Advances in Computer Science and Information Technology (ACSIT)
p-ISSN: 2393-9907; e-ISSN: 2393-9915; Volume 10, Issue 2; April-June 2023, pp. 85-92

http://www.krishisanskriti.org/Publication.html

Artificial Intelligence (AI) Systems: Challenges, Threats and Mitigating Strategies

Amaya Aggarwal

DSB International School
95, Ganpatrao Kadam Marg, Opposite Peninsula Business Park,
Lower Parel West, Mumbai, Maharashtra
E-mail: aggarwalamaya35@gmail.com

Abstract*—As the global community increasingly embraces digital technology across various sectors, cybersecurity has become an essential consideration. In 2020, the worldwide economy experienced a loss of nearly USD 1 trillion due to cybercrime, reflecting an alarming increase of over 50% compared to previous years. Since 2018, Artificial Intelligence (AI) has significantly impacted daily life, acting as a catalyst in digital transformation by offering automated decision-making capabilities. Although this emergent technology's advantages are profound, it has simultaneously given rise to serious concerns. AI may introduce novel arenas for manipulation and assault methodologies, as well as unprecedented challenges in privacy and data protection.*

This research paper systematically explores the AI cybersecurity ecosystem, focusing on its Threat Landscape. The primary objectives are threefold:

1. ***Definition of the Scope:*** *An in-depth analysis of AI within the context of cybersecurity, adopting a lifecycle approach that encompasses various stages from requirements analysis to deployment. Through this, the ecosystem of AI systems and applications is precisely mapped.*
2. ***Threat Mapping and Classification:*** *A detailed exploration and classification of the AI threat landscape, including potential attack scenarios and sectorial risk assessments. The listing of proportionate security controls, understanding threats to diverse AI lifecycle stages, and assessing their impact on various security properties.*
3. ***Solution:*** *Evaluation of potential legal and technical solution and prevention from various forms of cyberattacks.*

Emphasizing the need for a secure AI ecosystem, this study asserts that cybersecurity and data protection must be the focal points of innovation, capacity building, awareness raising, and research and development initiatives.

Introduction:

Artificial Intelligence (AI) has swiftly transcended into an indispensable technological tool, permeating a multitude of fields with an ever-expanding breadth of applications. Its centrality in our daily lives continues to grow, offering myriad benefits while simultaneously posing substantial challenges.

At its core, AI is the pursuit of developing computer systems that are designed to emulate, and in fact augment, human intelligence. It is, according to a widely accepted definition, "the capacity of a digital computer or computer-controlled robot to undertake tasks ordinarily necessitating human intelligence." This technological journey commenced in 1951 when Christopher Strachey penned the first successful AI program.

AI essentially involves the study and creation of systems that strive to mimic and utilize human mental faculties. Its capabilities extend to initiating intelligent human-like behaviours and finding solutions to complex problems without human intervention. These pursuits materialize through two fundamental approaches: machine learning and deep learning.

However, AI's vast potential is not without vulnerabilities. Its exposure to a unique category of cybersecurity threats, known as "artificial intelligence attacks," is of increasing concern. Malicious entities can manipulate AI systems, altering their behavior to serve nefarious objectives. Given AI's broad applicability in sectors like facial recognition and digital identities, space technologies, defense, healthcare, agriculture, transportation, and weather forecasting, these vulnerabilities can have dire consequences for personal and national security.

The novel forms of cyber assault exploit inherent limitations within the underlying AI algorithms, which are presently uncontrollable. Even data, traditionally seen as secure, can be manipulated and weaponized through these attacks. This necessitates a comprehensive rethinking of data collection, storage, and utilization.

Areas commonly affected by AI attacks include content filters, defense, law enforcement agencies, and the human tasks that AI is increasingly replacing. This complex landscape requires the development and implementation of AI security compliance programs to mitigate the risks. These programs should incorporate best practices in security systems, information technology reforms, and robust attack response strategies.

Regulation is paramount, with mandates for compliance across both governmental and private sectors. Striking a balance between security and innovation is crucial, with strict norms in place for vulnerable uses of AI and more flexible regulations for low-risk applications.

The intrinsic limitations of state-of-the-art AI methods render them susceptible to a catastrophic range of attacks that are as insidious as they are hazardous. These issues diverge from traditional cybersecurity problems and cannot be addressed with existing cybersecurity and policy toolkits. They require the development of novel techniques and resolutions, focusing on high-priority areas such as the military, content filtering, law enforcement, human task automation, and civil society.

Artificial intelligence attacks constitute a distinctive class of cyber threats, necessitating a departure from conventional cybersecurity paradigms. The urgent need for "AI security compliance" programs is apparent, and the pathway forward includes the cultivation of best practices, the deployment of IT reforms to impede AI attackers, and the creation of robust response plans.

