INL/CON-23-70978-Revision-0

Twenty Years and CountingWhere are they? Practical Recommendations for Commercializing AI/ML for Intrusion Detection in the Nuclear Industry

June 2023

Changing the World's Energy Future

Shannon Leigh Eggers, Robert S Anderson

Idaho National Laboratory

INL is a U.S. Department of Energy National Laboratory operated by Battelle 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/CON-23-70978-Revision-0

Twenty Years and CountingWhere are they? Practical Recommendations for Commercializing AI/ML for Intrusion Detection in the Nuclear Industry

Shannon Leigh Eggers, Robert S Anderson

June 2023

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

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

Twenty Years and Counting—Where are they? Practical Recommendations for Commercializing AI/ML for Cyber Intrusion Detection in the Nuclear Industry

U.S.A.

Shannon L. Eggers[1] and Robert S. Anderson[2]

Idaho National Laboratory

[1]shannon.eggers@inl.gov, [2]robert.anderson@inl.gov

Abstract

Research and development into applications for improving equipment condition monitoring programs at nuclear facilities has been around since the 1990s. However, while the field has moved from using data-driven machine learning (ML) algorithms for detection and prediction of equipment degradation and failure to prognostic capabilities, these applications are still not widely used in the U.S. nuclear industry. Additionally, there has been significant effort in designing both data-driven and physics-based artificial intelligence (AI) and ML models for many other potential applications in the nuclear industry, including cyber intrusion detection systems (IDS). However, as the last twenty years in condition-based maintenance research has shown us, there are significant hurdles that must be overcome for deployment of IDS on plant systems. This paper provides a discussion on the practical recommendations that researchers should consider for successful adoption of AI/ML IDS in the nuclear industry.

1. Introduction

Intrusion detection systems (IDS) have been used in information and control technology (ICT) since the early 1990s. Research into adopting artificial intelligence (AI) and machine learning (ML) techniques to detect intrusions in operational technology (OT) started in early 2000. However, twenty years later, there has been limited commercial success in AI/ML IDS in the OT industry and even less success in the nuclear industry. This paper provides a brief background on anomaly detection, ML for condition-based maintenance, and data-driven and physics-based ML for intrusion detection systems before providing a list of recommendations for researchers to consider when designing and developing AI/ML IDS applications for the OT environment in nuclear facilities.

2. Background

Anomaly detection

Anomaly detection is the ability to recognize patterns in data that are outside of expected or normal behavior. To detect an anomaly, a comparison is made between the target dataset and a "normal" dataset. This normal dataset may be defined by using predefined, hard-coded values or by techniques such as ML. In ML-based anomaly detection, mathematical algorithms are used for data analysis and decision making. ML applications perform based on what the algorithm is trained to do; it cannot adapt the learning process. In comparison, AI applications move one step further by using ML models to mimic human intelligence and adapt learning through voice recognition, natural language processing, computer vision, robotics and motion, planning and optimization, and knowledge capture [1]. Understanding and describing what caused the anomaly is termed 'explainability.' For example, consider an anomaly in data measured by a nuclear reactor process (e.g., pressure, temperature, flow). What caused the anomaly? Was it caused by normal process perturbations, human interaction, environmental changes, equipment degradation or failure, a cyber incident, or another irregularity?

Machine learning for condition-based maintenance

ML applications that first gained wider acceptance for use in nuclear power plants (NPP) used data-driven algorithms for online monitoring (OLM) [2-4]. The benefits of online sensor monitoring include better trending, identification of early degradation, improved performance evaluations, and reduced operations and maintenance costs [4]. Additionally, periodic instrumentation calibrations required by preventive maintenance programs often resulted in human performance errors leaving sensors in an error state rather than correcting instrumentation drift. Using performance data to identify issues reduces unnecessary maintenance and lowers potential for calibration errors or equipment damage [4]. Therefore, the goal for these early OLM applications in the nuclear industry was to analyze process or instrument data to identify when a sensor or device needed calibration, moving from preventive maintenance to condition-based maintenance programs. Early research in supervised ML was also used to identify and predict device degradation and/or failure by analyzing equipment data against known degradation and failure curves. As research in OLM continued in the nuclear industry, anomaly detection moved from diagnostic to predictive to prognostic, with capabilities to predict the remaining useful life of equipment [5].

