National Defence Défense nationale

Canadian Forces College
Collège des Forces Canadiennes

Machine Learning in the Military Intelligence Process

Major James Barr

JCSP 48

Exercise Solo Flight

Disclaimer

PCEMI 48

Exercice Solo Flight

Avertissement

Canada

CANADIAN FORCES COLLEGE – COLLÈGE DES FORCES CANADIENNES

JCSP 48 – PCEMI 48
2021 – 2022

Exercise Solo Flight – Exercice Solo Flight

Machine Learning in the Military Intelligence Process

Major James Barr

UNLOCKING THE SECRETS HIDDEN IN DATA:
MACHINE LEARNING IN THE MILITARY INTELLIGENCE PROCESS

Artificial Intelligence (AI) is the central focus of the fourth industrial revolution, and AI builds upon the foundation set by Automation of Data Processing (ADP) and Machine Learning (ML).[1] ADP and ML are no longer emerging fields. Thousands of companies globally already benefit from practical applications based on ADP and ML.[2] Canadians are becoming increasingly reliant on the daily support provided by machine learning technologies. TikTok learns their users' interests and provides individualised video recommendations. Facebook curates data, identifying the people and news stories most attractive to their users. Siri recognizes voice patterns and dialects, enabling every iPhone user to have their own personal assistant. Google can find a desired picture or video anywhere on the internet, through a simple written description. Advances in machine learning have driven increases in productivity, reducing the requirement for costly human intervention while improving the quality and granularity of information used to support decision-making.[3]

Private corporations that did not initially leverage the gains in productivity offered by machine learning, became less profitable and were driven to match the performance of their competitors. Despite Canadian corporations playing a leading role in the automation and machine learning industry, the Canadian Armed Forces (CAF) has not onboarded these advantages, and is now playing catch-up to our allies and

---

[1] Klaus Schwab et al., *Shaping the Future of the Fourth Industrial Revolution*, 1st American ed. (New York: Currency, 2018). viii.

[2] Nikunj C. Oza, Chad Stephens and Kaitlyn Dozier Fox, "Enabling Safety from Data: Machine Learning/Artificial Intelligence for in-Time Aviation Safety," (2021). 3.

[3] Schwab, *Shaping the Future of the Fourth Industrial Revolution*. 104.

adversaries. In the modern battlespace, gathering relevant, timely and accurate information is essential. Still, the best collection network in the world is rendered useless if the information cannot be refined into actionable intelligence. The CAF must embrace and fully integrate automated processes into its intelligence processing.

Private industry's primary goal is to maximize profits, seeking any opportunity to improve overall efficiency. The CAF is not driven in the same way as private industry, but is instead resource-constrained. It cannot rapidly adjust financial expenditures or personnel strength in reaction to changes in the threat situation. In the current environment, with personnel shortages being felt most acutely in specialist trades, expenditures should be made to maximise the use of limited personnel resources. Automated data processing and machine learning should be at the forefront of initiatives, reducing the inefficient and labour-intensive manual categorisation and processing of information. Personnel should instead focus on value-added tasks that require human expertise.

To avoid the potential for accidental disclosure of sensitive information, this paper will utilise only unclassified sources, and will not discuss the capabilities of current or future CAF or allied systems.

WHAT IS MACHINE LEARNING

The concepts of “Automated Data Processing”, “Machine Learning”, and “Artificial Intelligence” are listed in terms of increasing sophistication. Although these three concepts appear similar, they cannot be used interchangeably. Automated data processing is one possible application for machine learning, and machine learning is one of the ways that artificial intelligence can be implemented. It is essential to clearly define these concepts before discussing how automated data processing and machine learning can be of immediate benefit to the CAF.

- *Automated Data Processing* is “The Automation of business processes in which software performs tasks that can be codified by computers.”[4]
- *Machine Learning* is “An application of AI, capable of identifying patterns from a large set of data with the help of algorithms. It is self-learning in nature and becomes ‘smarter’ over time.”[5]
- *Artificial Intelligence* “Performs tasks that previously required human intelligence, such as extracting meaning from images, text or speech, detecting patterns and anomalies, and making recommendations, predictions or decisions.”[6]

Artificial Intelligence is typically grouped into two categories “Narrow” or “Weak” AI, and “General” or “Strong” AI.[7,8] All further discussion and recommendations will be limited to “Narrow” AI. Narrow AI is programmed to perform a single task, with explicit instructions on how the AI should learn. The Artificial Intelligence systems currently in widespread use are all narrow AI.[9] “General” or “Strong” AI is the future goal of AI researchers, but practical use will not become

---

[4] Sara Sikora, Blythe Hurley and Anya George Tharakan, *Automation with Intelligence : Reimagining the Organisation in the 'Age of with'*. 4.

[5] Shankar Bellam, *Robotics Vs Machine Learning Vs Artificial Intelligence: Identifying the Right Tools for the Right Problems*,[2018]). 3.

[6] Sikora, *Automation with Intelligence : Reimagining the Organisation in the 'Age of with'*. 4.

[7] Ben Goertzel and Cassio Pennachin, *Artificial Brains* (Berlin, Heidelberg: Springer Berlin Heidelberg, 2007). VI

[8] Tannya D. Jajal, "Distinguishing between Narrow AI, General AI and Super AI," *Mapping Out 2050,* May 21, 2018. 3.

[9] Ibid.

available for at least another ten years.[10,11] General AI is expected to be able to perform abstract thinking, reason, solve problems, and use judgement when making decisions.[12]

Machine Learning is not restricted to any specific set of hardware or software. It is a process whereby a computer uses a set of human-provided rules to learn patterns within the provided dataset.[13] Each machine learning system must be tailored to the dataset it is designed to analyse.[14] The individual machine learning systems can be

[10] Ibid.
[11] "General AI Vs Narrow AI," last modified Apr 20, https://levity.ai/blog/general-ai-vs-narrow-ai.
[12] Goertzel, *Artificial Brains*, VI
[13] Christopher M. Bishop, *Pattern Recognition and Machine Learning* (New York, NY: Springer, 2006). 193.
[14] Ibid. 32.

interconnected using different architectures to aggregate their results, enabling higher-level analysis to be performed by humans or other machine learning systems.