Overview of An Attack:

An AI attack involves the intentional alteration of an artificial intelligence system with the goal of causing it to malfunction. Machine Learning is a specialized branch that falls under the larger umbrella of artificial intelligence and computer science. Machine learning emphasizes the utilization of data and algorithms to simulate human learning processes, thereby enhancing its precision over time. When developing or programming an AI system, specific data must be input to train the system to recognize particular images and commands. During this manual process of data entry, incorrect information can be either intentionally or accidentally introduced to disrupt the system for individual benefit. This deliberate manipulation is known as an adversarial attack, which can be broken down into various subcategories such as;

- **Data Corrupting Threats**: These involve attacking and corrupting the inputted data during the creation of AI systems, leading to malfunctions. If this data is corrupted or altered intentionally, the AI system can produce incorrect or malicious outputs. An attacker could, for example, insert misleading data into a facial recognition training set, causing the system to misidentify individuals intentionally. The difference in white and black box attacks lie in the intention and the access of information the attacker has.

i. White-Box Attacks: In these attacks, the adversary has full access to the target model, including its architecture, parameters, and training data. This complete knowledge allows the attacker to craft specialized inputs to deceive the system. For example, an attacker might alter a digital image of a stop sign in a way that the human eye can't detect but causes an autonomous vehicle's AI system to misclassify it as a yield sign.

ii. Black-Box Attacks: Unlike white-box attacks, black-box attacks occur when the attacker has no knowledge of the underlying model's parameters or architecture. The attacker can only access the input and output of the AI system. The intent is to mislabel the output.[SEP]Example: An attacker could use trial and error to figure out an audio command that a voice-controlled assistant misinterprets, allowing unauthorized access to a secured system.

- **Input Attacks/Model Inversion Attack** : This type of attack aims to reverse-engineer an AI model. By feeding the model with numerous inputs and analyzing the outputs, attackers can infer sensitive information about the training data or even reconstruct the original data in order to fulfill the attacker's objectives. Due to this type of attack, facial recognition systems can give inaccurate results.

- **Hiding Information**: An example might be causing a content filter designed to block extremist content to malfunction, thereby allowing prohibited material to spread.

- **Downgrading Trust in a System**: An example could be a monitored system that triggers false alarms, leading to manual control being exerted over the system.

Challenges Faced by Users Regarding Cyberattacks on AI Systems:

1. **Rapid Technological Advancements**: AI systems are evolving at an unprecedented pace, making it challenging for users to stay updated with the latest technologies.
2. **Detection Difficulties**: There's no straightforward method to ascertain if a system has been compromised, leaving users in the dark about potential breaches.
3. **Lack of Understanding**: Many users lack a foundational understanding of how AI systems operate, making it difficult to diagnose and address issues.
4. **Complex Technical Jargon**: Information available on AI and cybersecurity is often laden with technical terms, making it hard for the average user to grasp quickly.
5. **Need for Simplified Explanations**: A more user-friendly approach, with explanations in layman's terms, would make the information more accessible and less overwhelming.
6. **User's Limited Preventive Capacity**: Given the complexities, it's unrealistic to expect individual users to prevent or resolve cyberattacks on their own.

This proves that it’s unlikely that the user themselves will be able to make much difference in protecting themselves against cyberattacks and hence It's imperative for governments and organizations to take the onus of preventing cyberattacks, rather than placing the burden on individual users.

Responsibilities of Governments and Organizations:

1. **Implementation of Technical Solutions**: Governments should mandate the deployment of advanced technical solutions at the organizational level to thwart potential threats.
2. **Swift Response to Breaches**: In the event of a cyberattack, prompt action is crucial to determine the root cause and address it.
3. **Future-Proofing Systems**: Post an attack, it's essential to ensure that measures are in place to prevent similar breaches in the future.
4. **Legal Recourse**: Adequate legal frameworks should be established to penalize perpetrators of cyberattacks, sending a strong deterrent message.
5. **Educational Initiatives**: Governments and organizations should invest in educational campaigns to raise awareness about AI systems and cybersecurity, making information more accessible to the general public.

Solutions:

TECHNOLOGICAL METHODS TO PREVENT & MITIGATE THREATS IN AI SYSTEMS:

Adversarial attacks is the most common form of cyberattacks and methods ot prevent it are as follows-

1. Adversarial Training:
Description: This involves training the model on adversarial examples to make it robust against such attacks. By exposing the model to these malicious inputs during training, it learns to recognize and correctly classify them during inference.

Strengths: Can significantly improve model robustness against known adversarial attacks.

Limitations: Can be computationally expensive and may not defend against all types of adversarial attacks.

2. Input Pre-processing:
Description: Before feeding an input to the model, it undergoes preprocessing to remove potential adversarial perturbations. Techniques like image denoising or feature squeezing can be used.

Strengths: Simple to implement and can be effective against certain types of attacks.

Limitations: May not be effective against sophisticated or previously unseen attacks.

3. Regularization Techniques:
Description: Regularization methods, such as dropout or L2 regularization, can be used to prevent overfitting and potentially increase model robustness against adversarial attacks.

Strengths: Can improve generalization and potentially increase adversarial robustness.

Limitations: Alone, may not be sufficient to defend against targeted adversarial attacks.

