ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal) | Impact Factor: 7.512 |

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

The Role of Explainable AI in Cybersecurity: Addressing Transparency Challenges in Autonomous Defense Systems

Sundar Tiwari[1], Vishal Sresth[2], Aakash Srivastava[3]

Independent Researcher[1]

Independent Researcher[2]

Independent Researcher[3]

ABSTRACT: Two major and rapidly evolving domains that are becoming more integrated into military and defense activity, are cyber security and Artificial Intelligence. Cyber threatening has become sophisticated thus making the application of AI crucial in identifying as well as preventing assaults. In this paper, a review of advances in the area of cyber security and AI and their application relating to war and defense is provided. The work begins by providing an outline of the kinds of AI models deployed in cybersecurity, such as machine learning, natural language processing, and anomaly detection. This work then goes further to explore how AI helps in the improvements of military and defence system security, especially in regard to the recognition of threats in the cyber space as well as protective the critical infrastructure. Some of the challenges associated with AI in cybersecurity, which we then discuss are; A lot of challenging attempts such as adversarial attacks, the problem of the difficulty of interpreting results produced by AI. Furthermore, general and specific ethical issues using AI within military and defense domains are reviewed, including issues relating to responsibility, bias, and audacity. Finally, the necessities for further investigation of this area of study and also proposing some research questions for future research to give specific attention to the followings are mentioned: the algorithm development for learning adaptive AI, and the consequences of AI in short and long-term operation of military/defense application. In conclusion, this paper gives a fairly good picture of where artificial intelligence (AI) and cyber security stands in military/defense today and what some of the big challenges and future directions for additional research are likely to be.

KEYWORDS: Explainable AI (XAI), Cybersecurity Transparency, Autonomous Defense Systems, AI Explainability Challenges, AI in Cyber Defense.

I. INTRODUCTION

Definition of Explainable Artificial Intelligence and Significance of Explainable Artificial Intelligence

The term XAI involve AI techniques and models that are capable of offering human understandable explanations of their reasoning processes. XAI aims at making AI Systems more explainable, trustworthy and comprehensible to the human side, particularly when AI integrated to sensitive fields such as health, finance, and law and auto mobile systems.

The Significance of Explainable Artificial Intelligence

Due to AI algorithms that are applied to important aspects of human life such as healthcare, credit scoring and loan processing, model explainability has become necessary. This is important for the purpose of maintaining high ethical, judicial and safety measures. While there are multiple reasons for the importance of XAI, this study highlights three primary concerns: Cultures for expertise, trust, transparency and bias and fairness in artificial intelligence systems.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal) | Impact Factor: 7.512 |

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

Interpretability in the current processes is conducted as a warm-up before deploying an ML model. Yet, this method is typically used only for quite simple models such as the models built by a method called decision trees. Due to the increased complexity of deep learning algorithms with millions of parameters, new XAI methods require a broad consideration of these challenges.

