Resilient Machine Learning: Advancement, Barriers, and Opportunities in the Nuclear Industry

Khadka, A., Sthapit, S., Epiphaniou, G. & Maple, C.

DOI 10.1145/3648608
ISSN 0360-0300
ESSN 1557-7341

Publisher: Association for Computing Machinery (ACM)

Resilient Machine Learning: Advancement, Barriers, and Opportunities in the Nuclear Industry

ANITA KHADKA, University of Warwick, Coventry, UK
SAURAV STHAPIT, Coventry University, Coventry, UK
GREGORY EPIPHANIOU, University of Warwick, Coventry, UK
CARSTEN MAPLE, University of Warwick, Coventry, UK

The widespread adoption and success of **Machine Learning (ML)** technologies depend on thorough testing of the resilience and robustness to adversarial attacks. The testing should focus on both the model and the data. It is necessary to build robust and resilient systems to withstand disruptions and remain functional despite the action of adversaries, specifically in the security-sensitive Nuclear Industry (NI), where consequences can be fatal in terms of both human lives and assets. We analyse ML-based research works that have investigated adversaries and defence strategies in the **NI**. We then present the progress in the adoption of ML techniques, identify use cases where adversaries can threaten the ML-enabled systems, and finally identify the progress on building **Resilient Machine Learning (rML)** systems entirely focusing on the NI domain.

CCS Concepts: • **Computing methodologies** → **Artificial intelligence;**

Additional Key Words and Phrases: Resilient machine learning, nuclear industry, adversaries, defences, resilience, robustness, survey

ACM Reference Format:
Anita Khadka, Saurav Sthapit, Gregory Epiphaniou, and Carsten Maple. 2024. Resilient Machine Learning: Advancement, Barriers, and Opportunities in the Nuclear Industry. *ACM Comput. Surv.* 56, 9, Article 224 (April 2024), 29 pages. https://doi.org/10.1145/3648608

1 INTRODUCTION

The **Nuclear Industry (NI)** is a highly complex and security-sensitive industry with strict safety regulations. It needs technologies that can be used in environments with no human accessibility such as small tunnels with radiation, reactors containing hazardous elements, and so on. Despite the regulations, it is reported that about 60% of major failures in the **Nuclear Power Plant (NPP)** are caused by human errors [70]. Consequently, there have been efforts to eliminate human errors by integrating automation in the industry [62]. Certain tasks and sectors within the NI can benefit from automation, as several studies have presented their attempts to automate tasks using **Machine Learning (ML)** techniques. These tasks range from *"Fault Diagnosis"* [132, 154], *"Remote*

The work presented has been funded by Grant **EP/R026084/1** Robotics and Artificial Intelligence for Nuclear (RAIN) through the Engineering and Physics Research Council (EPSRC).
Authors' addresses: A. Khadka, G. Epiphaniou, and C. Maple, University of Warwick, Coventry, UK, CV4 7AL; e-mails: anita.khadka@warwick.ac.uk, gregory.epiphaniou@warwick.ac, cm@warwick.ac.uk; S. Sthapit, Coventry University, Coventry, UK, CV1 5FB; e-mail: ae0066@coventry.ac.uk.

ACM 0360-0300/2024/04-ART224
https://doi.org/10.1145/3648608

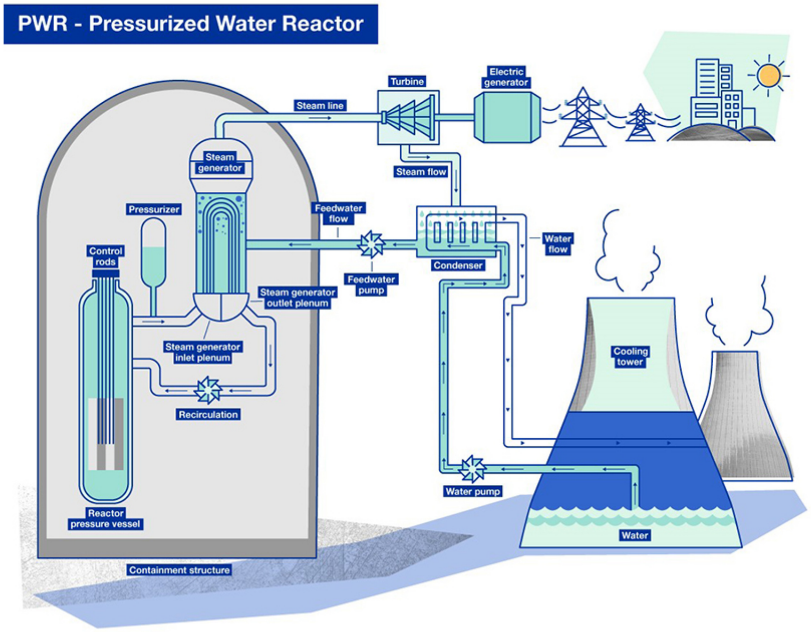

Fig. 1. Nuclear energy generator with pressurised water reactor. Source: Reference [51].

Inspection" [28], *"Nuclear Fuel Management"* [78] to *"Nuclear Decommissioning"* [162], to name a few. A comprehensive list of existing works, the benefits, and the risks of using ML in the industry is discussed in detail in Section 4.1. But first, we start the article by introducing the NI and **Artificial Intelligence (AI)**.

Nuclear Industry and the processes within it, revolve around the nuclear reaction that produces nuclear energy. Nuclear energy is a form of energy released from the nucleus of the atom as a result of *fission* or *fusion* [51]. In *fission*, the nuclei of atoms split into several parts and in *fusion* multiple nuclei fuse. For example, in *fission* reaction, an incident neutron splits *Uranium-235* into *Barium* nucleus and a *Krypton* nucleus and two or three neutrons [51]. Figure 1 depicts a simplified nuclear power generator using *fission* with basic components of a typical nuclear plant. A typical nuclear energy plant has a reactor along with a water supply, pressuriser, condenser, and so on. The nuclear plant and industry, in general, comprises both physical- and software-based infrastructures where monitoring, coordination, controlling, and integration of the operations are imperative and need to adhere to the safety regulations [136]. The faults can arise from sensor degradation to external attacks and its consequences can be from reduced performance to safety hazards [107]. Nuclear-related incidents can have severe consequences; we summarise some of the real-world nuclear disasters and their consequences due to various factors such as human error, safety test violation, and natural disasters (e.g., earthquake) in Table 1. To eliminate and reduce these incidents, one way is to build an intelligent and robust system in place.

Artificial Intelligence (AI) is a concept to create intelligent machines that can simulate human behaviour, and ML is the application that uses statistical methods and historical data to learn human behaviour without being programmed explicitly. The advancement in computational processing both in software (e.g., cloud computing, distributional computing, **Compute Unified Device Architecture (CUDA)**, Tensorflow, Pytorch) and in hardware such as **Graphical**

Table 1. Examples of Major Real-World Nuclear Disasters in History

Incident	When	Where	Explanation	Cause	Consequence	Reference
Three Mile Island Accident	1979	United States	Misinterpretation of reactor condition by operators.	Operator Misinterpretation and Misjudgement	Limited radioactive gas release; no direct fatalities	[163]
Chernobyl Disaster	1986	Soviet Union (now Ukraine)	Safety test violation leading to reactor explosion and fire.	Operational and Safety Protocol Violation	Extensive environmental contamination; large-scale human impact	[44]
Tokaimura Nuclear Accident	1999	Japan	Excessive addition of uranium solution to a tank.	Safety Protocol Violation	Two fatalities; radiation exposure to workers	[146]
Davis-Besse Incident	2002	United States	Severe corrosion problem on reactor vessel head due to maintenance issues.	Maintenance and Inspection Error	No fatalities; potential for severe accident prevented	[33]
Fukushima Daiichi Nuclear Disaster	2011	Japan	Tsunami and earthquake leading to equipment failures and meltdowns.	Underestimation of Natural Disaster Risk	Major radiation release; long-term environmental impact	[32]

Processing Unit (GPU) and **Tensor Processing Unit (TPU)** has led to the rapid progression in AI and ML. The benefits of adopting intelligent learning techniques in the NI can be highly rewarding. However, any malfunction in such automation can lead to economic loss as well as human lives. It is an extremely precarious environment that can, on one hand, benefit from automation but, on the other hand, can lead to a catastrophe.

In recent years, particularly with the introduction of deep learning, the performance of ML models has significantly improved and their adoptions have increased in various industries for tasks such as object classification, object recognition, natural language understanding, speech recognition and generation, and many more [67, 93]. This widespread adoption of ML techniques has also manifested the rise in malicious manipulation of ML algorithms. These manipulations, commonly known as adversarial attacks, can influence the decision process of ML techniques producing results favouring attackers' objectives [15, 56, 57]. Different types of adversarial attacks can be classified based on their goals and capabilities. The attacks could be white-box attacks, black-box attacks, evasive, poisoning, and exploratory attacks; they are explained in more detail in Table 3. In this article, we investigate the applications of ML and their security focusing on the NI. To comprehend the stage of resilient ML in the NI, we conducted a gap analysis on the adversarial attacks and defensive strategies in ML-enabled applications only focusing on the NI. The main findings of this work are as follows:

— We identify various applications adopting ML techniques in the NI.
— We observe that the adoption of ML techniques is slowly emerging in the NI, however, the study of the security and resilience towards the feared events in ML-enabled nuclear applications is still in the early stage.
— We identify various targeted nuclear use cases for the adversarial attacks and threats generated by the attacks.
— We identify the opportunities and barriers to the adoption of **Resilient Machine Learning (rML)** in the NI.
— We promote various open research issues and propose future research directions in ML-driven nuclear operations and processes.

Table 2. Acronyms Used in the Article

Acronym	Definition
AI	Artificial Intelligence
aML	Adversarial Machine Learning
API	Application Program Interface
BIM	Basic Iterative Method
CPS	Cyber Physical Systems
CW	Carlini-Wagner
CUDA	Compute Unified Device Architecture
DCN	Deep Contractive Networks
DNN	Deep neural networks
FGSM	Fast Gradient Simple Method
GAN	Generative Adversarial Network
GPU	Deep neural networks
JSMA	Jacobian-based Saliency Map Attack
MIM	Momentum Iterative Method
ML	Machine Learning
NPP	Nuclear Power Plant
PGD	Projected Gradient Descent
rML	Resilient Machine Learning
ISR	Intelligence, surveillance and reconnaissance
SVM	Support Vector Machine
TPU	Tensor Processing Unit
UAB	Universal Adversarial Perturbation

The article is structured as follows: We define acronyms used in the article in Table 2. Section 2 describes the methodology of collecting papers to review. Section 3 provides generic background on ML, Adversarial Machine Learning (aML), and rML. This section specifically presents the threat model in ML, various types of adversarial attacks, and discusses various defensive strategies against adversarial attacks on ML. Section 4 includes a discussion on the integration of ML in the NI, barriers and concerns of adopting ML techniques in the NI, identification of the nuclear sectors adopting ML techniques for different scenarios, and the uses of different types on ML techniques in the industry. This section also presents a discussion of **aML** and rML in the NI. Last, Section 5 reports various open research directions in the domain. Section 6 concludes the work.

2 RESEARCH METHODOLOGY

This work is related to the study of defence systems applied in aML techniques focusing on the NI. We used various combinations of words and phrases to find relevant research works from the plethora of research publications in the digital world. We combined words and phrases including adversar*, machine learn*, resilien*, artificial intelligen*, and nuclear to collect the literature in the field. We used two scientific databases, scopus[1] and web of knowledge,[2] to search the relevant articles. In total, we obtained more than 700 documents. We filtered a large number of papers by reading their title and abstracts and narrowed down the list to 120 papers systematically. These

[1]https://www.scopus.com/
[2]https://www.webofknowledge.com/

selected papers are strictly the study of AI in the NI. On reading each paper in detail, we found that there are not many works researching rML in nuclear systems. We obtained less than 10% of relevant papers that have studied adversarial attacks and defence mechanisms on ML algorithms focusing on the nuclear systems' scenarios. We also conducted backwards and forward citation trails from the narrowed-down papers to identify relevant papers and, at the end, we studied 186 publications.

3 BACKGROUND IN RESILIENT MACHINE LEARNING

In this section, we briefly discuss the background topics and relevant publications in rML for general purposes. For in-depth literature reviews on adversarial attacks and defensive strategies in ML algorithms, we direct the readers to References [3, 27, 102, 124, 137]. Specifically, the review presented in References [124] and [102] describes **Cyber Physical Systems (CPS)** and the progress of such systems towards their safety and resilience against adversarial attacks.

3.1 Adversarial Machine Learning (aML)

In simple terms, ML is the study of automated techniques that make use of large sets of data and algorithms to imitate and learn human behaviour [19, 71]. This field has undergone significant progress in the past few decades. As the technology towards AI matured, multiple learning methods have been proposed [67]. The learning method that requires labelled data to train learning algorithms is supervised learning. Some widely used supervised learning methods include but are not limited to Neural Networks, Naive Bayes, Linear regression, Logistic regression, and **Support Vector Machine (SVM)**. Unlike supervised learning, unsupervised ML does not need labelled data to learn. Unsupervised learning algorithms discover hidden patterns within data without the need for human intervention or labelling [19]. The progress of ML methods has led to the adoption of them in diverse domains. However, a rise in the adversarial attacks in the learning models is also undeniable [157]. Lately, researchers are exploring vulnerabilities of ML models and identified that the models may be susceptible even to a small perturbation [15, 57, 125, 157]. The perturbation can be applied via different mediums, e.g., on the training data (e.g., deliberated to cause incorrect classification) [35], on the learning model (e.g., manipulation in parameters or features of the model). These vulnerabilities can affect their trustworthiness and applying ML in security-sensitive environments such as nuclear facilities can have irreversible consequences and be dangerous. In the next section, we briefly discuss a threat model that comprises ML attack surface, adversarial capabilities, and adversarial goals.