4. Generative Models:
Description: As discussed with Defence-GAN, generative models can be used to reconstruct inputs, ensuring they lie on the data manifold and removing adversarial perturbations.

Strengths: Can effectively neutralize certain adversarial perturbations.

Limitations: Training generative models can be computationally intensive.

5. Model Ensembling:
Description: Using an ensemble of models can increase robustness as an attacker would need to deceive multiple models simultaneously.

Strengths: Can improve overall model performance and robustness.

Limitations: Increases computational overhead.

6. Monitoring and Detection:
Description: Continuously monitor the model's predictions to detect anomalies or patterns indicative of an adversarial attack. Once detected, appropriate countermeasures can be taken.

Strengths: Provides a real-time defense mechanism.

Limitations: Requires continuous monitoring and may have false positives.

7. Research and Collaboration:
Description: Engage in collaborative research to understand the evolving nature of AI attacks and develop effective countermeasures. Open platforms like OpenAI have been actively researching adversarial attacks and their defences.

Strengths: Collaborative efforts can pool resources and expertise to address the challenges collectively.

Limitations: Requires active participation and sharing of knowledge, which may not always be feasible due to competitive or security concern

8. Defense-Gan:
Description: The idea behind Defense-GAN is to use the generator to reconstruct inputs in a way that removes anu adversarial perturbations. In other words, for any input (whether it's clean or contains adversarial noise), Defense-GAN aims to find a "clean" version of that input by leveraging the GAN's generative capabilities.

Strengths:

1. **Versatility:** Defense-GAN seamlessly integrates with any classifier without altering its structure, acting as a pre-processing step before classification.

2. **Generality:** This tool is adaptable, defending against a wide range of attacks without being tied to a specific attack model. Instead, it harnesses the generative capabilities of GANs to reconstruct adversarial instances.
3. **Non-linearity:** Its highly non-linear nature impedes white-box gradient-based attacks, especially due to the gradient descent loop. Meaning this unpredictable counterattack path chosen makes it difficult for the original attacker to trace and counter preventing it from repeating.
4. **Consistency:** Demonstrated efficacy against a majority of prevalent attack methodologies.
5. **Projection:** During inference, Defense-GAN maps input images to the GAN's generator range before classification.
6. **Robustness:** It adeptly counteracts adversarial disturbances by aligning these samples with the learned data distribution.
7. **Generative Excellence:** Leveraging the prowess of contemporary GANs, Defense-GAN can generate high-fidelity images, enhancing its ability to modify adversarial inputs to mirror authentic data.

Limitations:

1. **Training Nuances:** GAN training, especially with vast datasets, demands significant computational resources and time. The effectiveness of Defense-GAN is closely tied to the GAN's generative capabilities. Perfecting GAN training remains a complex endeavour and a focal point of ongoing research.
2. **Performance Dependency:** The efficacy of Defense-GAN hinges on the meticulous training and tuning of the GAN. Inadequate training can compromise its performance.
3. **Specific Efficacy:** Defense-GAN excels against certain adversarial attacks but may falter against others.
4. **Adaptive Threats:** As defense strategies evolve, so do adversarial tactics. New attacks might emerge, targeting Defense-GAN's specific defense mechanisms.

While we looked at Adversarial attacks and its prevention here are a few more examples of potential attacks and their possible preventions-

Model Inversion and Membership Inference Attacks:

- **Differential Privacy**: Introduce random noise during training to obfuscate individual data points, making it harder to infer specifics about the training data.
- **Regularization**: Regularization techniques can prevent models from overfitting to specific training data points, thus providing protection against such inference attacks.

1. **Poisoning Attacks:**

- **Data Sanitization**: Rigorously clean and vet the training data to remove any malicious entries.
- **Outlier Detection**: Use outlier detection to identify and remove anomalous data points from the training set.
- **Model Regularization**: Helps in ensuring that the model doesn't overly rely on potentially poisoned data.

2. **Backdoor Attacks:**

- **Fine-pruning**: This involves retraining the model while ignoring certain suspicious neurons that could have been activated by backdoor triggers.
- **Neural Cleanse**: A technique to reverse-engineer potential backdoor triggers in the model and then mitigate them.
- **Regular Audits**: Regularly inspect and validate the training data, especially if sourced from third parties.

3. Data Privacy Attacks:

- **Federated Learning**: Train models on decentralized data sources, ensuring raw data doesn't leave its original device, thus protecting user privacy.
- **Homomorphic Encryption**: Allows for computations on encrypted data without needing to decrypt it first, offering strong privacy guarantees.

4. Misinformation and Deepfakes:

- **Watermarking**: Imprint AI-generated content with watermarks indicating its synthetic origin.
- **Deepfake Detection Models**: Develop and deploy models specifically designed to detect deepfake content.
- **Blockchain**: Use blockchain to track and verify the authenticity of digital content.