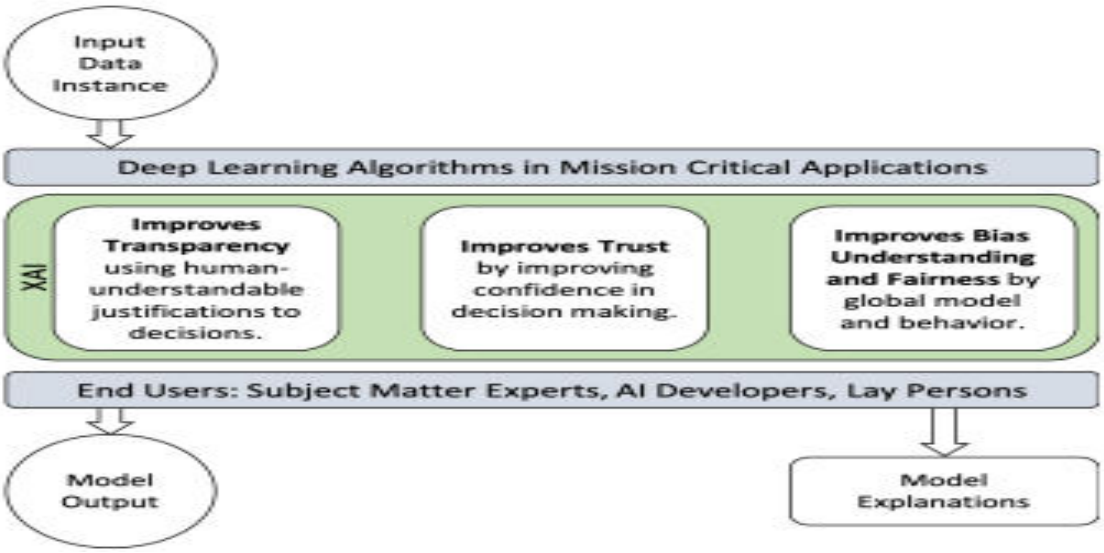

Figure 1: Significant expected improvements when using XAI techniques to support decision making of end-users. We believe XAI is important due to improvements in trust, transparency, and in understanding bias and fairness.

1. Enhancing Transparency

As for defenses, XAI can contribute to the amplification of model interpretability and provide human-read humans explanations of the actions taken by the model, which will help to see potential adversarial attacks. For instance, adversaries can fool a classifier into giving wrong results, for instance, categorizing a manipulated pictorial of a panda as that of a gibbon. AI systems do not hide their errors and allow the stakeholders to adjust according to the predictions made and protect themselves against such an attack.

The flow features include organizational transparency and structural transparency, where the former can be implemented within the algorithm or by other measuring techniques such as simulations or model decomposition. This is especially important as self-driving systems incorporate an even bigger part in day-to-day affairs, where providing purposeful explanations in visual or textual form is significant.

2. Building Trust

Trust in AI models means it has to give and can only give a reasonable and rationalization to why that decision may not be the best. The credibility of a model is the extent that end-users – developers, specialists in the topic field, policymakers and ordinary citizens have confidence in its ability to perform in the real world.

Considering the question of "why" a given decision was made is important to creating trust among users. Simple and easily understandable explanations of model predictions are particularly important because a confident acceptance of AI is expected as solutions that use artificial intelligence become more and more widespread.

3. Counteracting Prejudice and Fighting Discrimination

Through XAI, bias that is likely to come up due to data collection techniques or a wrong model design is also pointed out. Bias is a situation where an artificial model will favor or otherwise penalize certain groups within the inputs provided.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal) | Impact Factor: 7.512 |

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

In fact, using the explanation of AI models on various input data distributions, XAI reveals practical information about biases and skewness in datasets. As we will see in this paper, this understanding allows for the development of approaches that minimize biases and create fair and sound AI solutions. The second aspect of establishing justice in AI concerns treating the data under consideration and making decision fairly without discriminating any population group represented in that data.

1.2 Overview of Cybersecurity Challenges

Cyber threat environment is known for a wide variety of threats that only deepen the difficulties of protecting digital assets. Key challenges include:

Evolving Threats: The threat types ranging from ransomware, phishing, DDoS, and APTs have got more evolved, thereby can now no longer be prevented using normal security measures.

Volume of Data: Enterprises produce a massive amount of data, making monitoring for security incidents more of a challenge than a feasible process. This makes it difficult to develop and maintain an appropriate response mechanism to these threats since identifying viable threats from an ocean of noise often comes after the attacks have started.

Complexity of IT Infrastructure: The use of cloud computing, IoT, and remote working has resulted in raising the complexity and distribution of organizations' IT structures, and in turn, increases the risks that can compromise data security in the different environments.

Shortage of Skilled Security Professionals: Availability of qualified cybersecurity expertise aggravates the condition even while attempting to defend against such voyages. This lack of sufficient human resources in the security teams results in high workloads and likely hood of missing out on threats.

1.3 Role of AI in Cybersecurity

Cyber threats counter has been revolutionized by Artificial Intelligence. Leveraging AI in cybersecurity allows for:

Real-Time Threat Detection: Real-time applications can be solved with AI-driven systems because such systems can analyze millions of records per minute and determine a threat in an instant, providing short TTD.

Predictive Capabilities: Artificial intelligence models can review data of previous cyber-attacks, analyze and potentially forecast the next threats from hackers.

