

Advancing Cyber Assurance with Deep Learning

JOSEPH T. MATHEWS

*Center for High Assurance Computer Systems Branch
Information Technology Division*

DR. TRACY BRAUN

*Army Research Laboratory
Adelphi, MD*

April 20, 2022

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 20-04-2022		2. REPORT TYPE NRL Memorandum Report		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE Advancing Cyber Assurance With Deep Learning				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Joseph T Mathews and Dr. Tracy D Braun*				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER 6010	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Research Laboratory 4555 Overlook Avenue, SW Washington, DC 20375-5320				8. PERFORMING ORGANIZATION REPORT NUMBER NRL/5540/MR--2022/3	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Naval Research Laboratory 4555 Overlook Avenue, SW Washington, DC 20375-5320				10. SPONSOR / MONITOR'S ACRONYM(S) NRL	
				11. SPONSOR / MONITOR'S REPORT NUMBER(S) 5500 OH	
12. DISTRIBUTION / AVAILABILITY STATEMENT DISTRIBUTION STATEMENT A: Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES *DEVCOM Army Research Laboratory, 2800 Powder Mill Road, Adelphi, MD 20783					
14. ABSTRACT We summarize the results of a Research Specialists Workshop on Deep Machine Learning for Cyber Defence held virtually on 19-21 October 2021. We report to NATO Science and Technology Organisation on current and emerging cybersecurity applications of deep learning analysed at this forum via our paper. The workshop objectives were to introduce new perspectives, expose government research in related areas, illustrate how deep learning applies to cybersecurity, and introduce operational needs in the application of deep learning to military operations in cyberspace. Overall, the outcomes were greater awareness of the problem and opportunities, establishment of common needs across application domains, and documenting a path forward.					
15. SUBJECT TERMS Machine learning Deep learning Artificial intelligence Cybersecurity Cyber defense Military operations					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT U	18. NUMBER OF PAGES 24	19a. NAME OF RESPONSIBLE PERSON Joseph Mathews
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (include area code) (202) 404-8888

This page intentionally left blank.

ADVANCING CYBER ASSURANCE WITH DEEP LEARNING

Joseph Mathews

Naval Research Laboratory
4555 Overlook Ave SW Washington DC 20375
UNITED STATES

joseph.mathews@nrl.navy.mil

Dr. Tracy Braun

Army Research Laboratory
2800 Powder Mill Rd Adelphi MD 20783
UNITED STATES

tracy.d.braun.civ@army.mil

ABSTRACT

We summarize the results of a Research Specialists Workshop on Deep Machine Learning for Cyber Defence held virtually on 19-21 October 2021. We report to NATO Science and Technology Organisation on current and emerging cybersecurity applications of deep learning analysed at this forum via our paper. The workshop objectives were to introduce new perspectives, expose government research in related areas, illustrate how deep learning applies to cybersecurity, and introduce operational needs in the application of deep learning to military operations in cyberspace. Overall, the outcomes were greater awareness of the problem and opportunities, establishment of common needs across application domains, and documenting a path forward.

1.0 INTRODUCTION

Since the advent of the modern computer in the latter half of the 20th century, humans have programmed all software and served as principal drivers of computational and algorithmic advances. As of the early 21st century, however, practical advances in Deep Learning have shifted the software landscape. Deep learning enables computers to “program” their own software by training models which describe the relationships between inputs and outputs. The algorithmic breakthrough is accelerating advances in every industry with wild success. The most popular applications include those which can recognize objects [1] and translate speech [2] with human levels of precision in near real time. Experts ambitiously suggest that deep learning will eventually be able to “do everything” and perhaps even replicate human intelligence [3].

At the same time, a growing reliance on software reinforces the importance of protecting computer systems and networks from compromise or disruption to the services they provide. The rate and impact of data breaches over the first decades of the 21st century further illustrate how cyber intrusions can reshape the global security landscape. Demand for an increasingly resilient cyberspace, especially as it intersects with military systems, is therefore motivating many novel applications of deep learning. These applications may enhance strategic military positioning and create a resilient cybersecurity posture that keeps pace with evolving threats. Achieving this outcome, however, requires interdisciplinary applied research and experimentation in order to truly understand limitations and practical utility.

Manuscript approved April 15, 2022.

We therefore summarize the results of a research workshop on Deep Machine Learning for Cyber Defence held virtually on 19-21 October 2021¹. This NATO Science and Technology Organization (STO) forum focused on consolidating knowledge in the field of deep learning applications for cyber defence. Participants included research scientists and engineers from Australia, Belgium, Finland, France, Germany, Italy, Norway, Poland, Turkey, United Kingdom, and the United States. Organizational representation comprised a robust mix of perspectives from universities, civil research organizations, national defence establishments, and military research laboratories.

The purpose of the forum was to foster collaboration among NATO countries and allies in order to identify and pursue the most promising deep learning use cases and methods for the cyber domain, including computational techniques, architectures, and data sets or models. Towards this vision, it served to increase awareness of the symbiotic relationship between both topics: Deep learning stands to benefit hard problems in cybersecurity by automating laborious aspects of continuous monitoring. Cybersecurity, on the other hand, stands to benefit practical applications of deep learning with robust implementation design. Further, as deep learning applications proliferate and increasingly interact with the physical world (i.e., autonomous systems), the boundaries that have traditionally characterized and insulated cyberspace will erode. Consequently, realizing *cyber assurance* will necessitate new approaches beyond those traditionally applied for cybersecurity.

Dr. Frederica Free-Nelson, US Army Research Laboratory, noted in workshop-opening remarks that both the fields of deep learning and cybersecurity hold many open problems, and rather than relying on a few major contributors for solutions, it is important to share processes, approaches, and success stories in order to avoid wasting resources or impeding progress. Real world, user-driven problems appropriately matched to fundamental research and applied experimentation enable information dominance and decision advantage. It is therefore the desired vision of this workshop, in part, to capture those findings which inform leadership of the need to commit resources to ongoing challenges and connect scientists, practitioners, and end-users in a fashion amenable to repeatable success and continued advancement.

This paper is structured in accordance with the objectives of the workshop. Section 2 introduces terminology and perspectives that bound the problem space and shape potential solutions. Section 3 illustrates how deep learning is already applied to cybersecurity and proposes opportunities for further gain. Section 4 exposes related efforts within NATO STO and draws parallels among multiple application domains. Section 5 concludes by highlighting key findings and considerations for military environments. Ultimately, we aim to raise awareness of the value that deep learning provides to cybersecurity in the military context and identify opportunities ripe for exploration.

2.0 PERSPECTIVES FROM THE INTERSECTION OF CYBERSECURITY AND DEEP LEARNING

Cyberspace, as defined by US military doctrine [4], is the domain characterized by the use of electronics, electromagnetic spectrum, and software to store, modify, and exchange data via networked systems and associated physical infrastructure. This encompasses microelectronics, computing, communication, networking, and software technologies that include artificial intelligence, machine learning, and deep learning. Applications of cyberspace technologies underpin all economic sectors, critical infrastructures, and military operations. Technological trends that will continue evolving cyberspace include ubiquitous connectivity and sensing at the network edge, increased programmability and complexity of systems, applications of autonomy and accelerated decision loops,

¹ NATO STO, 'IST-183 Research Workshop on Deep Machine Learning For Cyber Defense (Virtual)', *NATO Science & Technology Organization Events*, <https://events.sto.nato.int/index.php/upcoming-events/event/18-workshop/369-ist-183-research-workshop-on-deep-machine-learning-for-cyber-defense>, (accessed 18 March 2022).

increasingly untrustworthy and opaque supply chains, and novel computing architectures (i.e., quantum and neuromorphic computing). Non-technological trends include a growing number of internet users, aggressive user metadata exploitation for consumer industry analytics, and international diplomacy or national defence considerations. Given the aggressive rate of technological change and uncertainty surrounding non-technological trends, cyberspace will continue evolving in ways that may be difficult to predict with precision.

