INL/RPT-22-66112
Revision 0

Artificial Intelligence for Digital Security and Protections

December 2020

A report prepared for the NNSA Office of International Nuclear Security Emerging Threats and Technologies Working Group

Char Sample, Shannon Eggers

Idaho National Laboratory

*INL is a U.S. Department of Energy National Laboratory
operated by Batelle Energy Alliance, LLC*

DISCLAIMER

This information was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness, of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. References herein to any specific commercial product, process, or service by trade name, trade mark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the U.S. Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the U.S. Government or any agency thereof.

INL/RPT-22-66112
Revision 0

Artificial Intelligence for Digital Security and Protections

A report prepared for the NNSA Office of International Nuclear Security Emerging Threats and Technologies Working Group

Char Sample, Shannon Eggers

December 2020

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
National Nuclear Security Administration
Office of International Nuclear Security
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

Page intentionally left blank

EXECUTIVE SUMMARY

Proper functioning of nuclear power plants relies on a mix of well-regulated human and machine-driven workflows. This regulation supports nuclear safety through a series of processes and many of the tasks that support these processes have a repetitive nature that make artificial intelligence (AI) informed by machine learning (ML) a potential aid in a variety of tasks. AI is being evaluated for activities that include inspections, fuel processing, monitoring, and other activities.

The introduction of any new technology presents a potential new attack vector. In the case of AI/ML, there are many attacks that have already been discovered and over time the attacks can be expected to follow the growth pattern observed in cyber security. While future planning is necessary, current efforts need to be established now to predict the threat emergence over the next year 10 years and mitigate potential threats. Based on these observations, AI/ML will need to become trustworthy, which corresponds to techniques and procedures that emphasize AI explainability along with resilience techniques to data, algorithms, models, and systems. This kind of system robustness is the foundation for defenses against AI/ML-specific attacks.

Attempting to look forward and take a broad view of capabilities provides input to research roadmaps and the ability to distill vulnerabilities into specific use cases may provide greater assistance in understanding the technology benefits while introducing new risks. The impact of current and future AI in three areas—capabilities, challenges, and recovery strategies—represents an initial attempt at balancing both.

The key findings of this study are as follows:

i. AI and ML are heavily dependent on data, which includes images, text, and control systems variables, and the protections extend beyond traditional information security controls into quality of data issues found in other disciplines such as information theory, resilience, and reliability.

ii. ML models are corruptible, not just by traditional hacking of the algorithms but also by manipulation of clean data quantities and groupings. Corruption of models can result in missing detection of cracks in materials (e.g., concrete), lower prioritization of tasks (e.g., maintenance), and procedures.

iii. In addition to data science expertise, nuclear security domain expertise is also needed. Both disciplines need to work in an integrated manner.

iv. Trustworthy solutions are needed that include explainable AI and resilient systems and data.

The core recommendation for INS is to engage with international partners to establish AI/ML-related regulations and best practices. While there is a wide set of cyber security focused standards, none of the current standards explicitly speaks to the unique vulnerabilities of AI/ML. The INS functional teams can work with partner countries and international organizations to establish such guidelines for nuclear security.

Page intentionally left blank

ACKNOWLEDGMENTS

This work was funded by the U.S. Department of Energy, National Nuclear Security Administration, Office of International Nuclear Security (INS) under DOE Idaho Operations Office, Contract DE-AC07-05ID14517. This report was initially published in December 2020 as part of the INS Emerging Threats and Technologies Working Group "*Artificial Intelligence—Applications to and Implications for Nuclear Security"* compendium.