3.1.1 Threat Model in Machine Learning. ML models are susceptible to adversarial actions; the security and privacy of the ML can be quantified by comprehending the adversarial capabilities and adversarial goals [57, 127, 128, 157]. We list some widely explored adversarial attacks in Table 3. We recommend research works [57, 127] for a detailed exploration of generic adversarial attacks. Based on the strength of adversarial attacks, a threat model has been proposed that comprises ML attack surface, adversarial capabilities, and adversarial goals [127, 128].

— **ML Attack Surface:** Broadly, an ML system has three main components: input, processing, and output. Adversaries can attack in any component, for example, they can attempt to manipulate the collection and processing of data, corrupt the model, or even tamper with the output [128]. Figure 2 shows an attack surface in a generic ML model [128]. An attack can happen at any stage, from the collection of data to the processing of the data and can also be present in the learning model. If an adversary is successful, then it will produce a wrong result that can have devastating consequences. If we consider autonomous driving, the input may be images of road signs and the task would be to classify the road sign into

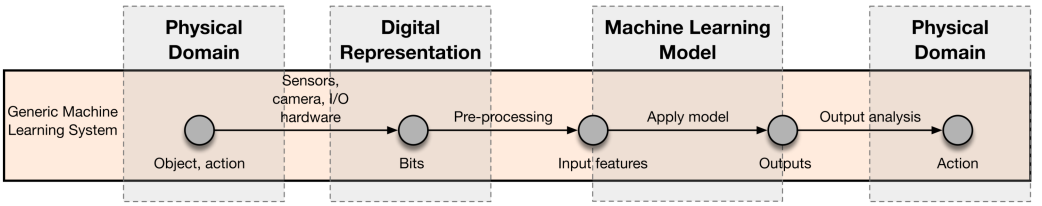

Fig. 2. System's attack surface: The generic ML model pipeline [128].

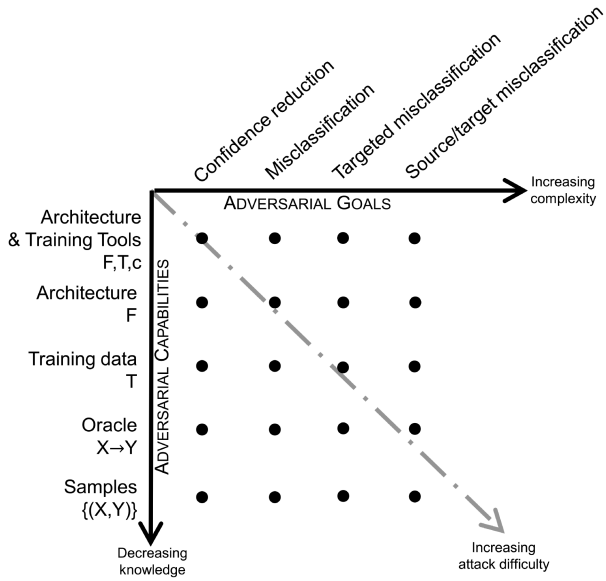

Fig. 3. A taxonomy of adversaries against ML models based on adversarial goals and adversarial capabilities [127].

one of the road signs (output). In case of an adversarial attack, if the "Go" sign is shown instead of the "Stop" sign in a traffic signal, then autonomous vehicles will continue to move when they should have stopped, which can lead to accidents costing human lives and assets.

— **Adversarial Capabilities:** A threat model can be defined by the actions and information based on the level of access to adversaries. Depending upon the type of information and access level, the adversaries can impose stronger or weaker adversarial attacks, as depicted in Figure 3. Adversarial capabilities try to identify **where** and **how** attackers can subvert the systems under attack [128].

— **Adversarial Goals:** One of the primary objectives of an adversary is to attack the ML model and generate the incorrect output. There can be various ways, such as by reducing the confidence of the model prediction and misclassifying the results. Therefore, modelling the security and safety of an ML model against adversaries can be structured around a well-defined taxonomy of adversarial capabilities and goals, proposed by Reference [127]. The taxonomy is shown in Figure 3.

Based on the threat model discussed in Section 3.1.1, we list some widely explored adversarial attacks in Table 3. We recommend readers to read References [57, 69, 127, 157] for a detailed exploration and discussion on adversarial attacks.

Table 3. List of Popular Adversarial Attacks

Attack	Description	Reference
Exploratory	It tries to gain access and modify ML models. For this attack, adversarial examples are crafted in such a way that the model passes them as real examples. Some of the popular types of exploratory attacks are Model Inversion, model extraction via **Application Program Interface (API)**, and membership inference attacks.	[18, 127]
Evasive	It is an adversarial attack that forces an ML model to provide false prediction and evade detection. **Generative Adversarial Network (GAN)**-based attacks, adversarial example generation, and adversarial classification are some of the notable methods to generate an evasive attack.	[15, 18, 127]
Poisoning	It tampers with training data, leading to predicting the correct output from the learning model. The goal of the attacker is to get their tampered data (adversarial examples) to be accepted in the training data.	[16–18, 127]
White-box	It is a type of attack where an attacker knows the architecture of ML models of the systems, e.g., the number of layers in a neural network, parameters' values, and algorithms such as gradient optimisation, activation function, and so on. With this information, the model can be exploited by altering the input by crafting perturbations.	[27, 43, 127]
Black-box	It is a type of attack where an attacker does not have any knowledge about the ML model except the input and the output of the model. Here, attackers may have access to the settings of past inputs to analyse the vulnerability of the model.	[126, 127]

3.1.2 Methods for Adversarial Attacks. The vulnerabilities of neural networks to adversarial examples were initially studied by Reference [157]. Szegedy et al. stated that imperceptible adversarial perturbations (examples) can be introduced to data to mislead ML models [157]. There are several research works carried out in the literature to minimise errors while calculating the adversarial sample. We present some of the notable works in aML literature in Table 4, however, for a more detailed discussion on adversarial attack-generating methods, we would like to direct the reader to References [4, 27, 128].

3.2 Resilient Machine Learning (rML)

ML models need to be secure, trustworthy, robust, and resilient. While **resilience** can be associated with the ability to return to normal operations over an acceptable period after the disruption in the operations, **robustness** is the ability to maintain operations during a crisis [21]. With the resilience and robustness in place on the systems, there can be increased trust and security towards the systems. Many efforts have been invested to achieve such salient features; some of them are understanding and generating different attacks [22, 25, 42, 56, 173], detecting adversarial examples [49, 108, 112], defending already trained models [61, 61], training robust models [92, 109, 114],

Table 4. Widely Used aML Methods

Methods	Definition	Reference
Fast Gradient Simple Method (FGSM)	It calculates the gradient of the cost function with respect to the input of the neural network.	[57]
Basic Iterative Method (BIM)	Reference [92] extended **FGSM** to improve the performance by running a small step size iterative optimiser multiple times, while clipping the intermediate adversarial samples after each step ensuring to be in the range of an original input.	[92]
Projected Gradient Descent (PGD)	It is similar to BIM, while **BIM** uses a negative loss function, the loss function is explored by re-starting the gradient descent from many points in the vector norm of infinity L_∞ around the input examples in **PGD**.	[109]
Momentum Iterative Method (MIM)	Gradient descent algorithms are accelerated by accumulating a velocity vector in the gradient direction of the loss function across iterations.	[42]
Carlini-Wagner (CW)	It creates an adversarial instance by finding the smallest noise added to an image that will change the classification to a class in such a way that the output is still in the valid range.	[25]
Universal Adversarial Perturbation (UAB)	It is computed to fool a network on all data in the dataset rather than a single input data with high probability.	[116]
Jacobian-based Saliency Map Attack (JSMA)	It uses the forward derivative to construct adversarial saliency maps, which show input features to include in perturbation to produce adversarial samples.	[127]
DeepFool	It aims at minimising the distance between perturbed samples and the original samples by iteratively adding the perturbations and estimating the decision boundaries between the classes.	[117]

understanding the weakness and vulnerability [48, 148], and more. Some of the defensive strategies are implemented during the training phase and on training data, while others are implemented during the testing phase. For example, while training the model, adding adversarial data in the training set can help the learning model become resilient to adversarial perturbations. This is one of the ways to tackle adversarial attacks. Table 5 presents several defensive methods proposed in the literature.

4 INTEGRATION OF INTELLIGENT OPERATIONS IN THE NUCLEAR INDUSTRY

The NI comprises activities that provide the equipment and services necessary for the construction, supply, and management of nuclear power plants [160, 161]. The industry is inherently complex and has technically challenging engineering systems that consist of numerous components and interdependent systems. Many hazardous elements are involved in the systems, such as Uranium and Plutonium [161]. Hence, they must operate safely and securely. With the

Table 5. Widely Used Defensive Strategies for Adversarial Attacks

Method	Definition	Reference
Brute-force adversarial training	Adversarial training is a standard brute force approach where the defending method generates adversarial examples and augments these perturbed data into the training set while training the targeted model.	[179]
Data Randomisation	This technique attempts to randomise the effects of adversarial perturbations. Reference [23] proposed the technique for randomly shuffling the order of data in memory, which makes it more difficult for attackers to exploit memory errors.	[23]
Deep Contractive Network	The **Deep Contractive Networks (DCN)** works by adding a smoothness penalty to the loss function. This penalty encourages the network to learn features that are smooth and invariant to small perturbations, which makes it more difficult for attackers to generate adversarial examples that will fool the network.	[58]
Gradient Masking	Adversarial examples generation techniques access the gradient of ML models to generate adversarial attacks on the model. Gradient masking is one of the techniques that denies access to the gradient details to the attackers.	[106, 127, 138]
Defensive Distillation	Distillation is used for distilling the knowledge of a more complex neural network into a smaller network. They used it as a technique for model compression, where a small model is trained to imitate a large and complex one to obtain computational savings.	[129]
DeepCloak	It is a defensive method that utilises a masking layer, where the layer is inserted just before the layer that handles classification. The added layer is explicitly trained by a forward-passing clean and adversarial pair of objects (e.g., images), and it encodes the differences between the output features of the previous layers for those pairs of objects.	[52]
Feature squeezing	It reduces the complexity of representing the data so the adversarial perturbations disappear because of low sensitivity, as shown in Figure 3. It reduces the search space available to an adversary by combining samples that correspond to many different feature vectors in the original space into a single sample.	[175]

involvement of radiation levels and extremely harsh and hazardous environments, human access is mostly restrained in many nuclear facilities. Building autonomous systems that can operate safely in such hazardous environments is preferable and beneficial. If the industry is to adopt ML techniques that can operate autonomously, then machines need to learn the operation as humans do and make decisions accordingly. Any wrong decision can be expensive in terms of human lives, as well as the economy. The NI covers a wide range of security-sensitive tasks, from nuclear decommissioning to nuclear fuel management. While Figure 4 shows the various nuclear sectors in the industry, Table 6 exhibits several applications mapped to the sectors presented in Figure 4. It can be seen that researchers have applied ML techniques to various applications. For example, cracks detection in underwater surfaces in NPPs [28, 40, 147], crack detection in the metallic surface in NPPs [29], identification of accidents like drop off a control rod [153], fault diagnosis in

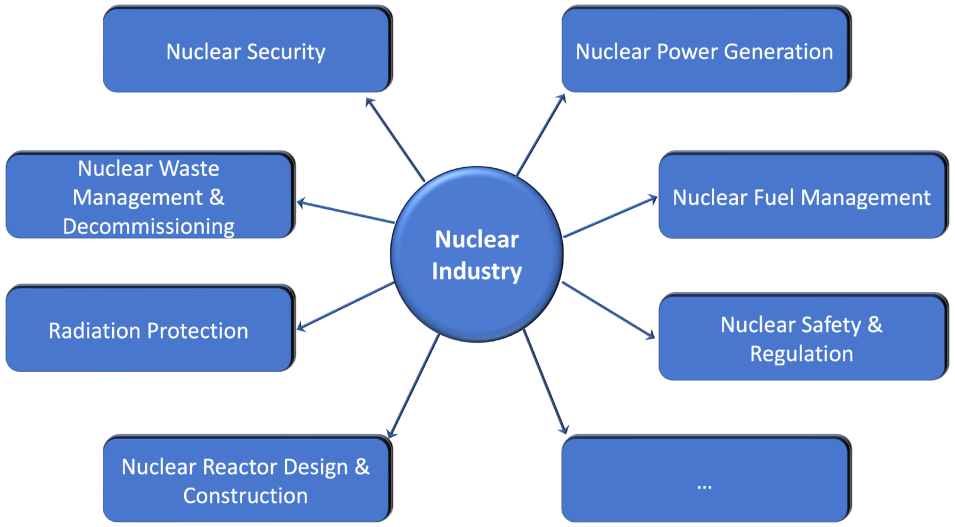

Fig. 4. Nuclear industry–related tasks that can leverage ML.

the nuclear facilities, fuel management, prediction of hydrogen concentration [30, 82], pressure vessel water level estimation [83, 89], loss of coolant accident in a nuclear reactor [120], and many more. NI is thriving to innovate, for example, to build safe and affordable energy, and existing NPP are committed to improving the safety of human lives and maintaining smooth operation in the facilities so there is a potential of high reward in embedding AI in the industry [72].

4.1 Machine Learning in the Nuclear Industry

In this section, we discuss existing applications of ML in the NI that have adopted ML techniques. We mention the references that have adopted ML for each task; the complete list of tasks and references is listed in Table 6.