Data-driven and physics-based machine learning intrusion detection systems

Real-time IDS for ICT systems have existed since the early 1990s [6]. IDS techniques are broadly separated into two categories—anomaly detection and knowledge detection. Knowledge (or misuse) detection uses known signatures to identify a specific pattern of misuse. For example, antivirus applications typically use knowledge, or signature-based, detection schemes. However, these traditional ICT IDSs are focused on monitoring network communications, traffic, and timing and are often insufficient for OT systems found in NPPs in part because they do not monitor the behavior of interconnected physical processes (e.g., laws of physics), are not designed for handling I&C systems with strict real-time communication requirements, and are ill-equipped for detecting highly sophisticated attacks [7, 8].

Historically, the breadth of research into IDS for OT systems began growing in the 2000s [7]. Early research included the use of host-based and network-based data-driven systems to detect anomalies in various aerospace, automotive, medical, and supervisory control systems [7]. Since these early methods were investigated, there has been significant research into the use of AI/ML data-driven, physics-based, and hybrid techniques for detecting cyber incidents in OT systems. However, twenty years later, there is still limited success for using AI/ML IDSs in the nuclear industry.

3. Practical Recommendations for Commercializing AI/ML IDSs in Nuclear Reactors

The use of AI/ML applications in OT systems is transformational technology. Despite the advantages of these applications, however, they are slow to be adopted in the nuclear industry. Table 1 provides a list of practical recommendations that researchers should evaluate as they develop new AI/ML IDSs for use in the nuclear industry.

Table 1. Practical recommendations for commercializing AI/ML intrusion detection technology in the nuclear industry.

Category	Discussion
Laws and regulations	Laws and regulations governing NPPs differ between nation-states. For example, while an OLM program has been used successfully with regulatory approval at the Sizewell B Nuclear Plant in the U.K. since 2005, it wasn't until 2021 when the U.S. Nuclear Regulatory Commission (NRC) determined that this same OLM methodology was acceptable for use with pressure, level, and flow transmitters for determining calibration surveillance intervals [9]. Similar regulatory constraints exist for implementing AI/ML IDS systems depending on where and how the technology will be implemented within an NPP. For instance, aside from other potential technical constraints, implementation of an AI/ML IDS on a safety-related instrumentation and control (I&C) system is improbable, due to the increased challenges in obtaining regulatory approval with these systems. *Recommendation: Researchers must understand the current laws and regulations in which the technology will be implemented. Can this technology by implemented as designed in an NPPs environment given the existing regulatory framework?*
Deployment options	IDSs can be designed to work in various locations. Are they standalone systems? Do they work in conjunction with another system? Are they part of a digital twin? Are they deployed on an I&C system, plant network, or business system? Do they need real-time data, near real-time data (i.e., from a data historian), or are there other time dependencies? Often, depending on the regulatory environment, systems installed on plant networks or I&C systems are considered 'systems, structures, and components' (SSC). In the U.S., SSCs are controlled by very strict engineering processes that differ based on the classification of the SSC (i.e., non-safety related, safety related, augmented quality). For instance, an IDS installed on a safety-related system or in such a manner that it will affect a safety-related system, will require extensive documentation and reviews prior to approval. And, as identified in the "laws and regulations" category, if NRC endorsement or approval is required, it may take years to be granted, if granted at all. On the other hand, IDSs installed on the business network are governed by corporate procedures, not plant engineering procedures. Thus, if a data-driven IDS uses near real-time data from a data historian located on the business network, the procedures and implementation process will be much simpler and faster. Furthermore, the design of plant network and/or I&C architectures differ greatly between NPPs so commercially viable products installed as SSCs must be amenable to different plant configurations. Additionally, it must be recognized that IDSs connected to I&C systems may unintentionally impact the operation of the system. I&C systems have very strict data and communication requirements; interruption of normal communications by an IDS may have an adverse impact on the I&C system. Generally, any type of active (versus passive) technique by an IDS is discouraged for use on an I&C system unless it is built into the system by the vendor (e.g., embedded IDS). Of course, incorporating a secure-by-design philosophy by designing IDS capabilities into a system during the systems engineering lifecycle is preferred rather than bolting on the capability after installation.