Page intentionally left blank

CONTENTS

TABLES

Page intentionally left blank

ACRONYMS

AI	artificial intelligence
INS	International Nuclear Security
ML	machine learning
NLP	natural language processing

Page intentionally left blank

Artificial Intelligence for Digital Security and Protections

1. INTRODUCTION

Nuclear power plants are well-regulated, high-target sites that rely on humans interacting with machines in order to maintain safe functioning of nuclear and radiologic assets. AI/ML offers the promise of workflows characterized by greater accuracy, efficiency, and reliability, but its introduction into this environment introduces new risks. AI/ML is enabling (i) a new generation of potential attacks against physical and cyber-physical systems including those used to protect and secure nuclear materials, transportation, facilities, and personnel; and (ii) a new generation of defenses against both traditional and emerging attacks.

The disruption brought about by AI/ML promises transformation to workflows that are more efficient and accurate, meaning decreased human workload through real-time decision-making or decision support. As mentioned above, AI/ML can also introduce new vulnerabilities into systems that were previously isolated or standalone. Some of these are traditional attacks applied to AI/ML; however, there are also attacks that are unique to the AI/ML environment. Both attack classes must be addressed. For example, attempts to hack the algorithms through buffer overflows or unhandled cases are examples of traditional cyber attacks used against AI/ML applications, whereas attacks that teach reinforcement learning systems to act in unintended ways are unique to the AI/ML domain [3]. Thus, a traditional cyber attack that impacts AI access control software may cause failure of the system, while an AI attack may cause the system to incorrectly identify an adversary as an employee and grant unauthorized facility access. While both groups of attacks are of interest, the unique nature of the second group creates opportunity and urgency for addressing and gaining deep understanding of AI/ML applications and threats for nuclear security. The initial focus on attacks is due in part to the nature of resilience,[a] which requires an understanding of attacks for prevention, detection, response, and recovery from adverse events impacting the nuclear environment safety [4].

This paper is divided into sections to present the emerging threat that AI tools pose when applied to digital security protections and recovery. A brief review will be discussed within the context of traditional protection methods followed by the role of AI in updating these methods introducing both improvements and new potential vulnerabilities. Finally, a way forward will be introduced presenting countermeasures that serve as recommendations for protection.

2. BACKGROUND

Cyber security has always recognized the equal value placed on securing the deployed devices with related data in addition to the actual organizational assets. Compromised security data, services, and devices cannot be expected to adequately defend assets; thus, self-security, while sometimes overlooked, takes on extra importance with AI/ML. When dealing with the nuclear security domain, the consequences of AI failures such as failure to detect problems with materials, access control, transportation of resources, and safeguards introduces a new set of risks that were previously controlled by two-person rules, segmentation, and other traditional mitigations.

[a] Resilient systems are characterized by robustness (the ability to withstand adversity), redundancy (substitutability to support requirements), resourcefulness (ability to prioritize problems and initiate solutions), and rapidity (capacity to restore functionality in a timely manner).

AI/ML is used widely across various domains. This breadth of use has resulted in rapid application and deployment of technology concurrently with the evolution of AI/ML security best practices. Many of the current approaches for securing AI/ML applications attempt to repurpose cyber security strategies. While certainly a good start, the AI/ML domain introduces new challenges that have yet to be discovered or anticipated. For this reason, the emphasis on protection and recovery begins with discussion of failures followed by recovery strategies.

3. IMPACT ON INTERNATIONAL NUCLEAR SECURITY

3.1 AI Capabilities and Challenges

Nuclear assets are of high interest to both government and non-government organizations, and while the threat actors may vary the technical vulnerabilities are consistent. Recognizing these vulnerabilities and threats with regard to safety aspects of nuclear assets provides the opportunity to identify AI/ML applications to protect and/or recover from attacks, so a discussion of potential use and challenges is provided in Table 1. The list is not exhaustive but rather represents examples of some of the security challenges, acting as a launching point for further discussion. In terms of regulations and inspections, AI/ML solutions can rapidly and accurately detect indicators of compromise. Supervised learning can be used to quickly analyze data for evidence of compliance or noncompliance [4]. Unsupervised learning can identify clusters that were previously unseen by human observers; however, with the positive actions come negative consequences [5]. Compliance can be observed but not the level or degree of compliance. Unsupervised ML algorithms can create false clusters when additional “chaff” that comprises the training data is injected into the “wheat.” While problematic in any domain, this issue carries additional weight when dealing with data that comprise nuclear security, where a false sense of safety and security results. As this section is dedicated to the use of AI in digital security protections and recovery, Table 1 provides a mapping of AI uses to the INS critical areas.