Automation of Security Tasks: AI assists security organisations to automate many functions in threat identification, assessment and mitigation, thus easing the pressure on human personnel and the entire system.

Adaptive Security: Second, the AI models themselves are dynamic and able to learn new strategies, unlike the 'Rule Based Systems' on which cyber criminals will focus, as a strategy they are less able to learn.

Mitigating Human Error: AI can eliminate the possibility of people making mistakes – human error, while still a major contributing factor to cyber threats, and can constantly monitor and quickly respond to emerging threats.

1.4 Problem Statement

Due to the increased integration of artificial intelligence (AI) in vital applications of human life, including medicine, finance, and law, their architectures become complex and their vulnerability to adversarial alteration increases thus masking their functioning. Out of these systems, the high-performing 'black-box' systems present deep ethical, judicial, and safety issues. More often, users and stakeholders struggle to trust and integrate these technologies because there is no justification for such forecasts.

Preset in the datasets or inherited by a chosen model architecture, prejudice deepens fairness problems that can be potentially discriminatory. Recurrent AI models, especially deep learning ones including those with millions of parameters, do not have built-in interpretability, thus it would be implausible to assess their behavior or ward off adversarial threats. Interpretability schemes presently remain accessible to more rudimentary machine learning algorithms leaving a fundamental gap in advanced intelligent systems.

Such lack of explainability risks hinder the AI adoption by preventing trust, transparency and fairness causes. As society shifts to using AI solutions in different fields today, there is a critical need for research on XAI to solve these challenges and foster the right, just, and safe utilization of AI technologies.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal) | Impact Factor: 7.512 |

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

1.5 Objectives of the Research

This study is main aim will focus on trying to solve the problems that occur due to the absence of interpretability in artificial intelligent systems. The study aims to:

a. Enhance Transparency: Organize methods for making AI models more Interpretable, constructing human-interpretable explanations for the automation made by models, and enabling the generation of prediction quality for users to gauge as well as adversarial tricks.

b. Build Trust: Develop guidelines that will enhance the recognition of the systems based on artificial intelligence by creating trust through the presentation of understandable models that drive the system and thus enhance acceptance among the developers, domain specialists as well as policy makers.

c. Promote Fairness and Mitigate Bias: Eliminate preconceptions that can be brought by the data or model's design so an AI system gives fair results across the demographic strata.

1.6 Scope of the Research

This research relates to explaining AI systems especially in hard-to-explain for high-risk applications including medical diagnoses, financial reports, and self-driving cars. The scope includes:

1. Development of XAI Techniques: Investigating and enhancing the techniques of fairness and hot local interpretabilities particularly for extreme nonlinearity and large models with millions of parameters.
2. Transparency and Adversarial Defense: Developing approaches to offer non-digital explanations of its actions so that adversarial manipulations can be identified and prevented.
3. Building Trust in AI Systems: Developing procedures and benchmarks that will help in evaluating user confidence in the ability of AI to make prediction while maintaining accuracy and tapering off on their complexity in practice.
4. Bias and Fairness Analysis: A closer look at the causes of bias in inputs and structure of AI models, and the special measures to minimizing biased outcomes.
5. Cross-Domain Applicability: Ensuring that the developed XAI methods can generally be applied across different domains, some of which include; health, banking and finance, and law where issues such as transparency, trusting and fairness matter paramount.

II. LITERATURE REVIEW

Existing Research on Application of Artificial Intelligence and Safety challenges of AI in Autonomous Systems.

2.1 Applications of AI in Cybersecurity

Artificial Intelligence (AI) is becoming valuable in cybersecurity to create new solutions to the existing security issues. This section discusses the various ways that AI has been used in cybersecurity and wants to recognize how these technologies help secure networks.

1. Real-Time Threat Monitoring

It also allows organizations to track security threats in their networks and within systems in real-time. This type of systems applies machine learning algorithms to data from various sources as network traffic, system logs or user's behavior and looks for patterns that could reveal an on-going attack.

Key Features:

Behavioral Analysis: AI systems set up patterns of users and systems' expected behavior which if breached will indicate a potential security breach.

