achieved huge success in everyday applications like Siri or Alexa. In the next Figure 2, using Venn diagrams, it is shown that Deep learning is a subset of Machine Learning.

Fig. 2: Artificial Intelligence and its Subset

Industry 4.0 is also bringing a lot of complexity due to the many n-to-n connections that are occurring between the systems. In the past, many of the now interconnected systems were isolated. However, this is now changing, and the attack surface of a typical ICS/SCADA is enlarged, making it more vulnerable and more exposed to different threat actors. Our previous paper discussed various classifications of threat actors, [1].

As previously mentioned, Industry 4.0 architectures are complex and hard to understand, which makes them difficult to protect. The production chain processes are also agile and dynamic. This also requires agile and quick self-learning cybersecurity solutions. The challenges around this have been researched already by, [2]. In summary, the design of CPS requires a deep understanding of system design, engineering, and sociology. The researchers have developed a framework that helps AI to be integrated with cyber risk analytics for CPS. Getting all the necessary data into account is key to success as described in the researchers' paper. Some of the key data required for AI are proper asset management and access control. The AI deep learning processes must be secured from manipulation by insiders, disgruntled employees, or state-sponsored actors. If an attacker can manipulate the deep learning process by doing data poisoning, they can potentially poison the dataset with anything they want, making it invisible during the next attack execution, [3]. As digital transformation reshapes industrial processes, a comprehensive analysis of practical machine learning implementations is indispensable. Our research seeks to provide insights into the role of machine learning in securing Industry 4.0 infrastructures. From anomaly detection to predictive analytics, this examination uncovers the diverse range of machine-learning scenarios that fortify cybersecurity measures in the era of Industry 4.0.

Another group of researchers has conducted research around the "ECHO Federated Cyber Range (E-FCR)" virtualization environment and the "ECHO Early Warning System (E-EWS)". In conclusion, the E-EWS and E-FCR systems offer valuable capabilities for national authorities and government agencies. They facilitate the sharing of information, identification of cyber security threats, and the development of mitigation tools and products. Additionally, the E-EWS can be utilized for educational purposes and procedural support. By promoting knowledge sharing, the E-EWS enhances community resilience by increasing awareness and understanding of cyber security issues, [4].

2 Machine Learning/ Deep Learning Standardization

While doing this research to understand the current posture in deep learning, we have observed the lack of clear definitions and standards around AI and Cybersecurity frameworks, including ML/DL.

Advanced Intrusion Detection Systems for ICS are already making use of Artificial Intelligence and, more specifically, Machine learning (ML). They are trained by using publicly available datasets or private network traces obtained from self-owned honeypots. Due to the fast pace of malware evolution and the constant introduction of new attack vectors, the datasets are not fully protected from new types of cyberattacks, [5].

In 2018, the Joint Research Centre (JRC) of the European Commission published a report on AI. The report addresses key aspects of the adoption of AI. It is made clear that AI is a twofold coin, and it can have potential dangers to the overall security of systems. The report suggests that "further research is needed in the field of adversarial ML to better understand the limitations in the robustness of ML algorithms and design effective strategies to mitigate these vulnerabilities", [6].

On 19 February 2020, the European Commission published a white paper on Artificial Intelligence. This move has outlined the strategy of the EU to

centrally align the AI ecosystems of all EU member states, [7].

On 14 December 2021, the ENISA published the Securing Machine Learning Algorithms report. This report mainly focuses on describing what the taxonomy is for machine learning algorithms. The report also describes what are the present threats to machine learning systems. Some of the threats that are identified but not limited to are data exfiltration and poisoning, adversarial attacks, etc., [8].

Artificial intelligence, as defined by ISO2382, pertains to a field within computer science that focuses on constructing data processing systems capable of performing tasks typically associated with human intelligence. These tasks include logical reasoning, learning, and self-enhancement.

Under the guidance of Fraunhofer IKS, an international consortium is actively formulating norms and standards addressing the safety aspects of artificial intelligence, exemplified by the ISO/PAS 8800 standard. This initiative aims to establish a comprehensive framework for standardizing the development, testing, and, where applicable, regulation of forthcoming AI systems employed in vehicles. Beyond solely providing specifications for neural networks, the standard encompasses a broader spectrum, incorporating more easily comprehensible AI approaches. These approaches, often better suited for safety functions, extend the applicability of the standards not only to autonomous driving but also to various domains within artificial intelligence, particularly those emphasizing safety considerations.

On the other hand, the United States National Institute of Standards and Technology (US NIST) has not yet issued an official process model concerning artificial intelligence, [9]. Nevertheless, the most recent framework provided by NIST, known as the NIST CSF framework, offers recommendations and actions that organizations can contemplate for the deployment of Cyber-Physical Systems (CPS), [10].

However, our research showed that the NIST team is already preparing the ground for the potential future release of an AI Risk Management framework that will standardize the usage of AI and support businesses to securely implement AI, [11].

It's important to note that standardization efforts should be collaborative, involving a wide range of stakeholders, including industry experts, researchers, policymakers, and ethicists. Additionally, standardization should be adaptable and flexible to accommodate the rapid advancements in the field of machine learning.

Ultimately, the goal of standardization is to ensure that machine learning technologies are safe, ethical, and accessible to all while promoting innovation and economic growth.

3 Machine Learning/ Deep Learning for Cybersecurity

As we have already mentioned in the introduction of this paper there are different types of DL approaches and DL architectures. We will explain them in the next chapters.

3.1 Machine Learning/ Deep Learning Approaches

As of today, the DL approaches are divided into three major categories – supervised, unsupervised, or semi-supervised.

Supervised learning, alternatively referred to as supervised machine learning, is a prominent subset of both machine learning and deep learning. In this methodology, the algorithm leverages labeled datasets during its training process, allowing it to acquire the capability to effectively classify data. This approach is widely employed when addressing classification and regression challenges.

Classification problems mostly consist of a function that maps the input to a final value. Whereas regression problems create a function that maps to a continuous variable, [12]. The process of supervised learning involves two main components: the input features (also known as independent variables) and the target labels (also known as dependent variables). The input features denote the traits or properties of the data points, while the target labels signify the sought-after output or the group to which the data points pertain. During the training stage of supervised learning, the algorithm scrutinizes the input features in conjunction with their corresponding target labels to unveil concealed patterns and associations. It aims to unearth a mapping or function capable of reliably predicting the target label for new input data points.

Typically, this learned function is often depicted as a model or hypothesis. The selection of a particular supervised learning algorithm hinges on the characteristics of the problem and the nature of the target variable. For instance, in classification tasks where the target variable is categorical, one can employ algorithms such as decision trees, logistic regression, support vector machines (SVM), and neural networks. These algorithms are designed to categorize the input data points into distinct predefined groups.

Unsupervised learning, sometimes referred to as unsupervised machine learning, is employed for the examination of untagged datasets. In unsupervised learning, the primary emphasis is placed on scrutinizing and comprehending the inherent structure and patterns within the data, devoid of any explicit target labels.