Assuring cyberspace in the military context comprises two distinct missions: cybersecurity and cyber defence. Cybersecurity aims to limit vulnerability by assuring critical system properties such as confidentiality, integrity, and availability. As cyber-physical systems such as critical infrastructure, smart manufacturing, weapons systems, and perhaps eventually biological-neurological interfaces proliferate, cybersecurity increasingly encompasses non-traditional properties that include safety, timeliness, and resilience. Further, the physical nature of these systems affords novel instrumentation and telemetry by which to assure their cyber posture [5]. Cyber defence, on the other hand, describes actions taken to respond to adversarial manoeuvres in cyberspace. Though these roles may overlap in some organizations, the distinction exists because of how military units organize and execute mission essential functions. Cybersecurity is the responsibility of those who design, develop, and operate a given system. Cyber defence, however, is the responsibility of some focused activity chartered specifically to monitor, react, and orchestrate responses to adversarial threat across the entire organization (i.e., Security Operations Centre).

Cyberspace is fundamental to the strategic military landscape, and NATO countries must mitigate cyber threats to their military systems, platforms, and missions. Deep learning is an emergent software technology with applications capable of bolstering such a resilient posture. For that reason, the Information Systems Technology Panel of NATO STO established a Research Task Group (RTG) on “Deep Machine Learning for Cyber Defence”². In workshop-opening remarks, Mr. Raphael Ernst, Fraunhofer FKIE (Germany), clarified the RTG’s charter was not to develop new deep learning techniques, but to consolidate NATO-wide knowledge in the applications of deep learning to cyber defence, identify gaps between civilian solutions and military needs, and collaborate with other NATO countries to use data processing, share data, and pursue the transfer of the most promising technologies and applications to the military domain. The RTG, comprising a mix of cybersecurity and machine learning experts, reviewed a thorough selection of technology standards, academic research, and commercial technology offerings to assess current state of the art. The results of that study on the current technological landscape is summarized in Section 3.

Ongoing advances in the field of artificial intelligence and proliferation of cyber-physical systems will, however, alter the cyber landscape and give way to new classes of cyber-attacks. Cyber-physical systems employ software to control mechanisms that intertwine with their physical environment, operate on mixed temporal scales, and interact in ways that change with context. Self-driving vehicles, for instance, apply deep learning to onboard cameras in order to see and make decisions about how to drive on roads. Studies show how physical world attacks on these software systems may cause harm [6]. Driverless vehicles further rely on persistent connectivity to share telemetry information with other devices, networks, and vehicles. Though intended as a mechanism for feedback, this and similar cyber-physical system designs expose attack surface [7].

Ultimately, assuring the operation of a cyber-physical system grows increasingly difficult, and a resilient cyber posture necessitates means beyond traditional cybersecurity methods. Therefore, characterizing the open challenges among a variety of deep learning applications is crucial for understanding cyber risk. The RTG initiated

² NATO STO, ‘Deep Machine Learning For Cyber Defense’, *NATO Science & Technology Organization*, <https://www.sto.nato.int/SitePages/newsitem.aspx?ID=3576&IsDlg=1>, (accessed 18 March 2022).

the workshop to advance this agenda among a group of diverse experts with insight into various military and civil applications. Given the engagement across distinct fields, a discussion of terminology will provide useful context.

Artificial Intelligence (AI) generally describes any technique enabling computers to mimic human intelligence. Early successes in AI stemmed from rule-based systems and systems that captured human expert knowledge. Though taxonomies by which to categorize artificial intelligence capabilities exist, we select a simple distinction between Machine Learning and Deep Learning to frame our discussion. The workshop's adoption of the "Deep Machine Learning" nomenclature signals a principle focus on deep learning that does not exclude traditional machine learning, yet acknowledges that continued advances will cement deep learning's stature as the most prominent technique in the field of Artificial Intelligence.

Machine Learning is the most important subset of AI techniques, and provides the ability to improve computational performance by discovering patterns in data without the need to follow explicitly programmed instructions. After decades of slow progress, Machine Learning has recently garnered wide adoption in a variety of applications that include consumer analytics and social media. Machine-learning algorithms use statistics to find patterns in massive quantities of data that encompass numbers, words, images, or other unique representations of digital information [8]. Machine learning applications employ a pipeline that can be generalized among four stages. First, data acquisition involves identifying and collecting datum elements. Second, feature engineering involves pre-processing or extracting statistics regarding that data. Third, the initial data or derived statistics are used to train a model capable of recognizing patterns and relationships. Lastly, the model is evaluated or deployed with input data that may or may not reflect the population of training data from the initial stages. Though characterized linearly, the stages are often implemented iteratively and with considerable feedback and tuning amongst them. Ultimately, data dependence and quality inform the effectiveness of each application.

Deep Learning refers to a specific subset of Machine Learning techniques that permit models to train themselves by exposing multi-layered neural networks to vast amounts of data. The distinction lies in specifically the second and third stages of the aforementioned pipeline. Neural networks are a simplified digital model of the neurons and synapses found in the human brain and comprise layers of simple computational nodes that process data. While early neural networks were limited to few layers of neurons, a breakthrough technique known as backpropagation enabled scaling those layers theoretically, thereby paving the way for "deep" neural networks made up of many more layers [9]. Decades after the initial discovery, increased computational power enabled a highly successful demonstration of deep neural networks' ability to classify imagery, subsequently establishing it as state of the art [10]. The increased computational power afforded by emergent Graphics and Tensor Processing Unit hardware have further accelerated exploitation of increasingly larger datasets, benefitting a wide range of pattern recognition and classification problems.

Both machine and deep learning can be applied in a number of ways. Supervised learning applications, often regarded as the most prevalent, exploit training data that is labelled to tell the computer exactly what patterns it should look for. Unsupervised learning applications, on the other hand, exploit training data without the benefit of labels. Reinforcement learning is an emergent frontier where algorithms learn how to achieve a clear objective by trial and error based on some prescribed reward function. Separately, "future learning" techniques encompass emergent approaches by which to realize applications under varied operating contexts. Transfer learning, for instance, uses knowledge gained from solving one problem in the application to a different but related problem. Federated learning, though still an active field of research, has been shown to reduce data dependency by

distributing training functions across a number of nodes. A thorough study of these and other related methods is detailed in a forthcoming technical report by the RTG³.

In discussing security implications of deep machine learning systems, Mr. Espen Hammer Kjellstadli, Norwegian Defence Research Establishment, expounded on considerations for data-driven cybersecurity that advance beyond traditional rules-based approaches. Deep machine learning introduces new vulnerabilities surrounding the training and testing of a given model. Training data garnered in the initial stages of a pipeline can be manipulated so as to influence model correctness. The model may also be exploited in the latter stages due to an imbalance among features which led to its construction. The study of these vulnerabilities, and how to defend against or weaponize them, is referred to as adversarial machine learning. A comprehensive overview can be found in work from the US National Institute of Standards (NIST) [11] and MITRE [12]. Both references serve as an excellent resource for the initial design or security assessment of any machine learning architecture.

Many defences have been proposed against adversarial machine learning attacks. For instance, training stage attacks, where an attacker deduces what type of knowledge a model might learn, can be mitigated by protecting training data with encryption, sanitization, deliberate selection, or by introducing other artificial constraints to the training corpus. The training corpus may further be generalized, either by increasing its population or by exploring alternative representations for its data. This approach has been shown to provide performance gains as well as security benefits at every stage of the pipeline. Lastly, synthetic adversarial data has been shown to complement traditional training data and bolster the resilience of resulting models. Ultimately, data quality plays a crucial role in the performance of a given model, and the more data employed to train a model, the more effective that model generally grows.