5. **Infrastructure Attacks** (attacks on the underlying systems supporting AI):

- **Regular Patching**: Keep all software updated to prevent known vulnerabilities.
- **Intrusion Detection Systems**: Monitor and detect suspicious activities on networks and systems.
- **Hardware-based Security**: Secure AI accelerators and other hardware components from physical tampering and other threats.

6. Model Accountability and Interpretability:

- **Interpretable Models**: Use models that offer better transparency and interpretability, allowing for easier inspection of their decision-making process.
- **Model Monitoring**: Continuously monitor models in deployment for drifts or changes in performance that might indicate compromises.

Legal Protection against Cyberthreats:
The Computer Fraud and Abuse Act (CFAA) is a U.S. federal statute that criminalizes unauthorized access to computer systems and networks its implemented in order to prevent a myriad of cyberthreats.

In order for the law to be enforced unauthorized access to a computer is a prerequisite and any of the following points after including:

Required-Unauthorized Access or Exceeding Authorized Access: The CFAA primarily targets those who intentionally access a computer without authorization or exceed authorized access. This means that the individual must not have permission to access the computer or system in question or must go beyond the permissions granted to them.

1. **Intent to Defraud**: Some provisions of the CFAA require the government to prove that the defendant accessed a computer with the intent to defraud. This means that there must be an intention to deceive or cheat.
2. **Obtaining Information**: The CFAA can be applied when someone intentionally accesses a computer without authorization and as a result, obtains information from a protected computer. A "protected computer" under the CFAA is broadly defined and can include government computers, financial institution computers, or computers used in interstate or foreign commerce or communication
3. **Causing Damage**: The CFAA can also be invoked when there is intentional access to a protected computer, and as a result, there is damage or impairment to the integrity or availability of data, a program, a system, or information
4. **Trafficking in Passwords/Information**: The act prohibits trafficking in passwords if such conduct affects interstate or foreign commerce or if the individual knowingly traffics in passwords for the purpose of unauthorized access to a computer
5. **Interstate or Foreign Commerce Requirement**: Many of the CFAA's provisions require that the offense affects interstate or foreign commerce. This is a jurisdictional element that ensures the federal government has the authority to prosecute the offense
6. **Financial Gain or Commercial Advantage**: Some sections of the CFAA require the act to be committed for the purpose of financial gain or commercial advantage

Evaluating the effectiveness of the CFAA-

Strengths;
1. **Private Right of Action**: The CFAA includes a private right of action, allowing any person to sue if they have incurred damages or losses due to a CFAA violation
2. **Protection Against Unauthorized Access**: The CFAA prohibits anyone who "intentionally accesses a computer without authorization" or "exceeds authorized access"
3. **Quick to Implement:** There are not many options other than the CFAA when it comes to computer hacking making it easy to know which law to enforce in a particular case. Furthermore, its efficient as it has a clear set prerequisite in order for it to be enforced when these conditions are met then only can the CFAA be applied.
4. **Potential for Better Security**: If the CFAA is interpreted narrowly, focusing on hacking and bypassing technological barriers, it could lead to better security outcomes in the long term. This would incentivize the development of more robust technological and code-based measures to protect against adversarial attacks

Limitations;
1. **Ambiguity in Scope:** The CFAA's has a broad scope meaning it can be applied in many cases including data protection and e-privacy laws, intellectual property laws, confidentiality laws, information security laws, and import/export controls, among others. In 2020, the Van Buren V US case argued on the matter of broad interpretation of this law and hence its low effectiveness however, the Court's clarified that the CFAA does not criminalize every violation of a computer use policy. Instead, it is more concerned with unauthorized access to information on a computer.
2. **Inadequate foreign defence**: The global nature of cybercrime poses jurisdictional challenges. Many cyberattacks originate from outside the U.S., making it difficult to prosecute offenders
3. **Outdated**: The CFAA was enacted in 1986 and while there has been a few modifications of it through the years the change in the laws has been slower than the rate of technological development.
4. **Inconsistent Application**: The CFAA has many different uses in different cases since it is broadly defined therefore there is some inconsistency in the application of it. It's interpretation described as "fragmented" and "unclear". This creates uncertainty and confusion regarding its application on how to proceed in the number of years of sentence and how severe the crime is with accordance to the law. Additionally, the US each state has different laws and therefore adds to the inconsistency.
5. **Lack of options**: Other than CFAA there is little to no law in place to tackle issues of cybersecurity.
6. **Terms of Service (TOS) Issues**: Expansive TOS may deter legitimate researchers from testing systems or reporting results due to fear of CFAA liabilities. However, truly bad actors or sophisticated adversaries are unlikely to be deterred by TOS

In order to organise the legal processing better the us government has a classifying system. Once a case falls under the CFAA it is then categorized into the sector the crime is applicable for example if it's a trade related concern it falls under the The Federal Trade Commission ("FTC") who then deal with the threat accordance to the laws of the state, similarly with The Health Insurance Portability and Accountability Act ("HIPAA").

Moreover, consider the state government of New York as an example. It has established a comprehensive legal framework for each industry. This framework mandates that companies within these industries not only implement robust cybersecurity measures but also regularly monitor and audit their systems to prevent any unauthorized intrusions.