- **Fault Diagnosis** identifies any abnormal activities in the nuclear facilities [154]. Faults can be sensor failure, sensor blockage, cracks in the walls, the reactor, and more. Since nuclear facilities can have multiple components including sensors, actuators, and controllers, manual identification of the precise location of faults can be a challenging task. Therefore, the adoption of ML techniques to identify faults autonomously can be hugely beneficial. ML techniques can help ease challenges such as learning adversaries from sensors, identifying patterns, diagnosing abnormality [132], identification of sensor failures in boiling water reactor, checking sensor condition, control systems to tolerating faults including loss of coolant accident, ejection of control rod, and so on [65].
- **Remote inspection** of the components of NPP is an important task of the NI and needs to be carried out at regular intervals, because the nuclear facilities can be extremely volatile and dangerous [154]. The preventive remote inspection helps avoid accidents by detecting a fault earlier and ensuring safety in the facilities, such as detecting crack patches in each recorded video clip by the remote inspecting devices [28].
- **Nuclear Fuel Management (NFM)** is a complex task that manages fuel-related tasks such as maintaining the quantity and quality of new fuel assemblies, reloading the partly burnt fuel assemblies, the core-loading pattern, and planning for control rod insertion for each reloaded cycle. The core objective of NFM is to minimise the cost while optimising energy

Table 6. Research Works Adopting ML Techniques in the Nuclear Industry

Focus	Task	Reference
Fuel Management	In-core fuel management	[47, 121, 122, 145]
	Fuel loading/reloading pattern	[10, 36, 39, 80, 81, 182]
	Reactor core parameters optimisation	[66, 90]
Fault diagnosis	Plant condition monitoring	[184]
	Fault diagnosis	[63, 91]
	Power control of reactor	[65]
	Identification of operational parameters	[65]
	Transducer and actuator condition	[54]
	Cracks detection	[28, 29]
Transient Identification	Identification of plant transients	[46]
	Classification of transients	[181]
	Plant transients diagnosis	[143]
	Transient type and severity	[115]
	Reactor plant transients	[103, 111]
	Identification of faults in transients	[8]
	Initiating event detection	[171]
Identification of accidents	Loss of coolant accident	[31, 120, 140, 142, 158]
	Maximum cladding temperature	[24]
	Power peaking factor	[11]
	Identification of nuclear accidents	[55, 133, 153]
	Prediction of hydrogen concentration	[30, 82]
	Pressure vessel water level estimation	[83, 89]
Radiation Protection	Radiation monitoring	[50, 53, 76]
	Radiation detection	[64]
	predict radiation dose levels	[13]
Nuclear Waste Management & Decomissioning	Waste material classification for nuclear decommissioning	[149]
	identify and remove waste material from nuclear facilities	[2]
	Safety of nuclear waste isolation repositories	[14]
	improve the safety of nuclear waste disposal	[176]
	Safety and efficiency of nuclear waste vitrification	[38, 60, 152]
	use of digitalisation for nuclear waste management	[88]
Nuclear Security	Investigation of security measure for NPP	[87, 96–98, 151]
	Nuclear infrastructure security modelling and simulation	[41, 139]
	Fault diagnostic system for online security assessment	[75]
	Nuclear energy security and sustainability	[134]
	Anomaly detection in nuclear security	[6, 7]
Nuclear Safety	Design of NPP safety systems	[1, 95, 110]
	Safety requirement analysis	[12, 85, 94, 131, 165, 167, 178, 183]
	Monitoring nuclear operator safety-relevant tasks	[77, 86, 141]
	Knowledge discovery for nuclear safety	[59, 164]
	Nuclear safety enhancement or assessment or management	[135, 150, 155, 166, 168, 172, 174]
	Safety parameters prediction in NPP	[84, 85]

demand and maintaining safety. Table 6 presents research focused on the use of AI in NFM.

— **Nuclear decommissioning** refers to the decommissioning of the nuclear facility after its operational life. A typical life span of NPP is approximately 25 years, however, there are hundreds of nuclear power plants still around that were built in the 1950s [70]. Most NPPs that were built in such period have well passed their due date to decommissioning [162]. Due to the higher cost of decommissioning a nuclear power plant, the Nuclear Regulatory Commission extended the operating licenses of several NPP that are over 40 years old. The decommissioning NPPs can cost about $3 billion and results in the loss of jobs [74, 104, 105, 162]. This resulted in most of the operating nuclear power plants around the world, having an average age of 25 years, being around for some time. This is considered one of the major and essential tasks of the NI to be completed [162].

— **Transient identification** is the process of identifying undesirable changes in the state of a NPP from normal to abnormal. It can be caused by a variety of factors, such as component failures, such as rupture in the steam-generating tube, disturbance in the flow of coolant of the reactor, and control systems sending a wrong signal [118, 154, 180]. ML techniques can improve the accuracy of transient identification. For example, identifying untagged transients [46], classification of U-tube steam generator [181], and use of resilient backpropagation for the diagnosis of the NPP [143], to name a few.
The use of ML for transient identification has several advantages. First, ML algorithms can learn to identify transients even if they are not tagged. This is important, because not all transients are tagged, and tagging can be a time-consuming and labour-intensive process. Second, ML algorithms can be trained on a large dataset of transients, which can improve their accuracy. Third, ML algorithms can be used to identify transients in real time, which can help operators take corrective action quickly. There can be some challenges associated with using ML for transient identification. For example, ML algorithms can be sensitive to the quality of the data they are trained on. If the data is noisy or incomplete, then the accuracy of the algorithm can drop. Additionally, ML algorithms can be computationally expensive to train and run. This can be a challenge for NPPs with limited computing resources.

— **Identification of accident scenarios** Accidents in nuclear facilities can be very expensive and devastating and can cost human lives. There can be several types of accidents in the facilities such as the breakdown of the pump (e.g., pump coolant to the nuclear reactor), ejection of the control rod, a burst of coolant pipe, and leakage of nuclear material. Researchers have applied ML techniques to various accidental scenarios, including loss of coolant accident [142], prediction of hydrogen concentration [30, 82], and many more. The accidental scenarios are listed in the Table 6. Due to the complexity and involvement of dangerous elements in the facilities, identifying or predicting accidents manually can be challenging [37]. If adversaries have access to either ML model or data in the nuclear facilities, then there may be dangerous consequences. For example, a pressure system in a nuclear power plant has the critical functions of maintaining the coolant level in the core and defining pressure in the primary heat transport systems. If the adversaries update the pressure in the system and the coolant level, then the coolant pipe may burst.

4.2 Barriers for Machine Learning Adoption

Nuclear facilities are among the most secure infrastructure in the world; however, their systems are not updated regularly, and most of them are still analogue [70]. With all the advancements in digital technologies, the industry is starting to implement new digital systems throughout its facilities. The progression towards modernisation has led to automating the processes in the domain,

Table 7. List of Machine Learning (ML) Techniques Adopted in Different Domains of the Nuclear Industry

Domain	ML model	Reference
Nuclear Radiation	Neural Network	[119]
	Gaussian Mixture Model	[7, 156]
	Least Square Support Vector Machine model	[53]
	Support Vector Machine	[76],
	Random Forest Regressor	[64]
	Gaussian Process Regressor	[13]
Nuclear Waste	Convolution Neural Network	[149]
	Support Vector Machine	[2, 38, 88, 152, 176]
	Artificial Neural Network	[14, 38, 68, 88]
	Random Forest regressor	[38, 88, 176]
	Linear Regression	[68]
	Principle Component Analysis	[68]
	Gaussian Process Regression mode	[60]
Nuclear Security	IBM Watson	[96]
	Monte Carlo–based learning	[41]
	G-Descent	[97]
	Convolution Neural Network	[97]
	Fuzzy Rules–based Gaussian processes	[6, 7]
Nuclear Fuel Management	Neural Network	[122]
	Support Vector Machine	[145]
	Long Short-Term Memory	[95]
Nuclear Safety and Regulation	Long Short-Term Memory	[95]
	k-Nearest Neighbour and MetaModeling	[170]
	AdaBoost and Random Forest	[131]
	k-Nearest Neighbour	[77, 86]
	Support Vector Machine	[77, 86]
	Evidential Reasoning	[174]
	Probabilistic Model	[168]
	Random Forest	[183]
	Artificial Neural Network	[141, 167, 183]
	Bidirectional Long Short Term Memory	[84, 85]
	Deep Rectifier Neural Network	[178]
	Reinforcement learning	[12]
	Multi-system deep learning network	[94]
	Active learning	[5]
	Probabilistic Analysis	[166]
	Interpretive Structural Modelling	[172]
	Multi-Layer Perceptron with Resilient Backpropagation	[63]

specifically using **Artificial Intelligence (AI)**—see Table 7. However, if a system gets attacked (e.g., cyber) in the **Nuclear Industry (NI)**, then the consequences can be dangerous. As the NI is a highly complex and security-critical infrastructure, any attack can be heavily damaging [124]. Therefore, the industry has strict protection policies and practices that lower the risks of physical and cyber-attacks in nuclear facilities [159]. However, the rise of **ML** techniques is still ongoing in different domains, and the building of regulations and guidelines for using AI around the world is still in preparation. The industry is one of the highly security-sensitive organisations and the regulations for implementing AI in NI are yet to be robust [123], leading to the slow adoption of ML techniques in the NI as compared to the other industries such as Marketing, Finance, and e-commerce.

As shown in Table 7, ML techniques have been adopted in the NI but they are mainly for research-based work [154]. To adopt the ML techniques in the live systems, ML systems must be robust, resilient, and secure against cyber attacks. For instance, cyber attacks on nuclear power plants and their control systems could expedite the theft of usable nuclear materials and malicious acts by adversaries [159]. Adversarial attacks such as physical attacks like intrusions, fault injection, and how malicious actors could navigate through isolated networks to disable physical protection systems and then take over control systems can be catastrophic. Therefore, it is important to study, understand, and test various adversarial attacks that can impact the industry and build defensive strategies against such attacks. It has been shown that even a small perturbation in strong ML techniques such as deep learning methods are susceptible to adversarial attack and are not robust [57, 127].

Research in **Adversarial Machine Learning (aML)** is still in its infancy and the NI is one of those areas that cannot afford systems that can be easily attacked. For example, if the temperature of a cooling rod is altered, it could lead to a reactor core meltdown. The implementation of autonomous decision-making techniques needs to be explored and tested thoroughly before relying completely on it. Hence, **transparency or understanding the behaviour of developed intelligent model** is critically necessary for the NI [26]. Advances in ML and autonomy could be beneficial to all the key areas of the nuclear systems architecture such as command and control, **Intelligence, surveillance and reconnaissance (ISR)**, nuclear weapon delivery, and non-nuclear counter-force operations (e.g., air defence, cyber security, and physical protection of nuclear assets). However, ML methods are yet to reach the stage where they could lead to maturity in the nuclear strategy. Presently, there are three main reasons for this. They are as follows:

— ML model is a **black-box model** where the knowledge of ML architecture is not known to all. Researchers and practitioners change models' parameters to fulfil their objectives [154]. These models are yet to obtain maturity, and it would be dangerous to rely on the safety and reliability from the immaturity of the technology from a perspective of command and control systems.
— Due to the **lack of transparency and explainability** of the ML models, there is uncertainty around the predictability and reliability of the output [26].
— ML model could be **compromised by adversarial attacks**, such as data poisoning to deceive, and spoofing the input data [101].

As discussed already, the ML model is yet to be at the stage where it can be comfortably applied to cyber-physical systems where safety is critical. For example, for sensitive matters like nuclear weapon control systems, there will always be the risk of a nuclear catastrophe if the weapons are mishandled. It would be dangerous and far-fetched to accept ML-enabled autonomy in nuclear facilities without testing factors such as security, resilience, and robustness. For instance, an accidental escalation resulting from incorrect information (e.g., regarding nuclear weapons) provided by an algorithm is a far more likely scenario that will have to be taken into account [20]. Even small malicious information relating to a nuclear weapon can have irreversible consequences. Furthermore, if attackers have access to model architecture or data where they can manipulate them, then this could be immensely risky, too. Next, we present the security concerns and ethical implications of adopting ML technologies in the NI.

4.2.1 Security Concerns. The NI is extremely security-sensitive, and any security concerns should be taken seriously. Adopting ML techniques could implicate various security concerns such as:

— Data security: The NI generates a large amount of sensitive data, which must be protected from unauthorised access. ML algorithms often require access to this data to train and operate, so it is important to ensure that the data is secure.
— Algorithmic bias: ML algorithms can be biased [57, 131], which means that they may not accurately reflect the real world. This could lead to problems in the NI, such as the misidentification of a security threat or the incorrect diagnosis of a problem with a reactor.
— Explainability: It is important to be able to explain how ML algorithms make decisions. This is especially important in the NI, where it is important to be able to understand why an algorithm has made a particular decision.

4.2.2 Ethical Implications. The use of machine learning techniques provides several benefits in the NI, including improved safety, increased efficiency, enhanced security, and improved decision-making. It can also be used to predict equipment failures, optimise fuel usage, detect anomalies, or even monitor plant security. It has the potential to revolutionise the way nuclear power plants are operated and maintained. However, the use of ML in the NI is still in its early stages and raises several ethical implications such as:

— Transparency and explainability: ML models are often complex and not transparent, making it difficult to understand how they make decisions. It is considered a black-box approach. This can make it difficult to assess the fairness and accuracy of the decisions.
— Bias: ML models can be biased, reflecting the biases in the data they are trained on. This can lead to discrimination against certain groups of people, also the production of inaccurate or misleading results.
— Privacy: ML models often require access to large amounts of data to learn. In terms of the NI, the use of the sensitive nature of the data and the potential risks associated with its exposure can be the major issues. This raises concerns about **privacy and data protection**.
— Security: ML models could be hacked, which can lead to major incidents such as the disruption of nuclear power plants or the release of radioactive material, and so on. This could have serious consequences for public safety.

To address these ethical implications, it is important to ensure that ML models are transparent, explainable, and fair following **Responsible AI** principles.

4.3 Machine Learning Attack Surface in the Nuclear Industry

In this section, we present an attack surface of an ML-based system built with data and ML model reflective of its purpose to the NI. Since there are different scenarios and tasks in the NI, we map the threat model from Section 3.1.1 to the NI scenario. As an example, we present an attack surface related to the NI through a use case: *crack detection in a **Nuclear Power Plant (NPP)** [79]*. We discuss **HOW** and **WHAT** kinds of attacks that can happen when ML is adopted to detect cracks. To detect cracks in nuclear power plants, traditionally a human operator goes through a video of plant inspection. They have to concentrate through the video frames. With an ML-based system, it can collect sensor inputs (e.g., video image, network events) from which intrinsic features (e.g., pixels, flows) are extracted and fed to the model to learn. The model learns a pattern and generates output (such as cracks and marks on the NPPs walls). The output is then interpreted and corrective action will be taken (such as shutting the plant or pausing the works in the plant). Here, adversaries can attempt to manipulate the collection and tamper with the data, corrupt the model, or even fabricate the outputs. To present this, we illustrate an attack surface during crack detection in a nuclear power station in Figure 5, which was studied in Reference [79].