Another security consideration impacting deep machine learning involves characterizing how explainable or interpretable the results of a given model are. Unlike traditional algorithmic systems, trust cannot be derived from a transparent understanding of the criteria upon which decisions were made. The complexity of these systems, scaling to millions of feature-weights, effectively transforms deep machine learning applications into black boxes. This is an important consideration, as humans must increasingly understand, verify, and act upon the judgements of these systems.

3.0 DEEP MACHINE LEARNING APPLICATIONS IN CYBERSECURITY

Deep machine learning holds many applications in cybersecurity. In separate presentations, Mr. Joseph Mathews, U.S. Naval Research Laboratory, and Dr. Tracy Braun, US Army Research Laboratory, elucidated on current state of the art by reporting the results of a recent study by the RTG. The study characterized its findings by adopting guidance from NIST [13], which assists organizations in implementing an information security continuous monitoring program for its assets, awareness of cyber threats and vulnerabilities, and visibility into the effectiveness of deployed security controls. Monitoring is defined as the ongoing checking, supervising, and critical observing in order to identify change from an expected or required performance. The terms “continuous” and “ongoing” in this context suggest organizational risks are assessed at some frequency sufficient to support risk management activities and adequately protect organizational information.

Specifically, [13] defines 11 security automation domains that address a set of security controls required to establish and maintain ongoing cybersecurity awareness. Each domain includes a grouping of tools, technologies,

³ NATO STO, “PRE-RELEASE: Deep Machine Learning for Cyber Defence”, DOI: 10.14339/STO-TR-IST-163. (accessed 5 April 2022).

and data that must be collected, analysed, and reported. Incidentally, these domains form a useful structure by which to characterize current and proposed cybersecurity applications of deep learning. In consideration of similarities and for ease of exposition, we borrow and collapse those 11 domains into 8 distinct categories reported herein. The complete findings of the study will be thoroughly reported in a forthcoming technical report.

3.1 Malware Detection

Malware is any malicious software that is intentionally designed to infiltrate, modify, disrupt, or damage a computer system without the owner's knowledge or consent. Malware assumes many forms of digital content that include executable code, scripts, and active objects embedded inside of interactive files. Malware detection mechanisms perform periodic or near-real-time scans of information systems with prior knowledge of malign content. Antivirus signatures and similar identification techniques (i.e., heuristics) are the product of exhaustive forensic analyses which necessitate a combination of static and dynamic methods.

Deep Machine Learning has been widely explored in attempts to improve malware detection. Traditional approaches rely on manually designed features derived from expert knowledge of the domain. These solutions provide an abstract view of software that can be used to generalize its traits. Feature engineering and feature extraction are key, time-consuming processes of the workflow. Following advances in other fields, malware detection capabilities are increasingly leveraging deep learning architectures.

Research suggests the application may overcome traditional challenges in malware detection. Industry trends point to a growing sector of companies offering AI-based cybersecurity solutions that implement some form of deep learning for malware detection. Proposed applications in academic work further demonstrate increased efficacy achieved with novel, distinct representations of program data [14]. These applications, however, generally continue to suffer from the unavailability of robust training data, overfitting of models, lack of explain-ability, and diminished longevity over time as malware tradecraft evolves.

3.2 Event and Incident Management

Event Management involves monitoring observable occurrences in and among information systems. Traditional intrusion detection systems [15], implemented on either the network or endpoint, employ signature-based and anomaly-based models which suffer drawbacks. Anomaly-based models have been shown capable of high false positive rates, whereas signature-based models have been shown capable of high false negative rates. Both stand to benefit from advances in deep machine learning that are not reliant on prior knowledge of specific attack patterns. Similarly, advances in e-mail filtering employ deep learning applications of Natural Language Processing (NLP) to identify message patterns indicative of spam. Google, for instance, has used TensorFlow to significantly enhance the spam detection capabilities of Gmail [16].

Incident Management tools similarly assist in detecting and responding to cyber-attacks. These tools rely on logs and audit records which capture the behaviour and state of information systems, often as they relate to system transactions, security controls, or performance. Tools that aid the process of generating, transmitting, storing, analysing, and disposing of log data have grown increasingly critical for many cybersecurity operations.

Current offerings among these domains (i.e., platforms or technology stacks for Extended Detection and Response; Security Information and Event Management; and Security Orchestration, Automation, and Response) collect data that is naturally suited for exploitation with deep learning. Many commercial tools already offer some native support for deep learning plug-ins. Most attempts to implement that support, however, bypass the native system

in favour of external pre-processing and post-processing pipelines, or by integrating with third-party frameworks. This suggests deep learning capabilities will evolve in tune with those capabilities native to their platforms.

It is important to note these applications can be limited by proprietary data formats, an inability to expose or add context to garnered data, and a lack of trust in the reported output. Experts further express concern that existing tools might not collect data with optimal fidelity (e.g., aggregate metadata) for solving practical problems. We suspect research for deep learning applications within event management will become increasingly popular once tools commonly deployed by security operations centre teams integrate deep learning frameworks or algorithms. At the same time, these fully featured platforms may introduce new challenges with respect to complexity and maintenance when architected to integrate and coexist with other tools.

3.3 Information Management

Information Management refers to managing the location and transfer of information that is essential to protecting the confidentiality, integrity, and availability of organizational data. Digital information is both deliberately and incidentally stored among a myriad of systems and devices. Consequently, data loss, theft, and leakage pose a considerable risk to an organization's information security posture. Data Loss Prevention (DLP) tools afford capabilities to inventory, classify, and track data creation, use, storage, transmission, and disposal.

Current DLP systems implement a mixture of data classification techniques that include labelled data, exact matching, partial matching, regular expressions, and machine learning. Much research in the field of DLP tools is currently centred around analysing the data and its classification. Deep learning excels in interpretation of complex data (e.g., texts, images, videos), and can consequently provide machine readable access to the information encoded within.

Accessibility to robust, organizationally-relevant training data is especially important for achieving positive results, yet zero trust information security principles generally prevent unfettered access to sensitive documents that might form the basis of some training corpus. The proliferation of end-to-end-encryption of data-at-rest and data-in-transit further introduce challenges to robust data acquisition. Business process modelling and increasingly multimodal data promulgation also present new challenges and opportunities [17].

3.4 Vulnerability and Patch Management

A vulnerability is a software defect that introduces a potential security exposure. A patch is a software fix that eliminates or abates that vulnerability. As the number of vulnerabilities and patches continue to increase, vulnerability and patch management tools allow organizations to identify, report, and remediate vulnerabilities in a coordinated manner. Vulnerability scanners, for instance, are commonly used to identify vulnerabilities on endpoints, networks, operating systems, and applications. Patch management tools similarly scan for vulnerabilities on systems and facilitate the application of necessary patches and other updates.

Many commercial product offerings in this space purport to implement machine learning, largely in order to prioritize remediation. Notably, the 2016 Cyber Grand Challenge competition sponsored by DARPA demonstrated the potential for automated cyber security systems that can discover, evaluate and patch vulnerabilities in real-time [18]. Among the methods used by competing teams were fuzzing tools augmented with deep learning. A future vision is to employ similar tools that can scan software for vulnerabilities and assist in autonomously patching them. Ongoing development of enabling capabilities persist and are extremely promising, but lack maturity.

3.5 Software Assurance

Software assurance is the planned set of activities that ensures software functions as intended and is free from defects. Common software assurance technologies include secure coding, source code analysis and application fuzzing tools. Ultimately, software assurance helps achieve trustworthiness where no exploitable vulnerabilities exist; and predictability where software confidently executes as intended. Three major types of tools and techniques for software analysis have been identified. Static analysis tools examine the system/software without executing it, including examining source code, bytecode, and/or binaries. Dynamic analysis tools examine the system/software by executing it, giving it specific inputs, and examining outputs. Hybrid tools integrate static and dynamic approaches; for example, test coverage analysers use dynamic analysis to run tests and then use static analysis to determine which parts of the software were not tested.