Automated Alerts: Potential threats, when detected, can also produce alerts to the cybersecurity teams, thereby giving them an easier time as they conduct analysis.

Integration with Security Information and Event Management (SIEM): AI improves SIEM systems in many ways including; advanced analytics since it helps correlate events from several sources in order to provide a holistic view of the threat space.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal) | Impact Factor: 7.512 |

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

2. Brief on Predictive Analytics for Cyber Security

Security analytics is an AI and machine learning application used to estimate future security events to happen. Because it involves using past data and designing patterns to forecast future problems that may be hazardous to an organization, AI can guide entities to be averse to such risks as well as assessing effective ways of handling detrimental events.

Key Features:

Threat Intelligence: Threat intelligence feeds can be used directly with AI systems to have the AI assimilate threat intelligence data into its predictive algorithms to look for potential trends of new threats and weaknesses.

Risk Assessment: Security managers can apply predictive algorithms in order to discover the levels of risk the organization is exposed to and allocate security defenses accurately to the most exposed items.

Scenario Simulation: Or, AI can reproduce different types of attacks on an organization in order to identify the flaws and improve the protection measures.

3. Artificial Intelligence for Cybersecurity Automation

Self-learning automation on the part of AI relieves the enormous pressure on staff specializing in cybersecurity by performing most of the repetitive tasks. The above automation can cause them to respond better and manage incidents faster.

Key Features:

Incident Response: AI can take response measures to many security incidents like isolating systems or blocking IPs which are malicious based on rule and behavior learned.

Threat Hunting: Implementing the AI decision-making system in conjunction with automated threat hunting procedures increases the general threat detection rate by targeting opportunities that were not previously detected by conventional security frameworks.

Security Policy Enforcement: AI systems can automate the effecting of security policies on network and endpoints and this will remove the element of human error.

4. AI in Phishing Detection

Phishing attack, specifically, is still a key threat to organizations, especially one which exploits the human element. AI provides superior algorithms to define cases of phishing and distinguishes between genuine and fake e-mail messages, users' actions and interactions.

Key Features:

Content Analysis: Automated systems can scan the content and form of the messages to look for such features as typical for phishing activity, for instance, the sender's addresses that are suspicious, or the links involved.

User Behavior Analytics: As for the specific measures, it has to be mentioned that AI systems can detect signs of phishing attempts based on analyzing user's behavior and activities as well as possible changes in users' access patterns, including abnormal login attempts.

Training and Awareness: AI is also useful for creating individual training for the users, which will help them to distinguish phishing and raise security level.

2.2 Safety Challenges of AI in Autonomous System

AI has shifted to many different safety related fields, including land transport, maritime, drone technology and many others. This trend has elicited interest in the recent scientific literature. Nevertheless, the issue of safety AI is still quite an emerging concept, let alone a successful one. It is therefore impossible to discuss safety in conjunction with AI without touching on ethics besides legal and regulatory, as well as the general admissibility of AI. In the literature, major safety barriers of AI applications for autonomous systems have been revealed in this research, They indicate deficiencies that are resisting the improvement of AI applications.

The special emphasis of this paper lies in several AI safety risks described in the works that are considered canonical. Even though, other challenges have been proposed in the literature, the ones explained here are among some of the most significant for safety-critical autonomous systems. These challenges are interpreted here within the context of autonomous systems as follows:

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal) | Impact Factor: 7.512 |

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

i. Minimizing adverse side effects
ii. A reference to keeping the adverse effects of the system on society and its ability to deliver its goals minimal.
iii. Avoiding Reward Hacking
iv. Safeguard against the system gaming its reward structure and work towards undesired results.

III. THEORETICAL FRAMEWORK

3.1 Challenges and Limitations of AI in Cybersecurity

The known benefits of AI-based cybersecurity systems are numerous, but so are the threats and considerations the solution has to be able to overcome if it is to be reliable and efficient.