Unlike supervised learning, where the algorithm is provided with labeled examples to learn from, unsupervised learning algorithms work with unlabeled data, [13]. Unsupervised learning aims to uncover valuable insights, unearth concealed patterns, and unveil inherent structures in data. It employs diverse techniques like clustering, dimensionality reduction, and density estimation to accomplish this.

Clustering algorithms gather data points that exhibit proximity or likeness, facilitating the detection of cohesive clusters or groups within the data. Through the grouping of akin data points, these algorithms establish a foundation for structuring and comprehending intricate datasets.

Semi-supervised learning, also referred to as semi-supervised machine learning, is employed in unique scenarios wherein the training dataset consists of a combination of labeled and unlabeled data. It represents a learning approach that falls between the domains of supervised and unsupervised learning. The goal is to harness the limited labeled data in conjunction with the wealth of unlabeled data to enhance the learning process.

The labeled data, encompassing input features paired with their associated target labels, is utilized in a manner akin to that of supervised learning. The algorithm gains knowledge from these labeled instances to make predictions and classify new cases with precision.

Nevertheless, what distinguishes semi-supervised learning is its integration of supplementary unlabeled data. While these instances lack specific target labels, they still hold valuable insights into the underlying data structure. Through the inclusion of unlabeled data, the algorithm strives for enhanced generalization, the discovery of more significant patterns, and overall performance improvement. Semi-supervised learning algorithms primarily operate by adhering to the consistency principle, striving to ensure that comparable instances in the input space yield similar predictions.

Through upholding uniformity in predictions across both labeled and unlabeled data points, the algorithm aims to transmit knowledge from labeled instances to unlabeled ones, thereby enhancing the model's capacity to generalize and produce precise predictions, [14].

3.2 Deep Learning Architectures

Deep learning architectures accommodate the three machine learning techniques mentioned earlier. Specifically, the discriminative model supports supervised machine learning methods, the generative model supports unsupervised machine learning methods, and a hybrid model underpins semi-supervised machine learning approaches.

According to ENISA, the research team was able to identify around the 40 most commonly used DL algorithms. In our research, we have additionally limited the number to 19 algorithms that are being used in the cybersecurity realm. The architectural diagram of different types of machine learning algorithms is presented in Figure 3.

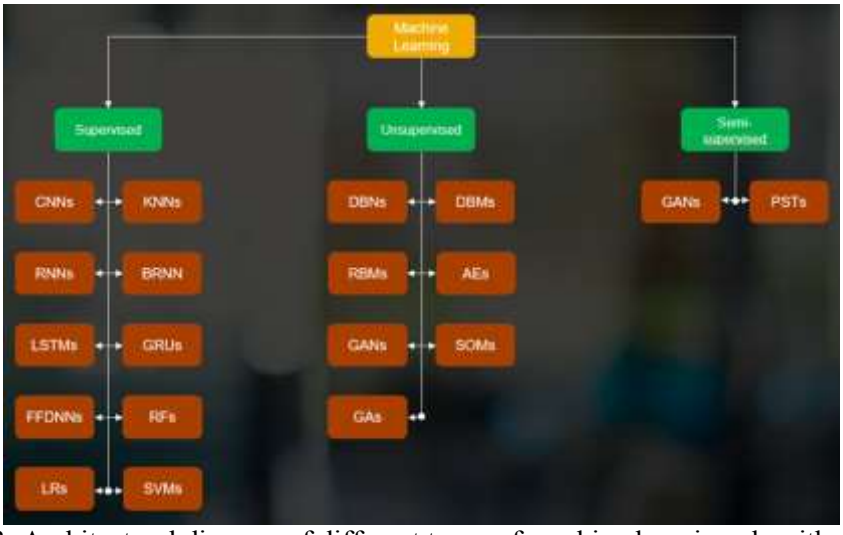

Fig. 3: Architectural diagram of different types of machine learning algorithms

3.2.1 Deep Convolutional Neural Networks (CNNs)

A deep convolutional neural network is a class of artificial neural networks that are used to identify patterns in images and videos. A deep CNN processes images as input and uses them to later on train a classifier. They are made up of multiple layers, including convolutional layers, pooling layers, and fully connected layers. Convolutional layers apply filters to the input image to detect features, such as edges, shapes, or patterns. Pooling layers reduce the size of the image, keeping only the most important information. Fully connected layers connect the output of the previous layers to the final output layer, which can be used for classification or regression.

3.2.2 K-Nearest Neighbors (KNNs)

K-Nearest Neighbors (KNN) constructs a model consisting of the training samples. It compares the new data point with the previously recorded samples, selecting the K nearest ones. The label assigned to the new data point is determined by the majority of those K samples. The choice of distance function for comparing data points is made by the data analyst, often employing Euclidean distance. KNN is a non-parametric algorithm used for classification and regression tasks in machine learning. KNNs are part of a family of "lazy learning" models, meaning that they only store a training dataset versus undergoing a training stage.

3.2.3 Recurrent Neural Networks (RNNs)

A recurrent neural network (RNN) is a machine learning model which uses consequent data. The most common usage of this algorithm is for processing language, captioning of images, or recognition of speech. RNN algorithms make use of training data to learn similar to CNN algorithms. RNNs are a type of artificial neural network that can process sequential data, such as text, speech, or video. Unlike feedforward neural networks, which only take the current input into account, RNNs have a memory that allows them to use information from previous inputs to influence the current output. One common difference from traditional deep neural networks is that RNN algorithms rely on the full sequence of elements to learn, [15].

3.2.4 Bidirectional Recurrent Neural Networks (BRNNs)

Bidirectional recurrent neural networks (BRNNs) are neural networks that process sequences in both forward and backward directions. Each hidden layer has a set of neurons that perform some computation on the input and produce an output. The output of each hidden layer is then combined and fed into a final output layer, which can be used for different tasks, such as language translation, text classification, or named entity recognition. They capture contextual information from past and future inputs, making them effective in tasks such as speech recognition and natural language processing.

3.2.5 Deep Belief Networks (DBNs)

Deep Belief Networks (DBNs) are multilayer neural networks composed of multiple layers of hidden units. They utilize unsupervised learning to pre-train each layer, allowing them to learn hierarchical representations of data. The bottom layers capture low-level features, while the top layers represent high-level abstractions. The learning process of DBNs consists of two phases: pre-training and fine-tuning. In the pre-training phase, DBNs learn to reconstruct the input data by using unsupervised learning algorithms, such as restricted Boltzmann machines (RBMs) or autoencoders. These algorithms train each layer of hidden units separately, by minimizing the reconstruction error between the input and the output of each layer. In this way, each layer learns to extract features from the previous layer's output. In the fine-tuning phase, DBNs learn to perform a specific task, such as classification or regression, by using supervised learning algorithms, such as gradient descent or backpropagation. These algorithms train all the layers together, by minimizing the error between the output and the target labels of the input data. DBNs are commonly used for tasks like classification, feature learning, and generative modeling, [16].

