Critical Infrastructure Security and the Role of AI: An Overview

Siphesihle Sithungu[1] and Christoph Lipps[2]
[1]University of Johannesburg, Johannesburg, South Africa
[2]German Research Center for Artificial Intelligence, Kaiserslautern, Germany

siphesihles@uj.ac.za
christoph.lipps@dfki.de

Abstract: Critical Information Infrastructures (CIIs) are an increasingly important focus area of industrial automation, particularly regarding the current developments towards Industry 5.0 and the industrial metaverse. Critical Information Infrastructure Protection (CIIP) is one of the fastest growing areas of cyber security primarily due to the expectations of both large companies and governments to protect their critical infrastructure in the interest of economic stability and citizen security. The critical infrastructures themselves are becoming increasingly automated due to the increasing availability and lower cost of Artificial Intelligence (AI) methods for downstream tasks such as predictive maintenance, load forecasting and anomaly detection. AI methods can also be used to protect critical information infrastructures, for example by implementing sophisticated algorithms for threat modelling and intrusion detection. The focus of this work is on the latter: The applications (and potential) of AI to secure CIIs in the presence of increasing amounts of cyberattacks. It is becoming ever more important to understand the current state of the art in using AI to protect CII. For example, it is imperative to understand the general capabilities of AI for downstream tasks such as intrusion detection and investigate the potential capabilities of AI for upstream tasks such as self-supervised learning, representation learning and generative modelling specifically for cybersecurity.

Keywords: Critical information infrastructure protection, Artificial Intelligence, Representation learning, Generative modelling

1. Introduction

Critical infrastructures (CIs), as “facilities of major importance for society whose failure or impairment would cause significant disruptions to public order, safety and security” (German Federal Office for Information Security), provide services upon which modern society depend. If these services are disrupted over a long period of time or across a significant territory, deaths (Lopez, et al., 2012) or serious economic disturbance will result (Wei, et al., 2022). In this regard, there are three main factors considered to be of primary importance when classifying an infrastructure as critical: (3) the symbolic significance of the infrastructure for a nation; (2) the amount of human dependence on the infrastructure; and (4) the interconnectedness of the infrastructure with other infrastructures (Herrera & Maennel, 2019).

As technology advances, traditional CI is becoming increasingly dependent on Information and Communication Technology (ICT) infrastructure, leading to the concept of Critical Information Infrastructure (CII) (Mbanaso & Kulugh, 2021; Herrera & Maennel, 2019). Among the reasons why protecting CII is difficult and challenging are the extremely different time scales in which the physical and virtual components must be viewed. The negative impact of a compromise of CII (physical element) usually has to be measured in decades, while the recovery time of an industrial control system (virtual element) after a cyber-attack must be in the millisecond range (Lopez, et al., 2012).

The evolution of modern technologies -towards the hyperconnectivity of tomorrow- (such as Industrial Internet of Things (IIoT), Beyond 5G (B5G) & Sixth Generation (6G) cellular networks, cloud-, fog- and edge computing, as well as industrial control systems and smart grids) will further broaden the attack surface for CII systems. Their dependency on ICT (Mbanaso & Kulugh, 2021) results in an associated increase in the need for ever more sophisticated protection mechanisms and resilience strategies (Lipps, et al., 2022).

One technology set to have a significant impact on this development is the methods of Artificial Intelligence (AI). Even though these have been the subject of intensive research for more than two decades, the technology is undergoing a significant acceleration in its development due to the availability of (inexpensive) computing power (Zhang, et al., 2022). However, although the majority of AI research is conducted in the field of Machine Learning (ML), AI methods can generally be understood as the study of various techniques for building intelligent systems that can learn without explicit programming (Russell & Norvig, 2020). The growing understanding of AI's capabilities has also led to efforts to use AI to defend against cyber threats, among other things. For example, AI models are used to classify cyber threats due to their ability to capture non-linear patterns in data (Li, 2018).

To take this current development into account, the aim of this work is to examine the role of AI specifically in Critical Information Infrastructure Protection (CIIP). Thereby, the focus is on examining attack vectors in CII,

threat mitigation techniques using AI, as well as State-of-the-Art AI approaches. The rest of the work is organized as follows: Section 2 provides a background on attack vectors commonly encountered in cyberattacks on CII. Section 3 addresses the role AI may have in defending against cyber threats to CII, whereas Section 4 examines current AI approaches to defending against CII threats. Finally, Section 5 provides insights from research, suggestions for CIIP practitioners, and possible future research directions.

2. Attack Vectors in Critical Information Infrastructures

Attacks on CII are usually aimed at the Operational Technology (OT) of the infrastructure as well as to control systems (Lehto, 2022; Makrakis, et al., 2021; Tsantikidou & Sklavos, 2024). Thereby attackers are typically interested in the physical connected assets supporting industrial processes than in stealing data. For example, 54% of US CI suppliers have reported attack attempts on their control systems while 40% report attempts to shut down systems (Allianz, 2016).

2.1 Human Vulnerabilities

Attack vectors primarily based on human vulnerabilities are one of the most common weaknesses in CII security. Over 70% of data breaches are a result of human vulnerabilities and most technology professionals consider human error a significant threat to control systems (Shakeel, 2023). The most common types of human vulnerabilities are: i) insider threats and ii) social engineering.