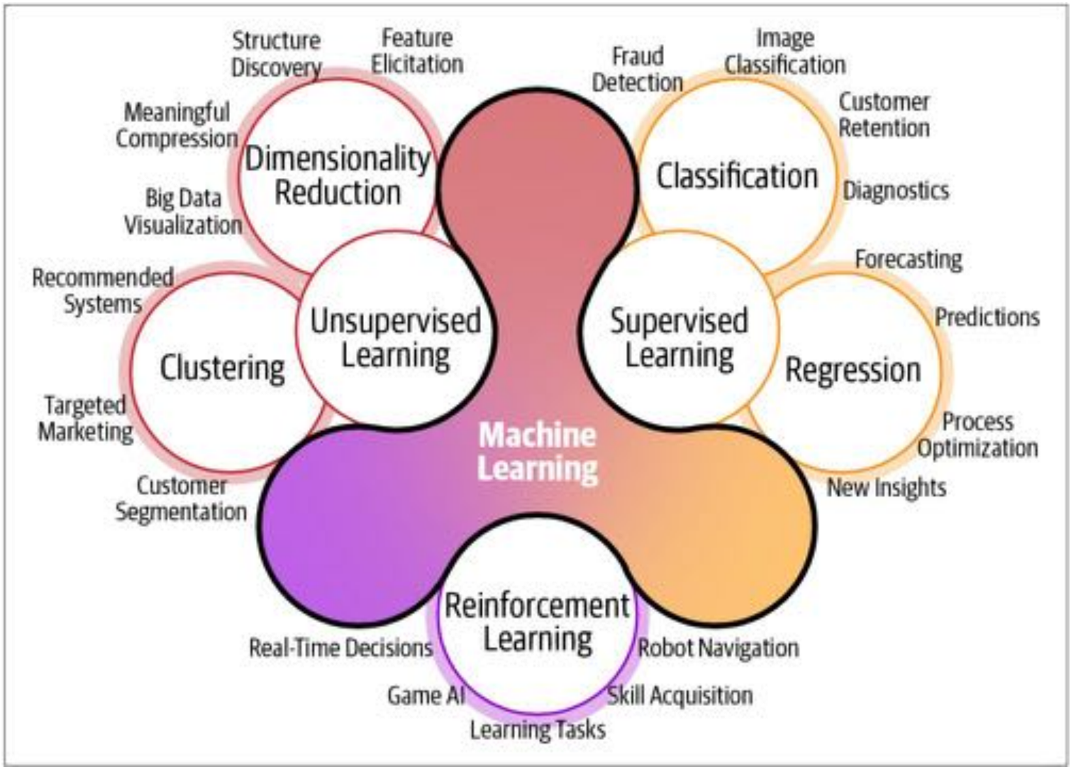

Figure 1 - The Machine Learning Model
Source: Norris, The Value of AI-Powered Business Intelligence, 5.

WHY AUTOMATION IS NECESSARY

Academia and private industry have been developing, and continue to develop machine learning technologies across a variety of disciplines. The most mature fields that will also have an immediate impact for military use relate to image processing, language, and the analysis of trends.

The People's Republic of China (CCP) has identified Artificial Intelligence as one of the main objectives in their 14th five-year plan (2021-2025).[15] Researchers at the National University of Defense Technology have developed a two-step system which processes massive quantities of imagery, and identifies key aspects of the images.[16] The first pass detects equipment of interest in the image, including aircraft, armoured vehicles and infrastructure. The result of the first pass is sent to a second system based on the type of equipment detected. All images tagged with "aircraft" are sent to a second engine designed only to analyse aircraft and determine the specific airframe within the image. This system allows for near real-time tracking of military equipment around the world. For aircraft, it can give intelligence on the current locations, composition, and concentration. In aggregation, the information can be extrapolated to determine aircraft maintenance schedules.

[15] Asian Development Bank, *The 14th Five-Year Plan of the People's Republic of China -Fostering High-Quality Development Observations and Suggestions 观察与建议* (Manila, Philippines: Asian Development Bank, 2021). 3.

[16] Yunsheng Xiong et al., "Non-Locally Enhanced Feature Fusion Network for Aircraft Recognition in Remote Sensing Images," *Remote Sensing* 12, no. 4 (Feb 19, 2020). 681-704. 685.

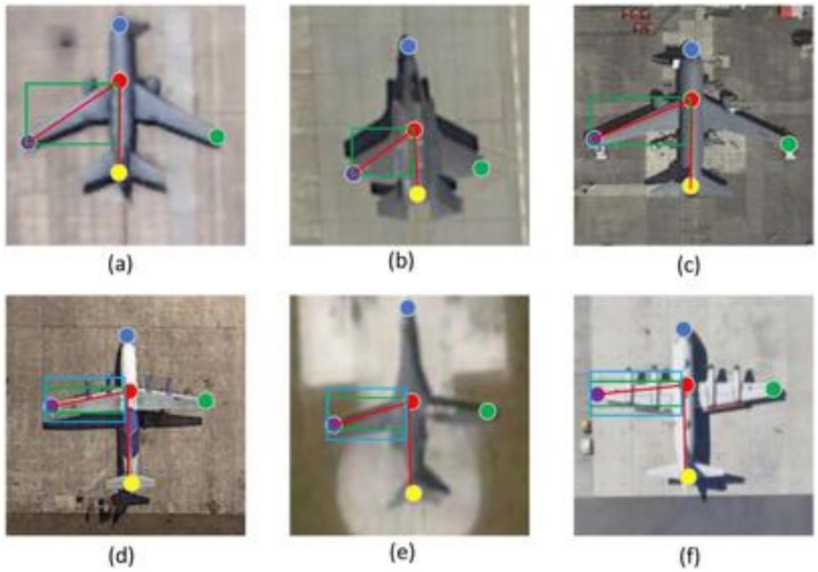

Figure 2 - Feature Extraction using Maps and Feature Fusion of Parts
Source: Xiong, *Non-locally Enhanced Feature Fusion Network for Aircraft Recognition in Remote Sensing Images*, 10.

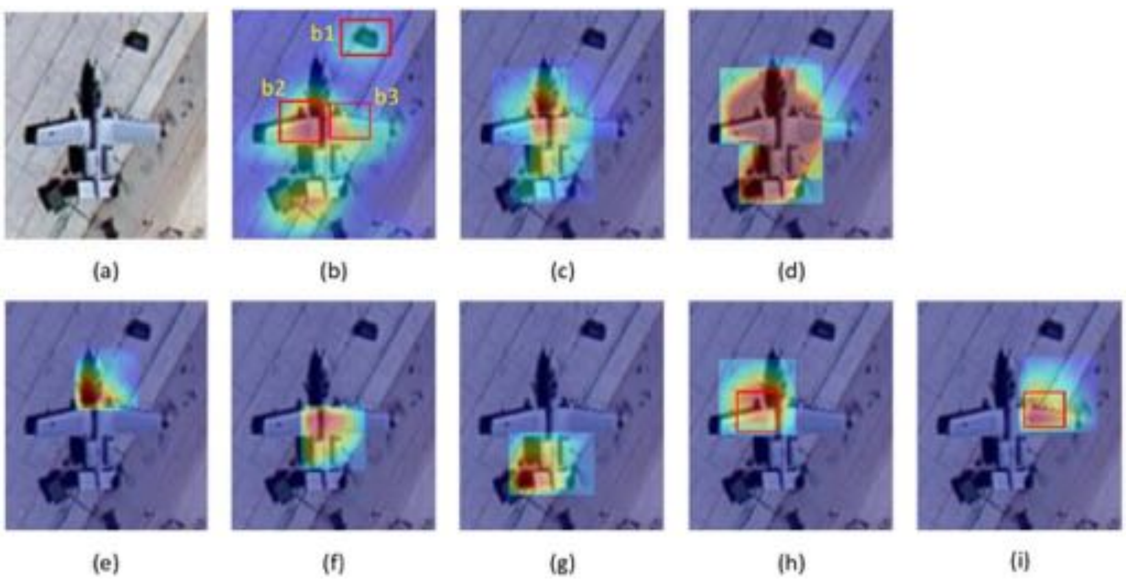

Figure 3 - Feature Extraction using Heatmaps on an A10
Source: Xiong, *Non-locally Enhanced Feature Fusion Network for Aircraft Recognition in Remote Sensing Images*, 15.