Table 1. Examples of digital security protection AI applications and their challenges in INS mission areas.

INS Area	AI Use Example(s)	Potential AI-related Challenges
National Infrastructure (Regulations)	Natural language processing (NLP) is being used to convert textual documents into data that can be synthesized to quickly pull together commonalities found in various national nuclear policy documents.	• NLP software is not fully accurate, requiring human verification, and can choose from multiple meanings for terms resulting in imprecise or incorrect language outcomes; context is problematic. • Weight manipulation can result in wrong tradeoffs or recommendations (e.g., speed over safety).
Physical Protection	• Automated and integrated security applications. • Increase target characterization/response planning capability or vulnerability assessment. Assess creation of physical protection systems.	• Minor changes in physical domain result in mischaracterization. • Mischaracterization of physical protection systems as safe when they are not or as secure when backdoors and other vulnerabilities exist.
NMAC	• Assess and perform nuclear materials accounting and control functions. Use case: accounting and control functions tend to be repetitive in nature, making AI an attractive solution.	• Disruption of accounting and controls via data manipulation or algorithm breaking resulting in inaccurate findings (findings on security controls introducing inaccuracies; data loss or exposure from accounting data providing insights to overall safety processes).

INS Area	AI Use Example(s)	Potential AI-related Challenges
Transport Security	• Automated transportation, route optimization (safety, distance, security), and personnel. • Provide practical scenario-based training simulations. • Promote INFCIRC 909 efforts in region.	• Improper tradeoffs in transportation results in weight manipulation that fails to balance safety, speed, and efficiency. AI/ML programs are good at running simulations, but unanticipated behaviors are problematic. False interpretation of landmark signs may result in mapping a less safe/secure transportation route.
Response	• Evaluate tactical response force strategies for primary and secondary responders. AI prioritizes responses based on probabilities when combined with Bayesian analysis accuracy for situational awareness increases. • The ability to provide an efficient, unified task enumeration for incident response provides consistency and efficiency for all responders by allowing everyone to share the same situational awareness.	• Adversaries could develop methods of tricking algorithms to prioritize one incident over another by creating diversions. Training the AI model to prioritize spills over access violations can assist the attacker.
Cyber Security	• Assist with hunt team to find security breaches that are not identified through traditional cyber security reviews. Environments such as national defense, nuclear security, and critical infrastructure are such attractive targets that adversaries will use newly discovered, unreleased attacks on those targets. Hunt teams are experts in cyber security who examine systems for indicators of compromise. Example use case: when applied to a dependency model, AI can be extremely effective in tracking indicators of compromise left behind by intruders. • Initiate and implement cyber regulations. • Assess cyber capabilities.	• When used with hunt teams AI can detect intrusions, etc., but AI is less effective at human decision-making. If the problem occurred due to human–human links, AI is less effective. • Supply chains or dependency models are not always well-mapped creating ambiguity for the algorithm and less certainty on output. Hunt teamwork is a mix of human-machine work, human sensing, and prediction.
Insider Threat Mitigation	• Behavior-based analytics for insider threats. AI has been used to build user behavior profiles, so when users behave outside the profile, the activity is flagged. One problem was high false positivity rates in early implementation. • Policy enforcement. One application deals with email, a common spear phishing vector. Mail software can interpret the site security policy and enforce well-written policy rules to determine if a link is safe or not.	• User behavior analytics often suffers from a high rate of false positives due to the variability of human behaviors. Benign behaviors are often misclassified. • Mail enforcement of site policies translate acceptable use policies and reconcile with user behavior. One possible problem deals with negative mining or the ambiguity of words used in NLP to define acceptable use, causing confusion where AI cannot make a recommendation or blocks a good, necessary attachment.
Sabotage Mitigation	• Conduct physical protection system effectiveness evaluation specific to sabotage threats	• Physical monitoring is susceptible to vulnerabilities associated with