An insider threat comes about when someone (i.e. employee, service provider or partner) with privilege access to sensitive enterprise resources or system intentionally or unintentionally misuses the resource in a way that is detrimental to the organisation. Naturally, this makes insider threats unpredictable and difficult to counteract using traditional software security measures. As such, organisations typically do not know how to avoid or handle insider threats (Gaidarski & Minchev, 2021).

Human vulnerability in CII scenarios can also be exploited through social engineering efforts. Social engineering is achieved through social interactions initiated with personnel within an organisation by malicious actors (Ghafir, et al., 2018). Social engineering is a systematic approach that follows the steps show in Figure 1.

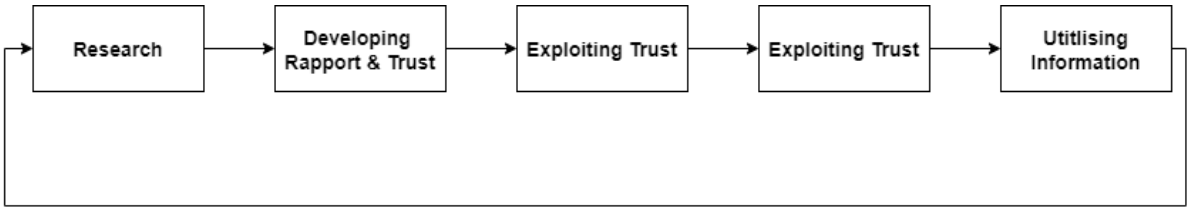

Figure 1: The main steps involved in a social engineering attack (Mitnick & Simon, 2003)

The research stage involves gathering information about the target, which is used in the second stage to gain their trust by developing a rapport. Thereafter, the malicious exploits the target's trust to obtain the desired information through manipulation or instruction. Finally, the malicious actor uses the information obtained to execute an attack on the CI or organisation. The strategies generally employed by social engineers are psychological manipulation, obedience to authority and exploiting naivety (Ghafir, et al., 2018).

2.2 Malware and Ransomware

Malware exploits target vulnerabilities in software systems for the purpose of allowing the attacker unauthorized access. Once an attacker has infiltrated the target system or network, they launch the intended cyberattack. Software vulnerabilities can exist at any layer of a CII system. For instance, database software can contain vulnerabilities, which could be exploited using specialised malware (Tsegaye & Flowerday, 2014). It is common for industrial control systems to contain a significant number of security vulnerabilities due to such systems not being patched timeously. Patches on ICS devices are usually deferred due to uptime being a priority in CIs (Wang, et al., 2017).

There have been several cases of malware being used as an attack vector on CIIs. Some of the notable examples are BlackEnergy3, Trisis, Crashoverride (Geiger, et al., 2020), Stuxnet (Baezner & Robin, 2017), Industroyer (Kozak, et al., 2023) and Triton (Altaleb & Rajnai, 2023). Malware attacks on CIIs are also in the form of ransomware. Notable ransomware attacks on CIIs are NotPetya (Bederna, et al., 2020), DarkSide (Beerman, et al., 2023) and WannaCry (Musluoglu, et al., 2024).

2.3 Unpatched Vulnerabilities

Industrial devices run real-time operating systems that are designed to provide deterministic performance. Such devices are designed to operate 24/7 under various external conditions. As such, they are only replaced after years of continuous operation due to constantly interacting with the physical environment. Therefore, it is very common for the OS running on such devices to remain unpatched, even if patches are available (Makrakis, et al., 2021). A malicious actor can exploit known vulnerabilities that have now been patched on operational technology (OT) infrastructure. For example, once delivered through other means (i.e. phishing), the WannaCry ransomware attack vector takes advantage unpatched vulnerabilities in the target's local network to spread quickly (Di Pietro, et al., 2021).

2.4 Supply Chain Attacks

Another attack vector in CII systems takes place through the intentional compromise of information technology (IT) supply chains through the exploitation of existing vulnerabilities (Lehto, 2022). Vulnerabilities can also be deliberately introduced into devices before they are shipped to the customer or when during firmware updates. Such attacks are challenging to identify due to their nature where in some instances, an attack is designed to autonomously launch at a specific stage or cause physical damage over an extended period (Duman, et al., 2019). Other supply chain attacks intend to cause interoperability issues between systems (Lehto, 2022).

It is especially difficult to mitigate supply chain attacks in CIs due to their sometimes geographically dispersed nature, which makes it difficult to ensure the authenticity of critical artefacts. Different types of supply chain attacks include stealing Intellectual Property (IP), design specifications or data, malicious substitution, design or specification alteration, development or programming alteration, malicious insertion and tampering or configuration manipulation (Eggers, 2021).

2.5 Distributed Denial of Service Attacks

The aim of a Distributed Denial of Service (DDoS) attack is to compromise the availability of a critical system by overwhelming it with requests to the point that legitimate clients receive the intended service. As such, the intention of a DDoS is to cause an outage of the CI (Hurst, et al., 2015), which could potentially have cascading effects on other CIs. For example, it is possible to exploit existing vulnerabilities in end-user IoT devices and create a botnet that could be used to launch a DDoS attack against a critical service. IoT devices are particularly attractive to malicious actors due to their low-cost and generally insecure design (Stellios, et al., 2018).

