications of machine learning has focused much more on *algorithmic decision-making* than on *data.* Yet without data, an algorithm cannot be trained, and even after training, an algorithm requires data in its processing to make decisions.

This primer explores data in the context of military decision-support tools and increasingly autonomous weapons systems by briefly discussing the links of the data chain (creation, collection, organization and use), potential challenges to data integrity in adversarial environments, and concludes with a few forward-looking questions for policymakers considering military applications of increasingly autonomous systems.

UNIDIR RESOURCES