3.2.6 Deep Boltzmann Machine (DBMs)

Deep Boltzmann Machines (DBMs) are neural networks with multiple layers of hidden units that employ undirected connections. They utilize a probabilistic approach, with units following a Boltzmann distribution, to learn hierarchical representations. DBMs are designed for unsupervised learning and excel in modeling intricate data dependencies. They find applications in tasks like feature learning, dimensionality reduction, and sample generation, [17].

3.2.7 Deep Autoencoders (AEs)

Deep autoencoders (AEs) are neural networks designed for unsupervised learning and data compression. They consist of an encoder and a decoder component. The encoder compresses the

input data into a lower-dimensional representation, while the decoder reconstructs the original data from the compressed representation. Deep AEs utilize multiple hidden layers to learn increasingly abstract features, [18].

3.2.8 Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) are neural network architectures comprising a generator and a discriminator. The generator generates new data samples, while the discriminator aims to distinguish between real and generated samples. The generator learns from the feedback of the discriminator, and the discriminator learns from the data itself. The goal of GANs is to reach a state where the generator can fool the discriminator with high probability, meaning that the fake data is indistinguishable from the real data. GANs are difficult to train, as they require finding a balance between the generator and the discriminator. If the generator is too weak, the discriminator will easily reject its output. If the generator is too strong, the discriminator will become confused and unable to learn. GANs are trained through a competitive process to produce high-quality data samples, [19].

3.2.9 Long Short-Term Memory (LSTMs)

Long Short-Term Memory (LSTM) is a specific type of recurrent neural network (RNN) that tackles the challenge of the vanishing gradient problem. LSTMs are composed of multiple layers, each of which has a set of neurons that perform some computation on the input and produce an output. The output of each layer is then passed to the next layer as well as fed back to the same layer as part of the next input. This feedback loop creates a recurrent connection that enables the network to store and access information from previous time steps. The output of the last layer is then used for the final prediction or classification. LSTMs are equipped with memory cells that enable the retention and utilization of information over time. This capability allows LSTMs to effectively capture long-term dependencies present in sequential data. LSTMs have proven to be highly valuable in various domains, including natural language processing, speech recognition, and time series analysis, [20].

3.2.10 Gated Recurrent Units (GRUs)

Gated Recurrent Units (GRUs) are recurrent neural network (RNN) variants that offer a simplified architecture compared to Long Short-Term Memory (LSTM) networks. GRUs incorporate gating mechanisms to regulate information flow, allowing them to capture and utilize relevant contextual information. The reset gate determines how much of the previous hidden state should be forgotten, while the update gate determines how much of the new input should be used to update the hidden state. The output of the GRU is calculated based on the updated hidden state. They are also easier to train than standard RNNs, as they avoid the problem of vanishing or exploding gradients, which occurs when the gradient of the error function becomes very small or very large as it propagates through time. GRUs are useful for many applications that involve sequential data, such as machine translation, text generation, speech synthesis, sentiment analysis, and video analysis, [21].

3.2.11 Ensemble of DL networks (EDLNs)

Ensemble of Deep Learning Networks (EDLNs) refers to a technique where multiple deep learning models are combined to make predictions. By aggregating the outputs of individual models, EDLNs can enhance performance, increase robustness, and improve generalization in various machine learning tasks, including classification, regression, and anomaly detection.

3.2.12 Self-Organized Maps (SOMs)

Self-organized maps (SOMs), while belonging to the category of artificial neural networks, follow a distinct training approach that involves competitive learning, in contrast to the error-correction learning methods, such as backpropagation with gradient descent, utilized by other artificial neural networks. The SOM architecture comprises a grid of nodes, often referred to as neurons, which establish connections with the input data. Each node possesses a weight vector that serves as a representation of an input data prototype or centroid. Through the SOM algorithm, these nodes are organized into a two-dimensional grid in a manner that clusters similar nodes near each other.

3.2.13 Genetic Algorithm (GAs) or Genetic Programming (GPs)

Genetic Algorithms (GAs) are a family of optimization algorithms inspired by the process of natural selection and evolution. They are a part of the broader field of evolutionary computation and are used to solve complex optimization and search problems. GAs are used for optimization in various fields, including engineering, finance, logistics, and operations research.

3.2.14 Federated Learning (FLs)

Federated Learning (FL) is a machine learning approach that allows a model to be trained across multiple decentralized edge devices (such as smartphones, IoT devices, or local servers) while keeping the data on those devices rather than sending it to a centralized server. This privacy-preserving approach has gained significant attention in recent years due to its potential to address data privacy concerns, reduce communication costs, and make machine learning more scalable. FL can be used for fraud detection and risk assessment while keeping sensitive customer data decentralized, [22].

3.2.15 Feedforward Deep Neural Network (FFDNNs)

Feedforward Deep Neural Networks (FFDNNs), also known as a feedforward artificial neural network or a multilayer perceptron (MLP), is a fundamental type of artificial neural network that consists of an input layer, one or more hidden layers, and an output layer. It is called "feedforward" because the information flows in one direction, from the input layer to the output layer, without cycles or feedback loops.

3.2.16 Probabilistic Suffix Trees (PSTs)

Prediction Suffix Trees (PSTs) serve as a powerful machine-learning model for sequence prediction tasks, providing an elegant and effective solution. PSTs find application in various domains such as compression and reinforcement learning. Notably, the advantage of PSTs lies in their ability to dynamically adjust the number of symbols used for prediction based on the context, utilizing the underlying suffix tree data structure. This feature ensures efficient storage and retrieval of string sets and their associated suffixes. Building PSTs for large datasets can be computationally expensive, and they may suffer from data sparsity issues when dealing with rare sequences. Researchers often use techniques like smoothing and back-off models to address these challenges. PSTs are extensively used in language modeling. Given a sequence of words or characters, a PST can predict the likelihood of observing a particular word or character as the next item in the sequence. This is crucial in applications like text generation, machine translation, and speech recognition, where predicting the most probable next symbol or word is essential for generating coherent and contextually relevant output.

3.2.17 Support Vector Machines (SVMs)

Support Vector Machines (SVMs) are versatile and robust machine learning algorithms used for classification and regression tasks. SVMs aim to find an optimal hyperplane that effectively separates different classes or predicts continuous values. By maximizing the margin between the hyperplane and the nearest data points, SVMs promote generalization and robustness. They handle both linearly separable and non-linearly separable data by leveraging kernel functions. SVMs exhibit remarkable performance in high-dimensional spaces, are resistant to overfitting, and can handle datasets with complex decision boundaries. They find applications in diverse domains, such as image recognition, text analysis, bioinformatics, and anomaly detection, offering a powerful and interpretable solution for various machine learning problems.