2.6 Zero-Day Exploits

A zero-day exploit compromises a vulnerability that has not yet been discovered and publicly reported in a system (Ibor, 2017). Zero-day exploits are a serious concern in CIIs due to the scale of the possible damage if carried out successfully. For example, zero-day exploits in CIIs could enable espionage or cyber-sabotage against a nation. In addition, zero-day exploits could be intentionally or unintentionally sold to extremist groups (Wilson, 2014). Traditional systems keep track of signatures of attacks to identify them, and zero-day attacks have no existing signatures, making it difficult or even impossible to identify them using traditional systems (Singh, et al., 2019). An illustration of the lifecycle of a zero-day vulnerability is shown in Figure 2.

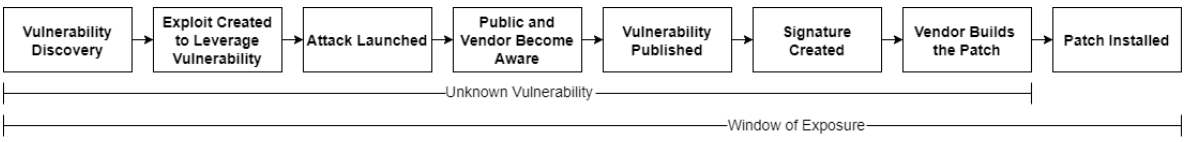

Figure 2: Zero-Day Vulnerability Life Cycle (Singh, et al., 2019)

As depicted in Figure 2, a system remains exposed to an attack from the moment a vulnerability is discovered by an attacker, which is before the vendor (or public community) is aware of it. This remains the case until the system is patched, which typically happens after an attack has concluded.

CIIs are prone to vulnerabilities due to the ongoing technological improvements to introduce efficiency and provide services to larger amounts of people. Therefore, there is a need for different approaches to ensure CIIP from various forms of attack vectors. The following section discusses the role of AI in mitigating threats on CIIs.

3. The Role of Artificial Intelligence in Threat Mitigation

AI methods are increasingly being used successfully in various areas, including cybersecurity. This Section focuses on the CII threats discussed in Section 2 and examines how AI methods can be used to mitigate these threats.

As mentioned above, the identified threats occur most frequently in CII, meaning that an understanding of the potential of AI to mitigate them may be beneficial for CIIP.

3.1 Mitigating Human Vulnerabilities through Pattern Recognition and Behavioural Biometrics

The key method used to exploit human vulnerabilities is through access to sensitive information. When a person's login credentials are used to access a sensitive resource on a network, the most logical way to detect this access is by identifying anomalies in usage behaviour. AI techniques can be used to formulate employee usage signatures over time. People tend to operate technological devices in unique and consistent ways, which makes it possible to learn and form representations of such behaviours. There are several approaches that can be used to encode usage patterns as sequences that can then be stored as signatures. For example, learning automata have previously been used in cloud computing to predict resource usage patterns based on historical data (Rahmanian, et al., 2018).

A similar approach can be used where historical individual usage patterns in a network are used to predict how individuals will use a certain resource. Significant deviations from learned usage patterns can be flagged as anomalies. Advances in the field of behavioural biometrics can benefit this domain to model expected user behaviour based on analysing resource usage over time. For example, Siamese neural networks have been applied to distinguishing legitimate users from intruders in Cis by converting keystroke data images more suitable for training the neural networks (Budžys, et al., 2023).

3.2 Detecting Malware and Ransomware through Machine Learning

The use of AI approaches for detecting malware in CIIs is an evolving area of research. The increasing availability of malware datasets are making it possible to benchmark different malware detection approaches. Some of the notable malware datasets are Critical Infrastructure Ransomware Attacks (CIRA) (Rege & Bleiman, 2023), SOREL-20M (Harang & Rudd, 2020), BODMAS (Yang, et al., 2021), CIC-MalMem-2022 (Carrier, et al., 2022) and Windows PE Malware (Catak & Yazı, 2021).

The use of deep learning with Control-Flow Graphs (CFGs) is also as a promising mechanism for identifying malware in CIIs. The advantage of analysing a CFGs for malware is that a CFG exposes every flow of execution in a program, which makes CFGs interesting tools for understanding the characteristics of executable files (Esmaeili, et al., 2024). Because a CFG is a graph, it is not ideal for training traditional ML and deep learning techniques. However, the emergence of Graph Neural Networks (GNNs) (Wu, et al., 2021) makes it possible to analyse graph-based examples.

3.3 Automated Vulnerability Assessment to Mitigate Unpatched Vulnerabilities

In the context of vulnerability management in CIIs, AI can be used to automate key processes, such as intelligence gathering, threat hunting and vulnerability assessments. Patch management in CIIs can be automated using specialised AI-enabled scanners and analysis tools to discover and categorise vulnerabilities based on severity. Such tools can also recommend appropriate patches or actions to eliminate vulnerabilities. Categorising vulnerabilities based on severity, can also help CII practitioners prioritize specific patches to minimise downtime (Sarker, 2024).

It is important to note that vulnerability management – like many other cybersecurity measures – cannot be achieved solely through automation. Automating vulnerability management is likely to succeed when the AI systems act as assistive technology to enhance human efforts. The key contribution of AI in vulnerability management is in ensuring accurate data handling, improving the speed and reliability of processes and reducing human errors (Saadallah, et al., 2024).