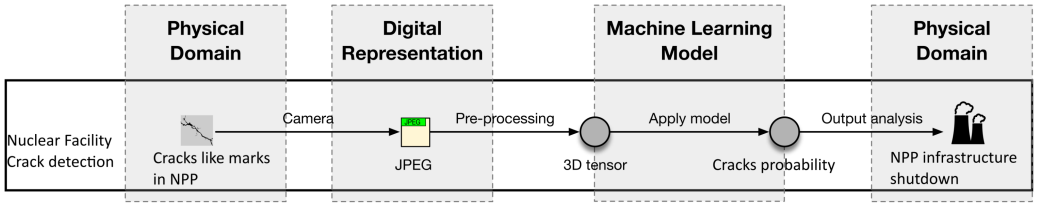

Fig. 5. aML pipeline in the context of cracks detection scenario in a NPP [79].

4.4 Adversaries in the Nuclear Industry

In this section, we discuss the consequences of adversarial attacks in the NI and the adopted ML models in the industry. Adversaries targeting the NI may have various motivations, including sabotage, espionage, or financial gains [9, 100]. Successful adversarial attacks in the industry can lead to severe consequences, including compromised safety systems, disruption of critical processes, and potential release of hazardous materials endangering human lives. Brundage et al. discuss in their report extensively the malicious use of AI and their consequences in critical industries like nuclear [113]. Such adversarial attacks can undermine the safety and integrity of nuclear systems, affecting the nuclear power plant operation and jeopardising the confidentiality of sensitive data [37, 113]. There can be several types of attacks including physical ones such as commando-like attacks on equipment that could lead to a reactor core meltdown or widespread dispersal of radioactivity; or cyber-attacks including power plant shutdown, wrong scheduling of temperature update on the coolant rod, and so on [70].

Only a small number of research works are available applying aML techniques in nuclear systems. Furthermore, they are limited to academic projects only and applied mostly in the simulators [20, 154]. The reasons can be the security concern regarding the use of autonomous systems on the live systems; we still do not have ML enabled technologies that provide complete safety against adversarial attacks in the NI. This gives a chance to study the security concerns of ML approaches before we rely on such technologies to handle highly complex tasks such as inspecting a nuclear power plant or protecting people from radiation dispersion autonomously. While conducting the literature review, we obtained less than 10 papers that have studied aML in the NI. References [99–101, 144] focused on injecting faulty data in the training dataset, which remains the most popular adversarial attack type in the NI.

Identifying vulnerabilities in ML models used in the NI is challenging, especially due to the complexity and opaqueness of the model design [70, 154]. The NI operates under various strict regulatory guidelines and frameworks that address cybersecurity concerns [73] such as the **International Atomic Energy Agency (IAEA)**, Nuclear Security Series, and national regulatory bodies requirements. However, the regulations on adopting AI in the security sensitive industries is still an ongoing topic [34, 123]. On the other hand, aML presents unique challenges that may require specific regulations and standards to address the risks and ensure the security of nuclear systems. Collaboration among industry stakeholders, government agencies, and researchers is essential for developing effective policies, guidelines, and standards that consider the nuances of aML in the NI [45]. Due to the lack of studies in aML in the NI, we explore some ML-based adversarial attack scenarios in Tables 8 to 10 that can potentially occur in the NI if ML techniques are to be implemented. For example, adversaries can target radiation monitoring ML models in a nuclear facility and inject manipulated data leading to inaccurate or delayed detection of radiation events; this can compromise safety and emergency response measures. We believe this list will encourage researchers on both the NI and cyber security in ML to further catalyse the aML study in the industry.

Table 8. Data Poisoning Attacks

Task	Scenario	Adversarial Outcome
Nuclear materials detection	Adversaries can aim to subvert the ML model used for nuclear materials detection, such as radioactive source identification or nuclear material tracking.	Adversaries can inject manipulated data into the training set, causing the ML model to misclassify, compromising the effectiveness of nuclear materials detection.
Radiation monitoring	Adversaries can target the ML model used for radiation monitoring in a nuclear facility.	By injecting manipulated data into the training set, the adversaries can manipulate the ML model's behaviour, leading to inaccurate detection of radiation events and compromising safety and emergency response measures.
Fault detection and diagnosis	Adversaries can try to disrupt the ML model used for fault detection and diagnosis.	Adversaries can poison the training data with carefully crafted samples that resemble specific anomalies, causing the ML model to provide incorrect fault diagnoses, potentially leading to undetected critical system failures.
Security monitoring	Adversaries can target the ML model used for security monitoring in a nuclear facility, such as intrusion detection or access control systems.	By injecting manipulated data into the training set, the adversaries can manipulate the ML model's behaviour, potentially bypassing security measures, gaining unauthorised access, or camouflaging their activities within the facility.
Radioactive waste classification	Adversaries can target the ML model used for classifying different types of radioactive waste for proper disposal and storage.	By injecting manipulated data into the training set, the adversaries can mislead the ML model into misclassifying waste materials, potentially resulting in incorrect handling, storage, or disposal of radioactive substances.
Nuclear material tracking	Adversaries can try to manipulate the ML model used for tracking the movement and inventory of nuclear materials within a nuclear facility or during transportation.	Adversaries can inject manipulated data into the training set, compromising the accuracy of the ML model's tracking capabilities and potentially enabling unauthorised diversion of nuclear materials.
Nuclear security event detection	Adversaries can target the ML model used for detecting security events, such as unauthorised access attempts or breaches, within a nuclear facility.	By injecting manipulated data into the training set, the adversaries can disrupt the ML model's ability to accurately detect security events, potentially allowing unauthorised individuals or malicious activities to go undetected.
Radiation hotspot identification	Adversaries can target the ML model used for identifying radiation hotspots in the vicinity of a nuclear facility, such as areas with increased radiation levels or potential contamination.	By injecting manipulated data into the training set, the adversaries can cause the ML model to misidentify radiation hotspots, leading to inaccurate response measures and potential safety risks.

4.5 Resilient Machine Learning (rML) in Nuclear Industry

Based on the research methodology explained in Section 2, we identified only six papers that applied defensive strategy on the attacks in ML-based applications in the NI. In Table 11, we present research works that investigated defence strategies against adversarial attacks on ML models in the industry. While References [100, 177] focused on **adversary training defensive mechanism** into the training set, Reference [130] restricted parameter values in selective layers of learning methods

Table 9. Model Inversion Attack

Task	Scenario	Adversarial Outcome
Critical system parameter inference	The adversaries can try to infer critical system parameters such as reactor core temperature or coolant flow rate.	By observing the responses of the ML model used in the control system, the adversaries can deduce sensitive information about the system's (e.g., NPP) infrastructure, potentially aiding in unauthorised access.
Process anomaly detection	Adversaries can aim to extract information about the internal operations of a nuclear facility by performing a model inversion attack.	By making specific queries to the ML model used for anomaly detection, the adversaries can infer details about the facility's processes, potentially revealing vulnerabilities, operational patterns, or critical information.
Security system bypass	Adversaries can attempt to bypass the ML model-based security system in a nuclear facility, such as a biometric access control system.	By exploiting model inversion attacks, the adversaries can extract information about the ML model's decision boundaries, potentially enabling them to deceive the system and gain unauthorised access to secure areas within the facility.
Environmental monitoring inference	Adversaries can try to infer sensitive environmental information about a nuclear facility's surroundings, such as air quality, radiation levels, or potential sources of contamination.	By observing the outputs of the ML model used for environmental monitoring, the adversaries can deduce details about the facility's surroundings, potentially aiding in planning unauthorised activities or compromising the facility's security.
Safety system analysis	Adversaries can aim to analyse the behaviour and vulnerabilities of safety systems in a nuclear facility by performing a model inversion attack.	By querying the ML model used for safety system analysis, the adversaries can gain insights into the system's decision-making process, potentially identifying weaknesses or finding ways to bypass safety measures.

(e.g., Deep Neural Network) to strictly **limit the range of parameter values** that attackers can exploit. Reference [177] proposed an rML ensemble method that utilises **Moving Target Defense (MTD)** mechanism. MTD makes it difficult for hackers to attack a system by constantly changing the position of targets. This creates challenges for hackers to find their targets, and even if they find them, they will only find decoys that will capture the information for further analysis. As a result, MTD successfully prevents damage, rather than simply mitigating it. When an input (either clean or adversarial input) enters the system, the rML controller pulls the required ML models from the rML repository and creates the environment for the resilient decision mechanism. Each of the ML models evaluates the input from the user and provides a prediction. Next, a voting mechanism using the Boyer-Moore majority vote algorithm determines if there is any different output from the ML models and the majority of the decisions is accepted as the true output. Reference [99] built a defensive strategy against false data injection (adversarial training) attacks following the concept of active monitoring of system behaviour. Active monitoring involves deliberately perturbing the data traffic in a digital control system (such as nuclear systems) based on an understanding of the systems' behaviour derived from physics. These perturbations are carefully crafted to be subtle, causing no noticeable impact on the system's behaviour. The primary advantage of this approach is its ability to detect threats at an early stage, particularly during the initial period when attackers typically test the system by introducing small disturbances such as commands to actuators, similar

Table 10. Model Extraction Attacks

Task	Scenario	Adversarial Outcomes
Radiation detection model extraction	Adversaries can attempt to extract the ML model used for radiation detection in a nuclear facility.	By interacting with the ML model and querying it, the adversaries can aim to clone the model for unauthorised analysis. This can potentially enable them to exploit weaknesses, develop countermeasures, or gain insights into the facility's radiation detection capabilities.
Nuclear material tracking model extraction:	Adversaries can target the ML model used for tracking the movement and inventory of nuclear materials within a nuclear facility.	By interacting with the ML model and querying it, the adversaries can try to extract the model's parameters. This allows them to replicate the model's behaviour, potentially aiding in unauthorised movement or diversion of nuclear materials.
Predictive maintenance model extraction	Adversaries can aim to extract the ML model used for predictive maintenance of critical equipment in a nuclear power plant.	By querying the ML model, the adversaries can try to clone the model. This can enable them to analyse the model's predictions, identify vulnerabilities in the maintenance process, or develop counterfeit models for malicious purposes.
Reactor core temperature prediction model extraction:	Adversaries can target the ML model used for predicting the temperature of the reactor core in a nuclear power plant.	By interacting with the ML model and probing it with specific inputs, the adversaries can aim to extract the model's parameters. This can provide them with insights into the reactor's behaviour, potential vulnerabilities, or critical operational information.
Control system model extraction	Adversaries can try to extract the ML model used in the control system of a nuclear facility, responsible for regulating various parameters and maintaining safe operation.	By querying with the ML model, the adversaries can attempt to extract the model's details, enabling them to replicate its behaviour. This can lead to unauthorised control actions, tampering with critical systems, or understanding the facility's control mechanisms.

to injecting **adversarial examples** in the training data to learn from the examples. It is important to detect attacks early on due to the requirement of fast and effective response time to critical incidents in the NI. Reference [101] believed off-the-shelf methods are not suitable to defend against adversarial attacks on ML models especially in critical systems like Nuclear, as design, operation, and safety are based on well-established practices. Therefore, Li et al. combined multiple techniques including **Fast Fourier Transform (FFT)**, Least Squares, **Alternating Conditional Estimation (ACE)** and Regularisation, and a physics-based model to protect the prediction outcomes [101].

5 OPEN RESEARCH DIRECTION

The adoption of AI in the NI brings both opportunities and challenges. While leveraging ML techniques accelerates the potential of various aspects of nuclear operations such as radiation monitoring, several open issues need attention for the successful integration of ML in the industry. This section discusses the key open issues.

— **Responsible AI:** As the implementation of ML algorithms continues to expand across diverse domains, there is a pressing need for in-depth studies on building responsible

Table 11. Research Works that Studied Adversarial Attacks and Defence Strategies in Terms of Nuclear Applications

Adversarial attack	Defence strategy	Case study	Reference
Fault data injection (BinFi)	Restricting parameter values in selective **Deep neural networks (DNN)** layers (Ranger)	Autonomous Vehicle in safety-critical domain	[130]
Fault data injection	-	Power and Gas grid	[144]
Fault data injection	Adversarial training	Nuclear reactor	[100]
Manipulate ML at testing phase	Ensemble Methods	Safety critical domain	[177]
Fault data injection	Ensemble methods	Nuclear reactor	[101]
Fault data injection	Active Monitoring	Nuclear reactor	[99]
Fault data injection	Distillation method	Nuclear Power plant	[63]

(including safe, secure, robust, and resilient principles) ML systems. The demand for responsible ML-enabled systems is crucial in high-stake industries such as Nuclear Industry, and Healthcare, as they involve situations where human lives can be at risk if the technologies are susceptible to easy attacks. Therefore, the interpretability and explainability of ML algorithms can be crucial in safety-critical domains like nuclear operations, where understanding the decision-making process is imperative for building trust. For instance, nuclear reactors experience many changes during their service time, causing updates in monitoring and operational guidelines [169]. Additionally, this sector is dynamic in nature with **continuously updating operational objectives** based on the dynamic needs. Due to such nature, the regulatory guidelines are also updated. Such continuous changes will make implementation of the AI and ML systems challenging [154], as models and data need to be updated accordingly.

— **Evaluation:** Even if the development leads to cutting edge AI techniques for the NI, there will still be the need for thorough evaluation in real-world scenarios, which can be challenging. This is due to the secure nature of nuclear facilities. Therefore, researchers and practitioners, both in academia and industry, need to **closely work together** by sharing resources such as real data, and scenarios. Systems need to be broken several times to achieve a robust and precise model; this can be challenging in the NI.