Among the tools and techniques listed above, deep machine learning has been applied to source code analysis and fuzz testing. Additionally, recent practical applications can be found among operating system development, where maintainers use machine learning to distinguish patches that fix bugs from those which do not [19]. Beyond these traditional approaches to increase software assurance, emerging capabilities in the commercial sector increasingly attempt to make AI-generated code through low/no code platforms. Whether or not this can reduce the number of bugs, and thus security vulnerabilities, is uncertain. We suspect this can be a direction worth pursuing, in future research.

3.6 Asset and License Management

Asset Management refers to the maintenance of an inventory of software and hardware systems within the organization. This can be accomplished via a combination of system configuration, network management, and license management tools, or with a special-purpose tool. Software asset and licensing information may be centrally managed by a software asset management tool to track license compliance, monitor usage status, and manage the software asset life cycle. Asset management is currently used by humans for inventory and configuration management of computing hardware, software, and devices.

Asset management tools can produce and record large quantities of data that afford insight into both cybersecurity and business operations. Recent work towards cyber defence of military systems argue for distributed autonomous agents can sense and adaptively defend their environment [20]. Among the common themes of these applications are the ability to reduce human oversight, adaptively manage technology consumption, optimize resource utilization, map dependencies between assets and digital workflows, and predict or respond to organic business risks. Trends towards smart, ubiquitous devices will drive the demand for innovative approaches to asset management, where not only the human operator, but all manner of devices themselves, are capable of adapting to their environment in order to continually optimize themselves.

Future applications, generally referred to as “Industry 4.0” [21], envision machine learning-enabled asset management on every device, enable by Edge Computing and next generation wireless technologies (i.e., 5G). Such device-to-device communications will catalyse and optimize smart factory processes, such that devices tune their configuration by dynamically sensing their environment. Scaling deep learning applications for asset management in this context will necessitate a comprehensive and interdisciplinary approach in tune with technological advances from complementary technologies that include mobile devices, IoT platforms, location detection technology (e.g., Radio Frequency Identification, Near-Field Communication), 3D printing, smart sensors, data analytics, augmented reality, wearable computing, and network-enabled robots and machines.

3.7 Network Management

Network management tools include host discovery, inventory, change control, performance monitoring, and other network device management capabilities. Recently, machine learning has been proposed as a mechanism by which to dynamically configure and orchestrate these tools for moving target defences which thwart adversarial manoeuvres. These applications are explored in literature, but many use cases are limited to simple scenarios such as bandwidth throttling and performance management. Adapting network management techniques for strategic and tactical assets in a military context also remains a considerable challenge.

Many emerging applications exist, including those involving user behaviour analytics [22], and vehicular networks [23]. On the former, suspicious user behavioural patterns may warrant network configuration changes. On the latter, clustering algorithms can efficiently characterize network traffic as suspicious or benign. Many deep machine learning applications, particularly those for event detection and malware detection, are already enabled by data which is collected or exposed by network management tools. Consequently, prior research has largely centred around network monitoring and event classification. Recent work, however, demonstrates the effectiveness of deep learning-based routing for traffic control in packet switched networks. Similarly, proposed applications of deep learning to network management aim to automate or optimize network management tasks without human oversight.

Ultimately, Moving Target Defence (MTD) is an area that stands to greatly benefit from deep learning. Where traditional network defences fail to account for the attacker's inherent advantage due to the static nature of the environment, MTD continuously shifts the configuration of that environment, in turn reducing the success rate of cyberattacks. Deep learning has already been shown to accurately classify applications with traffic that is naturally garnered by software-defined network controllers. Adapting network management techniques for strategic and tactical assets will be a considerable challenge, as military networks comprise considerable scale and diversity.

3.8 Configuration Management

Configuration management tools allow administrators to set, monitor, attest, and restore configuration settings. Managing the configurations among information systems has grown increasingly difficult as networks and devices have grown in complexity. Automated solutions enhance efficiency and improve reliability while generally lowering cost at scale. System configuration scanning tools provide the automated capability to audit and assess a target system to determine its compliance with a defined secure baseline configuration. Though few practical applications of deep learning in this domain exist, much potential remains in the context of moving target defences discussed in the prior section.

4.0 RELATED MILITARY APPLICATIONS AND OPEN CHALLENGES

Applications of deep learning can be generally characterized by problems stemming from large quantities of data upon which insights must be drawn, or a desire for greater automation. This sentiment was echoed among many complementary RTGs exploring an array of military information systems technology applications. Findings reported from those complementary charters are detailed herein.

4.1 Interoperability for Semi-Autonomous Unmanned Ground Vehicles

Towards Interoperability for Semi-Autonomous Unmanned Ground Vehicles⁴, Dr. Kim Mathiasen, Norwegian Defence Research Establishment, noted the challenges incurred by pursuing interoperability among military information systems technology. Interoperability, an important topic often taken for granted, is simply explained as the ability for distinct technologies of varied provenance to easily communicate and synchronize information. The development and adoption of standards by which to achieve this pose many hurdles.

NATO countries are investing in Unmanned Ground Vehicle (UGV) technology for a variety of warfighting missions (e.g., intelligence, surveillance, reconnaissance; chemical, biological, radiological, nuclear, high-yield explosive detection; etc.). In order to operate in coalition environments, information and perhaps even control of these platforms must be shared between nations. Prior experimentation has demonstrated some practical challenges towards achieving this that include the acquisition of distinct video and telemetry feed formats, as well as interference caused by distinct network and radio communication systems [24].

Current interoperability standards address how to transmit control data from operator control units, how the robots should transmit data back to the operator, and how to share data between vehicles. Challenges encountered during the construction of these standards include time synchronization, calibration, measurement accuracy, implicit assumptions, and format and representation of data (e.g., maps) and metadata. Feature overlap among similar or competing standards and development toolkits further necessitate deconflicting specific requirements and capabilities.

As interfaces and standards mature, military commanders envision adopting UGV technology in autonomous or semi-autonomous ways that will increasingly require them to perceive their surroundings. Such applications would prevent direct control of robots in favour of approaches such as transmitting intermediary waypoints for navigation. Consequently, deep learning is widely regarded as fundamental component of many UGV programs. Existing standards will need to accommodate cybersecurity considerations in addition to traditional telemetry and sensor information. Cyber posture may be introspectively attested and shared, as well as externally observed or queried. Situational awareness of cyber-physical assets with this degree of fidelity could inform mission and control decisions, particularly where platforms operate in contested environments and anticipate adversary manipulation through physical or electronic attack vectors.

4.2 Securing Unmanned and Autonomous Vehicles for Mission Assurance

Towards Securing Unmanned and Autonomous Vehicles for Mission Assurance⁵, Dr. Federico Mancini, Norwegian Defence Research Establishment, explained work to systematize the range of cybersecurity challenges of multi-domain missions conducted by autonomous platforms. Understanding the risk imposed by unmanned systems first necessitates a comprehensive understanding of all the factors that play a role in their security posture, including external threats.

Autonomous platforms come in many shapes and sizes that operate in land, sea, air, and space. These inherently cyber-physical systems depend on sensor inputs to collect data relevant to their mission at hand or perceive their surroundings in order to make decisions about how to navigate. Their mobile, network-connected nature of military

⁴ NATO STO, 'Interoperability for Semi-Autonomous Unmanned Ground Vehicles', *NATO Science & Technology Organization*, <https://www.sto.nato.int/Lists/test1/activitydetails.aspx?ID=16750>, (accessed 18 March 2022).