3.2.18 Random Forest (RFs)

Random Forest (RFs) is an ensemble learning method in machine learning that combines the predictive power of multiple decision trees to improve the overall accuracy and robustness of the model. It was introduced by Leo Breiman and Adele Cutler and has become one of the most popular and effective machine learning algorithms for classification and regression tasks.

3.2.19 Linear Regression (LRs)

Linear Regression (LRs) is a fundamental statistical modeling technique used for predicting a continuous target variable based on one or more predictor variables. It assumes a linear relationship between the predictors and the target variable, where the objective is to find the best-fitting line that minimizes the overall difference between the observed and predicted values. LRs provide interpretable insights into the relationship between predictors and the target variable. While traditional Linear Regression aims to find the best-fitting line without any constraints, there are advanced techniques like Ridge Regression and Lasso Regression that introduce regularization. These methods add penalty terms to the linear regression model, which helps prevent overfitting and improves the model's generalization to unseen data. Both Ridge and Lasso Regression are part of the broader family of Regularized Linear Regression techniques. They provide additional tools for handling complex datasets and improving the robustness of Linear Regression models, which is

especially important in high-dimensional data scenarios.

3.3 ML/DL for Industry 4.0

In our previous work, we have researched what are the potential threat vectors and threat attackers that are attacking Industry 4.0 systems. Based on that research and also the latest research, we have provided a short visualization in Figure 4 below of the potential threats and some of the vulnerabilities in wind turbine systems that could have a serious impact on wind farms.

After analyzing the DL methods that are currently being widely developed and used, we took the next step to research which of the DL methods are used for defending CPS systems and, more specifically, wind parks. The most used DL methods for protecting CPS systems are listed below:

- Using Auto Encoders was proved to be successful amongst other methods by the following group of researchers. The AR algorithm has shown the best performance for false positive rates, [23].
- Reinforcement learning as already stated in, [24], is also making its way and has some potential.
- Support Vector Machine algorithm was used for fault detection of wind turbines, [25].
- RBF, as part of the Artificial Neural Network (ANN) machine learning model, is usually used in forecasting. The algorithm has proved itself with a magnificent performance while forecasting. The RBF algorithm mostly adopts the radial basis function. As described in the following research, [26].

In the research, [27], the CNN DL method was used to discover if the wind turbine converter was faulty. The researchers are proposing an innovative CNN method that is called "AOC–ResNet50".

Another group of researchers has tried using Machine Learning for anomaly detection of the gearbox of a wind turbine, [28].

Another research endeavor involved implementing a monitoring system that tracks and identifies the state of IoT devices over time by employing a synchronized series of performance metrics, including processor usage, memory utilization, and network interface card activity. This system records synchronized performance metrics for both non-compromised IoT devices and those compromised as a result of well-known cyberattacks within the Internet of Things landscape, [29].

The incorporation of machine learning in cybersecurity within the context of Industry 4.0 presents both significant opportunities and challenges. The utilization of machine learning algorithms in Industry 4.0 systems for threat detection, anomaly identification, and real-time response has shown promise in enhancing cybersecurity measures. However, it is crucial to critically examine the implications and limitations of this approach.

We share the opinion of the authors [30] regarding the "Challenges in the Development of Artificial Intelligence". Developing companies face several challenges when it comes to embracing Industry 4.0 and adopting digital technologies. These challenges include the need to enhance capabilities, integrate existing production systems, improve infrastructure, bridge technology diffusion gaps, and address access disparities. The successful implementation of digital manufacturing and artificial intelligence (AI) can bring significant benefits to various sectors, such as finance, defense, and national security. AI systems have the potential to optimize decision-making, enhance threat detection and identification, and improve target selection and augmented reality applications in tactical systems.

A crucial consideration pertains to the inherent weaknesses that may emerge from machine learning models. Adversarial attacks and data poisoning present substantial threats, undermining the effectiveness of these models and potentially leading to security breaches. As machine learning algorithms become increasingly intricate and self-reliant, there is an urgent need to create robust methods to mitigate these vulnerabilities. Protecting cybersecurity infrastructure requires the adoption of strong measures to maintain reliability and integrity.

Furthermore, the dependence on machine learning algorithms gives rise to issues related to transparency, interpretability, and accountability. The opaqueness of certain machine learning models, often referred to as "black boxes," impedes the comprehension of their decision-making procedures, making it difficult to detect and correct possible biases or mistakes. This absence of interpretability presents difficulties in essential situations where building trust and confidence is necessary.

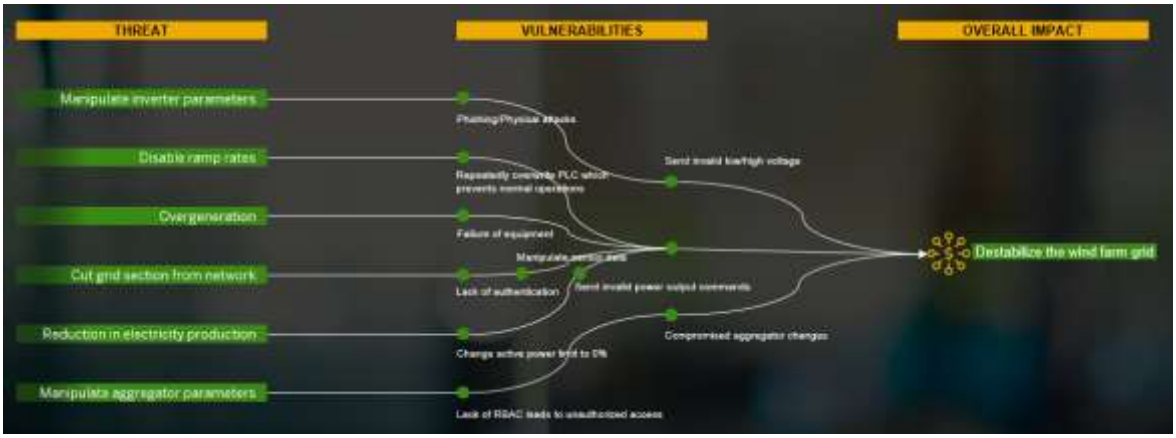

Fig. 4: Wind farm attack tree

Industrial control systems (ICSs) within wind farms are essential for the efficient operation of wind turbines (WTs). Nevertheless, in their pursuit of higher profits, unscrupulous wind power producers frequently employ false data injection (FDI) attacks to undermine the ICSs of competing wind farms. A group of researchers investigates a novel type of profit-oriented FDI (POFDI) attack targeting wind farms and explores corresponding detection mechanisms. The study introduces a unique attack model for POFDI attacks against wind farm ICSs, considering the power distribution within these facilities. Building upon the characteristics of the proposed attack model, an attack strategy focused on maximizing financial gain is discussed, which involves integrating attack detection methods. To identify POFDI attacks, countermeasures are developed based on differentiating the compromised data from normal data, leveraging the relationships between wind turbines in wind farms, [31].