As NI comprised a variety of applications as discussed in the earlier section (Section 4.1), it is possible for some components such as sensors to become faulty and give faulty readings leading to faulty output. However, it is equally possible that learning models may be attacked due to adversaries such as data poisoning. Therefore, it is necessary to learn to **differentiate what is adversarial attack and what is actual systems (e.g., hardware) failure**. The problem of distinguishing malicious attacks from systems' failures in any **Cyber Physical Systems (CPS)** can be challenging and needs to be solved as the cyber-physical industry continues to seek to build resilient ML systems [124].

6 CONCLUSION

The NI is one of the most challenging environments. With radiation levels and the involvement of hazardous elements and the environment, there are often restrictions placed on human access to the facilities. So, on paper, it is highly suitable for automation (using ML and robots). However,

even with the safety measures established, there can be high risk. Hence, understandably, the adoption of technologies is slow in comparison to other industries. In this article, we investigated existing ML applications and studies that explored the study of adversarial attacks in application to the NI.

From the review, we noted that **faulty data injection** and **adversarial training** are the most-studied adversarial attack and defensive mechanisms in NI, respectively. This demonstrates the lack of investigation for other attacks and defensive strategies. It is possible that a cyber-attack (e.g., spoofing, hacking, manipulation, and digital jamming) could infiltrate an NI, including nuclear weapons systems, threaten the integrity of its communications, and ultimately gain control of its possible command and control systems. For instance, a hacker might interfere with nuclear command-and-control systems, spoof and compromise warning systems, or in a worst-case scenario, trigger an accidental nuclear launch. Since the nuclear sector is vastly diverse from nuclear decommissioning to nuclear safety regulation, there can be numerous different scenarios. Therefore, it is highly essential to investigate diverse use cases. If AI is embedded in its applications, then they need to be aware of different angles of security in both physical and cyber systems.

REFERENCES

[1] M. B. Abbott, H. J. De Nordwall, and B. Swets. 1983. On applications of artificial intelligence to the control and safety problems of nuclear power plants. *Civil Eng. Syst.* 1, 2 (1983), 69–82. DOI: https://doi.org/10.1080/02630258308970321

[2] J. M. Aitken, S. M. Veres, A. Shaukat, Y. Gao, E. Cucco, L. A. Dennis, M. Fisher, J. A. Kuo, T. Robinson, and P. E. Mort. 2018. Autonomous nuclear waste management. *IEEE Intell. Syst.* 33, 6 (2018), 47–55. DOI: https://doi.org/10.1109/MIS.2018.111144814

[3] Naveed Akhtar, Jian Liu, and Ajmal Mian. 2018. Defense against universal adversarial perturbations. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 3389–3398. DOI: https://doi.org/10.1109/CVPR.2018.00357

[4] Naveed Akhtar and Ajmal Mian. 2018. Threat of adversarial attacks on deep learning in computer vision: A survey. *IEEE Access* 6 (2018), 14410–14430. DOI: https://doi.org/10.1109/ACCESS.2018.2807385

[5] A. M. Al-Bahi and A. Y. Soliman. 2015. Design of a project-based active cooperative course to develop and assess safety culture in undergraduate nuclear engineering programs. In *International Conference on Interactive Collaborative Learning (ICL'15)*. 832–837. DOI: https://doi.org/10.1109/ICL.2015.7318136

[6] M. Alamaniotis. 2021. Fuzzy integration of kernel-based Gaussian processes applied to anomaly detection in nuclear security. In *12th International Conference on Information, Intelligence, Systems and Applications (IISA'21)*. DOI: https://doi.org/10.1109/IISA52424.2021.9555524

[7] M. Alamaniotis and A. Heifelz. 2020. A machine learning approach for background radiation modeling and anomaly detection in radiation time series pertained to nuclear security. *Trans. Amer. Nuclear Societ.* 123, 1 (2020), 447–450. DOI: https://doi.org/10.13182/T123-33516

[8] R. M. Ayo-Imoru and A. C. Cilliers. 2018. Continuous machine learning for abnormality identification to aid condition-based maintenance in nuclear power plant. *Ann. Nuclear Ener.* 118 (2018), 61–70. DOI: https://doi.org/10.1016/j.anucene.2018.04.002

[9] Abiodun Ayodeji, Mokhtar Mohamed, Li Li, Antonio Di Buono, Iestyn Pierce, and Hafiz Ahmed. 2023. Cyber security in the nuclear industry: A closer look at digital control systems, networks and human factors. *Prog. Nuclear Ener.* 161 (2023), 104738. DOI: https://doi.org/10.1016/j.pnucene.2023.104738

[10] Davood Babazadeh, Mehrdad Boroushaki, and Caro Lucas. 2009. Optimization of fuel core loading pattern design in a VVER nuclear power reactors using Particle Swarm Optimization (PSO). *Ann. Nuclear Ener.* 36, 7 (2009), 923–930. DOI: https://doi.org/10.1016/j.anucene.2009.03.007

[11] I. Bae, M. Na, Y. Lee, and G. Park. 2009. Estimation of the power peaking factor in a nuclear reactor using support vector machines and uncertainty analysis. *Nuclear Eng. Technol.* 41 (2009), 1181–1190.

[12] C. H. Baek, K. B. Jang, and T. H. Woo. 2021. Analysis of systems thinking safety by artificial intelligence (AI) algorithm in South Korean nuclear power plants (NPPs). *Int. J. Emerg. Technol. Advanc. Eng.* 11, 10 (2021), 56–62. DOI: https://doi.org/10.46338/IJETAE1021_07

[13] S. A. Balanya, D. Ramos, P. Ramirez-Hereza, D. T. Toledano, J. Gonzalez-Rodriguez, A. Ariza-Velazquez, J. Vidal Orlovac, and N. Doncel Gutiérrez. 2022. Gaussian Processes for radiation dose prediction in nuclear power plant reactors. *Chemomet. Intell. Labor. Syst.* 230 (2022). DOI: https://doi.org/10.1016/j.chemolab.2022.104652

[14] M. Ben-Haim and D. D. Macdonald. 1994. Modeling geological brines in salt-dome high level nuclear waste isolation repositories by artificial neural networks. *Corros. Sci.* 36, 2 (1994), 385–393. DOI: https://doi.org/10.1016/0010-938X(94)90164-3

[15] Battista Biggio, Igino Corona, Davide Maiorca, Blaine Nelson, Nedim Šrndić, Pavel Laskov, Giorgio Giacinto, and Fabio Roli. 2013. Evasion attacks against machine learning at test time. In *Machine Learning and Knowledge Discovery in Databases*, Hendrik Blockeel, Kristian Kersting, Siegfried Nijssen, and Filip Železný (Eds.). Springer Berlin, 387–402.

[16] Battista Biggio, Blaine Nelson, and Pavel Laskov. 2011. Support vector machines under adversarial label noise. In *Asian Conference on Machine Learning (Proceedings of Machine Learning Research,* Vol. 20*)*, Chun-Nan Hsu and Wee Sun Lee (Eds.). PMLR, 97–112. Retrieved from http://proceedings.mlr.press/v20/biggio11.html

[17] Battista Biggio, Blaine Nelson, and Pavel Laskov. 2012. Poisoning attacks against support vector machines. In *29th International Coference on International Conference on Machine Learning (ICML'12)*. Omnipress, 1467–1474.

[18] Battista Biggio and Fabio Roli. 2018. Wild patterns: Ten years after the rise of adversarial machine learning. In *ACM SIGSAC Conference on Computer and Communications Security (CCS'18)*. Association for Computing Machinery, New York, NY, 2154–2156. DOI: https://doi.org/10.1145/3243734.3264418

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

[20] Vincent Boulanin, Anja Kaspersen, Chris King, S. M. Amadae, Jean-Marc Rickli, Shahar Avin, Frank Sauer, John Borrie, Dimitri Scheftelowitsch, Justin Bronk, Page O. Stoutland, Martin Hagström, Topychkanov Petr, and Michael C. Horowitz. 2020. *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk: Volume III South Asian Perspectives*. Technical Report. Stockholm International Peace Research Institute. Retrieved from http://www.jstor.org/stable/resrep24515

[21] Emma Brandon-Jones, Brian Squire, Chad W. Autry, and Kenneth J. Petersen. 2014. A contingent resource-based perspective of supply chain resilience and robustness. *J. Supply Chain Manag.* 50, 3 (2014), 55–73. DOI: https://doi.org/10.1111/jscm.12050

[22] Wieland Brendel, Jonas Rauber, and Matthias Bethge. 2018. Decision-based adversarial attacks: Reliable attacks against black-box machine learning models. In *6th International Conference on Learning Representations (ICLR'18)*. OpenReview.net. Retrieved from https://openreview.net/forum?id=SyZI0GWCZ

[23] Cristian Cadar, Periklis Akritidis, Manuel Costa, Jean-Phillipe Martin, and Miguel Castro. 2008. *Data Randomization*. Technical Report. Technical Report TR-2008-120, Microsoft Research.

[24] F. Cadini, E. Zio, V. Kopustinskas, and R. Urbonas. 2008. A model based on bootstrapped neural networks for computing the maximum fuel cladding temperature in an Rmbk-1500 nuclear reactor accident. *Nuclear Eng. Des.* 238, 9 (2008), 2165–2172. DOI: https://doi.org/10.1016/j.nucengdes.2008.01.018

[25] Nicholas Carlini and David A. Wagner. 2017. MagNet and "efficient defenses against adversarial attacks" are not robust to adversarial examples. *ArXiv* abs/1711.08478 (2017).

[26] Davide Castelvecchi. 2016. Can we open the black box of AI? *Nature* 538, 7623 (Oct. 2016), 20–23. DOI: https://doi.org/10.1038/538020a

[27] Anirban Chakraborty, Manaar Alam, Vishal Dey, A. Chattopadhyay, and Debdeep Mukhopadhyay. 2018. Adversarial attacks and defences: A survey. *ArXiv* abs/1810.00069 (2018).

[28] Fu-Chen Chen and Mohammad R. Jahanshahi. 2018. NB-CNN: Deep learning-based crack detection using convolutional neural network and Naïve Bayes data fusion. *IEEE Trans. Industr. Electron.* 65, 5 (2018), 4392–4400. DOI: https://doi.org/10.1109/TIE.2017.2764844

[29] Fu-Chen Chen, Mohammad R. Jahanshahi, Rih-Teng Wu, and Chris Joffe. 2017. A texture-based video processing methodology using Bayesian data fusion for autonomous crack detection on metallic surfaces. *Comput.-Aid. Civil Infrast. Eng.* 32, 4 (2017), 271–287. DOI: https://doi.org/10.1111/mice.12256

[30] Geon Pil Choi, Dong Yeong Kim, Kwae Hwan Yoo, and Man Gyun Na. 2016. Prediction of hydrogen concentration in nuclear power plant containment under severe accidents using cascaded fuzzy neural networks. *Nuclear Eng. Des.* 300 (2016), 393–402. DOI: https://doi.org/10.1016/j.nucengdes.2016.02.015

[31] Geon Pil Choi, Kwae Hwan Yoo, Ju Hyun Back, and Man Gyun Na. 2017. Estimation of LOCA break size using cascaded fuzzy neural networks. *Nuclear Eng. Technol.* 49, 3 (2017), 495–503. DOI: https://doi.org/10.1016/j.net.2016.11.001

[32] The Fukushima Nuclear Accident Independent Investigation Commission. 2012. The Fukushima Nuclear Accident Independent Investigation Commission. https://reliefweb.int/report/japan/official-report-fukushima-nuclear-accident-independent-investigation-commission

[33] U. S. Nuclear Regulatory Commission. 2018. Davis-Besse Improvement Activities. Retrieved from https://www.nrc.gov/docs/ML0925/ML092540336.pdf

[34] Geoffrey Currie and K. Elizabeth Hawk. 2021. Ethical and legal challenges of artificial intelligence in nuclear medicine. *Semin. Nuclear Med.* 51, 2 (2021), 120–125. DOI: https://doi.org/10.1053/j.semnuclmed.2020.08.001

[35] Nilesh Dalvi, Pedro Domingos, Mausam, Sumit Sanghai, and Deepak Verma. 2004. Adversarial classification. In *10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'04)*. Association for Computing Machinery, New York, NY, 99–108. DOI: https://doi.org/10.1145/1014052.1014066

[36] Alan M. M. de Lima, Roberto Schirru, Fernando Carvalho da Silva, and José Antonio Carlos Canedo Medeiros. 2008. A nuclear reactor core fuel reload optimization using artificial ant colony connective networks. *Ann. Nuclear Ener.* 35, 9 (2008), 1606–1612. DOI: https://doi.org/10.1016/j.anucene.2008.03.002

[37] Mauro Vitor de Oliveira and José Carlos Soares de Almeida. 2013. Application of artificial intelligence techniques in modeling and control of a nuclear power plant pressurizer system. *Prog. Nuclear Ener.* 63 (2013), 71–85. DOI: https://doi.org/10.1016/j.pnucene.2012.11.005

[38] B. J. Debusschere, D. T. Seidl, T. M. Berg, K. W. Chang, R. C. Leone, L. P. Swiler, and P. E. Mariner. 2023. Machine learning surrogates of a fuel matrix degradation process model for performance assessment of a nuclear waste repository. *Nuclear Technol.* 209, 9 (2023). DOI: https://doi.org/10.1080/00295450.2023.2197666

[39] Cecilia Martín del Campo, Miguel Ángel Palomera-Pérez, and Juan-Luis François. 2009. Advanced and flexible genetic algorithms for BWR fuel loading pattern optimization. *Ann. Nuclear Ener.* 36, 10 (2009), 1553–1559. DOI: https://doi.org/10.1016/j.anucene.2009.07.013

[40] Michael G. Devereux, Paul Murray, and Graeme M. West. 2020. A new approach for crack detection and sizing in nuclear reactor cores. *Nuclear Eng. Des.* 359 (Apr. 2020). DOI: https://doi.org/10.1016/j.nucengdes.2019.110464

[41] Dean Dominguez, Mancel Jordan Parks, Adam D. Williams, and Susan Washburn. 2012. Special nuclear material and critical infrastructure security modeling and simulation of physical protection systems. In *IEEE International Carnahan Conference on Security Technology (ICCST'12)*. IEEE, 10–14. DOI: https://doi.org/10.1109/CCST.2012.6393531

[42] Yinpeng Dong, Fangzhou Liao, Tianyu Pang, Hang Su, Jun Zhu, Xiaolin Hu, and Jianguo Li. 2018. Boosting Adversarial Attacks with Momentum. arXiv:1710.06081 [cs.LG]

[43] Vasisht Duddu. 2018. A survey of adversarial machine learning in cyber warfare. *Defence Sci. J.* 68, 4 (June 2018), 356–366. DOI: https://doi.org/10.14429/dsj.68.12371

[44] Anatoly Dyatlov. 1991. How it was: an operator's perspective. *Nuclear Engineering International* 36, 448 (1991), 43–44, 46, 48–50.