⁵ NATO STO, 'Securing Unmanned and Autonomous Vehicles for Mission Assurance', *NATO Science & Technology Organization*, <https://www.sto.nato.int/Lists/test1/activitydetails.aspx?ID=16524>, (accessed 18 March 2022).

applications further introduce unique challenges for cyber assurance. For instance, in traditional civilian applications, self-driving cars are designed to adhere to clearly defined traffic laws and roadway infrastructure. Autonomous ground, sea, and air platforms, on the other hand, may operate in the open environments without prescribed norms and under contested conditions.

To study this problem, researchers have proposed a theoretical framework addressing how platform behaviour changes with mission context. The framework, largely based upon threat, comprises three distinct layers. A mission layer defines the functions and outcomes necessary for mission success. A vehicle layer defines those platforms which are assigned to the mission and in need of protection. Finally, a component layer defines subsystems (i.e., actuators) within each platform that allows those assets to be used. At each layer, the framework employs a set of catalogues that define the relationship with other layers. For instance, generic mission security goals such as safety, reliability, and confidentiality can be mapped to platform components which play a part in realizing those properties. Some generic examples include the ability to navigate autonomously, collect and process sensor information, communicate and cooperate among their payload capability, and to securely store sensitive data. Each task will depend on certain components on the platform, and each component is susceptible to certain threats.

A major consideration for defence against these threats is the platform's ability to autonomously respond. Implementing security capabilities for unmanned systems is tricky because of the operating constraints imposed by the environment. Traditional information systems operate under conditions of constant connectivity and generally advantageous bandwidth, whereas tactical environments must be able to operate under disconnected, intermittent, latent, or covert conditions of connectivity. Cyber defence solutions for these environments, including those which implement deep learning, must adapt and reason under these conditions. This includes those mechanisms which address traditional cyber threats, as well as those which address physical attacks meant to disrupt the cyber-physical nature of its functionality. The identification of a problem may trigger a variety of responses that account for varied aspects of mission success (e.g., return to base, shut down, self-destruct, erase stored content). Deep learning's success in perceiving the physical environment may very well drive some of these algorithmic decisions.

4.3 Artificial Intelligence, Machine Learning, and Big Data for Hybrid Military Operations

Towards Artificial Intelligence, Machine Learning, and Big Data for Hybrid Military Operations⁶, Dr. Prithviraj Dasgupta, U.S. Naval Research Laboratory, spoke to the challenge of military scenarios that grow increasingly dominated by AI technology, and the role that adversarial AI and game theory can play in addressing the challenges of hybrid warfare. To align this view in the context of the workshop, a discussion of adversarial AI applications for malware detection followed.

Generative Adversarial Networks (GANs), a deep-learning based generative model, are a method for creating synthetic data that shares characteristics with training data. Though applications of GANs to fool AI systems have been widely explored, they have largely concentrated on imagery and text data. In more recent cybersecurity applications, GANs have been shown to effectively transform known malware in ways where it appears benign while remaining malicious, thereby fooling traditional detection methods which include machine learning classifiers. Doing this in practice incurs cost, however, because the adversary must discover how much and where to insert noise among training samples.

⁶ NATO STO, 'AI, ML and BD for Hybrid Military Operations (AI4HMO)', *NATO Science & Technology Organization*, <https://www.sto.nato.int/Lists/test1/activitydetails.aspx?ID=17026>, (accessed 18 March 2022).

Prior art is limited by its alteration of the feature space found in binary program data, which can prevent the resulting GAN-derived program from being executed. Therefore, current work has focused on modification of data at the byte level [25]. Three strategies, in particular, frame a series of reported experiments [26]. First, padding attacks add a number of blank “padding” bytes and then substitutes each added byte with a byte derived from the body of the training malware. Second, DOS header attacks modify portions of the malware executable program header since most machine learning classifiers limit inspection to that portion. Third, genetic attacks selectively replace bytes within the malware based on a genetic algorithm.

Assessing the results of these approaches involves quantifying the evasion rate, or how well the modified malware is able to fool a classifier; time required to perform the modification; and the number of modifications necessary to generate the modified malware. Experimental results confirm header attacks as the most effective due to its high evasion rate and low measure for time and alterations required. It was further noted that, as operating systems have grown more complex, so too have the number of alterations required to craft adversarial malware variants. Researchers continue this line of experimentation on how to use GAN techniques in order to craft adversarial malware samples that can fool the AI-based detectors. Ultimately, game theoretical approaches could be applied to characterize the trade-offs between the attacker-defender interactions that involving crafting adversarial samples.

4.4 Data Hiding in Information Warfare Operations

Towards Data Hiding in Information Warfare Operations⁷, Dr. Zbigniew Piotrowski, Military University of Technology (Poland), detailed techniques to realize a hidden data layer among extant communications channels, and the opportunities, threats, and challenges they impose. Hidden data layers are communication channels which conceal information and complement existing cryptographic and steganographic methods. Traditionally considered a challenge and threat, recent work exploring advances in covert channel methods have demonstrated how disparate branches of research can provide benefit to each other [27].

At the moment, academic research largely addresses innovative communication equipment, e.g., a steganographic filter to remove hidden transmission, a steganographic router, numerous methods of data hiding in multimedia based on the digital watermarking, new steganographic methods in radio communication (radio steganography) and also in computer networks (network steganography). Software-defined networks (SDN) can further support hidden communication in case of malware infection of the main SDN interface. At the same time, there are numerous descriptions of methods for analysing and detecting hidden data and also for identification of the attacks carried out with the use of hidden transmission [28].

Potential topics in this area ambitiously encompass methods of detecting and preventing hidden transmission of data in both wired and wireless connections; detecting and preventing multimedia content watermarking on the Internet and in documents; the standards of conduct in the case of identification of hidden communication; the use of hidden transmission and digital watermarking in military communications; NATO steganography applications and devices standardization with built-in of data transmission techniques stealth class; procedures of steganalysis of digital objects and data streams; the standardization and transparency of perceptual tests (e.g., speech, audio, video), robustness and transparency steganalysis; and alternative ways to protect data in the context of quantum technology.

⁷ NATO STO, ‘Data Hiding in Information Warfare Operations’, *NATO Science & Technology Organization*, <https://www.sto.nato.int/Lists/test1/activitydetails.aspx?ID=17017>, (accessed 18 March 2022).

Recent success in practical applications include data hiding among dedicated tactical radio communication handsets which can independently authenticate the speaking party (or voice broker) through a personnel identification number encoded within the channel. Many similarly creative applications are being proposed, and the application of deep learning is naturally suited to exploit those involving digital multimedia content and signals. Other examples include channel selection and orchestration that share characteristics with existing Network and Configuration Management capabilities.

4.5 Robustness and Accountability in Machine Learning Systems

Towards Robustness and Accountability in Machine Learning Systems, Dr. Douglas Lange, US Naval Information Warfare Center - Pacific, summarized challenges facing all of the applications discussed in the workshop. Though much research explores how Machine Learning systems can be manipulated by creative input, few efforts address how to make them more robust. Such systems may require a fundamental change in training, testing, validation, and production.

Robustness is often characterized in terms of an obstacle, such as an attack or enemy, and can comprise many distinct goals (i.e., performance, security). Understanding the necessity and utility of those goals in the context of the system and mission is crucial to creating a methodology for assuring robustness. Applied to machine learning systems, this can be expressed as a system's ability to produce predictable output and comparable performance with new data as with that from which it was trained.

Uncertainty exists in both operational data and training data, though the former is far more recognized and accounted for during the design and development of machine learning systems. Practitioners, however, must expect their systems to process input both within and outside the bounds of the training population. Complicating matters for the military context, the supply of training data for warfighting applications is often more limited than that of commercial applications where the actions of unknowing or unwilling users can be harvested (i.e., advertisement targeting). Warfare situations are often unobservable and novel, and models trained with peacetime or military exercise data will not always reflect the dynamic nature of conflict. Therefore, those systems capable of adapting the fastest are most likely to succeed.