3.4 Enhancing Supply Chain Security through Predictive Analytics and Blockchain

One of the ways in which AI can benefit CII supply chain security is through predictive analytics, particularly monitoring the behaviour and functionality of third-party software and physical devices. This means that AI can be used to look for anomalous or malicious behaviour in third-party components (Sarker, 2024). Blockchain is another emerging candidate solution for supply chain security (Govea, et al., 2024; Marjanović, et al., 2021; Mylrea & Gourisetti, 2018) and in some cases, it has been recommended for use in conjunction with AI approaches (Sarker, 2024; Charles, et al., 2023).

3.5 Early Detection of DDoS Attacks using Machine Learning

Anomaly detection through AI approaches is also relevant in the context of detecting DDoS attacks on CIIs. For example, ML algorithms can be trained to differentiate between normal and anomalous traffic patterns. Additionally, there is the potential to implement automated responses to such attacks (Sarker, 2024). Some example datasets containing DDoS examples for training ML models are WUSTL-IIoT-2021 (Zolanvari, et al., 2021) and NSL-KDD (Tavallaee, et al., 2009). Research on AI-based DDoS detection in CIIs has been ongoing with various algorithms having been proposed. Examples of proposed algorithms have been support vector machines (Panagiotis, et al., 2021), random forests (Santos, et al., 2020), convolutional neural networks (Hussain, et al., 2021; Saheed, et al., 2023) and long-short term memory networks (Saheed, et al., 2023).

3.6 Mitigating Zero-Day Exploits through Anomaly Detection

As mentioned in Section 2.6, zero-day exploits cannot be pre-empted due to them exploiting unknown vulnerabilities. AI has the potential to proactively mitigate zero-day exploits through behaviour-based anomaly detection. Once a model has learnt the baseline behaviour of a system or device, it can be used to detect anomalous behavioural patterns that could indicate compromise. Moreover, the ability to automate threat hunting and intelligence gathering can further minimise the chances of zero-day exploits. Finally, AI presents an opportunity to build models that can continuously update themselves in response to emerging threats (Sarker, 2024).

4. State of the Art AI Approaches in Threat Mitigation

The purpose of this section is to explore state of the art AI approaches that have significantly informed the direction of modern AI research. Specifically, the section focuses on three cutting-edge domains in AI research: generative AI, representation learning and self-supervised learning.

4.1 Generative Artificial Intelligence

Generative AI, which has made significant breakthroughs in the past 10 years, especially since the invention of variational autoencoders (VAEs) (Kingma & Welling, 2022) and Generative Adversarial Networks (GANs) (Goodfellow, et al., 2014), has the potential to enable more innovative and sophisticated CIIP (Yigit, et al., 2024). The ability of generative models to model data generating distributions presents several opportunities for CIIP. The most straightforward application of generative models in CIIP is generating synthetic data to alleviate class imbalance in cybersecurity datasets (Pinto, et al., 2023) or generating unexplored failure scenarios in predictive maintenance (Ćelić, et al., 2024).

Several state-of-the-art generative models have been proposed in various CIIP contexts. For example, large language models can be fine-tuned to handle nuances specific to CIIP using relevant datasets and thereafter used to identify and predict CII threats (Yigit, et al., 2024). Models such as GANs and VAEs can be used to generate synthetic attack samples in order to train discriminative models for attack detection (Shahriar, et al., 2021; Liu, et al., 2022). Artificial immune systems can also be used to generate synthetic data in order to train an intrusion detector. For example, GAAINet is an artificial immune network capable of generating synthetic attack samples for industrial IoT (IIoT) intrusion detection (Sithungu & Ehlers, 2022).

4.2 Representation Learning

Representation learning is a concept that is at the core of many generative models, and it is based on the idea of generating low-dimensional representations of high-dimensional data samples. For example, a VAE uses an encoder artificial neural network (ANN) to map data sample into a lower-dimensional latent or embedding space treated as a Gaussian distribution by estimating the mean and variance parameters of the distribution.

The above property of a VAE means that the encoder network can be used to learn low-dimensional representations of normal behavioural patterns or CII software, physical devices or network traffic. The learned latent representations can then be used to establish baseline behaviour for example, which when compared to anomalous behaviour will result in a significant reconstruction error (Pinto, et al., 2023).

4.3 Self-Supervised Learning

CIIs generate a significant amount of real-time data, which can be used for training ML models. However, the large number of data samples can often result in huge costs being incurred to label each sample (Wang & Shang, 2014). Self-supervised learning is an emerging technique for training models that can infer labels from data point, a concept referred to as data-drive labelling (Yellapragada, et al., 2022). Self-supervised learning can be

reconstructive (Chen, et al., 2024) or contrastive (Liu, et al., 2023). Contrastive self-supervised learning is succinctly known as contrastive learning.

Several works have already proposed the use of self-supervised learning techniques to enhance CIIP. For instance, time-series self-supervised learning to detect cyber-physical system attacks in water distribution infrastructure (Mahmoud, et al., 2022). In another work, self-supervised learning was used to implement a privacy-preserving intrusion detection system for 5G-V2X networks (Hossain, et al., 2025). A self-supervised learning model with transferrable techniques has also been proposed for intrusion detection in 5G industrial networks (Kim, et al., 2024). In order to conceptualise the main attack vectors in CII, their potential impact and possible mitigation strategies, Table 1 summarises the points made in Sections 2, 3 and 4.