Government approved systems in organisations to detect any interference in their systems are as follows-

- Beacons (i.e. imperceptible, remotely hosted graphics inserted into content to trigger a contact with a remote server that will reveal the IP address of a computer that is viewing such content)
- Honeypots (i.e. digital traps designed to trick cyber threat actors into taking action against a synthetic network, thereby allowing an organisation to detect and counteract attempts to attack its network without causing any damage to the organisation's real network or data)
- Sinkholes (i.e. measures to re-direct malicious traffic away from an organisation's own IP addresses and servers, commonly used to prevent DDoS attacks)

Conclusion

The rapid integration of Artificial Intelligence (AI) into various sectors has brought forth a multitude of benefits, revolutionizing numerous fields with its capabilities. However, this technological advancement is not devoid of challenges, particularly in the realm of cybersecurity. To prevent these attacks one must understand the nature of the attacks and then work towards preventing it. Not only is there more awareness amongst used needed but more accountability and responsibility required on the part of the government. The preventions of these challenges have a number of limitations which the government and firms must collaborate in order to foster research and minimize the limitation as much as possible. Lastly, to safeguard the users interests there needs to be adequate legal measures to prevent future attacks and penalize the attackers.

References:

[1]. Yao Jun, Alisa craig, Wasswa Shafik, and buleSharif. Artificial intelligent Application in cybersecurity and cyber defenceWiley Hindawi, wireless communications ard Mobile computing, volume 2021 , Article ID 3329581 http./ doi, org /110.1155/2021/3329581

[2]. Paola Breda, Rada Markova, Adam f Abdin, Devanghntha. Autoniocarlo and Nibilepelinmanti cyber vulnerabilities and risks of Al Technologies in space Applications in 73rdinternationalAstronantical congress (IAC) - paris, frand 18-22 September 2022: 1AC-22-D 5.4 x 70380

[3]. Yuchong Li and Qinghai Liu. A comprehending review study of cybir-attacks and cyber security; emergingtrends and recent developments. energy Reports 7(2.2) 8176-8186, https:11 doi.org 11011016 J. egyr. 2021 08 126

[4]. Lorenzo pupillo, Stefano Fantin, AfonsoFerrcira and carolinapolito. Artificiar intelligence and cybersecurity, Technolege Governance and policy challenges. centre, 2 for Europeanpulicystudies(CEPS). Brussels, may 2021 ISBN. 978 94-6138-785-1

[5]. Frank cremer, Barry shechen, michaelfortmann, Arash N. Kia, martin mullins, finbars murphy and stefanmaterne. cyber risk and cybersecurity: A systematic review of data availability. The Gene papers on Risk and Insurance - issues and practice (2022) 47: 698-736https:lldoi org110 1057/s 41288-022-0066-6

[6]. Rammanohar das and raghav sandman Artificial intelligence in security lop publishing ICACSE 2022. Formed of physics: conference series 1964 (2021)042072 doi : 10.108811742 – 659611964141042072 doi 10.108811742-659611964141042072

[7]. Ricardo Raimundo and andAlberico Rosario The impact of Artificial intelligence on Data system security: A Literature review Sensors 2021, 21, 7029 http:lldoi org 110.3390/s 21217029

[8]. Melville, N.; McQuaid, M. Generating shareable statistical databases for business value: Multiple imputation with multimodal perturbation. Inf. Sysl. Res.2012, 23,559-574. [Cross Ref]

[9]. Zhu, F.; Li, G. Study on Security of Electronic Commerce Information System. In Proceedings of the 20112 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC 2011), Zhengzhou, China, 8-10 August 2011; Institute of Electrical and Electronics Engineers: Piscataway, NJ, USA, 2011; pp. 1546-1549.

[10]. Sukhanova, N.V.; Sheptunov, S.A.; Glashev, R.M. The Neuron Network Model of Human Personality for Use in Robotic Systems in Security Information Technologies of the 2019 IEEE International Conference Quality Management, Transport and Information Security, Information Technologies IT and QM and IS Sochy, Rusia, 23-27 September 2019; Institute of Electrical and Electronics Engineers: Piscataway, NI, USA, 2014; pp. 11-16.

[11]. Ekenberg, L.; Danielson, M.; Boman, M. Imposing security constraints on agent-based decision support. Decis. Support Syst. 1997, 20, 3-15. [Cross Ref]

[12]. Bai, J., Wu, Y., Wang, G., Yang, S. X., & Qiu, W. (2006). A novel intrusion detection model based on multi-layer self-organizing maps and principal component analysis. Lecture Notes in Notes in Computer Science (Including Subseries Lecture Notes in artificial intelligence and lecture notes in Bioinformatics), 3973 LNCS, 255-260. https://doi.org/10.1007/11760191_37.

[13]. Bitter, C., north. J . Elizondo, D.A., & Watson T. (2012). An introduction to the use of neural networks for networks for network intrusion detection. Studies in computational Intelligence, 394, 5-24https://doi.org/10.1007/978-3-642-25237-2-3.