1. Issue of adversarial attacks on AI systems.

Adversarial attacks are considered the main threat to AI systems at the present stage, including cybersecurity. These attacks are specifically designed to modify input data so as to make the machine learning model make wrong decision or assignments. In the context of cybersecurity, the input can be designed by the adversaries in a specific way that human or any artificial intelligent based systems are not capable of detecting them and ultimately be exposed to the threats or in fact the adversary may be able to penetrate through the defenses. Therefore, adversarial authenticity's use reveals how AI models are vulnerable to such manipulations and important training and validation to improve AI resilience. Academics are now in the process of searching for approaches to mitigate these kinds of attacks because the bad guys are constantly mutating their strategies, making it a kind of game of cat and mouse.

2. Processing of Personal Data in Smart Spaces and Informatics: Data Privacy and Ethical Issues

AI penetration in the context of cybersecurity design also introduces severe data protection and ethic issues. AI depends on data; therefore, they require massive data, which could contain personal data of an individual. Mist implementation of such features might lead to violation of people's privacy especially if full reliance is made on the data collected. Also, issues of fairness are usually questioned, as to how AI algorithm arrives at certain decisions, especially where in its outputs could mean gains or losses by individuals based on their false positive/negative outcome. To attend to such apprehensions, one must embrace and implement transparency, accountability and mandatory compliance with rules governing data privacy while creating AI led cybersecurity systems.

3. See Integration and Scalability Issues

AI technology can be easily incorporated into existing structures of cybersecurity systems and applications. Large organizations have established architectures that may not seamlessly integrate with the new AI solutions therefore creating barriers to the effective use of the solutions. Third, and specifically as a function of data volume, it is increasingly important that AI systems be able to accommodate new data at a large scale without reducing their efficacy. To address these integration and scalability issues, organizations have no option but to adapt by coming up with a plan on how, when, and where they will invest in infrastructure and train their personnel, and in the process propagate the culture of innovation in cybersecurity practices.

3.2 key Concepts of Explainable AI Principles

Fig. 2: High-level illustration of deep learning model f. Generally, a single input instance x generates outputs $\bar{y}$. No other metadata or explanations are generated other than the output classification. Most model inference scenarios involve this method where model f is considered as a blob of information which takes an input x and generates an output $\bar{y}$.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal) | Impact Factor: 7.512 |

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

Several previous works discuss the subtleties of differentiation between Explainability and Interpretability of neural networks. We affirm the general idea of XAI as a set of methods and models intended for enhancing the reliability and explainability of artificial intelligence systems. Definitions refer to additional information from the AI model that provides understanding of an often decision made by the AI or the working of the AI model in general. The literature review of Explainability approaches used with Deep Neural Networks can be seen in the following sections of this paper. Figure 2 presents one type of deep learning model that consumes one input and produces one output prognosis. The goal of XAI when applied to deep neural networks is toward explaining these predictions using different techniques captured by this survey. Generally, for an input x ∈ Rd, a deep learning model function f (θ) describes f: From the problem of classification, the function is between R d and R C (C is the number of output classes and θ is the parameters of the model). In current proposition, the model inference can be stated as $\bar{y} = f(\theta, x)$ with the prediction of $\bar{y}$ as the output. We now provide a working definition of the main topics discussed in the survey, namely explainability of deep learning models. Subsequent parts of the survey elaborate on these definitions.

Definition 1: We define interpretability as the desirable quality or feature of an algorithm that yields sufficient expressive data relative to its functioning. Here it is possible to define interpretable domains as domains which are understandable by a human being, for example, images, text, etc. Cambridge Dictionary defines: "If something is interpretable, it is possible for meaning to be found in the case or in the particular case or in some case".

Definition 2: Interpretation is a mapping of the abstract domain like the outputs produced by a machine learning model to an understandable and reasonable concept. It is possible to interpret output predictions of a simple rule based model by traversing the rule-set. Likewise a small tree in decision tree can be easily interpreted. Or the Deep Convolution Networks (CNN) model that can explain according to the regions of the input image that lead to the decision.