[45] Shannon Leigh Eggers and Char Sample. 2020. Vulnerabilities in artificial intelligence and machine learning applications and data. Idaho National Laboratory, (12 2020). https://inldigitallibrary.inl.gov/sites/sti/sti/Sort_57369.pdf

[46] M. J. Embrechts and S. Benedek. 1998. Hybrid identification of unlabeled nuclear power plant transients with artificial neural networks. In *IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, Vol. 2. IEEE, 1438–1443. DOI: https://doi.org/10.1109/IJCNN.1998.685987

[47] Adem Erdoğan and Melih Geçkinli. 2003. A PWR reload optimisation code (XCore) using artificial neural networks and genetic algorithms. *Ann. Nuclear Ener.* 30, 1 (2003), 35–53. DOI: https://doi.org/10.1016/S0306-4549(02)00041-5

[48] Alhussein Fawzi, Omar Fawzi, and Pascal Frossard. 2016. Analysis of Classifiers' Robustness to Adversarial Perturbations. arXiv:1502.02590 [cs.LG]

[49] Reuben Feinman, Ryan R. Curtin, Saurabh Shintre, and Andrew B. Gardner. 2017. Detecting Adversarial Samples from Artifacts. arXiv:1703.00410 [stat.ML]

[50] M. Fukumoto. 2019. *Low-dose Radiation Effects on Animals and Ecosystems: Long-term Study on the Fukushima Nuclear Accident*. Springer Singapore, 1–264. DOI: https://doi.org/10.1007/978-981-13-8218-5

[51] Andrea Galindo. 2022. What Is Nuclear Energy? The Science of Nuclear Power. Retrieved from https://www.iaea.org/newscenter/news/what-is-nuclear-energy-the-science-of-nuclear-power

[52] Ji Gao, Beilun Wang, Zeming Lin, Weilin Xu, and Yanjun Qi. 2017. DeepCloak: Masking Deep Neural Network Models for Robustness against Adversarial Samples. arXiv:1702.06763 [cs.LG]

[53] Song Gao, Yaogeng Tang, and Xing Qu. 2012. LSSVM based missing data imputation in nuclear power plant's environmental radiation monitor sensor network. In *IEEE 5th International Conference on Advanced Computational Intelligence (ICACI'12)*. IEEE, 479–484. DOI: https://doi.org/10.1109/ICACI.2012.6463210

[54] Abu Bakar Ghazali and Maslina Mohd Ibrahim. 2016. Fault detection and analysis in nuclear research facility using artificial intelligence methods. *AIP Conf. Proc.* 1704, 1 (2016), 030010. DOI: https://doi.org/10.1063/1.4940079

[55] Carla Regina Gomes and Jose Antonio Carlos Canedo Medeiros. 2015. Neural network of Gaussian radial basis functions applied to the problem of identification of nuclear accidents in a PWR nuclear power plant. *Ann. Nuclear Ener.* 77 (2015), 285–293. DOI: https://doi.org/10.1016/j.anucene.2014.10.001

[56] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Networks. arXiv:1406.2661 [stat.ML]

[57] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. 2015. Explaining and Harnessing Adversarial Examples. arXiv:1412.6572 [stat.ML]

[58] S. Gu and Luca Rigazio. 2015. Towards Deep Neural Network Architectures Robust to Adversarial Examples. *CoRR* abs/1412.5068 (2015).

[59] J. W. Guan and D. A. Bell. 2000. Rough knowledge discovery for nuclear safety. *Int. J. Gen. Syst.* 29, 2 (2000), 231–249. DOI: https://doi.org/10.1080/03081070008960931

[60] LaGrande Lowell Gunnell, Kyle Manwaring, Xiaonan Lu, Jacob Reynolds, John Vienna, and John Hedengren. 2022. Machine learning with gradient-based optimization of nuclear waste vitrification with uncertainties and constraints. *Processes* 10, 11 (2022). DOI: https://doi.org/10.3390/pr10112365

[61] Chuan Guo, Mayank Rana, Moustapha Cisse, and Laurens van der Maaten. 2018. Countering Adversarial Images Using Input Transformations. arXiv:1711.00117 [cs.CV]

[62] Ezgi Gursel, Bhavya Reddy, Anahita Khojandi, Mahboubeh Madadi, Jamie Baalis Coble, Vivek Agarwal, Vaibhav Yadav, and Ronald L. Boring. 2023. Using artificial intelligence to detect human errors in nuclear power plants: A case in operation and maintenance. *Nuclear Eng. Technol.* 55, 2 (2023), 603–622. DOI: https://doi.org/10.1016/j.net.2022.10.032

[63] Kamal Hadad, Mojtaba Mortazavi, Mojtaba Mastali, and Ali Akbar Safavi. 2008. Enhanced neural network based fault detection of a VVER nuclear power plant with the aid of principal component analysis. *IEEE Trans. Nuclear Sci.* 55, 6 (2008), 3611–3619. DOI: https://doi.org/10.1109/TNS.2008.2006491

[64] Sherief Hashima and Imbaby Mahmoud. 2021. Efficient wireless sensor network for radiation detection in nuclear sites. *Int. J. Electron. Telecommun.* 67, 2 (2021), 175–180. DOI: https://doi.org/10.24425/ijet.2021.135961

[65] Ehsan Hatami, Nasser Vosoughi, and Hassan Salarieh. 2016. Design of a fault tolerated intelligent control system for load following operation in a nuclear power plant. *Int. J. Electric. Power Ener. Syst.* 78 (2016), 864–872. DOI: https://doi.org/10.1016/j.ijepes.2015.11.073

[66] Afshin Hedayat, Hadi Davilu, Ahmad Abdollahzadeh Barfrosh, and Kamran Sepanloo. 2009. Estimation of research reactor core parameters using cascade feed forward artificial neural networks. *Prog. Nuclear Ener.* 51, 6 (2009), 709–718. DOI: https://doi.org/10.1016/j.pnucene.2009.03.004

[67] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. 2015. Distilling the Knowledge in a Neural Network. arXiv:1503.02531 [stat.ML]

[68] Guang Hu and Wilfried Pfingsten. 2023. Data-driven machine learning for disposal of high-level nuclear waste: A review. *Ann. Nuclear Ener.* 180 (2023), 109452. DOI: https://doi.org/10.1016/j.anucene.2022.109452

[69] Qingyu Huang, Shinian Peng, Jian Deng, Hui Zeng, Zhuo Zhang, Yu Liu, and Peng Yuan. 2023. A review of the application of artificial intelligence to nuclear reactors: Where we are and what's next. *Heliyon* 9, 3 (2023), e13883. DOI: https://doi.org/10.1016/j.heliyon.2023.e13883

[70] IAEA. 1997. *Advanced Control Systems to Improve Nuclear Power Plant Reliability and Efficiency* (TECDOC Series, no. 952). International Atomic Energy Agency, Vienna. Retrieved from https://www.iaea.org/publications/5604/advanced-control-systems-to-improve-nuclear-power-plant-reliability-and-efficiency

[71] IBM. 2021. What is Machine Learning? Retrieved from https://www.ibm.com/uk-en/cloud/learn/machine-learning

[72] International Atomic Energy Agency. 2023. Nuclear Innovations for Net Zero. International Atomic Energy Agency. Retrieved from https://www.iaea.org/sites/default/files/nuclearinnovations_0.pdf

[73] International Atomic Energy Agency (IAEA). 2012. Safety of Nuclear Power Plants: Design. Retrieved from https://www-pub.iaea.org/MTCD/Publications/PDF/Pub1534_web.pdf

[74] Beth Jackson. 2020. Nuclear Decommissioning Authority Turns to Drones for Help with Toxic Sellafield Legacy. Retrieved from https://techmonitor.ai/technology/data/nuclear-decommissioning-authority-technology

[75] E. Jharko, V. Promyslov, and A. Iskhakov. 2019. Extending functionality of early fault diagnostic system for on-line security assessment of nuclear power plant. In *International Russian Automation Conference (RusAutoCon'19).* DOI: https://doi.org/10.1109/RUSAUTOCON.2019.8867790

[76] Hua Jin, Qingyu Yue, and Wenhai Wang. 1997. Continuous monitoring system of environmental radiation near nuclear facility. *Yuanzineng Kexue Jishu/Atomic Ener. Sci. Technol.* 31, 3 (1997), 204–210. https://www.scopus.com/inward/record.uri?eid=2-s2.0-0031130479&partnerID=40&md5=ce05b4f0ceec2aee5e6a0c6b675bbcd8

[77] Yong Hee Lee Jung Hwan Kim, Chul Min Kim and Man-Sung Yim. 2021. Electroencephalography-based intention monitoring to support nuclear operators' communications for safety-relevant tasks. *Nuclear Technol.* 207, 11 (2021), 1753–1767. DOI: https://doi.org/10.1080/00295450.2020.1837583

[78] Iurii Katser, Dmitriy Raspopov, Vyacheslav Kozitsin, and Maxim Mezhov. 2022. Machine Learning Methods for Anomaly Detection in Nuclear Power Plant Power Transformers. arXiv:2211.11013 [cs.LG]

[79] Anita Khadka, Gregory Epiphaniou, and Carsten Maple. 2023. Resilient machine learning in the nuclear industry: Crack detection as a case study. *Int. J. Nuclear Quant. Eng.* 17, 1 (Jan. 2023), 11–17. Retrieved from https://publications.waset.org/10012926/resilient-machine-learning-in-the-nuclear-industry-crack-detection-as-a-case-study

[80] F. Khoshahval, H. Minuchehr, and A. Zolfaghari. 2011. Performance evaluation of PSO and GA in PWR core loading pattern optimization. *Nuclear Eng. Des.* 241, 3 (2011), 799–808. DOI: https://doi.org/10.1016/j.nucengdes.2010.12.023

[81] F. Khoshahval, A. Zolfaghari, H. Minuchehr, M. Sadighi, and A. Norouzi. 2010. PWR fuel management optimization using continuous particle swarm intelligence. *Ann. Nuclear Ener.* 37, 10 (2010), 1263–1271. DOI: https://doi.org/10.1016/j.anucene.2010.05.023

[82] Dong Yeong Kim, Ju Hyun Kim, Kwae Hwan Yoo, and Man Gyun Na. 2015. Prediction of hydrogen concentration in containment during severe accidents using fuzzy neural network. *Nuclear Eng. Technol.* 47, 2 (2015), 139–147. DOI: https://doi.org/10.1016/j.net.2014.12.004

[83] Dong Yeong Kim, Kwae Hwan Yoo, Geon Pil Choi, Ju Hyun Back, and Man Gyun Na. 2016. Reactor vessel water level estimation during severe accidents using cascaded fuzzy neural networks. *Nuclear Eng. Technol.* 48, 3 (2016), 702–710. DOI: https://doi.org/10.1016/j.net.2016.02.002

[84] Hyojin Kim and Jonghyun Kim. 2021. Multi-step prediction algorithm for critical safety parameters at nuclear power plants using BiLSTM and AM. *31st European Safety and Reliability Conference (ESREL'21).* 818–824. DOI: https://doi.org/10.3850/978-981-18-2016-8_479-cd

[85] Hyojin Kim and Jonghyun Kim. 2023. Long-term prediction of safety parameters with uncertainty estimation in emergency situations at nuclear power plants. *Nuclear Eng. Technol.* 55, 5 (2023), 1630–1643. DOI: https://doi.org/10.1016/j.net.2023.01.026

[86] Jung Hwan Kim, Chul Min Kim, Eun-Soo Jung, and Man-Sung Yim. 2020. Biosignal-based attention monitoring to support nuclear operator safety-relevant tasks. *Front. Computat. Neurosci.* 14 (2020). DOI: https://doi.org/10.3389/fncom.2020.596531

[87] Seungmin Kim, Sangwoo Kim, Ki-haeng Nam, Seonuk Kim, and Kook-huei Kwon. 2019. Cyber security strategy for nuclear power plant through vital digital assets. In *International Conference on Computational Science and Computational Intelligence (CSCI'19).* IEEE, 224–226. DOI: https://doi.org/10.1109/CSCI49370.2019.00045

[88] Olaf Kolditz, Diederik Jacques, Francis Claret, Johan Bertrand, Sergey V. Churakov, Christophe Debayle, Daniela Diaconu, Kateryna Fuzik, David Garcia, Nico Graebling, Bernd Grambow, Erika Holt, Andrés Idiart, Petter Leira, Vanessa Montoya, Ernst Niederleithinger, Markus Olin, Wilfried Pfingsten, Nikolaos I. Prasianakis, Karsten Rink, Javier Samper, István Szöke, Réka Szöke, Louise Theodon, and Jacques Wendling. 2023. Digitalisation for nuclear waste management: Predisposal and disposal. *Environ. Earth Sci.* 82, 1 (Jan. 2023), 42. DOI: https://doi.org/10.1007/s12665-022-10675-4

[89] Young Do Koo, Ye Ji An, Chang-Hwoi Kim, and Man Gyun Na. 2019. Nuclear reactor vessel water level prediction during severe accidents using deep neural networks. *Nuclear Eng. Technol.* 51, 3 (2019), 723–730. DOI: https://doi.org/10.1016/j.net.2018.12.019

[90] Akansha Kumar and Pavel V. Tsvetkov. 2015. A new approach to nuclear reactor design optimization using genetic algorithms and regression analysis. *Ann. Nuclear Ener.* 85, C (2015), 27–35. DOI: https://doi.org/10.1016/j.anucene.2015.04.028

[91] Yong kuo Liu, Chun li Xie, Min jun Peng, and Shuang han Ling. 2014. Improvement of fault diagnosis efficiency in nuclear power plants using hybrid intelligence approach. *Prog. Nuclear Ener.* 76 (2014), 122–136. DOI: https://doi.org/10.1016/j.pnucene.2014.05.001

[92] Alexey Kurakin, Ian Goodfellow, and Samy Bengio. 2017. Adversarial Machine Learning at Scale. arXiv:1611.01236 [cs.CV]

[93] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521, 7553 (2015), 436.