[14]. Chang. R. I., Lai, L. Bin, &Kouh, J. S. (2009). Detecting network intrusions using signal processing with query-based sampling Filter. Eurasip Journal on Advances in Signal Processing, 2009. https://doi.org/10.1155/2009/735283.

[15]. Chatzigiannakis, V., Androulidakis, G., &Maglaris, B. (2004). A Distributed Intrusion Detection Prototype using Security Agents. HP Open View University Association, June 2014.

[16]. Corral, G., Llull, U. R., Herrera, A. F., Management, H., Ignasi, S., &Llull, U. R. (2007). Innovations in Hybrid Intelligent Systems (--) Proceedings of the 2nd International Workshop on Hybrid Artificial Intelligence Systems (HAIS'07).

44/2008(June 2014). https://doi.org/10.1007/978-3-540-74972-1.

[17]. Ghosh, A. K., Michael, C., & Schatz, M. (2000). A real-time intrusion detection system based on learning program behavior. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 1907, 93-109. https://doi.org/10.1007/3-540-39945-3_7.

[18]. Kotenko, 1. V, Konovalov, A., &Shorov, A. (2010). Agend-based Modeling and Simulation of Botnets and Botnet Defense. In Conference on Cyber Conflict (pp. 21-44). http://ccdcoe.org/229.html.

[19]. Rajani, P., Adike, S., & Abhishek, S. G. K. (2020). ARTIFICIAL INTELLIGENCE: THE NEW AGE. 8(2), 1398-1403.

[20]. Venkatesh, G. K., Nadarajan, R. A., Venkatesh, G. K., Nadarajan, R. A.. Botnet, H, Using, D., Learning, A. (2017). HTTP Botnet Detection Using Adaptive Learning Rate Multilayer Feed-Forward Neural Network To cite this version: HAL Id: hal-01534315 HTTP Botnet Detection using Adaptive Learning Rate Multilayer Feed forward Neural Network.

[21]. Dyson, B. '2020. COVID-19 crisis could be 'watershed' for cyber insurance. says Swiss Re exec. https:// www:spglobal.com/matketintelligence/en/news-insights/lates-news-headlines/corid-19-crisis-could-be-watershed-for-cyber-insurance-says-swish-re-exec-59197154. Accessed 7 May 2020.

[22]. Fang, Z.J., M.C. Xu, S.H. Xu, and T.Z. Hu. 2021. A framework for predicting data breach risk: Leveraging dependence to cope with sparsity. IEEE Transactions on Information Forensics and Security 16: 2186-2201. htps://doi.org/10.1109/tifs.2021.3051804.

[23]. Field, M. 2018. WannaCry cyber attack cost the NHS £92m as 19,000 appointments cancelled. https:// www.tclegraph.co.uldtcchnolo.4/2018/10/11/wannacry-cybcr-attack-cost-nhs-92in-19000-appointments-cancelled/. Accessed 9 May 2018.

[24]. Hemo, B., T. Gafni, K. Cohen, and Q. Zhao. 2020. Searching for anomalies over composite hypotheses. IEEE Transactions on Signal Processing 68: 1181-1 196. https://doi.org/10.1109/ISP.2020.2971438

[25]. Husak, M., M. Zadnik. V. Bartos. and P. Sokol. 2020. Dataset of intrusion detection alerts from a sharing platform. Data in Brief33: 106530.

[26]. Khan. 1.A.., D.C. Pi, A.K. Bhatia, N. Khan, W. Haider, and A. Wallab. 2020. Generating realistic loT¬based MS dataset centred on fuzzy qualitative modelling for cyber-physical systems. Electronics Letters 56 (9): 441-443. https://doi.or010.10,19/cl.2019.4158.

[27]. Report Concerning Space Data System Standards - Security Threats Against Space Missions. The Consultative Committee for Space Data Systems, CCSDS 350.1-G-3, February 2022.

[28]. The future of the European space sector - How to leverage Europe's technological leadership and boost investments for space ventures, European Investment Bank, 2019.

[29]. INCOSE - International Council for Systems Engineering Systems Security Engineering: Mission and Objectives, July 2021.

[30]. Jaekel, S., B. Seholz. Utilizing Artificial Intelligence to Achieve a Robust Architecture for Future Robotic Spacecraft, 2015 IEEE Aerospace Conference, 07-14 March 2015.

[31]. Manning, J., Langerman, D., Ramesh, B., Gretok, E. Wilson, C.M, George, A.D., Mackinnon, J. and Crum. G., Machine-Learning Space Applications on SmallSat Platforms with TensorFlow, 32nd Annual AIAA USU Conference on Small Satellites, 04-09 August 2018.

[32]. Qiu, S., Liu, Q., Zhou, S. and Wu, C., Review of artificial intelligence adversarial allack and defense technologies, in Applied Sciences, 9(5), 2019, .909.