Table 1: A summary of attack vectors, impact and mitigation strategies

Threat	Probability of Occurrence	Impact	Mitigation Strategy
Human Vulnerabilities	High	Reputational damage Data breaches	Behavioural Biometrics Representation Learning Pattern Recognition
Malware and Ransomware	High	Financial extortion Operational shutdowns	Machine Learning Graph-Based Approaches Generative AI Representation Learning
Unpatched Vulnerabilities	High	Regulatory penalties System compromises	Automated Scanning and Vulnerability Assessments Generative AI
Distributed Denial of Service Attacks	Medium	Cascading failures Service unavailability	Machine Learning Self-Supervised Learning Generative AI
Supply Chain Attacks	Medium	Third-party risks Widespread disruption	Predictive Analytics Blockchain
Zero-Day Exploits	Medium	Delayed mitigation Espionage	Anomaly Detection Self-Supervised Learning Representation Learning Generative AI

5. Discussion

CIIP, when characterized as a problem domain, is a complex, dynamic and real-time environment where disruptions can cause large scale instability due to the infrastructure constituting a sensitive backbone of modern society. This is one of the reasons why such a significant number of resources and effort is currently being invested in research into protecting CI. However, we are more dependent and vulnerable than perhaps ever before. Recent attacks, such as those on the Colonial Pipeline in the US, current reports on vulnerabilities in solar infrastructure, cyberattacks on French hospitals, the British water supply, and acts of sabotage on the German power grid clearly demonstrate this. In parallel, the field of AI has seen major advancements resulting in algorithms capable of solving complex problems in diverse domains, such as finance, gaming, biochemistry, medicine and cybersecurity. It is therefore natural to dedicate efforts into discovering and understanding the possible applications of state-of-the-art AI techniques to CIIP.

The impressive and continuing growth of generative AI and representation learning has resulted in new ways in which AI can be used to solve downstream tasks more efficiently. Therefore, techniques involving better separation between types of observations to better defend against adversarial attacks (contrastive learning), compressing high-dimensional data into low-dimensional latent representations (representation learning), generating possible future threats and synthetic training data (generative AI) should be explored in great depth. The ability of self-supervised learning algorithms to derive class labels independently also presents a significant

opportunity to leverage the real-time data generated by IIoT devices in CIIs to create better baselines for normal functionality.

6. Conclusion

With the rapid growth of digitalization, our CI is more vulnerable than ever before, and at the same time, the value of sensitive information and the impact and dependency of CI are also stronger than never. This is why this work explored the role of AI in CII security by first identifying the most common attack vectors in this area. It was noted that DDoS, insider threats, social engineering, malware, ransomware, unpatched vulnerabilities, supply chain attacks and zero-day exploits were some of the most common attack vectors. Thereafter, the work explored how AI can be applied to handle the attack vectors identified. The discussion also noted some of the AI techniques proposed in the literature.

Current state-of-the art AI techniques can also be used in the context of CIIP. Generative AI can be used to forecast future threats by generating synthetic attacks from learned attack patters (adversarial training). Representation learning can be used to better understand data points by mapping them to latent spaces which are useful for a variety of purposes. Self-supervised learning can also be implemented in context where annotating data is an expensive and time-consuming process.

Acknowledgement

This work has been supported by the University of Johannesburg and the Federal Ministry of Education and Research of the Federal Republic of Germany (Förderkennzeichen 16KIS2239K SUSTAINET_guarDian). The authors alone are responsible for the content of the paper.

References

Allianz ”Cyber attacks on critical infrastructure”, 2016. Available at: https://commercial.allianz.com/news-and-insights/expert-risk-articles/cyber-attacks-on-critical-infrastructure.html [Accessed 29 01 2024].

Altaleb, H. & Rajnai, Z., “Malware Attacks on SCADA Systems: Assessing Risks and Strengthening Cybersecurity Measures“, *IEEE,* pp. 625-630 2023.

Baezner, M. & Robin, P., ”Stuxnet”, Technical Report, *ETH Zurich*, 2017.

Bederna, Z., Rajnai, Z. & Szadeczky, T., “Attacks against energy, water and other critical infrastructure in the EU“, *2020 IEEE 3rd International Conference and Workshop in Óbuda on Electrical and Power Engineering (CANDO-EPE),* pp. 125-130, 2020.

Beerman, J., Berent, D., Falter, Z. & Bhunia, S., “*A* Review of Colonial Pipeline Ransomware Attack“, *2023 IEEE/ACM 23rd International Symposium on Cluster, Cloud and Internet Computing Workshops (CCGridW)*, pp. 8-15, 2023.

Budžys, A., Kurasova, O. & Medvedev, V., ”Behavioral Biometrics Authentication in Critical Infrastructure Using Siamese Neural Networks”, *HCI for Cybersecurity, Privacy and Trust*, Springer Nature Switzerland*,* pp. 309–322, 2023.

Carrier, T., Victor, P., Tekeoglu, A. & Lashkari, A. H., “Detecting Obfuscated Malware using Memory Feature Engineering“, *2023 IEEE Colombian Conference on Applications of Computational Intelligence (ColCACI)*, pp. 177–188, 2022.

Catak, F. O. & Yazı, A. F., “A Benchmark API Call Dataset for Windows PE Malware Classification“, *arXiv*, 2021.