Another research aims to investigate the effects of a stealthy false data injection attack on critical sensor measurements in a wind farm. To achieve this, an analysis framework is developed to model the attack, considering the wind farm's parameters, power generation properties, and attack constraints as inputs. The objective of an adversary in such an attack is to disrupt the balance of power generation within the wind farm and induce system instability. To stay stealthy, the adversary strategically manipulates the power equations and modifies the sensor measurements. The research focuses on quantifying the impact of this attack and proposes countermeasures to mitigate its effects, [32].

During an extensive two-year study, a team of researchers from the University of Tulsa conducted rigorous penetration tests on five distinct wind farms, excluding the one mentioned in this context. This comprehensive investigation aimed to assess the security vulnerabilities present within these wind energy installations. The researchers employed meticulous methodologies and techniques to simulate real-world cyberattacks, allowing them to identify potential weaknesses and evaluate the overall resilience of the wind farm systems. The results of this research offer valuable perspectives on the cybersecurity environment within wind farms and support continuous initiatives aimed at bolstering the safeguarding of vital infrastructure in the renewable energy industry.

The researchers executed three proof-of-concept attacks to illustrate the potential exploits that malicious actors could employ on compromised wind farms. Among the tools they created, Windshark stood out, as it allowed the sending of commands to other turbines within the network, resulting in their disruption or continuous activation of braking systems, leading to wear and potential damage. Another malicious software named Windworm went beyond this, utilizing telnet and FTP protocols to propagate from one programmable automation controller to another, ultimately infecting all the computers within a wind farm's infrastructure. These demonstrations highlight the critical importance of addressing cybersecurity vulnerabilities in the wind energy sector to safeguard against such damaging attacks, [33].

By leveraging various machine learning approaches, such as the Support Vector Machine (SVM) algorithm, it becomes possible to detect potential attacks aimed at damaging wind turbine equipment. The application of SVM in fault detection has proven effective in analyzing turbine data and identifying anomalous patterns indicative of malicious activity. With the utilization of this algorithm, we can proactively identify and mitigate threats posed by threat actors targeting wind turbines. By monitoring and analyzing data from these advanced machine learning models, in the

future, the security posture of wind farms can be enhanced.

The contribution of our research paper on different machine learning algorithms lies in the exploration, evaluation, and comparison of various algorithms in a specific domain or problem context. It aims to provide a comprehensive understanding of the strengths, weaknesses, and applicability of each algorithm. This research paper serves as a valuable resource by consolidating knowledge of different machine learning techniques, their performance metrics, and their suitability for specific tasks.

The subsequent research paper called "Analyzing Attacks on ICS/SCADA Wind Farm Physical Testbed with ML" was built upon this foundation by leveraging the knowledge gained from previous research. It utilizes the insights, findings, and recommendations presented in this paper to inform the selection and application of machine learning algorithms in a new context or problem domain. This subsequent research paper extends the knowledge base by showcasing practical implementations, new experiments, and potential advancements built upon the understanding established in this work.

The contribution of the later research paper lies in its ability to demonstrate the practical utilization and real-world implications of the knowledge gained from the exploration of different machine learning algorithms. It showcases how the understanding of these algorithms can be translated into tangible outcomes. Our physical model is designed with a modular architecture, enabling seamless extensions through the incorporation of additional OT components. This flexible framework ensures adaptability and scalability to accommodate future research contributions from fellow Ph.D. students.

3.4 Generating Datasets for Industry 4.0

Obtaining datasets from industrial facilities poses a significant challenge, making it difficult to conduct a comprehensive analysis of Supervisory Control and Data Acquisition (SCADA) systems. Addressing security concerns in SCADA systems is also hindered by the scarcity of openly accessible datasets that could facilitate research into various types of attacks. Recent years have witnessed a rise in sophisticated attacks. Researchers seeking to develop intrusion detection systems have often relied on available datasets for experimentation.

However, it's worth noting that most datasets, while valuable, lack the authenticity of real-world attack scenarios. Consequently, bridging this gap between simulated and real-world SCADA system attacks remains a crucial challenge in the field of cybersecurity.

3.5 Why Datasets are Limited

Most SCADA systems are under the ownership of either private businesses or governmental entities. However, obtaining authentic real-world data for research purposes is a significant challenge due to the hesitancy of affected companies to share their datasets. Furthermore, since SCADA systems are integral to industrial infrastructure, there are valid concerns about the security risks associated with sharing such sensitive data. As a result, the limited availability of datasets can be primarily attributed to these factors.

3.6 Attacks against SCADA Systems that could have been Prevented

In previous years, SCADA systems were relatively isolated and encountered fewer threats compared to the present. The significant increase in attacks on SCADA systems can be attributed to their current level of interconnectivity, a characteristic that was not as prevalent in the past.

Few of the most recent attack types and their objectives of targeting SCADA systems are shown in the in Table 1 below.

Table. 1 Attack types and their objectives

Attack Type	Objective
Eavesdropping	Authorization, confidentiality
Denial of Service	Availability
Malware	Availability, Integrity
Data Integrity Attacks	Integrity

On May 7, 2021, a cyber assault was initiated against the American colonial gas pipeline, which is responsible for transporting petroleum products from Texas to New York. This pipeline plays a crucial role, supplying approximately 45% of the petroleum consumed on the East Coast. Consequently, this ransomware attack led to the closure of one of their primary pipelines for refined products, [34].

At the beginning of March, Enercon GmbH, a wind turbine manufacturer, experienced a loss of remote connectivity to approximately 5,800 turbines due to a hack on Viasat's satellite network, [35].

As reported by The Wall Street Journal, Deutsche Windtechnik, which had approximately 2,000 turbines under its control compromised during the incident, did indeed suffer a ransomware attack. However, the company successfully restored its

systems without needing to engage with the attackers, [36].

Cyber adversaries can focus on control systems and the essential infrastructure they manage by capitalizing on the interconnections between operational technology (OT) and information technology (IT) networks.

3.6 Crating Own Dataset and Training A Model

Creating your own dataset for research or machine learning purposes can be a valuable endeavor, but it comes with several potential caveats and challenges:

- Time and Resources: Collecting and annotating data can be a time-consuming and resource-intensive process. It may involve hiring personnel, setting up data collection infrastructure, and investing in hardware and software tools.
- Bias and Quality: Ensuring that your dataset is representative and unbiased can be challenging. If the data collection process is not carefully designed, it can introduce biases that affect the performance of machine learning models.
- Privacy Concerns: If your dataset contains personal or sensitive information, you need to be cautious about privacy and legal issues. You may need to comply with data protection regulations and take measures to anonymize or secure the data.
- Data Labeling and Annotation: Manually labeling and annotating data can be error-prone and subject to human biases. It may require domain expertise and careful quality control.
- Scale and Diversity: Depending on your research goals, you may need a large and diverse dataset. Collecting enough data to train robust machine-learning models can be challenging, especially in niche domains.
- Data Maintenance: Data may become outdated over time, and you may need to continually collect and update it to keep your models relevant and accurate.
- Cost: Collecting and maintaining a dataset can be costly, especially if it involves purchasing data or hardware resources.
- Data Storage and Backup: Safeguarding your dataset is essential. Data loss can be a significant setback, so implementing proper storage and backup procedures is crucial.