Definition 3: An explanation may be just another meta-information which is computed outside of the application of the algorithm or through the machine learning model about the feature contribution or the relevance of an input record for a given classification. For a deep learning model f with input x and output prediction of $\bar{y}$ of class c, an explanation g can be generated, generally as an explanation map E, where E: R d → R d. Here, g is an object of the same shape as the input which gives the relative importance, or relevance, of that particular dimension relative to class output. In the case of image, the explanation map might also be an equally sized pixel map for text it could be word by word influence score.
Definition 4: When a deep learning model f is white-box, implies that the model parameters θ and the model architecture information are obvious. As such, with white-box model intermediate results are more comprehensible, and debugging is enhanced, promoting trust. However, knowing the architecture and the parameters of the model would not be enough to explain that model. Definition 5: A model f is said to be a black-box if the parameters and the structure of the deep learning network is concealed from the end user. Usually deep learning models deployed on web applications or limited business environments are interfaced with APIs that accept an input form the user and deliver the model's output as formatted into the type of presentation expected from '$\bar{y}$'.

IV. METHODOLOGY

Given the concern to make the review most rigorous and auditable, it was necessary to develop a detailed review protocol. The availability of review protocols minimize the imbalance of power in the review process by checking the freedom of the researcher and offers a guide to the whole process. This systematic review study adapted a method which has also been used in a separate field of research known as software engineering research. The main review process consists of three phases: three sections that we have in the research process which includes planning, conducting and documenting. The protocol described the justification of the study, review questions, approach to search, databases, and inclusion and exclusion factors and approach to collecting, reviewing, and synthesizing data. Questions to review were developed in the planning phase to seek the study objectives that informed the whole study. In use was the Goal Question Metric, to which evidence exists that it is useful in identifying goals of systematic reviews.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal) | Impact Factor: 7.512 |

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

In the light of the study, the terms to be used for searching the articles, keywords, databases, search engines and duration for search were decided. The inclusion and exclusion criteria were also defined. To minimize bias, a priori review methodology was established before four researchers (one principal investigator and two co-investigators, including a methodologist) separately constructed the review protocol. In order to identify the potential databases and the terms to use in the database search, each study member conducted a preliminary search. Each of the individual protocols was then adjusted to meet the needs of this study and then consulted an advisor on the protocol that would be used at this combined stage. The agreed protocol was then forwarded to another external reviewer who operates in the CyberSec domain to get her inputs of corrections. After subsequent corrections the final protocol was developed.

Table 1: Research Questions, Motivations, and Methodologies for AI in Cybersecurity

Research Questions	Motivation	Research Methods
RQ1: What are the Publication trends of AI methods in cybersecurity	To classify studies and assess interest, dominating venues and contributions.	Quantitative (Biometric analysis, trend analysis)
RQ2: What AI methods are used in cybersecurity and what types of issues are they applied for?	To identify various AI methods currently used in cybersecurity. This will identify the most dominant method used.	Quantitative (content analysis, classification)
RQ3: What impacts have these AI methods had on cybersecurity	To identify current impacts of AI on cybersecurity and to analyze and classify existing AI Approaches for cybersecurity with respect to specific reasons they seek to address.	Qualitative (Thematic analysis, case studies)

4.1 Data Collection

RQ1: That is the overall publication trend of AI methods in cybersecurity?

• Data Collection Approach: A quantitative technique will be used for the collection of literature trends data. To conduct this study, a bibliometric analysis will be performed with an aim of identifying the output of the publications over time, the journal and conference where papers were published, and the level of contribution toward AI in the cybersecurity domain. The study data will be retrieved from several academic databases using keywords such as 'AI in cyber security' 'Application of AI in security' among others.

RQ2: Which Artificial Intelligence techniques are utilized in cybersecurity, and for what sort of problems?

• Data Collection Approach: This paper will thus adopt a quantitative content analytical research design in making a systematic conclusion on the methods of AI implemented in cybersecurity. This will include a screening of a few recent papers and sorting them by AI techniques used (machine learning, deep learning, expert systems) and threats covered (intrusion detection, malware analysis).

RQ3: There are questions that can weighted the implications of the AI methods on the field of cybersecurity.

• Data Collection Approach: Exploration research design will be employed in understanding the effects of AI on cybersecurity. Data will be collected from systematic literature review through case descriptions, online questionnaires and semi structured interviews with cybersecurity specialists. Importantly, the following analysis will be devoted to examining the impact of AI on cybersecurity with reference to the practices, efficacy, and solutions of specific concerns in this domain

.