Ćelić, J., Bronzin, T., Horvat, M., Jović, A., Stipić, A., Prole, B., Maričević, M., Pavlović, I., Pap, K., Mikota, M. & Jelača, N., ”Generative AI in E-maintenance: Myth or Reality?“, *2024 47th MIPRO ICT and Electronics Convention (MIPRO)*, 2024.

Charles, V., Emrouznejad, A. & Gherman, T., “A critical analysis of the integration of blockchain and artificial intelligence for supply chain“, *Annals of Operations Research,* Vol 327, pp. 7–47, 2023.

Chen, X., Ding, M., Wang, X., Xin, Y., Mo, S., Wang, Y., Han, S., Luo, P., Zeng, G. & Wang, J., “Context autoencoder for self-supervised representation learning“, *International Journal of Computer Vision,* Vol 132, pp. 208–223, 2024.

Di Pietro, R., Raponi, S., Caprolu, M. & Cresci, S., “Critical Infrastructure“, *New Dimensions of Information Warfare,* Springer International Publishing, pp. 157–196, 2021.

Duman, O., Ghafouri, M., Kassouf, M., Atallah, R., Wang, L. & Debbabi, M., ”Modeling Supply Chain Attacks in IEC 61850 Substations”, *2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm),* pp. 1-6, 2019.

Eggers, S., ”A novel approach for analyzing the nuclear supply chain cyber-attack surface”, *Nuclear Engineering and Technology,* Vol 53, pp. 879-887, 2021.

El Husseini, F., Noura, H., Salman, O. & Chehab, A., ”Advanced Machine Learning Approaches for Zero-Day Attack Detection: A Review“, *2024 8th Cyber Security in Networking Conference (CSNet)*, pp. 297-304, 2024.

Esmaeili, B., Azmoodeh, A., Dehghantanha, A., Srivastava, G., Karimipour, H. & Chun-Wei Lin, J., ”A GNN-Based Adversarial Internet of Things Malware Detection Framework for Critical Infrastructure: Studying Gafgyt, Mirai, and Tsunami Campaigns”. *IEEE Internet of Things Journal,* Vol 11, pp. 26826-26836, 2024.

Federal Office for Information Security, ”What are Critical Infrastructures?”*,* Available at: https://www.bsi.bund.de/EN/Themen/Regulierte-Wirtschaft/Kritische-Infrastrukturen/Allgemeine-Infos-zu-KRITIS/allgemeine-infos-zu-kritis_node.html [Accessed 25 04 2025].

Gaidarski, I. & Minchev, Z., "Insider Threats to IT Security of Critical Infrastructures", In: T. Krassimir, K. Vyacheslav, K. J. T. Todor & Atanassov, eds., *Digital Transformation, Cyber Security and Resilience of Modern Societies*, Springer International Publishing, pp. 381-394, 2021.

Geiger, M., Bauer, J., Masuch, M. & Franke, J., "An Analysis of Black Energy 3, Crashoverride, and Trisis, Three Malware Approaches Targeting Operational Technology Systems", *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, pp. 1537-1543, 2020.

Ghafir, I., Saleem, J., Hammoudeh, M., Faour, H., Prenosil, V., Jaf, S., Jabbar, S. & Baker, T., "Security threats to critical infrastructure: the human factor". *The Journal of Supercomputing,* Vol 74, pp. 4986–5002, 2018.

Goodfellow, I. J., "Generative Adversarial Networks", *Advances in Neural Information Processing Systems*, Vol 27, 2014.

Govea, J., Gaibor-Naranjo, W. & Villegas-Ch, W., "Securing Critical Infrastructure with Blockchain Technology: An Approach to Cyber-Resilience", *Computers,* Vol 13, 2024.

Harang, R. & Rudd, E. M., "SOREL-20M: A Large Scale Benchmark Dataset for Malicious PE Detection"*, Arxiv,* 2020.

Herrera, L.-C. & Maennel, O., "A comprehensive instrument for identifying critical information infrastructure services", *International Journal of Critical Infrastructure Protection,* Vol 25, pp. 50-61, 2019.

Hossain, S., Senouci, S.-M., Brik, B. & Boualouache, A., "A privacy-preserving Self-Supervised Learning-based intrusion detection system for 5G-V2X networks", *Ad Hoc Networks,* Vol 166, pp. 103674, 2025.

Hurst, W., Shone, N. & Monnet, Q., "Predicting the Effects of DDoS Attacks on a Network of Critical Infrastructures", *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing,* pp. 1697-1702, 2015.

Hussain, B., Du, Q., Sun, B. & Han, Z., "Deep Learning-Based DDoS-Attack Detection for Cyber–Physical System Over 5G Network", *IEEE Transactions on Industrial Informatics,* Vol 17, pp. 860-870, 2021.

Ibor, A. E., "Zero day exploits and national readiness for cyber-warfare", *Nigerian Journal of Technology,* Vol 36, pp. 1174–1183, 2017.

Kim, H., Lee, J. & Park, J.-G., "SITRAN: Self-Supervised IDS With Transferable Techniques for 5G Industrial Environments", *IEEE Internet of Things Journal,* Vol 11, pp. 35465-35476, 2024.

Kingma, D. P. & Welling, M., "Auto-Encoding Variational Bayes", *Arxiv*, 2013.