[33]. J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (IoT): a vision, architectural elements, and future directions," Future Generation Computer Systems, vol. 29, no. 7, pp. 1645-1660,2013.

[34]. M. Yildirim, "Artificial intelligence-based solutions for cyber security problems," in Artificial Intelligence Paradigms for Smart Cyber-Physical Systems, pp. 68-86, IGI Global, 2021.

[35]. K. K. Patel and S. M. Patel, "Internet of things-1OT: definition. characteristics, architecture, enabling technologies, application & future challenges," International Journal of Engineering Science and Computing, vol. 6 , no. 5,2016.

[36]. G. Misra, V. Kumar, A. Agarwal, and K. Agarwal, "Internet of things (iot)-a technological analysis and survey on vision, concepts, challenges, innovation directions, technologies, and applications (an upcoming or future generation computer communication system technology)," American Journal of Electrical and Electronic Engineering, vol. 4, no. 1, pp. 23-32, 2016.

[37]. P. Sethi and S. R. Sarangi, "Internet of things: architectures, protocols, and applications," Journal of Electrical and Computer Engineering, vol. 2017, Article ID 9324035, 2017.

[38]. D. Zeng, S. Guo, and Z. Cheng, "The web of things: a survey, journal of Communications, vol. 6, no. 6, pp. 424-438, 2011.

[39]. H. Liu, C. Zhong, A. Alnusair, and S. R. Islam, "FAIXID: a framework for enhancing ai explainability of intrusion detection results using data cleaning techniques," journal of Network and Systems Management, vol. 29, no. 4, pp. 1-30, 2021.

[40]. A. Gupta, R. Gupta, and G. Kukreja, "Cyber security using machine learning: techniques and business applications," in Applications of Artificial Intelligence in Business, Education and Healthcare, pp. 385-406, Springer, Cham, 2021.

[41]. S. Soderi, "Acoustic-based security: a key enabling technology for wireless sensor networks," International Journal of Wireless Information Networks, vol. 27, no. 1, pp. 45-59, 2020.

[42]. Dash, P. Karimibuiki, M. Pattabiraman, K: Stealthy attack against robotic vehicles protected by control-based intrusion detection techniques, J. Digit. Threats Res. Pract. 2(1). 1-25 (2021)

[43]. Mitchell. R.. Chen. I-R.: A survey of intrusion detection techniques for cyber-physical systems. ACM Comput. Surv. (CSUR) 46(4), 55(2014)

[44]. Guiochet, J., Machin. M..Waeselynck, H.: Safety-critical advanced robots: a survey. Robot. Auton. Syst. 94, 43-52(2017)

[45]. Rubio, J.E.. Alcaraz, C., Roman, R., Lopez, J.: Current cyberdefense trends in industrat control systems. Comput. Secur 87, 101561(2014)

[46]. Kamel, M.A., Yu. X., Zhang, Y.: Formation control and coordination of multiple unmanned ground vehicles in normal and faulty situations: a review. Annu. Rev. Control 49, 128-144(2020)

[47]. Hong, J.H. Maison, E.T. Taylor, J.M. Design of knowledge based communication between human and robot using ontological semantic technology in firelighting domain. In: Robot Intelligence Technology and Applications, vol. 2, pp. 311-325. Springer (2014)

[48]. Wilson, C.. Improvised explosive devices in Iraq: effects and countermeasures. In: CRS Report for Congress. Library of Congress Washington DC Congressional Research Service (2005)

[49]. Khan, N.. Fahad, S., Naushad, M.. Faisal, S.: Analysis of Arminia and Azerbijan war and its impact on both countries economies. Alaitible at SSRN 3709,32(2020)

[50]. Zeng, Z. Chen, P.-J.. Lew, A.A.: From high-touch to high-tech Covid-19 drives robotics adoption. Tour. Geogr. 22,1-11(2020)

[51]. Wang, C., Carzaniga, A. Evans, D., Wolf, A.: Security issues and requirements for internet-scale publish-subscribe systems. In HICSS, p. 303. IEEE (2002)

[52]. Wang. X. Mal-Sarkar, T. Krishma. A.. Narasimhan. S., Bhunia. S. Soltware exploitable hardware Trojans in cmbedded processor. In: 2012 IEEE International Sy mposium on Defect and Fault Tolerance in V1.S1 and Vanotechnolosy Systems (DFT), pp. 55-58. IEEE (2012)

[53]. Bencsíth. B..Péh. G. Buttyan. L., Felegyhazi. M.: The cousins of stuxnet: Duqu, flame, and gauss. Future lnternet 4(4), 971-1003 (2012)

[54]. Navas. R.E., Le Bunder. H Cuppens, N. Cuppens, F Papadopoulos, G.Z.; Do not trust your neighbors! a small IoT platform illustrating a man-in-the-middle attack. In. International Conterence on Ad-Hov Networks and Wireless. pp. 120-125 Springer (2018)