Below is an illustrative methodology as an example.

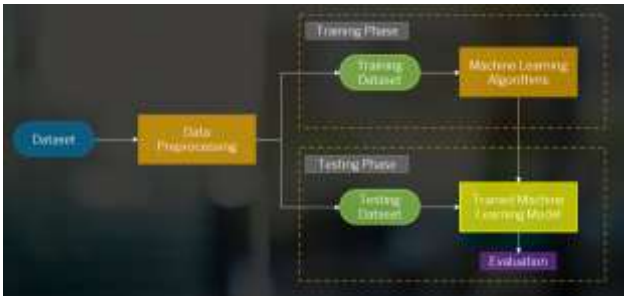

Fig. 5: Graphical Representation of Methodology

4 Challenges of Utilizing AI as Mitigation Controls and How to Defend AI

AI research remains in a constant state of evolution, and despite notable progress, it is not exempt from its own set of flaws. Many challenges have been recognized, and there might be unanticipated hurdles on the horizon. The significance of utilizing top-tier datasets is underscored in AI research. Nonetheless, even with these high-quality datasets, achieving favorable outcomes can prove to be a formidable task for machine learning algorithms. It's crucial to recognize that algorithm performance may not consistently reach its peak potential and may occasionally even deteriorate.

In addition to the previously well-known challenges associated with algorithms, there is another critical threat vector that necessitates attention - attacks targeting AI and ML systems, as well as the potential misuse of AI by adversaries, [37]. These AI attacks differ from conventional cyberattacks, such as ransomware campaigns or DDoS attacks. They can be classified into two types: attacks that exploit inherent architectural limitations within core AI algorithms, which may be unfixable or require significant time and resources to address, and attacks that exploit the training datasets of AI algorithms by injecting falsified data. These unique attack vectors pose distinct challenges and call for specific countermeasures to ensure the security and integrity of AI systems.

To provide evidence supporting the argument regarding the first type of AI attacks that target the algorithm itself, notable examples include instances of algorithm poisoning, [38], have successfully demonstrated a method to introduce a backdoor into a federated learning model. The demonstration raises concerns as threat actors who have control over the data generated by their mobile devices

could manipulate the federated learning algorithm on their devices, thereby exploiting it to their advantage. These findings highlight the vulnerability of AI algorithms to potential manipulation and emphasize the importance of implementing robust defenses to mitigate the risks posed by such attacks.

About the second category of AI attacks, it is pertinent to mention that a group of researchers in, [39], have effectively carried out experiments on data manipulation assaults targeting 26 datasets commonly employed for regression learning. This underscores the importance of considering how database models are constructed and how data flows within these models. Essentially, malicious actors can weaponize the data that is collected or stored when they are aware that it directly contributes to the creation of machine learning datasets. Consequently, it becomes crucial to implement robust protective measures and employ stringent data validation techniques to safeguard the integrity and security of machine learning datasets.

An avenue for prospective research lies in building upon the ongoing investigations conducted by Mina Todorova regarding the development of protective shells. This research direction suggests the potential necessity of thoroughly cleansing all data utilized in datasets to eliminate any malicious insertions. Furthermore, it proposes the implementation of a protective shell encompassing the entire training process of machine learning algorithms. This approach aims to ensure the preservation of data integrity and mitigate potential security risks, [40].

Furthermore, active research is being conducted in the realm of enhancing AI robustness against deception, [41]. The research team is focused on establishing the fundamental principles required for identifying and addressing broader categories of system vulnerabilities, as well as developing strategies to mitigate the potential exploitation of these vulnerabilities.

5 Future Contributions

At present, there is a dearth of research on wind energy systems and their potential application of artificial intelligence for safeguarding against intricate cybersecurity threats. Drawing from the deep learning insights presented in this paper, we will identify the most appropriate deep learning algorithms to establish a basic framework aimed at enhancing the security of wind turbines through deep learning techniques. We have already established the groundwork for our forthcoming test environment, with the intent to assess three distinct types of attacks and effectively mitigate them through the application of the most appropriate machine learning algorithms. The outcomes of this experiment are detailed in a separate research named "Analyzing Attacks on ICS/SCADA Wind Farm Physical Testbed with ML". The research was successfully presented at the International Conference on Electronics, Engineering Physics, and Earth Science on June 21-23, 2023 in Kavala, Greece.

As we delve into the practical implementation of ML models on our physical testbed, it becomes evident that there exists potential for future research in this domain. To further advance the field of cybersecurity for wind turbines, we propose other research to explore the following directions:

- Adversarial ML Defense: Investigate and develop robust defenses against adversarial attacks on ML models deployed in wind turbine cybersecurity, ensuring their resilience in dynamic threat landscapes.
- Explainability and Interpretability: Enhance the transparency of ML models to improve the understanding of their decision-making processes, facilitating better integration with existing wind turbine control systems and enabling effective human oversight.
- Real-time Threat Detection: Explore the integration of real-time monitoring and anomaly detection using ML algorithms to swiftly identify and respond to evolving cyber threats, minimizing potential damage to wind turbine operations.
- Edge Computing for Security: Investigate the feasibility of deploying ML models on edge devices within wind turbines to enhance local cybersecurity measures, reducing dependency on external networks and improving overall system resilience.
- Collaborative Security Frameworks: Develop collaborative and federated learning approaches that enable different wind turbines to share threat intelligence without compromising sensitive data, fostering a collective defense against emerging cyber threats.

6 Conclusion

The digital transformation of cybersecurity, empowered by artificial intelligence, reshapes the landscape of defensive strategies and organizational structures. It paves the way for novel cybersecurity models, innovative approaches to revenue

generation, expanded protection for a wider consumer base, and elevated levels of threat detection and response. Embracing artificial intelligence within the cybersecurity domain blurs traditional sector boundaries and gives rise to dynamic digital platforms.

Implementing a diverse range of digital technologies in practical cybersecurity applications yields exceptional results, including enhanced threat intelligence, proactive risk mitigation, robust defense mechanisms, optimal resource allocation, and resilient cyber operations. By harnessing the power of artificial intelligence, cybersecurity in Industry 4.0 attains unparalleled efficiency, adaptive protection, and secure digital ecosystems.

Integrating Secure-by-Default principles with the concept of least privilege, wherein users are granted access only to the essentials needed for their roles, stands as a critical aspect in enhancing the authorization process. Enhancing resistance against potential exploits resulting from end-user compromise contributes to a decrease in the occurrence of successful incidents affecting operational technology (OT).