The CAF could use a similar image identification system to identify equipment type, quantity, and location within a battlespace. Monitored over time, the movement of equipment could be used to predict probable enemy courses of action. Strategically, it could provide automated alerts if there is movement above a specific threshold, such as the movement of an inter-continental ballistic missile (ICBM) mobile launcher in North Korea or a new concentration of amphibious units near the Taiwan Strait. There is a growing prevalence and availability of civilian and military high-quality imagery and video. The current image processing methodology, heavily reliant on Intelligence Analysts using rudimentary software toolsets, does not even meet the current demands of the operational community. The importance of rapid, reliable analysis and categorisation of imagery and video is, and will continue to become increasingly critical for success in future conflicts.

The field of Natural Language Processing has seen sustained development, leading to the creation of Siri, Alexa and Google Translate. It includes both written and verbal communication across all commonly used languages. Natural Language Processing for written communication requires the system to read the input through cameras if input is not entered by a keyboard, and deduce its meaning. Currently available methods can scan an image for all text, printed and handwritten, detect the language and perform automated translation.[17] Systems designed for verbal communication must understand speech patterns, accents and dialects to interpret the intended message. For the military, these systems could take foreign language audio input, perform automated translation and add a complete transcription in both languages to a searchable database. Automated transcription and translation systems would allow valuable interpreters and linguists to focus on refining or validating the automated

[17] Darrell Young and Kevin Holley, "Classification of Handwriting," *Technology Today* 1, no. 1 (2018), 42-51.

transcription, instead of performing the entire translation. Each document or conversation in itself may yield information of value, but in aggregation or combined with other forms of information, it could provide valuable intelligence. In a theatre of operations, an open-source intelligence (OSINT) social media monitoring system could track the general sentiment of a population. Automated translation could give insight into any changes in how allied forces are perceived or the efficacy of information operations (IO). The transcription and translation of captured information, combined with entry into a searchable database, will significantly increase the availability of information to intelligence operators. Automation of this process will substantially reduce the time delta between when information is captured, and when it becomes available to the intelligence community in a useable format. Resources do not need to be expended to acquire intelligence if the adversary freely provides the information via social media. All it requires is the ability to gather and interpret the data to create the requisite intelligence.

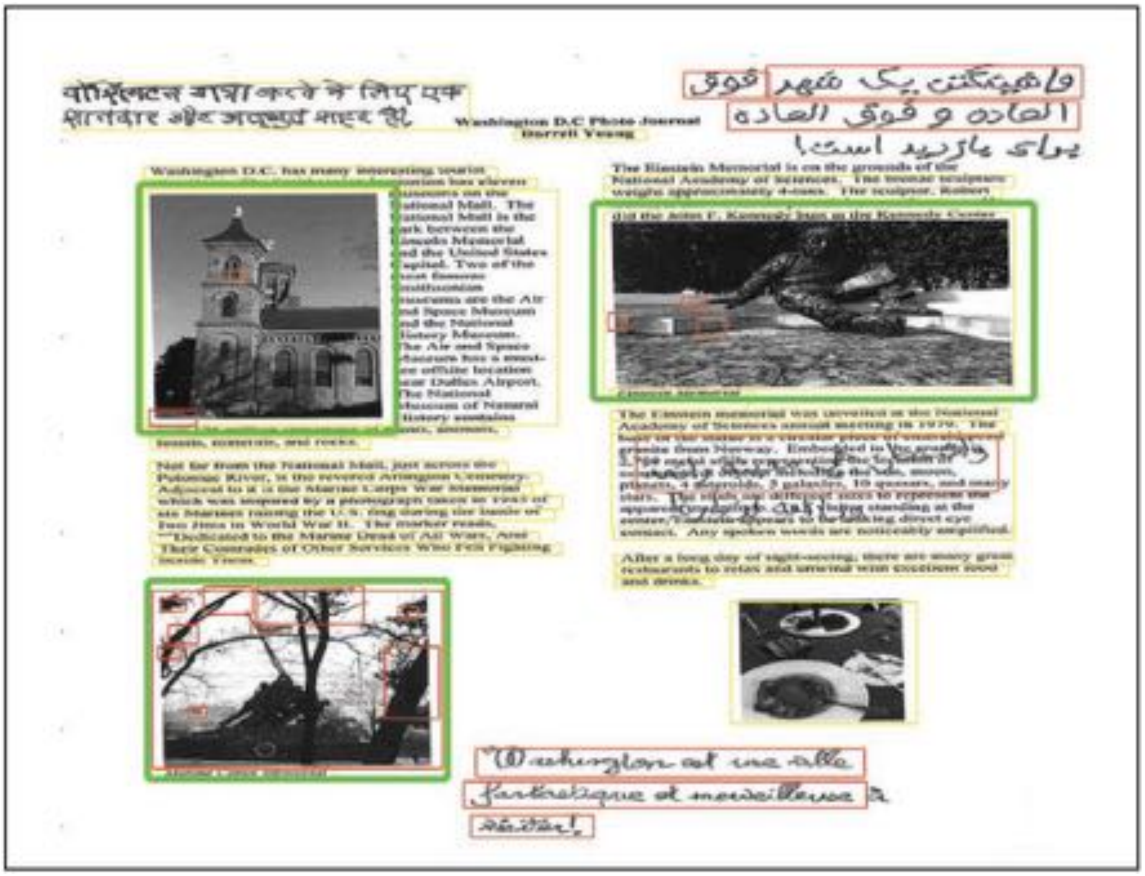

Figure 4 - Deep Learning Text Recognition
Source: Young, *Classification of Handwriting*, 46

Computers are exceptional at organising and processing large quantities of information, but are not yet sufficiently advanced to view the data in aggregation and see the bigger picture. For the foreseeable future, humans will still need to interpret the information stored and analysed by the machine learning systems and make the appropriate decisions. People, in general, are not good at understanding vast quantities of highly detailed textual information. Big data visualisation is the science of taking textual data stored in computer systems, and providing a visual representation of the data. A well-designed visualisation system enables decision-makers to understand what the information means, quickly.[18] The recent invasion of Ukraine has demonstrated the strength of visualising crowd-sourced information to give an accurate picture of the changing situation in a fluid battlespace. LiveUAMap and MapHub are two examples of systems that actively monitor social media applications for relevant updates and information.[19,20] An automated machine learning system scans each post for various features including metadata, textual information and imagery content. If the post meets all criteria to be included as an update, it is flagged for manual review before being added to the map.[21] Oryx uses social media to document Russian equipment losses and track Battle Damage Assessments (BDA).[22] Integration of open-source intelligence (OSINT) with military intelligence and other classified sources will be increasingly crucial for accurate, detailed and timely situational awareness in increasingly connected battlespaces. The ability to filter these sources automatically, while retaining a human in-the-loop, is necessary to manage the volume of information and maintain the system's quality and accuracy.