INS Area	AI Use Example(s)	Potential AI-related Challenges
	• Example: Work performed as a joint effort by the University of Pittsburgh and Idaho National Laboratory uses AI to detect reactor anomalies and sensor drift. Sabotage can take many forms and the ability to sense anomalies in processing is often hindered by high false positivity. Checking sensor drift can detect when drift has exceeded the supportable level (suggesting that processing may be faulty or failing without warnings being issued). Detection of deviations that are significant provides an early warning. This work along with uncertainty quantification work being performed at North Caroline State University can provide decision support to existing techniques.	physical domain changes or common corruption (tilting and zoom). • Triggered events remaining invisible until attack time. • Sabotage has a history of human targeting and phishing; customized phishing schemes will become more sophisticated and harder to detect.

3.2 AI Attacks and/or Failures

Discussing AI/ML applications with digital security protections is relatively straightforward; the discussion is more complicated with regard to attacks and failures. Recalling from earlier that ML powers AI, the coexistence of both provides a larger attack surfaces for adversaries. Attacks work in support of the goals of deny, deceive, disrupt, deter, and destroy. Depending on the attacker's goal, the methods vary and often coexist in an attack campaign. All five of these attack strategies are enhanced by AI. Coincidentally, AI is also vulnerable to these attacks.

Countering attack campaigns aimed at reducing the effectiveness of AI relies on strengthening training data, systems that process AI/ML software, ML models, and human/human-machine processes. For instance, data poisoning is an attractive attack vector partly due to AI's inability to quickly recover. When an algorithm learns the wrong lessons during training, it impacts result characterizations; in a supervised ML algorithm, there is limited opportunity for recovery. Reinforcement learning, the common response to AI failures, allows for modification of AI responses; but reinforcement learning techniques can also be compromised to respond with the wrong reward system, thereby further exacerbating the original problem.

AI applications require consideration in terms of anticipated use and projected use, including the possibility of misuse for both data and functions [4]. Thus, trustworthy AI obtained in part through explainable AI takes on higher importance [5]. The "black box" nature of AI creates a risk factor in general compounded by the existing examples where AI/ML has been tricked, and explainable AI would help to remove some of the associated mystery. Additionally, the inability to explain how algorithms reach their solutions puts AI in direct conflict with the European Union's General Data Protection Requirements [3], thus creating data privacy and exposure issues that vary from country to country.

The list that follows contains the ML failures resulting from attacks to AI digital security protections [1]. Not all of these attacks are viable in operational environments, but all have been demonstrated in academic settings. Some are of greater concern than others, but none is considered higher than standard cyber security risks.

1. Perturbation attack – Query modification to disrupt the learning and classification process, resulting in misclassification of data.
2. Poisoning attack – Contaminated training data resulting in misclassifications.