Looking ahead to Industry 5.0, the fusion of virtual and real-world cybersecurity through artificial intelligence-driven approaches holds tremendous promise. It not only safeguards critical assets but also enables the exploration of emerging threats, the development of advanced defense mechanisms, and the attainment of cyber resilience in an ever-evolving digital landscape. As highlighted in our research, the integration of AI as a crucial component in the implementation of advanced security controls within Industry 4.0 holds tremendous potential. However, it also introduces significant risks that must be carefully managed to maintain the overall security posture of Industry 4.0 systems. Also, adhering to data privacy and industry-specific regulations is crucial. Ensuring that AI implementations comply with these regulations can be complex and resource-intensive.

In our ongoing and future research, we aim to delve deeper into the examination of specific AI methods that can be utilized to safeguard wind turbines from potential cyberattacks. Wind turbines play a vital role in renewable energy generation and ensuring their resilience against cyber threats is of paramount importance.

To achieve this objective, we will explore various AI-based approaches, such as anomaly detection algorithms, machine learning models, and deep neural networks. These techniques can be employed to analyze the data generated by wind turbines, including operational parameters, performance metrics, and network traffic, to identify any abnormal or malicious activities.

In conclusion, our research endeavors are dedicated to an exhaustive exploration of AI methodologies, specifically leveraging machine learning (ML) models, to bolster the cybersecurity defenses of wind turbines. We intend to subject ML models to thorough examination and testing, utilizing a dataset gathered from our physical testbed. This empirical approach ensures the relevance and efficacy of our proposed solutions, contributing valuable insights to the development of robust and adaptive cybersecurity measures. By bridging the gap between theory and practical implementation, our research strives to fortify wind turbines against cyber attacks.

References:

[1] E. Sabev, G. Pavlova, R. Trifonov, K. Raynova and G. Tsochev, "Analysis of practical cyberattack scenarios for wind farm SCADA systems," *2021 International Conference Automatics and Informatics (ICAI)*, 2021, pp. 420-424, doi: 10.1109/ICAI52893.2021.9639550.

[2] Radanliev P., De Roure D., Walton R., Van Kleek M., Mantilla Montalvo R., La'Treall Maddox, Santos O., Burnap P. & Anthi E.. Artificial intelligence and machine learning in dynamic cyber risk analytics at the edge. SN Appl. Sci. 2, 1773 (2020). https://doi.org/10.1007/s42452-020-03559-4.

[3] Poremba, S. (2021) *Data Poisoning: When Attackers Turn AI and ML Against You,* [Online]. https://securityintelligence.com/articles/data-poisoning-ai-and-machine-learning/ (Accessed Date: October 10, 2021).

[4] Harri Ruoslahti, Bríd Davis, "Societal Impacts of Cyber Security Assets of Project ECHO," *WSEAS Transactions on Environment and Development*, vol. 17, pp. 1274-1283, 2021, http://dx.doi.org/10.37394/232015.2021.17.116

[5] Sahli Nabila, Benmohammed Mohamed, Artificial Intelligence and Cyber Security: Protecting and Maintaining Industry 4.0 Power Networks, [Online]. http://ceur-ws.org/Vol-2748/IAM2020_paper_35.pdf (Accessed Date: December 5, 2023).

[6] Annoni, A., Benczur, P., Bertoldi, P., Delipetrev, B., De Prato, G., Feijoo, C.,

Fernandez Macias, E., Gomez Gutierrez, E., Iglesias Portela, M., Junklewitz, H., Lopez Cobo, M., Martens, B., Figueiredo Do Nascimento, S., Nativi, S., Polvora, A., Sanchez Martin, J.I., Tolan, S., Tuomi, I. and Vesnic Alujevic, L., Artificial Intelligence: A European Perspective, Craglia, M. editor(s), EUR 29425 EN, *Publications Office of the European Union*, Luxembourg, 2018, ISBN 978-92-79-97217-1, doi:10.2760/11251, JRC113826.

[7] CEPS Task Force Report (2021) *Artificial Intelligence and Cybersecurity, Technology, Governance and Policy Challenges,* [Online]. https://www.ceps.eu/wp-content/uploads/2021/05/CEPS-TFR-Artificial-Intelligence-and-Cybersecurity.pdf (Accessed Date: December 15, 2021).

[8] European Union Agency for Cybersecurity (ENISA) (2021) *Securing Machine Learning Algorithms,* [Online]. https://www.enisa.europa.eu/publications/securing-machine-learning-algorithms (Accessed Date: January 20, 2022).

[9] Barrett, Matt., Marron, Jeff., Yan Pillitteri, Victoria., Boyens, Jon., Witte, Greg., and Feldman, Larry, "Draft NISTIR 8170, The Cybersecurity Framework: Implementation Guidance for Federal Agencies," Maryland, 2017.

[10] NIST (2018) *Cybersecurity Framework Version 1.1,* [Online]. https://www.nist.gov/news-events/news/2018/04/nist-releases-version-11-its-popular-cybersecurity-framework (Accessed Date: June 17, 2018).

[11] NIST (2021) *NIST Requests Information to Help Develop an AI Risk Management Framework,* [Online]. https://www.nist.gov/news-events/news/2021/07/nist-requests-information-help-develop-ai-risk-management-framework (Accessed Date: August 1, 2021).

[12] IBM (2021) S*upervised learning,* [Online]. https://www.ibm.com/cloud/learn/supervised-learning#toc-what-is-su-d3nKa9tk (Accessed Date: April 10, 2023).

[13] IBM (2021) *Unsupervised learning,* [Online]. https://www.ibm.com/cloud/learn/supervised-learning#toc-unsupervis-Fo3jDcmY (Accessed Date: April 10, 2023).

[14] IBM (2021) *Semi-supervised learning,* [Online]. https://www.ibm.com/cloud/learn/supervised-learning#toc-unsupervis-Fo3jDcmY (Accessed Date: April 10, 2023).

[15] IBM (2021) *What are recurrent neural networks,* [Online]. https://www.ibm.com/cloud/learn/recurrent-neural-networks (Accessed Date: April 10, 2023).

[16] ScienceDirect (2023) *Deep Belief Network,* [Online]. https://www.sciencedirect.com/topics/engineering/deep-belief-network (Accessed Date: April 20, 2023).

[17] ScienceDirect (2023) *Deep Boltzmann Machine,* [Online]. https://www.sciencedirect.com/topics/engineering/deep-boltzmann-machine (Accessed Date: April 20, 2023).

[18] Derlat, A. (2017) *Applied Deep Learning - Part 3: Autoencoders,* [Online]. https://towardsdatascience.com/applied-deep-learning-part-3-autoencoders-1c083af4d798 (Accessed Date: February 10, 2023).