Kozak, P., Klaban, I. & Šlajs, T., "Industroyer cyber-attacks on Ukraine's critical infrastructure", *2023 International Conference on Military Technologies (ICMT)*, pp. 1-6, 2023

Lehto, M., "Cyber-Attacks Against Critical Infrastructure", In: Martti, L. & Neittaanmäki, P., eds., *Computational Methods in Applied Sciences*, Vol 56, Springer International Publishing, pp. 3-42, 2022.

Li, J.-h., "Cyber security meets artificial intelligence: a survey", *Frontiers of Information Technology & Electronic Engineering, Vol* 19(12), pp. 1462-1474, 2018.

Lipps, C., Baradie, S., Herbst, J., Armistead, L. & Schotten, H.D., "Cybersecurity in Industrial Automation and Control Systems: The Recent Attack of the Colonial Pipeline", in: Van Niekerk, B., Ramluckan, T. and Kushwaha, N., eds., *Modelling Nation-state Information Warfare and Cyber-operations,* United Kingdom: Academic Conferences and Publishing International Limmited, pp. 215 – 236, 2022.

Lipps, C., Duque Antón, S. & Schotten, H. D., "Enabling Trust in IIoT: An PhySec Based Approach", *14th International Conference on Cyber Warfare and Security,* Stellenbosh, South Africa, *2019*.

Liu, C., Antypenko, R., Sushko, I. & Zakharchenko, O., "Intrusion Detection System After Data Augmentation Schemes Based on the VAE and CVAE". *IEEE Transactions on Reliability,* Vol 71, pp. 1000-1010, 2022.

Liu, X., Zhang, F., Hou, Z., Wang, Z., Mian, L., Zhang, J. & Tang, J., "Self-Supervised Learning: Generative or Contrastive", *IEEE Transactions on Knowledge and Data Engineering,* Vol 35, pp. 857-876, 2023.

Lopez, J., Setola, R. & Stephen, D. W., "Overview of Critical Information Infrastructure Protection", In: Roberto Setola, J. L. & Wolthusen, S. D., eds., Springer Berlin Heidelberg, pp. 1-14, 2012.

Mahmoud, H., Wu, W. & Gaber, M. M., "A Time-Series Self-Supervised Learning Approach to Detection of Cyber-physical Attacks in Water Distribution Systems", *Energies,* Vol 15, 2022.

Makrakis, G. M., Kolias, C., Kambourakis, G., Reiger, C. & Benjamin, J., "Industrial and Critical Infrastructure Security: Technical Analysis of Real-Life Security Incidents", *IEEE Access,* Vol 9, pp. 165295-165325, 2021.

Marjanović, J., Dalčeković, N. & Sladić, G., "Improving Critical Infrastructure Protection by Enhancing Software Acquisition Process Through Blockchain", *ECBS 2021: 7th Conference on the Engineering of Computer Based Systems*, New York, NY, USA, Association for Computing Machinery, 2021.

Mbanaso, U. M. & Kulugh, V. E., "Empirical Findings of Assessment of Critical Infrastructure Degree of Dependency on ICT", *Cybersecurity in Emerging Digital Era*, Vol 1436, Springer International Publishing, pp. 3-23, 2021.

Mitnick, K. D. & Simon, W. L., "The art of deception: Controlling the human element of security", John Wiley & Sons, 2003.

Musluoglu, M., Kunicina, N. & Caiko, J., "Vulnerability Assessment of Industrial Control Systems for Colonial Pipeline and WannaCry Ransomware", *2024 IEEE 65th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON)*, pp. 1-7, 2024.

Mylrea, M. & Gourisetti, S. N. G., "Blockchain for Supply Chain Cybersecurity, Optimization and Compliance", *2018 Resilience Week (RWS)*, pp. 70-76, 2018.

Panagiotis, F., Taxiarxchis, K., Georgios, K., Maglaras, L. & Ferrag, M. A., "Intrusion Detection in Critical Infrastructures: A Literature Review", *Smart Cities,* Vol 4, p. 1146–1157, 2021.

Pinto, A., Herrera, L.-C., Donoso, Y. & Gutierrez, J. A., "Survey on Intrusion Detection Systems Based on Machine Learning Techniques for the Protection of Critical Infrastructure", *Sensors,* Vol 23, 2023.

Rahmanian, A. A., Ghobaei-Arani, M. & Tofighy, S., "A learning automata-based ensemble resource usage prediction algorithm for cloud computing environment", *Future Generation Computer Systems,* Vol 79, pp. 54-71, 2018.

Rege, A. & Bleiman, R., "A Free and Community-Driven Critical Infrastructure Ransomware Dataset", *Proceedings of the International Conference on Cybersecurity, Situational Awareness and Social Media*, Singapore, Springer Nature Singapore, p. 25–37, 2023.

Russell, S. & Norvig, P., "Artificial Intelligence: A Modern Approach", 4 ed., Pearson, 2020.

Saadallah, M., Shahim, A. & Khapova, S., "Multi-method Approach to Human Expertise, Automation, and Artificial Intelligence for Vulnerability Management", *ICT Systems Security and Privacy Protection,* Springer Nature Switzerland, pp. 410–422, 2024.