3. Model inversion – Careful queries revealing secret features of model, allowing the attacker to determine which model is being used and making possible accurate predictions of how the model will respond to other events.
4. Membership inference – Queries to infer if data record is part of training set, resulting in creating use cases that may confuse the decision-making process.
5. Model stealing – Model recovery through carefully crafted queries and accurate prediction of model response in other areas making it possible to identify asset priorities and weaknesses.
6. ML system reprogramming – Repurpose ML system to perform an unanticipated activity, harder to detect than data-driven attacks, allows an adversary to control part of the safety processes.
7. Adversarial example in physical domain – Introduction of adversarial physical example resulting in mischaracterization of data.
8. Malicious ML provider recovering training data – Query of model to recover customer training data, resulting in provider understanding and accurately predicting responses.
9. Attacking the ML supply chain – Compromise of model while being downloaded.
10. Backdoor ML – Algorithmic backdoors that activate with a specific trigger (temporal or logical), this class of attacks result in software behaving as anticipated until a specific set of critical events occur, then the fault is activated. For example, if a rapid response is needed to shut down a resource, the corrupted software may cause the system to hang briefly or to fail.
11. Software dependencies exploitation – Traditional software exploits resulting in vulnerabilities in libraries, etc.
12. Reward hacking – Mismatch between stated and true reward to attain unintended outputs.
13. Side effects – Disrupts environment attempting to change the goal of the learning software.
14. Distributional shifts – Training environment differs from execution environment and system cannot adjust, resulting in unanticipated behaviors.
15. Natural adversarial examples – Hard negative mining invoked failures, resulting in missing data due to inaccurate clusters being formed when entries fit more than one grouping, causing low probabilities of events occurring when those events are more frequent.
16. Common corruption – Inability to handle tilting, zooming, or noisy images, resulting in missing relevant findings.
17. Incomplete testing – System not tested in realistic operating conditions, overlooked use cases or misuse cases are not tested.
18. Malicious use of AI – Using algorithms in a legitimate but unintended manner (e.g., creating filter bubbles that result in attempting to verify questionable findings and receiving a larger quantity or support for erroneous findings).
19. Overfit models – Models that work well in a well-known constrained environment but lose accuracy in new environments.

The last two failures modes, malicious use of AI and overfitted models, were not identified in [1] as they are directly related to AI, not ML. The relationship between ML and AI is such that many of the misuse or malicious use of AI issues are solved by securing the underlying ML algorithm. Similar to software, many AI solutions have dual-use capability and therefore are often used outside of their intended purpose leading to unverified or invalid results.

Overfitting results from training a model too specifically. If the use case is very specific and constrained, the most appropriate solution is use of an expert system. The second, more common solution is to properly size the training dataset. In this case, the dataset must be (a) large enough to cover a variety of lessons to create the groupings and (b) small enough to allow for distinct cluster formation and appropriate classification.

This list can be condensed into four groupings:

1. Attacks against data – Change aspects of the data resulting in models deriving the wrong answers and insights.

2. Attacks against algorithms or systems – Result in wrong answers and insights, introduction of new vulnerabilities.
3. Attacks against models – Take advantage of design flaws or assumptions that are part of the design, resulting in unanticipated outputs where the problem origin is difficult to trace.
4. Unknown-unknowns – Attacks that are not part of the previous three groupings but are recognized as well-known security and AI issues, results vary from algorithm confusion, wrong answers, and failure to introduction of new vulnerabilities.

4. THREAT COUNTERMEASURES AND AI RECOVERY

Identification of challenges, particularly security challenges that AI/ML introduce to nuclear security, provides an opportunity to define countermeasures that can be used to mitigate them. The following list identifies and describes potential countermeasures to AI/ML threats:

- Testing – Extensive testing of all AI software with extensive use and misuse cases, long before deployment. Testing should include red team tests of AI, ML, and all associated dependencies (e.g., libraries). The higher the consequence, the more extensive, the higher the requirement for testing and layered solutions.
- Tabletop exercises – These events involve pulling together both domain and operational experts to create unanticipated scenarios in order to determine the automated responses.
- Explainable AI – Requires foundational understanding of ML algorithms that inform AI decisions, making them easily understood by humans. Information gained from this effort, along with tabletop exercises, can be used to inform secure-by-design decisions made by developers and integrators.
- System resilience – Widely recognized as an underserved aspect of security. Resiliency recognizes that no solution is perfect, but when problems occur they are quickly addressed while maintaining safe operation with partial functionality, until full functionality can be restored. Made up of:
  - Robustness deals with how well the AI software can identify and withstand adversity without significant performance degradation. Robust AI and ML will have security built in, making many of the attacks listed in this paper obsolete [4].
  - Redundancy measures the degree in which parts or functions are interchangeable and refers to the substitutable nature of elements of the system including software libraries and functions [4]. This involves creating alterative processing when robustness is impaired.
  - Resourcefulness details the ability to diagnose and prioritize problems and solutions in accordance with the mission [4]. This requires additional priorities learned from ML to use for processing algorithms.
  - Rapidity measures the time needed to return to a normal operational state [4]. Resilient systems maintain state awareness and an acceptable level of operational normalcy in response to disturbances [4].
- Data Resilience – Captures contextualization, descriptiveness, and temporal components. A relatively new concept [6,7,8] that reflects the growing awareness of the role of data veracity in the data-centric world of AI/ML. This initiative attempts to close gaps between information security constructs of confidentiality, integrity, availability, and non-repudiation [8] and information theory constructs of time, descriptiveness, and context [8]. These tenants of information security work well with data that have been entered by trusted identities but no mechanism exists that verifies data input into the system.
  - Confidentiality is the assurance that data are accessed only by the intended recipient.
  - Integrity ensures that data sent or stored are free from tampering.
  - Availability is the promise that data will be obtainable when needed.
  - Non-repudiation assures the identity of the user who sent or created the data.
- Resilient data [8] are created by binding data objects to their required environmental variables, changes in either the environment or the object leaves digital perturbations. The environmental

variables include timestamps (temporal component), object description (characteristics of the data object), and environmental variables that existed when the data were created or modified. Similar work has been performed by Adobe [9] to address the problem of deepfakes by adding metadata to existing data. Additionally, operational use profiles that create quantitative pictures of the data and processing environment can be built then used to assist in measurement deviations from the base profile.

- Maintain human-in-the-loop override capabilities – AI is a work in progress and presently offers strong decision-support capabilities. AI sensing outperforms AI judging, which outperforms AI prediction on human behaviors [10]. Thus, humans making the final decision will remain a critical part of the AI equation, especially where nuclear safety issues apply.

While testing, tabletop exercises, and hardware resiliency represent traditional means of addressing challenges, establishing software or data resiliency is more recent. Table 2 provides a cross-reference between the failure modes previously identified with potential mitigations.

Table 2. ML failure modes [3] and mitigation strategies.

Failure	**Mitigation**
Data attacks	Data resilience, resilient systems, explainable AI
System attacks	Resilient systems, explainable AI
Model attacks	Testing, resilient systems, explainable AI, tabletop exercises, human in the loop
Miscellaneous: Unknown-unknowns	Testing, resilient systems, tabletop exercises, human-in-the-loop, explainable AI

5. RECOMMENDATIONS AND CONCLUSIONS

A recent report of the malicious use of AI [11] delivered four high-level recommendations relevant to forecasting, prevention, and mitigation. These recommendations, while general in nature, are applicable to this effort with regard to nuclear safety.

1. Policy makers need tighter coupling with researchers to investigate, prevent, and mitigate malicious use of AI [11] where AI is being considered and deployed for all aspects of nuclear safety.
   a. Researchers and engineers should take seriously dual-use and be proactive [8] when designing nuclear safety solutions.
   b. Researchers and engineers need to evaluate existing security solutions for vulnerabilities listed above in specific deployed environments.
2. Best practices need to be identified in research areas with more mature methods for addressing dual-use concerns [11].
3. Expand the range of stakeholders and domain experts involved in discussions of challenges [11].

More specifically there are initiatives in various areas of government and across national laboratories and academia that deal with the various aspects of the identified failure modes. For instance, ML has been used to automate the detection of malware binaries in library functions; a similar strategy can be applied to AI/ML software [12,13,14]. In addition, other efforts to support data resiliency are underway [15,16,17]. However, an overall strategy pulling together all of these issues with potential countermeasures is lacking. Furthermore, while many unknown areas of AI may eventually be understood through explainable AI, there is a high likelihood for introduction of new vulnerabilities. Adversarial machine learning will continue to grow leading to discovery of new vulnerabilities impacting AI use in nuclear safety environments.