[18] Daniel Keim, Huamin Qu and Kwan-Liu Ma, "Big-Data Visualization," *IEEE Computer Graphics and Applications* 33, no. 4 (2013), 20-21.

[19] "Live UA Map," last modified May 02, accessed May 02, 2022, https://liveuamap.com.

[20] "Russia-Ukraine Monitor Map," last modified May 02, accessed May 02, 2022, https://maphub.net/Cen4infoRes/russian-ukraine-monitor.

[21] "Follow the Russia-Ukraine Monitor Map," last modified Feb 27, accessed May 02, 2022, https://www.bellingcat.com/news/2022/02/27/follow-the-russia-ukraine-monitor-map/.

[22] "Attack on Europe: Documenting Russian Equipment Losses during the 2022 Russian Invasion of Ukraine," last modified May 02, accessed May 02, 2022, https://www.oryxspioenkop.com/2022/02/attack-on-europe-documenting-equipment.html.

Figure 5 - LiveUAMap Social Media Update
Source: “Live UA Map”, May 03, 2022, https://liveuamap.com.

Private industry has a long history of using computers to identify and analyse trends. Walmart uses its vast logistics and retail network to identify shopping trends, calling it *predictive technology*, to ensure the efficient and timely distribution of products, maximising profits.[23] The underlying reason for certain trends is unknown in some instances, but the correlation has been identified and proven through data analytics. Prior to the use of predictive technology, Walmart could accurately reason that batteries and flashlights would see increased sales volumes in the week before a hurricane, but did not know that Strawberry PopTarts would also see a 700% increase in sales volume.[24] Predictive technology enables Walmart to order and preposition the correct products, at the right locations, to improve profitability. The quantity and detail of information

[23] Constance L. Hays, "What Wal-Mart Knows about Customers' Habits," *New York Times,* Nov 14, 2004.
[24] Ibid.

available to the military continues to grow at a faster pace than the increase in intelligence analysts. Similar to the advantages gained by private industry, the use of machine learning systems in the military would allow for the transformation of an exponentially increasing volume of information into actionable intelligence. Humans are creatures of habit; the sooner those habits can be identified, the sooner they can be exploited for friendly advantage.

Automated data processing and machine learning systems reduces, but does not eliminate the requirement for manual processing, while vastly increasing the quantity of data processed. Within the CAF, the vast majority of information processing is performed by the Intelligence branch, the Intelligence Operator (Int Op) trade in particular. Unfortunately, as of 26 April 2022 the Int Op trade is a stressed or “red” trade, with only 59% of its approved strength filled with trained personnel.[25] The use of automated data processing would provide the limited numbers of military specialists with the opportunity to focus on higher-level tasks, while leveraging a wider variety and larger quantity of processed data when considering intelligence questions.

[25] Department of National Defence, *Military Command Software - Personnel Dashboard*, 2022.

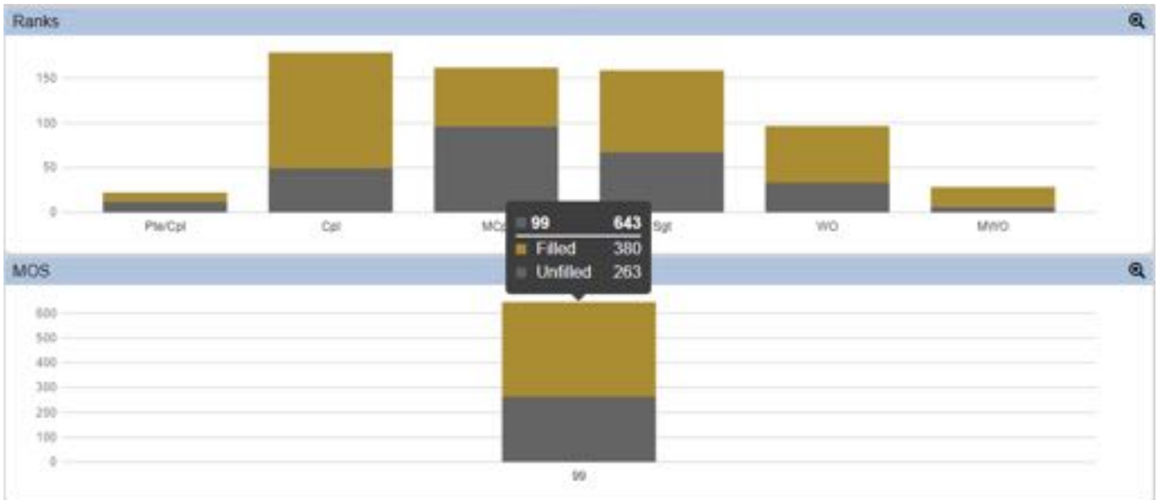

Figure 6 - Intelligence Operator Trade
Source: Military Command Software – Personnel Dashboard, 26 April 2022

RISKS AND MITIGATIONS

Integrating automated data processing and machine learning in a previously manual process is not possible without assuming some risk. The systems themselves must be developed and integrated into existing workflows. As computers, machine learning systems have similar issues and vulnerabilities as other computer systems. The systems themselves can be brought offline or corrupted by well-meaning but poorly trained users, or intentionally damaged by an adversary. The biggest barriers to most government initiatives are the costs and timelines associated with major projects. The overall cost of implementation for a complete machine learning solution is significant, impacting the procurement strategy and the associated timelines for approval, definition and implementation. Based on the history of success for government information technology (IT) procurements, it can be argued that limited financial resources would be better spent on other capabilities.

Machine learning applications recommended for adoption are widely used by private industry and provide reliable results. It is well-understood by private industry that the output of the machine learning systems will only be as good as the quality of the input data and self-learning algorithms.[26] Private industry, therefore, places significant effort on ensuring the quality and accuracy of the input data, and cross-validating the results of the algorithm with known results, using historical data as input.[27] For DND to achieve reliable and consistent results from its input data, curation of input data for each desired machine learning system would require similar levels of effort. A machine learning system designed to perform detection and categorisation of imagery requires a minimum of 1,000 manually labelled images are required to accurately train a system.[28,29] A class dedicated to the identification of vehicles could be trained to classify vehicles as aircraft, armoured vehicles or civilian vehicles. The second stage of the classification system would have individual machine learning systems dedicated to each of the vehicle sub-classes. Its output would further refine the previous label to include the specific vehicle within the image. For aircraft, images could be tagged with airframes, such as A10, F18, or Su35. Recent developments in machine learning systems are expected to reduce the requirement for manually labelled data by an order of magnitude, to 100 labelled images using pre-trained image classification models that can more rapidly learn the specific features necessary for classification.[30] However, the reduction in manual image classification effort does not reduce the importance of quality input data when training a

---

[26] Jonathan Grandperrin, *Bad Data: A $3T-Per-Year Problem with a Solution*, 2022.
[27] Bishop, *Pattern Recognition and Machine Learning*, 33.
[28] Muhammad Imran, *How Many Images are Required for Deep Learning Classification?*, 2020.
[29] Pedro Marcelino, "Transfer Learning from Pre-Trained Models," *Towards Data Science* (Oct 23, 2018). https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751.
[30] Ibid.

machine learning system. Just as human effort is required to perform manual labelling of input data to train the system, human effort is required to review and validate the output of the self-trained system. If output validation is not performed, confidence in the output accuracy will be questioned. Thus, processes and procedures must be put in place to ensure the confidence in the system, and achieve the intelligence multiplying effect of the system itself in order for this to be effectively implemented within the Canadian Armed Forces.