[55]. Haldernan, J. A. Schuen, S.D. Heninger, N. Clatken. W Pant. W. Calandrino, J. A. Feldinn, A.J. Appelbatm, J. Feiten, E b Lest we remember: cold-bout attacks on encryption hess. Cum. mun. ACM 52(5), 91-98 (2009)

[56]. Barcena, M.B., Wueest. C.: lnsecurity in the internet of things. Security Response, Symantec (2015)

[57]. Jiang, D., Omote, K.: An approach to detect remote access trojan in the early stage of communication. In: 2015 IEEE 29th International Conference on Advanced Information Networking and Applications (AINA), PP. 706-713. IEFE (2015)

[58]. Turner, A.. Glantz, K., Gall, J.: A practitioner-researcher partnership to develop and deliver operational value of threat, risk and vulnerability assessment training to meet the requirements of emergeney responders. J. HomelSecur. Emerg. Manag. 10(1) 319-332(2013)

[59]. Moalla, R., Labiod. H. I.me. B.. Simoni, N.: Risk analysis study of its communication architecture. In: 2012 Third International Conference on the Network of the Future (NOF). pp. 1-5. IEEE (2012)

[60]. Diab. M..Pomarlan. M. Beßker. D.. Akbari. A..Rosell, J. Bateman, J.. Beetz. M.: Skilman-a skill-based robotic manipulation framework based on perception and reasoning. Robot. Auton. Syst. 134, 103653(2020)

[61]. Choi, H., Kate, S., Aaffer. Y. Thang, X.. Xu, D.: Software-based realtime recovery from sensor attacks on robotic vehicles. In: 23rd International Symposium on Research in Attacks, Intrusions and Defenses (R A I D 2(020), pp. 349-364(2020)

[62]. Beaudoin, L..Avanthey, L.. Villard. C : Porting ardupilet to esp32: lowards a universal open-source architerime for agile and easily replicable multi-domains mapping robots In. Arch. Photogramm. Remote Sens Spat Inf. Sci. 43,933-939(2020)

[63]. Samangouei, P., Kabkab, M., Chellappa, R.: DEFENSE-GAN: PROTECTING CLASSIFIERS AGAINST ADVERSARIAL ATTACKS USING GENERATIVE MODELS.

[64]. Sun, H., Zhu, T., Zhang, Z., Jin, D., Xiong, P., Zhou, W.: Adversarial Attacks Against Deep Generative Models on Data: A Survey. IEEE Transactions on Knowledge and Data Engineering. 35, 3367–3388 (2023). https://doi.org/10.1109/tkde.2021.3130903.

[65]. Jaiswal, S.: Machine Learning Tutorial | Machine Learning with Python - Javatpoint, https://www.javatpoint.com/machine-learning. guest_blog: Machine Learning: Adversarial Attacks and Defense, https://www.analyticsvidhya.com/blog/2022/09/machine-learning-adversarial-attacks-and-defense/.

[66]. IBM: What is Machine Learning?, https://www.ibm.com/topics/machine-learning.

[67]. Anant Jain: Breaking neural networks with adversarial attacks, https://towardsdatascience.com/breaking-neural-networks-with-adversarial-attacks-f4290a9a45aa.

[68]. Kabkab, M.: Defense-GAN: Protecting Classifiers Against Adversarial Attacks Using Generative Models, https://github.com/kabkabm/defensegan, last accessed 2023/10/02.

[69]. Short, A., La Pay, T., Gandhi, A.: Defending Against Adversarial Examples., https://www.osti.gov/servlets/purl/1569514.

[70]. Goodfellow, I., Papernot, N.: Attacking machine learning with adversarial examples, https://openai.com/research/attacking-machine-learning-with-adversarial-examples.

[71]. Comiter, M.: Attacking Artificial Intelligence: AI's Security Vulnerability and What Policymakers Can Do About It, https://www.belfercenter.org/publication/AttackingAI.

[72]. 莫战A.M.: AI Safety — How Do you Prevent Adversarial Attacks?, https://towardsdatascience.com/ai-safety-how-do-you-prevent-adversarial-attacks-ede17480a24d.

[73]. U.S. Department of Justice, https://www.justice.gov.Law, C.: Computer and Internet Fraud, https://www.law.cornell.edu/wex/computer_and_internet_fraud.

[74]. McNicholas, E., Angle, K.: Cybersecurity Laws and Regulations USA, https://iclg.com/practice-areas/cybersecurity-laws-and-regulations/usa.

[75]. EES: Cybersecurity laws and regulations in US 2021, https://www.eescorporation.com/cybersecurity-laws-and-regulations-in-us/.
Bharara, P., Anderson, J., Jackson, R., Waxman, H., Cucinella, B.: BRIEF FOR THE UNITED STATES OF AMERICA. (2015).

[76]. SUPREME COURT OF THE UNITED STATES. (2020).

[77]. Shankar, R., Kumar, S., Penney, J., Schneier, B., Albert, K., School, H.: Legal Risks of Adversarial Machine Learning Research Legal Risks of Adversarial Machine Learning Research Legal Risks of Adversarial Machine Learning Research. (2020).