Saheed, Y. K., Misra, S. & Chockalingam, S., "Autoencoder via DCNN and LSTM Models for Intrusion Detection in Industrial Control Systems of Critical Infrastructures", *2023 IEEE/ACM 4th International Workshop on Engineering and Cybersecurity of Critical Systems (EnCyCriS)*, pp. 9-16, 2023.

Santos, R., Souza, D., Santo, W., Ribeiro, A. & Moreno, E., "Machine learning algorithms to detect DDoS attacks in SDN", *Concurrency and Computation: Practice and Experience,* Vol 32, pp. e5402, 2020.

Sarker, I. H., AI for Critical Infrastructure Protection and Resilience. In: *AI-Driven Cybersecurity and Threat Intelligence: Cyber Automation, Intelligent Decision-Making and Explainability,* Springer Nature Switzerland, pp. 153–172, 2024.

Shahriar, M. H., Khalil, A. A., Rahman, M. A., Manshaei, M. H. & Chen, D., "iAttackGen: Generative Synthesis of False Data Injection Attacks in Cyber-physical Systems", *2021 IEEE Conference on Communications and Network Security (CNS)*, pp. 200-208, 2021.

Shakeel, I., "When humans are the weak link in critical infrastructure cybersecurity", 2023, Available at: https://www.securitymagazine.com/articles/99867-when-humans-are-the-weak-link-in-critical-infrastructure-cybersecurity [Accessed 8 6 2025].

Singh, U. K., Joshi, C. & Kanellopoulos, D., "A framework for zero-day vulnerabilities detection and prioritization", *Journal of Information Security and Applications,* Vol 46, pp. 164-172, 2019.

Sithungu, S. P. & Ehlers, E. M., "Gaainet: A generative adversarial artificial immune network model for intrusion detection in industrial iot systems", *Journal of Advances in Information Technology,* Vol 13, 2022.

Stellios, I., Kotzanikolaou, P., Psarakis, M., Alcaraz, C. & Lopez, J., "A Survey of IoT-Enabled Cyberattacks: Assessing Attack Paths to Critical Infrastructures and Services", *IEEE Communications Surveys & Tutorials,* Vol 20, pp. 3453-3495, 2018.

Tavallaee, M., Bagheri, E., Lu, W. & Ghorbani, A. A., "A detailed analysis of the KDD CUP 99 data set", *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, pp. 1-6, 2009.

Tsantikidou, K. & Sklavos, N., "Threats, Attacks, and Cryptography Frameworks of Cybersecurity in Critical Infrastructures", *Cryptography*, Vol 8(1), 2024.

Tsegaye, T. & Flowerday, S., "Controls for protecting critical information infrastructure from cyberattacks", *World Congress on Internet Security (WorldCIS-2014)*, pp. 24-29, 2014.

Wang, B., Li, X., de Aguiar, L. P., Menasche, D. S. & Shafiq, S., "Characterizing and Modeling Patching Practices of Industrial Control Systems", *Proceedings of the ACM on Measurement and Analysis of Computing Systems,* Vol 1(1), 2017.

Wang, D. & Shang, Y., "A new active labeling method for deep learning", *2014 International Joint Conference on Neural Networks (IJCNN)*, pp. 112–119, 2014.

Wei, F., Koc, E., Li, N., Soibelman, L. & Wei, D., "A data-driven framework to evaluate the indirect economic impacts of transportation infrastructure disruptions", *International Journal of Disaster Risk Reduction,* Vol 75, pp. 102946, 2022.

Wilson, C., "Cyber Threats to Critical Information Infrastructure", In: Chen, T. M., Jarvis, L. & Macdonald, S., eds., *Cyberterrorism: Understanding, Assessment, and Response*, Springer New York, pp. 123–136, 2014.

Wu, Z., Pan, S., Chen F., Long, G., Zhang, C. & Yu, P. S., "A Comprehensive Survey on Graph Neural Networks", *IEEE Transactions on Neural Networks and Learning Systems,* Vol 32, pp. 4-24, 2021.

Yang, L., Ciptadi, A., Laziuk, I., Ahmadzadeh, A. & Wang, G., "BODMAS: An Open Dataset for Learning based Temporal Analysis of PE Malware*", 2021 IEEE Security and Privacy Workshops (SPW),* 2021.

Yellapragada, B., Hornauer, S., Snyder, K., Yu, S. & Yiu, G., "Self-Supervised Feature Learning and Phenotyping for Assessing Age-Related Macular Degeneration Using Retinal Fundus Images", *Ophthalmology Retina,* Vol 6, pp. 116-129, 2022.

Yigit, Y., Ferrag, M. A., Sarker, I. H., Maglaras, L. A. & Chrysoulas, C., Moradpoor, N., Janicke, H., "Critical Infrastructure Protection: Generative AI, Challenges, and Opportunities", *Arxiv*, 2024.

Zhang, Z., Ning, H., Shi, F., Farha, F., Xu, Y., Xu, J., Zhang, F. & Choo, K.-K. R., "Artificial intelligence in cyber security: research advances, challenges, and opportunities", *Artificial Intelligence Review*, Vol 55(2), pp. 1029-1053, 2022.

Zolanvari, M., Teixeira, M. A., Gupta, L., Khan, K. M. & Jain, R., "WUSTL-IIOT-2021 Dataset for IIoT Cybersecurity Research", 2021, Available at: http://www.cse.wustl.edu/~jain/iiot2/index.html [Accessed 01 02 2025].