A skilled adversary can still exploit the best-trained machine learning system by using specially crafted input. As Figure 7 shows, Tesla hackers determined that a slight modification to a speed limit sign, which poses no difficulty for humans to read correctly, can cause a Tesla self-driving car to drive at more than twice the posted speed limit.[31]

Figure 7 - Black tape attack causing a 35-mph sign to be incorrectly read as an 85-mph sign
Source: Povolny, "Model Hacking ADAS to Pave Safer Roads for Autonomous Vehicles", McAfee Advanced Threat Research (ATR), 19 February 2020,

[31] "Model Hacking ADAS to Pave Safer Roads for Autonomous Vehicles," last modified Feb 19, accessed May 02, 2022, https://www.mcafee.com/blogs/other-blogs/mcafee-labs/model-hacking-adas-to-pave-safer-roads-for-autonomous-vehicles/.

https://www.mcafee.com/blogs/other-blogs/mcafee-labs/model-hacking-adas-to-pave-safer-roads-for-autonomous-vehicles/

TensorFlow is software frequently used as a machine learning framework. TensorFlow can be perform natural language processing of audio, but specially crafted audio input files can cause the entire machine learning system to crash.[32] An adversary could play these files at the start of any transmission to disrupt the monitoring of communications on open channels. An adversary will likely look to target every possible military system, and machine learning is no exception. Similar to all military systems, physical and procedural security measures must be emplaced to ensure the continuity of operations.

In addition to the technical vulnerabilities, institutional challenges present further difficulties in the adoption of new technologies. Failures of the Canadian government to effectively introduce new Information Technology (IT) systems have been well-publicized. The Phoenix project was introduced to modernise the existing public servant pay system, which had been in use for over 40 years.[33] Introduction of the Phoenix pay system did not follow industry best practices. Of the industry best practices, the most critical oversights were the lack of a phased implementation schedule incorporating a small number of trial users and the associated restore procedures. The Lessons Learned report highlighted these issues, along with 15 other major topic areas.[34] The Phoenix project's poor project management practices were also very costly. The initial estimated

---

[32] Qixue Xiao et al., "Security Risks in Deep Learning Implementations" IEEE, May 2018. 3.

[33] Kevin Sorenson, *Report 1, Building and Implementing the Phoenix Pay System, of the 2018 Spring Reports of the Auditor General of Canada,* House of Commons, 2018. 1.

[34] "Lessons Learned from the Transformation of Pay Administration Initiative," last modified Oct 04, accessed 02 May, 2022, https://www.canada.ca/en/treasury-board-secretariat/corporate/reports/lessons-learned-transformation-pay-administration-initiative.html.

cost for the Phoenix project was $309 million in 2009, but by 2018 the actual cost had tripled to $954 million, and the final project cost could be as high as $2.6 Billion.[35] DND has also not escaped the failures of government IT procurements. The Defence Resource Management Information System (DRMIS) supports DND financial and logistical functions. The DND project to extend the life of the system was initiated in 2018, expects to begin the definition phase in 2023/24, with an overall cost estimate of $250-499 million.[36] According to the project plan itself, initial delivery of the updated DRMIS system is not expected until 2033/2034, a full *ten years* after project initiation.[37] Canada is certainly not alone in its difficulties in integrating modern IT into its military. US DoD, in partnership with Google, launched project MAVEN. Project MAVEN sought to develop "…computer-vision algorithms needed to help military and civilian analysts encumbered by the sheer volume of full-motion video data that DoD collects every day."[38] High-profile protests at Google led to the cancellation of project MAVEN, but not the partnership between DoD and Google.[39,40] US DoD understands the benefit of AI in military applications and has sought an allocation of (USD) $874 million specifically to enable AI in defence operations.[41]

---

[35] CBC News, *Phoenix Pay System Cost could Total $2.6B before Cheaper Replacement Ready*, 2019. https://www.cbc.ca/news/canada/ottawa/phoenix-pay-system-cost-report-1.5138036

[36] "Defence Resource Management Information System," last modified Jan 09, accessed 02 May, 2022, http://dgpaapp.forces.gc.ca/en/defence-capabilities-blueprint/project-details.asp?id=1912.

[37] Ibid.

[38] Cheryl Pellerin, *Project Maven Industry Day Pursues Artificial Intelligence for DoD Challenges* (Washington: Federal Information & News Dispatch, LLC, [2017]). 2.

[39] *Letter to Google CEO - Project Maven*, 2018).

[40] Scott Shane and Daisuke Wakabayashi, "'The Business of War': Google Employees Protest Work for the Pentagon," *NYTimes,* Apr 4, 2018.

[41] Office of the Under Secretary of Defense Chief Financial Officer, *Defense Budget Overview - United States Department of Defense Fiscal Year 2022 Budget Request*, [2021]. 3-2.

Labour is one of the largest costs for most organisations, averaging between 50% to 60% of expenditures, and the military is no exception.[42,43] Despite the challenges and high-profile failures of government IT procurements, and the monetary cost of implementing a machine learning system within DND, manual processing of exponentially increasing quantities of information is increasingly untenable. A streamlined solution aligned with the current institutional procedures must be sought in order to incorporate technological efficiencies into the current intelligence process. If there is either a failure in the adoption of the technological enhancement, or an integration failure, the intelligence processing chain will not be able to adequately support the operational community in the face of increasing information access.

DND cannot continue to be risk-averse in its adoption of a mature technology. Private industry, our allies, and our adversaries are all adopting machine learning technologies to increase their productivity. As with all military systems, equipment, processes, and procedures must be established to mitigate the risk and minimize the impact of any disruption. If DND wishes to retain operational relevance in a rapidly changing threat environment that is increasingly leveraging technological efficiencies, it must seek alternative methods to process data and generate useable intelligence.

HOW TO IMPLEMENT MACHINE LEARNING

The benefits of automation and machine learning will require significant effort and investment in the modernization of infrastructure, centralisation of data, and

---

[42] Deloitte, *LabourWise*, 2017.

[43] Jon Harper, *Pentagon Personnel Costs at Historic High*, Vol. 106 (Arlington: National Defense Industrial Association, 2021), 10.

interconnectivity of networks. Attempts to reduce costs by scoping or scaling the solution will lead to inferior results, interoperability failures, and higher overall monetary and opportunity costs.