The aim of machine learning applications is to learn the proper set of model parameters that fit the training population. That necessitates developing a cost function that measures how much error is imposed by changing the model and the impact of noise on that model. The resulting uncertainty can be generally characterized as epistemic or aleatoric. Epistemic, or systematic uncertainty, defines the reducible portion of total uncertainty. Aleatoric, or statistical uncertainty, defines the irreducible portion of total uncertainty. Additionally, inputs may exhibit noise and corruption, or exhibit meaningful variations from training data. The former reflects physical robustness while the latter reflects semantic robustness. Ultimately, machine learning systems must properly express what the classifier does before attempting to characterize robustness.

Lastly, bias is an important consideration for all deep machine learning applications. In one well-studied application, researchers employed image-to-image translation using cycle-consistent adversarial networks by demonstrating the image of a zebra projected on to a horse [29]. Inspection of this demonstration shows that the pixels in the immediately vicinity of the animal are also translated from a grassy field, where horses are commonly found, to that of a savannah, where zebras are commonly found. The environments portrayed in the training population therefore confirm bias, as those are not the only environments horses and zebras may be found.

Many deep machine learning applications exhibit difficulty simply identifying those inputs which fall outside of their training population. Achieving this alone would be a tangible improvement in quality and robustness. Often

however, models are fooled while they purport high confidence in their findings. ML applications tend to fare better at things they see very often and poorer at the things they don't. Self-driving cars, for instance, are trained with millions of hours of driving footage under real and simulated conditions. Data collected from the real world is often used to improve simulations. This approach is lacking in the military context, as adversaries may employ tactics otherwise unobserved during peacetime.

5.0 CONCLUSION

In the workshop-opening remarks, Dr. Nikolai Stoianov, Chair of the Information Systems Technology Panel in NATO STO, noted the strength of NATO's research network stems from its collaborative business model. NATO nations and partners elect to use their national resources to define, conduct, and promote cooperative research and information exchange. By bringing soldiers and researchers together in a common forum, participants benefit from each other's expertise, increase overall efficiency, and enhance the collective strength of the alliance. Further advantages are reaped by exposing the commonalities and finding balance among national efforts, industrial perspectives, and resident expertise. These themes are clearly evident among the findings reported herein. Through discourse spanning cybersecurity, computer science, artificial intelligence, autonomy, and military operations, workshop participants shared current and trending applications of deep machine learning poised to bolster the cybersecurity posture of military networks. Parallels discovered among complementary RTGs reflect many similar challenges and opportunities.

Deep machine learning can improve almost all digital technologies and applications that form the cyber landscape by enhancing data driven decision making and minimizing the role of human experts in the loop. Example applications include automated software development (including malware); automated protocol and architectural designs, including those derived from high level specifications; human-machine teaming for managing network operations, including virtualization, containerization, and cloud services; autonomous orchestration of network features such as spectrum management, QoS management, and network slicing; and autonomy among cyber-physical systems and system of systems. These applications necessitate a comprehensive and interdisciplinary approach in tune with advances in digital technology.

Understanding the cyber landscape further encompasses situational awareness of coalition or friendly networks, as well as characterization of adversarial threat. In complementary work from the IST-129 RTG⁸, findings show areas where deep machine learning could bolster prediction of adversarial operations and the analysis of attack and defence scenarios. Such understanding could lead to implementations of distributed autonomous agents which sense, react, and adapt to their environments and emergent threats [20]. Ultimately, deep machine learning allows more efficient exploitation of resources and better use of human experts' time.

The adoption of deep machine learning in safety-critical applications remains a subject of intense scrutiny [30]. Algorithms have been shown to demonstrate a tendency for erroneous function through instances such as misdirection with innocuous signage. They have demonstrated vulnerability to data poisoning and data sparsity attacks, leading to embarrassment and damage. This has naturally led the military to explore ways in which to leverage the technology while keeping their function predictable and reliable. Ultimately, there exists a critical need for strategies to design, build, deploy, and sustain cyber-physical systems that are trustworthy, secure, and dependable [31].

⁸ NATO STO, 'Predictive Analysis of Adversarial Cyber Behaviour', NATO Science & Technology Organization, <https://www.sto.nato.int/Lists/test1/activitydetails.aspx?ID=16546>, (accessed 18 March 2022).

The acceleration of deep machine learning over the last decade has been fuelled by decades of advances in computational power. Performant hardware made it possible to construct networks with vastly more layered connections and neurons giving way to an ability to model complex phenomena. This trend, however, has recently shown diminishing returns [32]. Despite continued advances on the hardware cost-performance curve, only marginal improvements in model performance can be made before computational demand grows infeasible. Novel hardware accelerative architectures have been proposed to partially overcome this challenge [33]. On the other hand, it has given way to “low risk” applications of autonomy where the action space can be explicitly defined and yield minimal impact of failure. This approach makes model deployment amenable to size, weight, and power constrained platforms.

A common theme among the discourse generated by the workshop participants was the criticality of data for the success of any application. Data acquisition in cyberspace is, counterintuitively, a challenging proposition. Though an organization’s cyberspace is often brimming with rich data, exposing that data in a form and manner amenable to rapid exploitation often presents many practical challenges. Data provenance, ownership, classification, stewardship, sensitivity, regulations, architectures, models, transport, federation, and other considerations often impede the fielding of capabilities. The added sensitivity and multi-classification nature of military operations and coalition networks further complicate matters. Extant challenges suggest alternative approaches, such as transfer and federated learning approaches, are worth pursuing and may enable model-sharing without data transfer or semantic interoperability concerns. Alternatively, GANs and similar emergent tools grow increasingly capable of generating synthetic data at scale.

We see a path forward advancing among all of these dimensions. Successful applications of deep machine learning necessitate comprehensive and interdisciplinary approaches synchronized with advances in hardware and other digital technologies, including improved techniques for data capture, data generation, and data sharing. The security of the deep machine learning system itself, at all levels, must be maintained. This includes data sets, classifiers, models, and learned responses which must be protected from manipulation. Reliability and explainability are particularly important areas of consideration important to building trustworthy systems. Lastly, a desire to push deep machine learning out to the tactical edge will require improvements in autonomy, and hardware size, weight, and power.

In this paper, we introduced terminology and perspectives that help bound the cybersecurity problem space and shape potential deep machine learning solutions. We illustrated how deep machine learning is already applied to cybersecurity and proposed opportunities for further advances. We exposed related efforts within NATO STO and drew parallels among multiple application domains. Finally, we highlighted some key considerations and findings for successful applications among military and coalition operating environments.