[94] Daeil. Lee and Jonghyun Kim. 2018. Autonomous algorithm for safety systems of the nuclear power plant by using the deep learning. *Advanc. Intell. Syst. Comput.* 599 (2018), 72–82. DOI: https://doi.org/10.1007/978-3-319-60204-2_8

[95] Daeil Lee, Poong Hyun Seong, and Jonghyun Kim. 2018. Autonomous operation algorithm for safety systems of nuclear power plants by using long-short term memory and function-based hierarchical framework. *Ann. Nuclear Ener.* 119 (2018), 287–299. DOI: https://doi.org/10.1016/j.anucene.2018.05.020

[96] Sangdo Lee and Jun-Ho. Huh. 2019. An effective security measures for nuclear power plant using big data analysis approach. *J. Supercomput.* 75, 8 (2019), 4267–4294. DOI: https://doi.org/10.1007/s11227-018-2440-4

[97] Sangdo Lee, Jun-Ho Huh, and Yonghoon Kim. 2020. Python tensorflow big data analysis for the security of Korean nuclear power plants. *Electronics* 9, 9 (2020). DOI: https://doi.org/10.3390/electronics9091467

[98] Jason K. Levy. 2009. Nuclear non-proliferation and international security: A drama theoretic approach. *J. Syst. Sci. Syst. Eng.* 18, 4 (2009), 437–460. DOI: https://doi.org/10.1007/s11518-009-5117-y

[99] Yeni Li, Hany Abdel-Khalik, Elisa Bertino, and Arvind Sundaram. 2018. Development of defenses against false data injection attacks for nuclear power plants. Sandia National Laboratories, https://www.osti.gov/biblio/1761347

[100] Yeni Li, Hany S. Abdel-Khalik, and Elisa Bertino. 2018. Analysis of adversarial learning of reactor state. In *IEEE International Symposium on Technologies for Homeland Security (HST'18).* 1–6. DOI: https://doi.org/10.1109/THS.2018.8574137

[101] Yeni Li, Elisa Bertino, and Hany S. Abdel-Khalik. 2020. Effectiveness of model-based defenses for digitally controlled industrial systems: Nuclear reactor case study. *Nuclear Technol.* 206, 1 (2020), 82–93. DOI: https://doi.org/10.1080/00295450.2019.1626170

[102] Fan Liang, William Grant Hatcher, Weixian Liao, Weichao Gao, and Wei Yu. 2019. Machine learning for security and the internet of things: The good, the bad, and the ugly. *IEEE Access* 7 (2019), 158126–158147. DOI: https://doi.org/10.1109/ACCESS.2019.2948912

[103] Chaung Lin and Hung-Jen Chang. 2011. Identification of pressurized water reactor transient using template matching. *Ann. Nuclear Ener.* 38, 7 (2011), 1662–1666. DOI: https://doi.org/10.1016/j.anucene.2010.11.027

[104] Béla Lipták. 2016. Béla Lipták on Safety: Cyber Security and Nuclear Power. Retrieved from https://www.controlglobal.com/articles/2016/bela-liptak-on-safety-cyber-security-and-nuclear-power/

[105] Matt Luckcuck, Michael Fisher, Louise Dennis, Steve Frost, Andy White, and Doug Styles. 2021. Principles for the Development and Assurance of Autonomous Systems for Safe Use in Hazardous Environments. DOI: https://doi.org/10.5281/zenodo.5012322

[106] Chunchuan Lyu, Kaizhu Huang, and Hai-Ning Liang. 2015. A unified gradient regularization family for adversarial examples. In *IEEE International Conference on Data Mining*. 301–309.

[107] Jianping Ma and Jin Jiang. 2011. Applications of fault detection and diagnosis methods in nuclear power plants: A review. *Prog. Nuclear Ener.* 53, 3 (Apr. 2011), 255–266. DOI: https://doi.org/10.1016/j.pnucene.2010.12.001

[108] Xingjun Ma, Bo Li, Yisen Wang, Sarah M. Erfani, Sudanthi Wijewickrema, Grant Schoenebeck, Dawn Song, Michael E. Houle, and James Bailey. 2018. Characterizing Adversarial Subspaces Using Local Intrinsic Dimensionality. arXiv:1801.02613 [cs.LG]

[109] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. 2019. Towards Deep Learning Models Resistant to Adversarial Attacks. arXiv:1706.06083 [stat.ML]

[110] Rober P. Martin. 2019. Science-based nuclear design and safety in the emerging age of data based analytics. *Nuclear Eng. Des.* 354 (2019). DOI: https://doi.org/10.1016/j.nucengdes.2019.110155

[111] Jose Medeiros and Roberto Schirru. 2008. Identification of nuclear power plant transients using the particle swarm optimization algorithm. *Ann. Nuclear Ener.* 35 (2008), 576–582.

[112] Jan Hendrik Metzen, Tim Genewein, Volker Fischer, and Bastian Bischoff. 2017. On Detecting Adversarial Perturbations. arXiv:1702.04267 [stat.ML]

[113] Miles Brundage, Shahar Avin, Jack Clark, Helen Toner, Peter Eckersley, Ben Garfinkel, Allan Dafoe, Paul Scharre, Thomas Zeitzoff, Bobby Filar, Hyrum Anderson, Heather Roff, Gregory C. Allen, Jacob Steinhardt, Carrick Flynn, Seán Ó hÉigeartaigh, Simon Beard, Haydn Belfield, Sebastian Farquhar, Clare Lyle, Rebecca Crootof, Owain Evans, Michael Page, Joanna Bryson, Roman Yampolskiy, and Dario Amod. 2018. The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. Website. (2018). https://maliciousaireport.com/

[114] Matthew Mirman, Timon Gehr, and Martin Vechev. 2018. Differentiable abstract interpretation for provably robust neural networks. In *35th International Conference on Machine Learning (Proceedings of Machine Learning Research,* Vol. 80*)*, Jennifer Dy and Andreas Krause (Eds.). PMLR, 3578–3586. Retrieved from https://proceedings.mlr.press/v80/mirman18b.html

[115] Kun Mo, Seung Jun Lee, and Poong Hyun Seong. 2007. A dynamic neural network aggregation model for transient diagnosis in nuclear power plants. *Prog. Nuclear Ener.* 49, 3 (2007), 262–272. DOI: https://doi.org/10.1016/j.pnucene.2007.01.002

[116] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, Omar Fawzi, and Pascal Frossard. 2017. Universal Adversarial Perturbations. arXiv:1610.08401 [cs.CV]

[117] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, and Pascal Frossard. 2016. DeepFool: A Simple and Accurate Method to Fool Deep Neural Networks. arXiv:1511.04599 [cs.LG]

[118] Khalil Moshkbar-Bakhshayesh and Mohammad B. Ghofrani. 2013. Transient identification in nuclear power plants: A review. *Prog. Nuclear Ener.* 67 (2013), 23–32. DOI: https://doi.org/10.1016/j.pnucene.2013.03.017

[119] Antônio Carlos A. Mól, Cláudio Márcio N. A. Pereira, Victor Gonçalves G. Freitas, and Carlos Alexandre F. Jorge. 2011. Radiation dose rate map interpolation in nuclear plants using neural networks and virtual reality techniques. *Ann. Nuclear Ener.* 38, 2 (2011), 705–712. DOI: https://doi.org/10.1016/j.anucene.2010.08.008

[120] Man Gyun Na, Sun Ho Shin, Dong Won Jung, Soong Pyung Kim, Ji Hwan Jeong, and Byung Chul Lee. 2004. Estimation of break location and size for loss of coolant accidents using neural networks. *Nuclear Eng. Des.* 232, 3 (2004), 289–300. DOI: https://doi.org/10.1016/j.nucengdes.2004.06.007

[121] Ephraim. Nissan. 2019. An overview of AI methods for in-core fuel management: Tools for the automatic design of nuclear reactor core configurations for fuel reload, (re)arranging new and partly spent fuel. *Designs* 3, 3 (2019), 1–45. DOI: https://doi.org/10.3390/designs3030037

[122] Ephraim Nissan, Hava Siegelmann, Alex Galperin, and Shuky Kimhi. 1997. Upgrading automation for nuclear fuel in-core management: From the symbolic generation of configurations, to the neural adaptation of heuristics. *Eng. Comput.* 13, 1 (1997), 1–19. DOI: https://doi.org/10.1007/BF01201857

[123] Office for Nuclear Regulation. 2021. *The Impact of AI/ML on Nuclear Regulation.* Research Report ONR-RRR-121. Office for Nuclear Regulation. Retrieved from https://www.onr.org.uk/documents/2021/onr-rrr-121.pdf

[124] Felix O. Olowononi, Danda B. Rawat, and Chunmei Liu. 2021. Resilient machine learning for networked cyber physical systems: A survey for machine learning security to securing machine learning for CPS. *IEEE Commun. Surv. Tutor.* 23, 1 (2021), 524–552. DOI: https://doi.org/10.1109/comst.2020.3036778

[125] Nicolas Papernot, P. McDaniel, and I. Goodfellow. 2016. Transferability in Machine Learning: From Phenomena to Black-box Attacks Using Adversarial Samples. *ArXiv* abs/1605.07277 (2016).

[126] Nicolas Papernot, P. McDaniel, I. Goodfellow, S. Jha, Z. B. Celik, and A. Swami. 2017. Practical black-box attacks against machine learning. In *ACM on Asia Conference on Computer and Communications Security*.

[127] Nicolas Papernot, Patrick McDaniel, Somesh Jha, Matt Fredrikson, Z. Berkay Celik, and Ananthram Swami. 2015. The Limitations of Deep Learning in Adversarial Settings. arXiv:1511.07528 [cs.CR]

[128] Nicolas Papernot, P. McDaniel, Arunesh Sinha, and Michael P. Wellman. 2016. Towards the science of security and privacy in machine learning. *ArXiv* abs/1611.03814 (2016).

[129] Nicolas Papernot, Patrick McDaniel, Xi Wu, Somesh Jha, and Ananthram Swami. 2016. Distillation as a Defense to Adversarial Perturbations against Deep Neural Networks. arXiv:1511.04508 [cs.CR]

[130] Karthik Pattabiraman, Guanpeng Li, and Zitao Chen. 2020. Error resilient machine learning for safety-critical systems: Position paper. In *IEEE 26th International Symposium on On-Line Testing and Robust System Design (IOLTS'20)*. IEEE, 1–4. DOI: https://doi.org/10.1109/IOLTS50870.2020.9159749

[131] Michael E. Rising, Pavel Grechanuk, and Todd S. Palmer. 2018. Using machine learning methods to predict bias in nuclear criticality safety. *J. Computat. Theoret. Transport* 47, 4-6 (2018), 552–565. DOI: https://doi.org/10.1080/23324309.2019.1585877

[132] Bin-Sen Peng, Hong Xia, Yong-Kuo Liu, Bo Yang, Dan Guo, and Shao-Min Zhu. 2018. Research on intelligent fault diagnosis method for nuclear power plant based on correlation analysis and deep belief network. *Prog. Nuclear Ener.* 108 (2018), 419–427. DOI: https://doi.org/10.1016/j.pnucene.2018.06.003

[133] Victor Henrique Cabral Pinheiro and Roberto Schirru. 2019. Genetic programming applied to the identification of accidents of a PWR nuclear power plant. *Ann. Nuclear Ener.* 124 (2019), 335–341. DOI: https://doi.org/10.1016/j.anucene.2018.09.039

[134] Yuliya Pranuza. 2022. Capacity to build artificial intelligence systems for nuclear energy security and sustainability: Experience of Belarus. *IFIP Advan. Inf. Commun. Technol.* 637 IFIP (2022), 128–140. DOI: https://doi.org/10.1007/978-3-030-96592-1_10

[135] Julwan Hendry Purba, Jie Lu, Da Ruan, and Guangquan Zhang. 2010. Probabilistic safety assessment in nuclear power plants by fuzzy numbers. In *9th International FLINS Conference: Computational Intelligence Foundations and Applications (FLINS'10)*. 256–262. DOI: https://doi.org/10.1142/9789814324700_0037

[136] Ragunathan Rajkumar, Insup Lee, Lui Sha, and John Stankovic. 2010. Cyber-physical systems: The next computing revolution. In *Design Automation Conference*. 731–736. DOI: https://doi.org/10.1145/1837274.1837461

[137] Kui Ren, Tianhang Zheng, Zhan Qin, and Xue Liu. 2020. Adversarial attacks and defenses in deep learning. *Engineering* 6, 3 (2020), 346–360. DOI: https://doi.org/10.1016/j.eng.2019.12.012

[138] Andrew Slavin Ros and Finale Doshi-Velez. 2018. Improving the adversarial robustness and interpretability of deep neural networks by regularizing their input gradients. In *32nd AAAI Conference on Artificial Intelligence and 30th Innovative Applications of Artificial Intelligence Conference and 8th AAAI Symposium on Educational Advances in Artificial Intelligence (AAAI'18/IAAI'18/EAAI'18)*. AAAI Press, Article 203, 10 pages.