For this discussion, we will assume the use of an intelligence processing chain that primarily operates on digital products, but does not employ any automation during intelligence data processing. This digitized, but labour-intensive organisation would benefit from automating routine tasks, leveraging mature technologies developed by private industry. This intelligence organisation operates on digital products in disparate locations and in proprietary formats. More than 80% of the raw information received by the organisation originates from UNCLAS data sources, but is typically hosted on higher classification systems. We will also assume that this organisation operates Level I (DESIGNATED), Level II (SECRET) and Level III (TOP SECRET) systems, and the number of users drops by an order of magnitude for every increase in the security classification of the network, such that there are 100,000 Level I users, 10,000 Level II users, and 1,000 Level III users.

For this intelligence organisation to maximise the opportunities for machine learning to assist in their processing chain, physical and virtual changes must occur.

1. Proprietary data formats must be converted to more open, accessible formats before being centrally hosted as unstructured data inside of a Data Lake.[44]
2. Raw data must be stored at the lowest possible classification.

---

[44] Fatemeh Nargesian et al., "Data Lake Management," *Proceedings of the VLDB Endowment* 12, no. 12 (Aug, 2019), 1986-1989. 1987.

3. Transfer Cross-Domain Solutions (Transfer CDS) must be installed to permit queries from higher-level networks to lower-level networks. Data Diodes must be installed to return the requested data from the lower-level networks to the higher-level networks.

4. Processed data, intelligence, must be stored in databases tuned for their specific function.

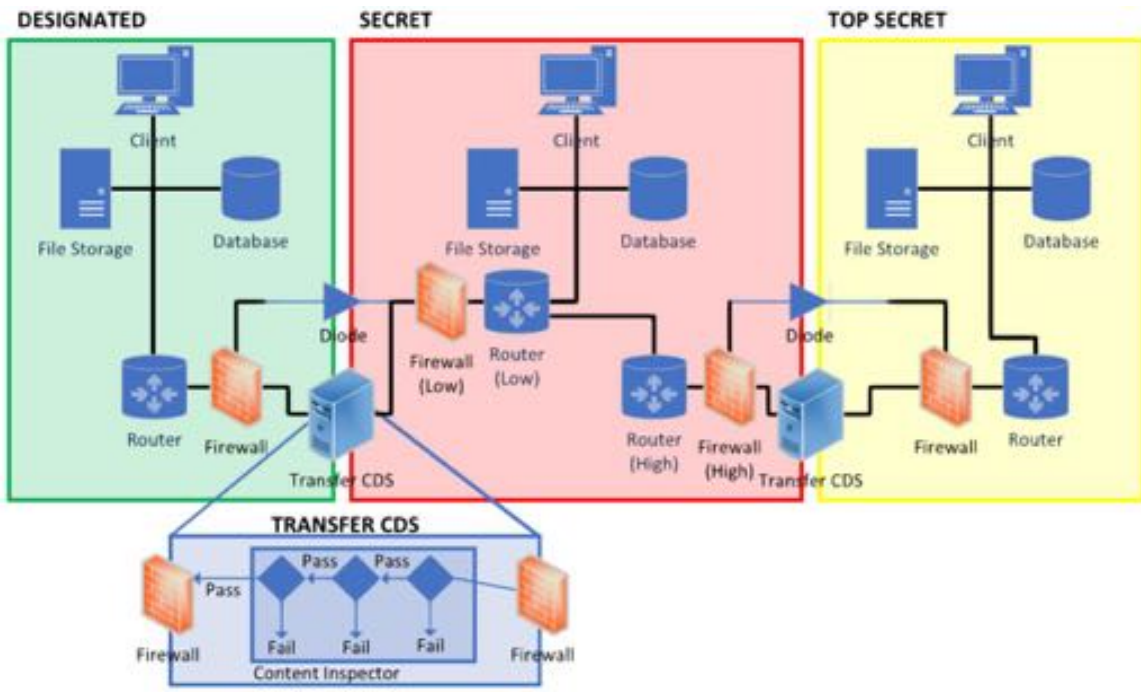

Figure 8 - Cross-Domain Transfers
Source: Diagram by author, Adapted from Smith, *Shedding Light on Cross Domain Solutions*, Fig 3.

Commercial software often relies on proprietary data formats, restricting access to applications outside the vendors' ecosystem. Notable examples of applications with strong controls due to proprietary data formats include ESRI for Geo Products and

AutoDesk for Engineering. Proprietary data formats do not allow for the synthesis and processing of a variety of data sources, inhibiting the benefits of a large data repository.[45]

Storage of raw data at the lowest possible classification permits access to the information for the greatest number of users, while also enabling cost efficiencies. It is cheaper to store data at a lower classification level, retaining more expensive processing and storage capacity on higher networks for information and intelligence requiring additional safeguards.

Although the raw data itself may originate from an UNCLAS data source, once processed, the additional context given to the data may result in intelligence of a higher classification. Therefore, intelligence analysts operating on higher classification networks will require access to the raw data stored on lower classification networks. Transfer Cross-Domain Systems (CDS) are most reliable when used with structured data, such as eXtended Markup Language (XML) and Structured Query Language (SQL).[46] A CDS will permit a user on a higher classification network to query the data lake located on a lower classification network. A data diode would be used to return the results to the user. A data diode differs from a Transfer Cross-Domain System, as it only allows data to travel from lower to higher-level networks. This restriction dramatically reduces the cost and complexity of the device. Although the Transfer CDS could be used to return the data, it is not recommended for use in low latency (streaming video) or high bandwidth (large data sizes) applications due to the intensive data processing requirements.

---

[45] Paul Sawers, *Breaking 'bad Data' with Machine Learning*, 2021.

[46] Raymond Haakseth et al., "A High Assurance Cross-Domain Guard for use in Service-Oriented Architectures" Military University of Technology, 2015. 4.

Returns of multiple queries would be processed by the tools used by the analyst, and compiled into intelligence. The contextual information, but not the raw data, would be stored in a tuned database on the higher classification network. As shown in Table 1, reading binary files “BLOBs” (images, video) from within a database is 5-20 times slower than a database accessing files stored in the filesystem. Therefore, it is preferable to reference the raw data sources in the filesystem by location or unique hash, to segregate the intelligence database from the raw data store.

	350 MB data	4.5 MB data
file system	46 ms	1 ms
Large Object	950 ms	8 ms
bytea	590 ms	6 ms

Table 1 - PostgreSQL Database Performance
Source: Laurenz Albe, Binary Data Performance in PostgreSQL, 2020

Following the consolidation of application data into data lakes, and establishing controlled connectivity between network classifications, the machine learning systems can be integrated into the networks. An overall design should be created to facilitate phasing and integration. Even if the end state is never reached, the standardisation of system components and documentation of each machine learning system will enable future upgrades and expansion of the network.