6.0 REFERENCES⁹

- [1] R. Steinberg, "Venturebeat," 8 December 2017. [Online]. Available: <https://venturebeat.com/2017/12/08/6-areas-where-artificial-neural-networks-outperform-humans/>. [Accessed 2022 March 18].
- [2] J. Hsu, "Starkey's AI Transforms Hearing Aids into Smart Wearables," IEEE Spectrum, 27 August 2018. [Online]. Available: <https://spectrum.ieee.org/starkeys-ai-transforms-hearing-aid-into-smart-wearables>. [Accessed 18 March 2022].
- [3] H. Karen, "MIT Technology Review," 3 November 2020. [Online]. Available: <https://www.technologyreview.com/2020/11/03/1011616/ai-godfather-geoffrey-hinton-deep-learning-will-do-everything/>. [Accessed 18 March 2022].
- [4] U.S. Department of Defense, "Joint Publication 3-12, Cyberspace Operations," 8 June 2018. [Online]. Available: https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp3_12.pdf. [Accessed 18 March 2022].
- [5] J. S. Mertoguno, R. M. Craven, M. S. Mickelson and D. P. Koller, "A physics-based strategy for cyber resilience of CPS," in *Proceedings Volume 11009, Autonomous Systems: Sensors, Processing, and Security for Vehicles and Infrastructure 2019*, Baltimore, 2019.
- [6] K. Eykholt, I. Evtimov, E. Fernandes, B. Li, A. Rahmati, C. Xiao, A. Prakash, T. Kohno and D. Song, "Robust Physical-World Attacks on Deep Learning Visual Classification," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1625-1634, 2018.
- [7] S. Garfinkel, "MIT Technology Review," 22 August 2017. [Online]. Available: <https://www.technologyreview.com/2017/08/22/242124/hackers-are-the-real-obstacle-for-self-driving-vehicles/>. [Accessed 18 March 2022].
- [8] K. Hao, "MIT Technology Review," 17 November 2018. [Online]. Available: <https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart/>. [Accessed 2022 March 18].
- [9] D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, pp. 533-536, 1986.
- [10] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.
- [11] E. Tabassi, K. J. Burns, M. Hadjimichael, A. D. Molina-Markham and J. T. Sexton, "A Taxonomy and Terminology of Adversarial Machine Learning (Draft NISTIR 8269)," National Institute of Standard and Technology, 2018.
- [12] MITRE, "Adversarial Threat Landscape for Artificial-Intelligence Systems," [Online]. Available: <https://atlas.mitre.org/>. [Accessed 18 March 2022].
- [13] K. Dempsey, N. S. Chawla, A. Johnson, R. Johnston, A. C. Jones, A. Orebaugh, M. Scholl and K. Stine, "Information Security Continuous Monitoring (ISCM) for Federal Information Systems and Organizations," 30 September 2011. [Online]. Available: <https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-137.pdf>. [Accessed 18 March 2022].

⁹ References [34] through [64] reflect works cited among research workshop presentations and discourse.

- [14] L. Nataraj, T. M. Mohammed, T. Nanjundaswamy, S. Chikkagoudar, S. Chandrasekaran and B. Manjunath, "OMD: Orthogonal Malware Detection using Audio, Image, and Static Features," in *IEEE MILCOM21*, San Diego, 2021.
- [15] D. Denning, "An Intrusion-Detection Model," *IEEE Transactions on software engineering*, vol. 2, pp. 222-232, 1987.
- [16] N. Kumaran, "Spam Does Not Bring Us Joy—Ridding Gmail of 100 Million More Spam Messages with TensorFlow.," Google Cloud Blog, February 6 2019. [Online]. Available: <https://cloud.google.com/blog/products/g-suite/ridding-gmail-of-100-million-more-spam-messages-with-tensorflow>. [Accessed 2022 April 6].
- [17] T. Allard, P. Alvino, L. Shing, A. Wollaber and J. Yuen, "A dataset to facilitate automated workflow analysis," *PLOS One*, vol. 14.2, no. e0211486, 2019.
- [18] N. Lee, "DARPA's Cyber Grand Challenge (2014–2016)," in *Counterterrorism and Cybersecurity*, 2015.
- [19] J. Edge, "Machine learning and stable kernels," LWN.net, 12 September 2018. [Online]. Available: <https://lwn.net/Articles/764647/>. [Accessed 18 March 2022].
- [20] P. Theron, A. Kott, M. Drašar, K. Rządca, B. LeBlanc, M. Pihelgas, L. Mancini and A. Panico, "Towards an active, autonomous and intelligent cyber defense of military systems: The NATO AICA reference architecture," in *2018 International Conference on Military Communications and Information Systems (ICMCIS)*, Warsaw, 2018.
- [21] H. Lasi, P. Fettke, H.-G. Kemper, T. Feld and M. Hoffmann, "Industry 4.0," in *Business & information systems engineering*, 2014.
- [22] M. Shashanka, M.-Y. Shen and J. Wang, "User and entity behavior analytics for enterprise security," *2016 IEEE International Conference on Big Data (Big Data)*, pp. 1867-1894, 2016.
- [23] N. Leslie, "An Unsupervised Learning Approach for In-Vehicle Network Intrusion Detection," *55th Annual Conference on Information Sciences and Systems (CISS)*, vol. IEEE, pp. 1-4, 2021.
- [24] K. Mathiassen, F. E. Schneider, P. Bounker, A. Tiderko, G. De Cubber, M. Baksas, J. Głowka, R. Kozik, T. Nussbaumer, J. Röning, J. Pellenz and A. Volk, "Demonstrating interoperability between unmanned ground systems and command and control systems," *International Journal of Intelligent Defence Support Systems*, vol. 6, no. 2, pp. 100-129, 2021.
- [25] D. Luca and B. Biggio, "Secml-malware: Pentesting Windows malware classifiers with adversarial EXEmples in Python," in *arXiv preprint arXiv:2104.12848*, 2021.
- [26] P. Dasgupta and Z. Osman, "A Comparison of State-of-the-Art Techniques for Generating Adversarial Malware Binaries," *arXiv preprint arXiv:2111.11487*, 2021.
- [27] O. Chen, C. Meadows and G. Trivedi, "Stealthy protocols: Metrics and open problems," in *Concurrency, Security, and Puzzles*, 2017.
- [28] A. Houmansadr, C. Brubaker and V. Shmatikov, "The Parrot Is Dead: Observing Unobservable Network Communications," in *2013 IEEE Symposium on Security and Privacy*, Berkeley., 2013.
- [29] J.-Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017.
- [30] M. Mitchell, "Why AI is harder than we think," *arXiv preprint arXiv:2104.12871*, 2021.
- [31] R. Bharadwaj, "Assuring autonomy," *Disruptive Technologies in Information Sciences IV*, vol. 11419, 2020.