Category	Discussion
	Recommendation: Researchers must consider the implications of placing an IDS on the plant network. Will IDS operation adversely impact existing I&C systems? Is it deployable on current I&C architectures? Will regulatory approval be required? Can it be installed and qualified as part of a new digital I&C system?
Performance	Many research studies attempt to prove low uncertainties in a model by using visual aids. While an anomaly might clearly be present on a static image, in a real environment operators will not continuously monitor an IDS. The IDS must provide an informational output in the format of a notice, warning, and/or alert based on exceedance of score or predefined threshold, whether it is point-based, conditional-based, or contextual-based. Threshold optimization is difficult and may vary depending on which system the application is deployed. IDS performance is often evaluated by false positive, false negative, and true positive rates. While a 98% false positive rate may seem adequate, the researcher must consider what that metric means from an operational perspective. For example, if the IDS monitors one-second real-time data, then plant personnel may be required to evaluate 1,728 (or 2%) false alarms daily. As that level of response is operationally prohibitive, this system would likely be ineffective and quickly result in operator disuse and/or complacency. Similarly, if false negative rates are too low, cyber incidents may remain undetected. *Recommendation: Researchers must understand the operational impacts of performance metrics and optimize the algorithm and alerting to minimize false positives. Can an NPP operator effectively respond to the anticipated number of alerts? Will the IDS require augmentation of staff to monitor and respond to the system or can it be managed by existing personnel?*
Depth and breadth of application	Researchers are very imaginative when creating potential OT attack scenarios, and the literature is full of complex, stealthy attacks. Researchers are also often equally creative in developing AI/ML algorithms to detect these sophisticated attacks. However, while detecting stealthy attacks via physics-based algorithms may be cutting edge research, detection capabilities for simple OT attacks is still limited. Further, complex attacks are much less likely to successfully occur, rendering the need to detect them less of a priority for an NPP. Thus, similar to the defense-in-depth philosophy, effective detection for the entire NPP requires more than one tool in the toolbox—implementing multiple overlapping detection technologies is a more complete solution. *Recommendation: Researchers must understand the current limitations of existing OT IDSs and focus on improving OT detection, starting with simple attack scenarios. Keep it simple. Multiple IDS technologies may be necessary to provide full coverage for an NPPs OT systems, but perfect the basics first. Is a complex, physics-based detection application that only identifies anomalies in a narrow application, such as a pressurizer, a viable commercial product?*
Trustworthiness of model	Trustworthiness of AI/ML models is developed by improving interpretability, explainability, and transparency. • Interpretability is the ability for humans to understand the basis for decision making. Is the output of the application (e.g., notice, warning, alert) designed so that the human-in-the-loop can understand and make a decision?

Category	Discussion
	• Explainability is the ability to determine the cause of system behavior. How is the cause of an anomaly differentiated? Is the disturbance caused by normal, anticipated load variations or system interactions, or is it caused by unplanned external events, such as equipment degradation or failure, environmental changes, or unintentional or deliberate cyber incidents? • Transparency is the ability to understand what has been learned and why it has been learned. Is the AI/ML model a “black box” in which operators have no insight into how it functions, or can operators easily understand how the IDS works? *Recommendations: Researchers should work towards enhancing trustworthiness in the IDS and ensuring operators and/or responders can be trained to understand how the IDS works. Is the IDS interpretable, explainable, and transparent? Can the system be expanded to identify and explain all anomalies, not just cyber incidents?*
Data quality	While NPP processes are closed loop systems that normally have regular and predictable behavior, process data is often noisy. Additionally, planned reactivity maneuvers, plant startups, plant shutdowns, and unexpected plant transients affect the process data and plant’s physical behavior (i.e., physics). AI/ML applications require high quality datasets—it must be correct, complete, and unbiased, otherwise incorrect training data may result in inaccurate model design and/or learning, leading to unexpected behavior and confidence reduction in the model. Incorrect or corrupt data during operation can skew results and confidence in the system over time. Model bias can result from under- or over-sampling of data. In data-driven models, the data may be univariate or multivariate time series data that may not always be sequential (e.g., data may be skipped), especially in data historians that are near real-time mimics of the original process. *Recommendations: Researchers must recognize the limitations of NPP process data and the underlying physics in the processes. Can the data be used “as is” or is preprocessing necessary? If necessary, can these steps be easily performed in a production environment?*
AI/ML vulnerabilities	Interestingly, while it may fall into the category of a complex, stealthy attack as described in the “depth and breadth of application” category, AI/ML IDSs designed to detect cyber-attacks are themselves vulnerable to cyber-attacks. Similar to data quality issues, data corruption attacks can skew model learning and cause misclassifications or inaccurate predictions. Similarly, the models themselves can be manipulated or reprogrammed, leading to alteration of the learning process or introduction of backdoor trigger conditions. Attacks against IDSs can enable OT attacks to remain undiscovered. Researchers should ensure that cybersecurity is implemented throughout the development process and designed into the product and deployment environment. *Recommendation: Researchers should use Cyber-Informed Engineering principles [10] throughout the IDS systems engineering lifecycle to ensure security-by-design in their final product.*
Computing expense	Many of these AI/ML models are computationally expensive, requiring significant processing power and memory. Additionally, IDS applications on real-time OT systems require extensive storage capacity. Depending on the deployment location,