Plans Officers within the Intelligence organisation should create a prioritised list of the datasets to be processed. This prioritised list will help guide the order in which machine learning systems are configured and implemented, to provide the largest benefit to the intelligence community, at the earliest possible time.

The intelligence organisation must not permit the machine learning specialists to independently install and configure the system. To create useable intelligence from the input data, the machine learning system parameters must be created. The computer needs to be told what is important, and how to categorise the input data. Canada is home to several global companies whose principal product lines leverage machine learning. Kinaxis is an Ottawa-based company that improves supply-chain management for corporations by incorporating AI-enhanced software. Folio3 is a Toronto-based company that delivers an image processing engine used to detect and categorise videos and imagery for multinationals, including Honda and Colgate. The CAF can leverage Canadian technical expertise, by contracting machine learning specialists to configure and optimise the system for the analysis Intelligence experts require. If the machine learning specialists and the intelligence specialists do not work together as a cohesive team, the system will not work as intended.

Assume the intelligence analysts provide input data containing pictures of household pets and want the system to group the pets into categories of cat, dog, fish, bird, and other. Suppose the machine learning and intelligence specialists do not communicate. In that case, the machine learning specialists may look at the raw input data and perceive that the intelligence specialists want to know how many animals are depicted in each image. The same set of images is given as input, but the system output does not match the intelligence analysts' desired output. Additionally, as the system was configured to detect quantities within an image, and not group images by characteristics within the image, the effort expended in programming and training the system is entirely wasted.

Once the initial machine learning systems are created and conduct initial analysis on the raw input data, subsequent layers of machine learning can be leveraged to develop higher levels of understanding. Higher-level analysis is typically more abstract, and therefore requires a clear and complete description of the desired result. Trend analysis is frequently implemented as a higher-level machine learning system, taking the output from multiple first-level machine learning systems to identify commonalities or factors relating to the desired result. The finance sector has long led the adoption of computer technologies, from automated stock market trading programs in the early 1980s to the increasingly sophisticated personal credit risk analysis software that consolidates all individual financial information into a single number – A credit score.

Credit underwriters use increasingly detailed and sophisticated machine learning systems to identify trends and reduce their exposure to risks.[47] Lenders previously knew the income, debts and credit rating of the individual or organisation requesting the business loan. With the advent of machine learning, business loan requests can now be overlaid with the physical geography, demographics, loan-specific factors like the number of competing businesses, and regional economic forecasts for the area before approving a loan.[48] The increasing quantity and fidelity of the information allows businesses to reduce their risk level, and increase transparency, all while removing human error and bias from the application review.

Ideally, the integration of machine learning systems into the intelligence organisations process enables accurate analysis of large quantities of data, while reducing the requirement for analysts to manually process the information. The increase in the information processed, coupled with the reduced requirement for manual processing,

---

[47] Kelly Thompson Cochran et al., *The use of Machine Learning for Credit Underwriting,* FinRegLab, [2021]. 9.

[48] Ibid. 97.

provides the experienced analysts with the opportunity to query the massive intelligence datasets and more rapidly respond to intelligence requests.

CONCLUSION

Automated data processing and machine learning are not new fields. These systems are in active use by private industry to increase their productivity and leverage their information to assist in business decision-making. Allies and adversaries, in partnership with industry, are adopting machine learning to automate the analysis of growing volumes of data.[49] China's 14th five-year plan (2021-2025) prioritised the development of multiple key technologies, including AI, to become the "world's major AI innovation center" by 2030.[50] The United States has created the Joint Artificial Intelligence Center (JAIC) to assist in the integration AI within DoD.[51]

The Canadian military does not have the size or scale to embrace revolutionary AI technologies like larger nations, but this does not preclude all advancements in this domain. The Canadian Armed Forces must transition from a manually driven intelligence process to one supported by automation to efficiently process increasing quantities of data. Manual processing of information is not scalable or responsive to the organization's needs, limiting the opportunity for skilled Intelligence Operators to add value with higher-level analysis, and becoming an operational liability in the modern threat environment.

---

[49] Nizam Uddin Ahmed, *Integrating Machine Learning in Military Intelligence Process: Study of Futuristic Approaches Towards Human-Machine Collaboration* (Dhaka, Bangladesh, [2022]).67.

[50] Department of Defense, *Military and Security Developments Involving the People's Republic of China,* Office of the Secretary of Defense, [2021]. 146.

[51] Department of Defense, *Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance our Security and Prosperity,* U.S. Department of Defense, [2019]. 9.

The Canadian Armed Forces has severe personnel shortages across all trades, but these shortages are most acute in the skilled trades, including Intelligence Operators. Not only is manual processing of information human-resource intensive, but the ability for the CAF to perform manual processing of information is further reduced by the limited number of personnel available to complete the task.

Therefore, the CAFs lethargic adoption of automation strategies has two compounding, negative effects on the organisation. It cannot maintain the pace of converting information into intelligence at the same pace as our allies and adversaries, in an era when volume and access to raw data is increasing. The CAF is also not using its limited personnel resources effectively, splitting focus between analysis of raw data and high-value analysis of processed intelligence. Adoption of well-established automation systems would simultaneously increase the quantity of processed information while creating the time for skilled Intelligence Operators to further refine the information into highly actionable intelligence.

If the CAF was a private corporation, its board of directors would likely be questioning its efficiency, referencing the gap between the CAF and competing organisations, before mandating that programs be instituted to improve its effectiveness. However, the CAF is not a private industry, and if programs are not instituted, it will not go bankrupt. In defence, a capability gap between two nations can be so broad, as to become operationally insurmountable. The CAF must institute efficiencies in its intelligence processing, better identifying threats, and reducing the intelligence gap between ourselves and our increasingly sophisticated adversaries.

BIBLIOGRAPHY

Letter to Google CEO - Project Maven, 2018.

Ahmed, Nizam Uddin. *Integrating Machine Learning in Military Intelligence Process: Study of Futuristic Approaches Towards Human-Machine Collaboration*. Dhaka, Bangladesh, 2022.

Albe, Laurenz. "Binary Data Performance in PostgreSQL." Accessed May 02, 2022. https://www.cybertec-postgresql.com/en/binary-data-performance-in-postgresql/.

Ammann, Manuel, Kristian Blickle, and Christian Ehmann. "Announcement Effects of Contingent Convertible Securities: Evidence from the Global Banking Industry." *European Financial Management : The Journal of the European Financial Management Association* 23, no. 1 (Jan, 2017): 127-152.

Asian Development Bank. *The 14th Five-Year Plan of the People's Republic of China - Fostering High-Quality Development Observations and Suggestions 观察与建议*. Manila, Philippines: Asian Development Bank, 2021.

Bellam, Shankar. *Robotics Vs Machine Learning Vs Artificial Intelligence: Identifying the Right Tools for the Right Problems*, 2018.

Bendett, Samuel, Mathieu Boulègue, Richard Connolly, Margarita Konaev, Pavel Podvig, Katarzyna Zysk Russia, and Eurasia Programme. *Advanced Military Technology in Russia Capabilities and Implications* 2021.