4.2 Case Studies of AI -Driven Cybersecurity Solutions

To illustrate, this paper offers several examples to explain the use of AI in improvement of cybersecurity in various industries. The case studies explored here shed light on AI solutions and their deployment, their impact on security issues, and the results delivered. All the case studies discussed presents a different facet of AI security in relation to computer networks and fraud.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal) | Impact Factor: 7.512 |

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

Case Study 1: AI-Powered Network Security
The present paper focuses on the experience of a vast company that decided to integrate its security system with AI-based network security. The machine learning algorithms are employed to process flow information of the computer network in order to extract features that mark intrusions. This means that the new data incoming into the system is used to teach the model new strategies applied by the attackers thus cutting down on time needed to identify threats. This section explains how implementation occurred, what technologies were used to accomplish it, and what tangible benefits were gained in terms of organizational security.

Case Study 2: AI in Endpoint Protection
Due to the growth of the numbers of employees working from home and use of mobile devices, endpoint protection then proved essential. This case study looks at a mid–sized financial services firm that implemented an AI powered endpoint protection product. Endpoint activity is autonomously analyzed based on behavioral modeling, and there is a search for signs of compromise such as, for example, suspicious attempts to enter a user's credentials. It also worked for the improvement of the institution's capacity to avoid getting into breaches of data and increase capacity for response to incidents. Instead, it focuses on the problems encountered during implementation, the technology applied, and the results obtained.

Case Study 3: AI for Cloud Security
Since organization have moved their applications to cloud, it becomes difficult to secure these structures. This paper focuses on a technology firm that implemented an AI-driven cloud security solution in order to protect its cloud-stored information. The AI system has the ability to analyse risks of user activity, data access patterns and changes made to the cloud configuration. Through integrating the threat spotting and response the company enhanced the surveillance by decreasing the risk of having its information stolen. Here, the information about the technologies used, the plan for the implementation and the expected advantages are given.

Case Study 4: What is Artificial Intelligence and its use in the localization of the fraud and prevention of fraud?
Fraud detection is a must for sectors like finance, e-commercial business, and insurance company. This paper focuses on one which investigates the use of AI in the prevention and identification of fraudulent activities for an e-commerce company. This automated solution is figured by using artificial intelligence to audit programme transactions in real time to detect fraud within certain patterns and contradictions. The use of machine learning escalates the detection techniques thus minimizing fraud in the system but at the same time does not interfere with the user experience. This section concerns the decisions made in the model's architecture, integration's process and the outcomes in fraud minimization and customer satisfaction.

V. RESULTS AND DISCUSSIONS

In this work, five important details were gradually collected: the year of the publication, authors' address and country, the publication that was being used, AI methods that had been applied, and the type of security application that has been described. In addition, those papers that aimed at enhancing the given AI approaches were examined and classified into the following categories. The following section provides a discussion of the results obtained in this study.

A. Publication Trends: Analyzing the obtained results, one can judge on the fact that the number of articles devoted to AI methods to provide cybersecurity has grown considerably during the last few years. AI techniques in cybersecurity started appearing about 2015. Primary articles have comprised only 26 % of the total selected primary studies and all these studies have been published between 2008 and 2015. But there was a leap in the volume of publications after 2016; the number of publications has increased nearly twice a year. This increase started with seven articles in 2015 and increased to fourteen in 2016, and then increase even more in 2017. However, as seen in Fig. 3, 2018 alone had fifty-seven publications which is more than double that of 2017 showing a shift to the use of AI in cybersecurity. This increase in publications can be attributed towards the growing tendency of exploring AI based approaches to cybersecurity issues. A variety of AI approaches including machine learning, deep learning,

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal) | Impact Factor: 7.512 |

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

reinforcement learning, etc. have been studied to enhance cybersecurity measures against ever increasing threat landscape.