- [32] N. C. Thompson, K. Greenewald, K. Lee and G. F. Manso, "Deep Learning's Diminishing Returns," *IEEE Spectrum*, 24 September 2021. [Online]. Available: <https://spectrum.ieee.org/deep-learning-computational-cost>. [Accessed 18 March 2022].
- [33] R. Elnaggar, K. Basu, K. Chakrabarty and R. Karri, "Runtime Malware Detection Using Embedded Trace Buffers," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 41, no. 1, pp. 35-48, 2021.
- [34] Y. Wang, P. Jia, L. Liu, C. Huang and Z. Liu. "A systematic review of fuzzing based on machine learning techniques." *PLOS ONE* 15(8): e0237749, 2020. <https://doi.org/10.1371/journal.pone.0237749>
- [35] P. Zeng, G. Lin, L. Pan, Y. Tai and J. Zhang. "Software Vulnerability Analysis and Discovery Using Deep Learning Techniques: A Survey." *IEEE Access*, vol. 8, pp. 197158-197172, 2020, doi: 10.1109/ACCESS.2020.3034766.
- [36] M. Pradhan, A. Tiderko, and D. Ota. "Approach towards achieving interoperability between military land vehicle and robotic systems." 2017 International Conference on Military Communications and Information Systems (ICMCIS). IEEE, 2017.
- [37] B. Kolosnjaji, A. Demontis, B. Biggio, D. Maiorca, D. Giacinto, C. Eckert, and F. Roli. "Adversarial malware binaries: Evading deep learning for malware detection in executables." *2018 26th European signal processing conference (EUSIPCO)* (pp. 533-537). IEEE, 2018.
- [38] L. Demetrio, B. Biggio, G. Lagorio, F. Roli, and A. Armando. Explaining vulnerabilities of deep learning to adversarial malware binaries. *arXiv preprint arXiv:1901.03583*. (2019)
- [39] Demetrio, L., Biggio, B., Lagorio, G., Roli, F. and Armando, A., 2021. Functionality-preserving black-box optimization of adversarial windows malware. *IEEE Transactions on Information Forensics and Security*, 16, pp.3469-3478.
- [40] Raff, E., Barker, J., Sylvester, J., Brandon, R., Catanzaro, B. and Nicholas, C.K., 2018, June. Malware detection by eating a whole exe. In *Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence*.
- [41] Yang, W., Kong, D., Xie, T. and Gunter, C.A., 2017, December. Malware detection in adversarial settings: Exploiting feature evolutions and confusions in android apps. In *Proceedings of the 33rd Annual Computer Security Applications Conference* (pp. 288-302).
- [42] Anderson, H.S., Kharkar, A., Filar, B., Evans, D. and Roth, P., 2018. Learning to evade static PE machine learning malware models via reinforcement learning. *arXiv preprint arXiv:1801.08917*.
- [43] Castro, R.L., Schmitt, C. and Dreo, G., 2019, August. Aimed: Evolving malware with genetic programming to evade detection. In *2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)* (pp. 240-247). IEEE.
- [44] Ebrahimi, M., Zhang, N., Hu, J., Raza, M.T. and Chen, H., 2020. Binary black-box evasion attacks against deep learning-based static malware detectors with adversarial byte-level language model. *arXiv preprint arXiv:2012.07994*.
- [45] Rosenberg, I., Meir, S., Berrebi, J., Gordon, I., Sicard, G. and David, E.O., 2020, July. Generating end-to-end adversarial examples for malware classifiers using explainability. In *2020 international joint conference on neural networks (IJCNN)* (pp. 1-10). IEEE.
- [46] Boutsikas, J., Eren, M.E., Varga, C., Raff, E., Matuszek, C. and Nicholas, C., 2021. Evading Malware Classifiers via Monte Carlo Mutant Feature Discovery. *arXiv preprint arXiv:2106.07860*.
- [47] Gómez Gálvez, F. J. Static PE antimalware evasion by using Reinforcement Learning. <http://openaccess.uoc.edu/webapps/o2/handle/10609/127010>

- [48] Fang, Z., Wang, J., Geng, J., Zhou, Y. and Kan, X., 2021. A3CMal: Generating adversarial samples to force targeted misclassification by reinforcement learning. *Applied Soft Computing*, 109, p.107505.
- [49] Jin, B., Choi, J., Kim, H. and Hong, J.B., 2021, March. FUMVar: a practical framework for generating Fully-working and Unseen Malware Variants. In *Proceedings of the 36th Annual ACM Symposium on Applied Computing* (pp. 1656-1663).
- [50] Lucas, K., Sharif, M., Bauer, L., Reiter, M.K. and Shintre, S., 2021, May. Malware Makeover: breaking ML-based static analysis by modifying executable bytes. In *Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security* (pp. 744-758).
- [51] Jouppi, N.P., Young, C., Patil, N., Patterson, D., Agrawal, G., Bajwa, R., Bates, S., Bhatia, S., Boden, N., Borchers, A. and Boyle, R., 2017, June. In-datacenter performance analysis of a tensor processing unit. In *Proceedings of the 44th annual international symposium on computer architecture* (pp. 1-12).
- [52] Zhang, J.J., Gu, T., Basu, K. and Garg, S., 2018, April. Analyzing and mitigating the impact of permanent faults on a systolic array based neural network accelerator. In *2018 IEEE 36th VLSI Test Symposium (VTS)* (pp. 1-6). IEEE.
- [53] Kundu, S., Banerjee, S., Raha, A., Natarajan, S. and Basu, K., 2021. Toward functional safety of systolic array-based deep learning hardware accelerators. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 29(3), pp.485-498.
- [54] Han, S., Pool, J., Tran, J. and Dally, W., 2015. Learning both weights and connections for efficient neural network. *Advances in neural information processing systems*, 28.
- [55] Ashmore, R., Calinescu, R. and Paterson, C., 2021. Assuring the machine learning lifecycle: Desiderata, methods, and challenges. *ACM Computing Surveys (CSUR)*, 54(5), pp.1-39.
- [56] Tomsett, R., Chan, K. and Chakraborty, S., Investigating the robustness of peer-to-peer machine learning. Annual Fall Meeting of the DAIS ITA, 2019
- [57] Alzantot, M., Chakraborty, S. and Srivastava, M., Mitigating Adversarial Examples in Neural Networks. 1st Annual Fall Meeting of the DAIS ITA, 2017
- [58] Mitchell, M., 2021. Why AI is harder than we think. *arXiv preprint arXiv:2104.12871*.
- [59] Carlini, N. and Wagner, D., 2017, May. Towards evaluating the robustness of neural networks. In *2017 IEEE Symposium on Security and Privacy (SP)* (pp. 39-57). IEEE.
- [60] Tramer, F., Carlini, N., Brendel, W. and Madry, A., 2020. On adaptive attacks to adversarial example defenses. *Advances in Neural Information Processing Systems*, 33, pp.1633-1645.
- [61] Ducoffe, M. and Precioso, F., 2018. Adversarial active learning for deep networks: a margin based approach. *arXiv preprint arXiv:1802.09841*.
- [62] Melrose, J.S.J., Madahar, B., Aktaş, M., Martinel, N., de Marchi, J., Solberg, E., Lange, D.S., Tanik, G.O., Kurth, F. and Luotsinen, L., Robustness of Artificial Intelligence for Hybrid Warfare.
- [63] Singh, Y., Kuruvila, A.P. and Basu, K., 2021, February. Hardware-assisted Detection of Malware in Automotive-Based Systems. In *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)* (pp. 1763-1768). IEEE.
- [64] Hoffman, W., 2021. Making AI Work for Cyber Defense. *Center for Security and Emerging Technology*.

ANNEX A

For reference, we include references to tools exposed by workshop participants.

Tool	Description	Website
DetGen	A cyber security research project at University of Edinburgh's School of Informatics in collaboration with Edinburgh Parallel Computing Centre, British Telecom Labs in Adastral Park, Norwegian Defence Research Establishment (FFI) and The Alan Turing Institute. By combining methods from machine learning and formal verification, we plan to learn precise semantic models of software and devices which describe normal traffic patterns and logging behaviours. Then anomalous, potentially malicious behaviours stand out as being different to these learned behaviours. The eventual aim is to build an adaptive anomaly detection framework, to raise alarms to help manage and automatically configure application-level firewalls.	https://detlearsom.github.io/ https://github.com/detlearsom/DetGen
Malpedia	The primary goal of Malpedia is to provide a resource for rapid identification and actionable context when investigating malware. Openness to curated contributions shall ensure an accountable level of quality in order to foster meaningful and reproducible research.	https://malpedia.caad.fkie.fraunhofer.de/
DGArchive	DGArchive is a service offered by Fraunhofer FKIE. The service allows resolving or calculating domain names that are dynamically created by malware using Domain Generation Algorithms (DGAs).	https://dgarchive.caad.fkie.fraunhofer.de

ANNEX B

For reference, we include references to datasets exposed by workshop participants.

Dataset	Reference	Website
Ember	Anderson, Hyrum S., and Phil Roth. "Ember: an open dataset for training static pe malware machine learning models." arXiv preprint arXiv:1804.04637 (2018).	https://github.com/elastic/ember
Sorel	Harang, Richard, and Ethan M. Rudd. "SOREL-20M: A large scale benchmark dataset for malicious PE detection." arXiv preprint arXiv:2012.07634 (2020).	https://github.com/sophos-ai/SOREL-20M
Vx-underground	Dasgupta, Prithviraj, and Zachariah Osman. "A Comparison of State-of-the-Art Techniques for Generating Adversarial Malware Binaries." arXiv preprint arXiv:2111.11487 (2021).	https://www.vx-underground.org