Bishop, Christopher M. *Pattern Recognition and Machine Learning*. New York, NY: Springer, 2006.

CBC News. *Phoenix Pay System Cost could Total $2.6B before Cheaper Replacement Ready,* 2019.

Cen4infoRes. "Russia-Ukraine Monitor Map." Accessed May 02, 2022. https://maphub.net/Cen4infoRes/russian-ukraine-monitor.

Cochran, Kelly Thompson, Colin Foos, Hilary Griggs, Tess Johnson, Mashrur Khan, Duncan McElfresh, Kerrigan Molland, P-R Stark, YaYa Sun, and Sormeh Yazdi. *The use of Machine Learning for Credit Underwriting*: FinRegLab, 2021.

Congressional Research Service. *AI and National Security* 2020.

Deloitte. *LabourWise* 2017.

Department of Defense. *Military and Security Developments Involving the People's Republic of China*: Office of the Secretary of Defense, 2021.

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

Department of National Defence. *Canadian Forces Joint Publication (CFJP) 2-0 - Intelligence* 2011.

———. *Military Command Software - Personnel Dashboard*, Accessed 03 May, 2022.

Explorium. *AI is Making BI Obsolete, and Machine Learning is Leading the Way*.

Gilroy Report. "Lessons Learned from the Transformation of Pay Administration Initiative." Accessed 02 May, 2022. https://www.canada.ca/en/treasury-board-secretariat/corporate/reports/lessons-learned-transformation-pay-administration-initiative.html.

Goertzel, Ben and Cassio Pennachin. *Artificial Brains*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

Government of Canada. "Defence Resource Management Information System." Accessed 02 May, 2022. http://dgpaapp.forces.gc.ca/en/defence-capabilities-blueprint/project-details.asp?id=1912.

Grandperrin, Jonathan. *Bad Data: A $3T-Per-Year Problem with a Solution* 2022.

Haakseth, Raymond, Nils Agne Nordbotten, Oyvind Jonsson, and Bengt Kristiansen. "A High Assurance Cross-Domain Guard for use in Service-Oriented Architectures."Military University of Technology, 2015.

Harper, Jon. *Pentagon Personnel Costs at Historic High*. Vol. 106. Arlington: National Defense Industrial Association, 2021.

Hays, Constance L. "What Wal-Mart Knows about Customers' Habits." *New York Times,* Nov 14, 2004.

Imran, Muhammad. *How Many Images are Required for Deep Learning Classification?* 2020.

Jajal, Tannya D. "Distinguishing between Narrow AI, General AI and Super AI." *Mapping Out 2050,* May 21, 2018.

Kaniyar, Sanjay, Kapil Bhushan Srivastava, and Ross Tisnovsky. *Driving Impact at Scale from Automation and AI* 2019.

Keim, Daniel, Huamin Qu, and Kwan-Liu Ma. "Big-Data Visualization." *IEEE Computer Graphics and Applications* 33, no. 4 (2013): 20-21.

Larkin, Zoe. "General AI Vs Narrow AI." https://levity.ai/blog/general-ai-vs-narrow-ai.

Marcelino, Pedro. "Transfer Learning from Pre-Trained Models." *Towards Data Science* (Oct 23, 2018). https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751.

Marlin, Elizabeth. "Using Artificial Intelligence to Minimize Information Overload and Cognitive Biases in Military Intelligence." Fort Leavenworth, KS : US Army Command and General Staff College.

Mitzer, Stijn. "Attack on Europe: Documenting Russian Equipment Losses during the 2022 Russian Invasion of Ukraine." Accessed May 02, 2022. https://www.oryxspioenkop.com/2022/02/attack-on-europe-documenting-equipment.html.

Nargesian, Fatemeh, Erkang Zhu, Renée J. Miller, Ken Q. Pu, and Patricia C. Arocena. "Data Lake Management." *Proceedings of the VLDB Endowment* 12, no. 12 (Aug, 2019): 1986-1989.

Norris, Michael. *The Value of AI-Powered Business Intelligence*: O'Reilly Media, Inc, 2019.

Office of the Under Secretary of Defense Chief Financial Officer. *Defense Budget Overview - United States Department of Defense Fiscal Year 2022 Budget Request*, 2021.

Oza, Nikunj C., Chad Stephens, and Kaitlyn Dozier Fox. "Enabling Safety from Data: Machine Learning/Artificial Intelligence for in-Time Aviation Safety." (2021).

Pellerin, Cheryl. *Project Maven Industry Day Pursues Artificial Intelligence for DoD Challenges*. Washington: Federal Information & News Dispatch, LLC, 2017.

Povolny, Steve. "Model Hacking ADAS to Pave Safer Roads for Autonomous Vehicles." Accessed May 02, 2022. https://www.mcafee.com/blogs/other-blogs/mcafee-labs/model-hacking-adas-to-pave-safer-roads-for-autonomous-vehicles/.

Qixue Xiao, Kang Li, Deyue Zhang, and Weilin Xu. "Security Risks in Deep Learning Implementations."IEEE, May 2018.

Rozhkovsky, Rodion and Alexander Bil'chenko. "Live UA Map." Accessed May 02, 2022. https://liveuamap.com.

Sawers, Paul. *Breaking 'bad Data' with Machine Learning* 2021.

Schwab, Klaus, Nicholas Davis, Inc OverDrive, and OverDrive ebook. *Shaping the Future of the Fourth Industrial Revolution*. 1st American ed. New York: Currency, 2018.

Shane, Scott and Daisuke Wakabayashi. "'The Business of War': Google Employees Protest Work for the Pentagon." *NYTimes,* Apr 4, 2018.

Sikora, Sara, Blythe Hurley, and Anya George Tharakan. *Automation with Intelligence : Reimagining the Organisation in the 'Age of with'*.

Smith, Scott D. *Shedding Light on Cross Domain Solutions* Department of National Defence, 2015.

Sorenson, Kevin. *Report 1, Building and Implementing the Phoenix Pay System, of the 2018 Spring Reports of the Auditor General of Canada*. Report of the Standing Committee on Public Accounts. House of Commons, 2018.

Strick, Benjamin. "Follow the Russia-Ukraine Monitor Map." Accessed May 02, 2022. https://www.bellingcat.com/news/2022/02/27/follow-the-russia-ukraine-monitor-map/.

Sungrim Cho, Woochang Shin, Neunghoe Kim, Jongwook Jeong, and Hoh Peter In. "Priority Determination to Apply Artificial Intelligence Technology in Military Intelligence Areas." *Electronics (Basel)* 9, no. 2187 (Dec 01, 2020).

Tekin, Muhammet. *New Perspectives on Intelligence Collection and Processing*, 2016.

Xiong, Yunsheng, Xin Niu, Yong Dou, Hang Qie, and Kang Wang. "Non-Locally Enhanced Feature Fusion Network for Aircraft Recognition in Remote Sensing Images." *Remote Sensing* 12, no. 4 (Feb 19, 2020).

Young, Darrell and Kevin Holley. "Classification of Handwriting." *Technology Today* 1, (2018a): 42-51.