[19] Brownlee, J. (2019) *A Gentle Introduction to Generative Adversarial Networks (GANs),* [Online]. https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/ (Accessed Date: March 15, 2023).

[20] Understanding LSTM Networks (2015), [Online]. https://colah.github.io/posts/2015-08-Understanding-LSTMs/ (Accessed Date: March 15, 2023).

[21] Jordan, Ian D et al. "Gated Recurrent Units Viewed Through the Lens of Continuous Time Dynamical Systems." Frontiers in *Computational Neuroscience*, vol. 15 678158. 22 Jul. 2021, doi:10.3389/fncom.2021.678158.

[22] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, Blaise Agüera y Arcas: Communication-Efficient Learning of Deep Networks from Decentralized Data. AISTATS 2017: 1273-1282.

[23] D. L. Marino, C. S. Wickramasinghe, V. K. Singh, J. Gentle, C. Rieger, and M. Manic, "The Virtualized Cyber-Physical Testbed for Machine Learning Anomaly Detection: A Wind Powered Grid Case Study," in *IEEE Access*, vol. 9, pp. 159475-159494, 2021, doi: 10.1109/ACCESS.2021.3127169.

[24] Graf, P. (2019) *Innovative Optimization and Control Methods for Highly Distributed Autonomous Systems,* [Online].

https://www.nrel.gov/docs/fy19osti/74798.pdf (Accessed Date: February 20, 2023).

[25] Nassim Laouti, Nida Sheibat-Othman, Sami Othman, Support Vector Machines for Fault Detection in Wind Turbines, *IFAC Proceedings Volumes*, Vol. 44, Issue 1, 2011, pp. 7067-7072, https://doi.org/10.3182/20110828-6-IT-1002.02560.

[26] Farhad Elyasichamazkoti, Abolhasan Khajehpoor, Application of machine learning for wind energy from design to energy-Water nexus: A Survey, *Energy Nexus*, Vol. 2, 2021, pp. 100011, https://doi.org/10.1016/j.nexus.2021.100011.

[27] Xiao C., Liu Z., Zhang T., Zhang X.. Deep Learning Method for Fault Detection of Wind Turbine Converter. *Appl. Sci*. 2021, vol.11, 1280. https://doi.org/10.3390/app11031280.

[28] Rashid, H.; Batunlu, C. Anomaly Detection of Wind Turbine Gearbox based on SCADA Temperature Data using Machine Learning. *Preprints* 2021, 2021010356, doi: 10.20944/preprints202101.0356.v1.

[29] A. Hristov and R. Trifonov, "A Model for Identification of Compromised Devices as a Result of Cyberattack on IoT Devices," *2021 International Conference on Information Technologies (InfoTech)*, Varna, Bulgaria, 2021, pp.1-4, doi: 10.1109/InfoTech52438.2021.9548556.

[30] Kateryna Kraus, Nataliia Kraus, Mariia Hryhorkiv, Ihor Kuzmuk, Olena Shtepa, "Artificial Intelligence in Established of Industry 4.0," *WSEAS Transactions on Business and Economics*, vol. 19, pp. 1884-1900, 2022, https://doi.org/10.37394/23207.2022.19.170.

[31] W. Bi, G. Chen and K. Zhang, "Profit-Oriented False Data Injection Attack Against Wind Farms and Countermeasures," in *IEEE Systems Journal*, vol. 16, no. 3, pp. 3700-3710, Sept. 2022, doi: 10.1109/JSYST.2021.3107910.

[32] Amarjit Datta and Mohammad Ashiqur Rahman. 2017. Cyber Threat Analysis Framework for the Wind Energy Based Power System. In *Proceedings of the 2017 Workshop on Cyber-Physical Systems Security and Privacy (CPS '17)*. Association for Computing Machinery, New York, NY, USA, 81–92. https://doi.org/10.1145/3140241.3140247.

[33] Greenberg, A. (2017) *Researchers Found They Could Hack Entire Wind Farms,* [Online]. https://www.wired.com/story/wind-turbine-hack/ (Accessed Date: January 30, 2020).

[34] CISA (2021) *The Attack on Colonial Pipeline: What We've Learned & What We've Done Over the Past Two Years,* [Online]. https://www.cisa.gov/news-events/news/attack-colonial-pipeline-what-weve-learned-what-weve-done-over-past-two-years (Accessed Date: May 25, 2023).

[35] Viasat (2022) *KA-SAT Network cyber attack overview,* [Online]. https://news.viasat.com/blog/corporate/ka-sat-network-cyber-attack-overview (Accessed Date: May 20, 2023).

[36] Securityweek (2022) *German Wind Turbine Firm Hit by 'Targeted, Professional Cyberattack',* [Online]. https://www.securityweek.com/german-wind-turbine-firm-discloses-targeted-professional-cyberattack/ (Accessed Date: April 30, 2023).

[37] Comiter, M. (2019) *Attacking Artificial Intelligence: AI's Security Vulnerability and What Policymakers Can Do About It,* [Online]. https://www.belfercenter.org/publication/AttackingAI (Accessed Date: January 10, 2023).

[38] Eugene Bagdasaryan, Andreas Veit, Yiqing Hua, Deborah Estrin, *Vitaly Shmatikov Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, PMLR 108:2938-2948, 2020.

[39] Müller, Nicolas & Kowatsch, Daniel & Böttinger, Konstantin. (2020). Data Poisoning Attacks on Regression Learning and Corresponding Defenses, Publication at PRDC2020, IEEE, https://doi.org/10.48550/arXiv.2009.07008.

[40] Todorova, M. (2022) *Research and Evaluation of a "Protective Shell" for two Industrial Cyber Physical System (CPS) Incidents,* [Online]. https://tu-dresden.de/ing/informatik/smt/st/die-professur/mitarbeiter/?person=4&embedding_id=1e8badf421c649b89f65fb84a05ecbb0&set_language=de (Accessed Date: September 20, 2022).

[41] DARPA GARD program (2021) *Holistic Evaluation of Adversarial Defenses,* [Online]. https://www.gardproject.org/#gard (Accessed Date: March 20, 2023).

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

Roumen Trifonov and Evgeni Sabev carried out the conceptualization - the introduction, the future contribution, and the conclusion.
Galya Pavlova and Kamelia Raynova have investigated machine learning/ deep learning standardization.
Evgeni Sabev has analyzed the application of machine learning/ deep learning for cybersecurity in Industry 4.0.
Evgeni Sabev, Galya Pavlova, and Kamelia Raynova have researched the challenges of utilizing AI as mitigation controls and how to defend AI.
Roumen Trifonov was responsible for the project administration and funding acquisition.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

The presented research is funded by National Science Fund under the Ministry of Education and Science in Bulgaria with contract КП-06-Н 47/7 for the scientific-research project "Possibility Investigation of Increasing the Cybersecurity of the Systems in Industry 4.0 using Artificial Intelligence".

Conflict of Interest

The authors have no conflicts of interest to declare.