Category	Discussion
	these resources may be unavailable. For example, an algorithm requiring a high-performance computer to run has limited deployment options in the nuclear industry as the expense to run and maintain these computers is likely prohibitive. *Recommendation: Researchers should investigate optimization strategies early in the project to reduce the computing resources required for operation of the IDS. Will the application be capable of running in the nuclear facility?*
End-user	Many IDSs work well in a constrained lab environment in specific applications. Not only is it challenging to develop AI/ML anomaly detection models for one system, but it is challenging to apply the same model to additional systems within the same NPP. It may be even more challenging to apply the model to systems at another NPP. For an OT IDS to be commercially deployed in the nuclear industry, it must work in many situations. Additionally, the IDS must be user friendly to a range of personnel with different skillsets (e.g., engineers, maintenance, operators). They must be able to install, configure, monitor, and respond to the system. *Recommendation: To develop a commercially viable product for use throughout the nuclear industry, researchers should work towards developing robust models that are transferable between systems and plants, independent of differences in actual training and operational data. Researchers should also provide capabilities for educating the end-user and allow them to configure the system. If a non-computer expert is responsible for responding to an alert, is it straightforward enough for them to use or do they need to call an expert?*

Conclusion

This paper presents recommendations for research and development of AI/ML IDSs for use within OT environments in a nuclear facility. While specifically addressing IDS, many of these concepts can be broadly applied to other AI/ML applications currently under consideration for the nuclear fleet, such as semi-autonomous, autonomous, or anticipatory reactor control; integrated energy systems; load-following grid designs; emergency response; environmental monitoring; and other various concepts in the expansive field of digital twins. Additionally, these recommendations are broadly applicable to both the existing fleet and the future fleet. While it is understood that low technology readiness level (TRL) research is focused on basic principles and proof of concept, as applications move further into development and deployment TRLs, the likelihood of continued commercial success will be improved if these recommendations are considered.

Acknowledgments

This work was supported by the U.S. DOE National Nuclear Security Administration Office of International Nuclear Security under the DOE Idaho Operations Office, Contract DE-AC07-05ID14517.

References

[1] van Duin, S. and N. Bakhshi. *Part 1: Artificial intelligence defined: The most used terminology around AI*. Deloitte, Accessed on: March 28, 2017. Available: https://www2.deloitte.com/nl/nl/pages/data-analytics/articles/part-1-artificial-intelligence-defined.html

[2] Garvey, J., D. Garvey, R. Seibert, and J.W. Hines, "Validation of on-line monitoring techniques to nuclear plant data," *Nuclear Engineering and Technology,* vol. 39, no. 2, p. 133, 2007.

[3] Uhrig, R.E. and J. Hines, "Computational intelligence in nuclear engineering," *Nuclear Engineering and Technology,* vol. 37, no. 2, pp. 127-138, 2005.

[4] EPRI, "On-line monitoring of instrument channel performance," Electric Power Research Institute, 2000, Available: https://www.epri.com/research/products/000000000001000604.

[5] Coble, J.B., P. Ramuhalli, L.J. Bond, W. Hines, and B. Upadhyaya, "Prognostics and Health Management in Nuclear Power Plants: A Review of Technologies and Applications," United States, 2012, Available: https://www.osti.gov/biblio/1047416.

[6] Kemmerer, R.A. and G. Vigna, "Intrusion detection: a brief history and overview," *Computer,* vol. 35, no. 4, pp. supl27-supl30, 2002.

[7] Mitchell, R. and I.-R. Chen, "A survey of intrusion detection techniques for cyber-physical systems," *ACM Computing Surveys (CSUR),* vol. 46, no. 4, pp. 1-29, 2014.

[8] Cárdenas, A.A., S. Amin, B. Sinopoli, A. Giani, A. Perrig, and S. Sastry, "Challenges for securing cyber physical systems," in *Workshop on future directions in cyber-physical systems security*, 2009, vol. 5, no. 1: Citeseer.

[9] NRC, "ML21179A062, Safety evaluation by the U.S. Nuclear Regulatory Commission for Analysis and Measurement Services Corporation topical report AMS-TR-0720R1, "Online monitoring technology to extend calibration intervals of nuclear plant pressure transmitters" by the Office of Nuclear Reactor Regulation," 2021.

[10] DOE. *The U.S. Department of Energy (DOE) National Cyber-Informed Engineering (CIE) Strategy Document*. U.S. Department of Energy Office of Cybersecurity, Energy Security, and Emergency Response, Accessed on: September 2022. Available: https://www.energy.gov/ceser/articles/us-department-energys-doe-national-cyber-informed-engineering-cie-strategy-document