6. REFERENCES

[1] National Security Commission on Artificial Intelligence. Available: https://www.nscai.gov/about/faq.

[2] A. Parisi, August 2019. *Hands-on: Artificial intelligence for cybersecurity*, Packt Publishing Ltd., Birmingham, UK, ISBN: 978-1-78980-402-7.

[3] R. Shankar Siva Kumar, J. Snover, D. O'Brien, K. Albert and S. Viljoen, November 2019. "Failure modes in machine learning". Available: https://docs.microsoft.com/en-us/security/engineering/failure-modes-in-machine-learning.

[4] M. Bishop, M. Carvalho, R. For and L.M. Mayron, 2011. "Resilience is more than availability", In Proceedings of the 2011 New Security Paradigms Workshop, pp. 95 – 104, ACM.

[5] L. Wall, September 17, 2017. "Some financial and regulatory implications of artificial intelligence", Federal Reserve Bank of Atlanta.

[6] C. Sample, T. Watson, S. Hutchinson, B. Hallaq, J. Cowley and C. Maple, "Data fidelity: Security's soft underbelly", In Proceedings of the 11th IEEE International Conference on Recent Challenges for Information Science, pp. 315 – 321, May 10-12, 2017.

[7] M. DeLucia, S. Hutchinson and C. Sample. "Data fidelity in the post truth era part 1: Network data", in proceedings of the *International Conference on Cyber Warfare and Security,* pp. 149 – 159, March 2018.

[8] C. Sample, S.M. Loo and M. Bishop, "Resilient data: An interdisciplinary approach", In proceedings of IEEE Resilience Week, October 2020.

[9] N. Kobie, August 14, 2020. Adobe battles fake photos with editing tags. https://www.itpro.com/software/356786/adobe-battles-fake-photos-with-editing-tags.

[10] A. Narayanan, 2019. "How to recognize AI snake oil", https://www.cs.princeton.edu/~arvindn/talks/MIT-STS-AI-snakeoil.pdf.

[11] M. Brundage, S. Avin, J. Clark, H. Toner, P. Eckersley, B. Garfinkel, A. Dafoe, P. Scharre, T. Zeitzoff, B. Filar, H. Anderson, H. Roff, G.C. Allen, J. Steinhardt, C. Flynn, S. OhEigeartaigh, S. Beard, H. Belfield, S. Farquhar, C. Lyle, R. Crootof, O. Evans, M. Page, J. Bryson, R. Yampolskiy, D. Amodei, "The Malicious Use of Artificial Intelligence Forecasting, Prevention and Mitigation. February 2018.

[12] B. Beckman and J. Haile. "Binary analysis with architecture and code section detection using supervised machine learning", In proceedings of *Cyber Resilient Supply Chain Technologies (CReSCT) Workshop,* May 2020.

[13] J. Haile and S. Havens. "Identifying ubiquitous third-party libraries in complied executables using annotated and translated disassembled code with supervised machine learning", in proceedings of *Cyber Resilient Supply Chain Technologies (CReSCT) Workshop,* May 2020.

[14] E. Peterson, "Flexible adaptive malware identification using techniques from biology", *Summer Security Seminar Series Purdue University,* August 19, 2020.

[15] H.S. Che, A.S. Abdel-Khalik, O. Dordevic and E. Levi. "Parameter estimation of asymmetrical six-phase induction machines using modified standard tests. *IEEE Transactions on Industrial Electronics,* 64(8), pp. 6075-6085, 2017.

[16] K. Chan, K. Marcus, L. Scott, and R. Hardy. "Quality of information approach to improving source selection in tactical networks. In *IEEE 18th International Conference on Information Fusion,* pp. 566-573, 2015.