[139] Emir Roumili, Jean-François Bossu, Vincent Chapurlat, Nicolas Daclin, Robert Plana, and Jérôme Tixier. 2021. Collaborative safety requirements engineering: An approach for modelling and assessment of nuclear safety requirements in MBSE context. In *Smart and Sustainable Collaborative Networks 4.0*, Luis M. Camarinha-Matos, Xavier Boucher, and Hamideh Afsarmanesh (Eds.). Springer International Publishing, Cham, 227–236.

[140] Mahdi Saghafi and Mohammad B. Ghofrani. 2019. Real-time estimation of break sizes during LOCA in nuclear power plants using NARX neural network. *Nuclear Eng. Technol.* 51, 3 (2019), 702–708. DOI: https://doi.org/10.1016/j.net.2018.11.017

[141] Kaur Sandhu, Saran Srikanth Bodda, and Abhinav Gupta. 2022. Deep learning framework for post-hazard condition monitoring of nuclear safety systems. In *13th International Workshop on Structural Health Monitoring*. Retrieved from https://api.semanticscholar.org/CorpusID:252787393

[142] T. V. Santosh, A. Srivastava, V. V. S. Sanyasi Rao, A. K. Ghosh, and H. S. Kushwaha. 2009. Diagnostic system for identification of accident scenarios in nuclear power plants using artificial neural networks. *Reliab. Eng. Syst. Safety* 94, 3 (2009), 759–762. DOI: https://doi.org/10.1016/j.ress.2008.08.005

[143] T. V. Santosh, Gopika Vinod, R. K. Saraf, A. K. Ghosh, and H. S. Kushwaha. 2007. Application of artificial neural networks to nuclear power plant transient diagnosis. *Reliab. Eng. Syst. Safety* 92, 10 (2007), 1468–1472. DOI: https://doi.org/10.1016/j.ress.2006.10.009

[144] Abdullah M. Sawas, Hadi Khani, and Hany E. Z. Farag. 2021. On the resiliency of power and gas integration resources against cyber attacks. *IEEE Trans. Industr. Inform.* 17, 5 (2021), 3099–3110. DOI: https://doi.org/10.1109/TII.2020.3007425

[145] Evert Schlünz, Pavel Bokov, and Jan Vuuren. 2016. An optimisation-based decision support system framework for multi-objective in-core fuel management of nuclear reactor cores. *South Afric. J. Industr. Eng.* 27 (11 2016), 201–209. DOI: https://doi.org/10.7166/27-3-1650

[146] Siegfried D. Schmid and Hideaki Kanekiyo. 2001. The Tokaimura nuclear accident: A tragedy of human errors. *J. Nuclear Sci. Technol.* 38, 10 (2001), 865–874.

[147] Stephen J. Schmugge, Lance Rice, N. Rich Nguyen, John Lindberg, Robert Grizzi, Chris Joffe, and Min C. Shin. 2016. Detection of cracks in nuclear power plant using spatial-temporal grouping of local patches. In *IEEE Winter Conference on Applications of Computer Vision (WACV'16)*. IEEE, 1–7. DOI: https://doi.org/10.1109/WACV.2016.7477601

[148] Ali Shafahi, W. Ronny Huang, Christoph Studer, Soheil Feizi, and Tom Goldstein. 2020. Are Adversarial Examples Inevitable? arXiv:1809.02104 [cs.LG]

[149] Affan Shaukat, Yang Gao, Jeffrey A. Kuo, Bob A. Bowen, and Paul E. Mort. 2016. Visual classification of waste material for nuclear decommissioning. *Robot. Auton. Syst.* 75 (2016), 365–378. DOI: https://doi.org/10.1016/j.robot.2015.09.005

[150] Yong Shi, Xiaodong Xue, Yi Qu, Jiayu Xue, and Linzi Zhang. 2021. Machine learning and deep learning methods used in safety management of nuclear power plants: A survey. In *International Conference on Data Mining Workshops (ICDMW'21)*. 917–924. DOI: https://doi.org/10.1109/ICDMW53433.2021.00120

[151] Galen M. Shipman, Jason Pruet, David Daniel, Josh Dolence, Gary Grider, Brian M. Haines, Aimee Hungerford, Stephen Poole, Tim Randles, Sriram Swaminarayan, and Chris Werner. 2023. The future of HPC in nuclear security. *IEEE Internet Comput.* 27, 1 (2023), 16–23. DOI: https://doi.org/10.1109/MIC.2022.3229037

[152] Natalie J. Smith-Gray, Irmak Sargin, Scott Beckman, and John McCloy. 2021. Machine learning to predict refractory corrosion during nuclear waste vitrification. *MRS Advan.* 6, 4-5 (2021), 131–137. DOI: https://doi.org/10.1557/s43580-021-00031-2

[153] Thiago Juncal Souza, Jose A. C. Medeiros, and Alessandro Gonçalves. 2017. Identification model of an accidental drop of a control rod in PWR reactors using thermocouple readings and radial basis function neural networks. *Ann. Nuclear Ener.* 103 (2017), 204–211. DOI: https://doi.org/10.1016/j.anucene.2017.01.004

[154] Siddharth Suman. 2021. Artificial intelligence in nuclear industry: Chimera or solution? *J. Clean. Product.* 278 (2021), 124022. DOI: https://doi.org/10.1016/j.jclepro.2020.124022

[155] Bill K.-H. Sun. 1988. Control and diagnostics for nuclear power plant performance and safety enhancement. In *IEEE Conference on Human Factors and Power Plants*. 13–21. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-0024028052&partnerID=40&md5=0e1a6e9bc5128430b2f7f990514dbdf8

[156] Dajie Sun, Haruko M. Wainwright, Carlos A. Oroza, Akiyuki Seki, Satoshi Mikami, Hiroshi Takemiya, and Kimiaki Saito. 2020. Optimizing long-term monitoring of radiation air-dose rates after the Fukushima Daiichi nuclear power plant. *J. Environ. Radioact.* 220-221 (2020), 106281. DOI: https://doi.org/10.1016/j.jenvrad.2020.106281

[157] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. 2014. Intriguing Properties of Neural Networks. arXiv:1312.6199 [cs.CV]

[158] Xiange Tian, Victor Becerra, Nils Bausch, T. V. Santhosh, and Gopika Vinod. 2018. A study on the robustness of neural network models for predicting the break size in LOCA. *Prog. Nuclear Ener.* 109 (Nov. 2018), 12–28. DOI: https://doi.org/10.1016/j.pnucene.2018.07.004

[159] Monirangsey Touch. 2019. Cybersecurity and the Nuclear Industry. Retrieved from https://jsis.washington.edu/news/cybersecurity-nuclear-industry/

[160] U.S. Department of Energy. 2018. 5 Incredible Ways Nuclear Powers Our Lives. Retrieved from https://www.energy.gov/ne/articles/5-incredible-ways-nuclear-powers-our-lives

[161] U. S. Energy Information Administration. 2022. Nuclear Power and the Environment. Retrieved from https://www.eia.gov/energyexplained/nuclear/nuclear-power-and-the-environment.php

[162] Andrew Wade. 2019. UK to Tackle Nuclear Waste with Robots and AI. Retrieved from https://www.theengineer.co.uk/nuclear-waste-ncnr-robots/

[163] J. Samuel Walker. 2004. *Three Mile Island: A Nuclear Crisis in Historical Perspective*. University of California Press.

[164] T. Washio, M. Kitamura, K. Kotajima, and K. Sugiyama. 1986. Automated generation of nuclear power plant safety information (qualitative simulation and derivation of failure symptom knowledge). *Proc. Power Plant Dynam., Contr. Test. Sympos.* 1 (1986), 39. 01–39. 17. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-0022941828&partnerID=40&md5=e83ffb47699e80011f5272f4744696b7

[165] John G. Williams and W. C. Jouse. 1993. Intelligent control in safety systems: Criteria for acceptance in the nuclear power industry. *IEEE Trans. Nuclear Sci.* 40, 6 (1993), 2040–2044. DOI: https://doi.org/10.1109/23.273446

[166] Taeho Woo. 2012. *Nuclear Safety Assessment for the Passive System of the Nuclear Power Plants (NPPs) in Safety Margin Estimation*. Springer London, London, 61–73. DOI: https://doi.org/10.1007/978-1-4471-4030-6_6

[167] Tae Ho Woo, Chang Hyun Baek, and Kyung Bae Jang. 2022. Safety analysis for integrity enhancement in nuclear power plants (NPPs) in case of seashore region site. *Kerntechnik* 87, 3 (2022), 271–277. DOI: https://doi.org/doi:10.1515/kern-2022-0013

[168] Tae-Ho Woo and Un-Chul Lee. 2010. Safety assessment for the passive system of the nuclear power plants (NPPs) using safety margin estimation. *Energy* 35, 4 (2010), 1799–1804. DOI: https://doi.org/10.1016/j.energy.2009.12.034

[169] World Nuclear Association. 2021. Safety of Nuclear Reactors. Retrieved from https://www.world-nuclear.org/information-library/safety-and-security/safety-of-plants/safety-of-nuclear-power-reactors.aspx

[170] Clarence Worrell, Louis Luangkesorn, Joel Haight, and Thomas Congedo. 2019. Machine learning of fire hazard model simulations for use in probabilistic safety assessments at nuclear power plants. *Reliab. Eng. Syst. Safety* 183 (2019), 128–142. DOI: https://doi.org/10.1016/j.ress.2018.11.014

[171] Shun-Chi Wu, Kuang-You Chen, Ting-Han Lin, and Hwai-Pwu Chou. 2018. Multivariate algorithms for initiating event detection and identification in nuclear power plants. *Ann. Nuclear Ener.* 111 (2018), 127–135. DOI: https://doi.org/10.1016/j.anucene.2017.08.066

[172] Yunna Wu, Qing Bian, Wei Luo, Xinliang Hu, and Lingshuang Xu. 2012. The research on the affecting factors of safety management of nuclear power based on improved ISM. In *International Conference on Automatic Control and Artificial Intelligence (ACAI'12)*. 2179–2182. DOI: https://doi.org/10.1049/cp.2012.1431

[173] Chaowei Xiao, Jun-Yan Zhu, Bo Li, Warren He, Mingyan Liu, and Dawn Song. 2018. Spatially transformed adversarial examples. In *International Conference on Learning Representations*. Retrieved from https://openreview.net/forum?id=HyydRMZC-

[174] Dong-Ling Xu, Da Ruan, and Jian-Bo Yang. 2011. Supporting nuclear safety culture assessment using Intelligent Decision System software. In *IEEE Symposium on Computational Intelligence in Multicriteria Decision-Making (MDCM'11)*. IEEE, 67–72. DOI: https://doi.org/10.1109/SMDCM.2011.5949281

[175] Weilin Xu, David Evans, and Yanjun Qi. 2018. Feature squeezing: Detecting adversarial examples in deep neural networks. In *Network and Distributed System Security Symposium*. DOI: https://doi.org/10.14722/ndss.2018.23198

[176] Xinyi Xu, Taihao Han, Jie Huang, Albert A. Kruger, Aditya Kumar, and Ashutosh Goel. 2021. Machine learning enabled models to predict sulfur solubility in nuclear waste glasses. *ACS Appl. Mater. Interf.* 13, 45 (2021), 53375–53387. DOI: https://doi.org/10.1021/acsami.1c10359

[177] Likai Yao, Cihan Tunc, Pratik Satam, and Salim Hariri. 2020. Resilient machine learning (rML) ensemble against adversarial machine learning attacks. In *Dynamic Data Driven Applications Systems*, Frederica Darema, Erik Blasch, Sai Ravela, and Alex Aved (Eds.). Springer International Publishing, Cham, 274–282.

[178] Ceyhun Yavuz and Senem Şentürk Lüle. 2022. The application of artificial intelligence to nuclear power plant safety. In *Artificial Intelligence for Knowledge Management, Energy, and Sustainability*, Eunika Mercier-Laurent and Gülgün Kayakutlu (Eds.). Springer International Publishing, Cham, 117–127.

[179] Sicong Zhang, Xiaoyao Xie, and Yang Xu. 2020. A brute-force black-box method to attack machine learning-based systems in cybersecurity. *IEEE Access* 8 (2020), 128250–128263. DOI: https://doi.org/10.1109/ACCESS.2020.3008433

[180] Shuqiao Zhou, Duo Li, and Xiaojin Huang. 2017. Transient identification for nuclear power plants based on the similarity of matrices. In *8th International Conference on Intelligent Control and Information Processing (ICICIP'17)*. IEEE, 225–230. DOI: https://doi.org/10.1109/ICICIP.2017.8113946

[181] Enrico E. Zio and Piero Baraldi. 2005. Identification of nuclear transients via optimized fuzzy clustering. *Ann. Nuclear Ener.* 32 (2005), 1068–1080.

[182] A. K. Ziver, C. C Pain, J. N. Carter, C. R. E. de Oliveira, A. J. H. Goddard, and R. S. Overton. 2004. Genetic algorithms and artificial neural networks for loading pattern optimisation of advanced gas-cooled reactors. *Ann. Nuclear Ener.* 31, 4 (2004), 431–457. DOI: https://doi.org/10.1016/j.anucene.2003.08.005

[183] Rehan Zubair, Atta Ullah, Asifullah Khan, and Mansoor H. Inayat. 2022. Critical heat flux prediction for safety analysis of nuclear reactors using machine learning. In *19th International Bhurban Conference on Applied Sciences and Technology (IBCAST'22)*. IEEE, 314–318. DOI: https://doi.org/10.1109/IBCAST54850.2022.9990190

[184] Serhat Şeker, Emine Ayaz, and Erdinç Türkcan. 2003. Elman's recurrent neural network applications to condition monitoring in nuclear power plant and rotating machinery. *Eng. Applic. Artif. Intell.* 16, 7 (2003), 647–656. DOI: https://doi.org/10.1016/j.engappai.2003.10.004

Received 8 August 2022; revised 12 January 2024; accepted 31 January 2024