B. Techniques and Field of Uses in Cybersecurity; In each of the studies that formed the basis of the review, it was possible to identify over 50 different algorithms. The most dominant algorithms were: Supervised: ANN, CNN, DT, KMeans, KNN, AdaBoost, q-Learning, RF, RNN, SVM. A total of fourteen studies did not clarify the algorithm used to adopt the security measure. The most applied algorithm was SVM, which has been used in 24 studies, while only 14 studies came out using CNN. There were 10 researches employed ANN followed by Random Forest in 9, K-means in 8, and Decision Trees in 7. For convenience and simplicity, the domains of application were grouped into six categories: The types of threats analysed are: (i) intrusion detection and prevention systems (IDPS), (ii) traffic classification systems, (iii) imaging and captcha, (iv) encryption and certification, (v) denial of service attacks, (vi) malware, virus, phishing, etc. This research confirmed that domain or field of IDPS received the most focus with 38(41%) of research problems/issues related to Issues on IDPS while Malware, Virus, Phishing etc. received the focus with 27 (21%) of the studies. Other topics that were also grouped under the above themes include power systems security assessment; covert channels; security games; fingerprint liveness detection; biometric verification including facial verification; malicious webpages; cyberbullying; false data injection; online deception review; information risk assessment; and iris recognition were all grouped as 'others' because each of them was the subject of less than 2 studies. Of all the studies reviewed, 47 of them (36%) could be categorized into this category. To see how each algorithm has been used in the respective areas, the AI methods used were categorized based on domain. The findings showed that amongst the categories focused on IDPS and malware, the ensembling of techniques – the use of two or more AI types – dominated. Interestingly, the method of encryption and certification, and denial of service (DoS), and imaging as well as captcha were often not studied using ensemble methods.

C. Effectiveness of AI Methods on Cyberspace Security: In detection and prevention systems the most notable impact of AI has been the optimization of false positives in IDPS. In the improvement of its detection algorithm, all studies on IDPS cited a lower incidence of false alerts. For instance, one research has revealed that their IDPS approach have a very low false alarm rate of approximately 0.012% which is the lowest on record for impersonation attack detection. One of the issues affecting IDPS was established to be the computational complexity. Evaluations regarding IDPS identified almost all of them reported computational complexity as a problem, and it was inherited from the general AI techniques, especially the machine learning method. Overloading is well known in machine learning as one of the major barriers to computational speed and efficiency. Nonetheless, several studies documented a relatively diminished computational complexity, and the authors also found that feature selection by the genetic algorithm can successfully alleviate this problem in IDPS. In the case of network intrusion detection, the most effective approach that emerged from the use of the ensemble methods was the ability to detect new traffic patterns and anomalies. They also resulted in higher IDPS accuracies of detected malicious activities with comparatively few false alarms and low energy consumption. The AI methods have also shown the ability to identify both traditional and emerging/unknown attacks such as the Controller Area Network (CAN) attacks. Also, the AI more specifically has played a key role of identifying the bad actors. For instance, the clickstream models, where no a priori knowledge or assumptions can be made about the user classes, was shown to accurately predict both the systemic and occasional deviations of users from the assumed pattern as well as their typical behavior. Current systems can determine if an account is malicious by looking at coloration of similar click streams. AI has also helped advance ways of enhancing the privacy of images. For example, deep CNN and SVM have shown enhanced precision and analysis speeds and image manipulation detection and malware image detection. Similarly, more sophisticated methods of detecting malicious Web sites have been developed based on spectral clustering of Web site graphs.

Discussions

AI is the disruptive technology today that has made multiple impacts on cybersecurity solutions or tools; algorithms such as SVM, CNN, and ANN are key drivers of detection accuracy or false alarms, especially in the IDPS and malware detection subfields. This is especially the case since a large number of studies underlines the benefits and versatility of ensemble methods in regard to security issues. Location-wise, the information available dominating the research work

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal) | Impact Factor: 7.512 |

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

since most of the technological advancement has occurring in Asia also spearheading the increased Cybersecurity demands.

The antisubmarine warfare has been enhanced through the use of AI in detection and in this aspect there are some gaps in computation that hinders the performances. That is why optimization techniques, including the use of genetic algorithm in selecting the most important features, may help to overcome these problems. The future of AI in cybersecurity appears bright in light of increased use in threat detection which expands to more unique and accurate identification of new threats; however, more research is needed, as follows, to overcome current challenges and address emerging ethical concerns.

VI. CHARTS, DIAGRAMS, GRAPHS AND FORMULAS

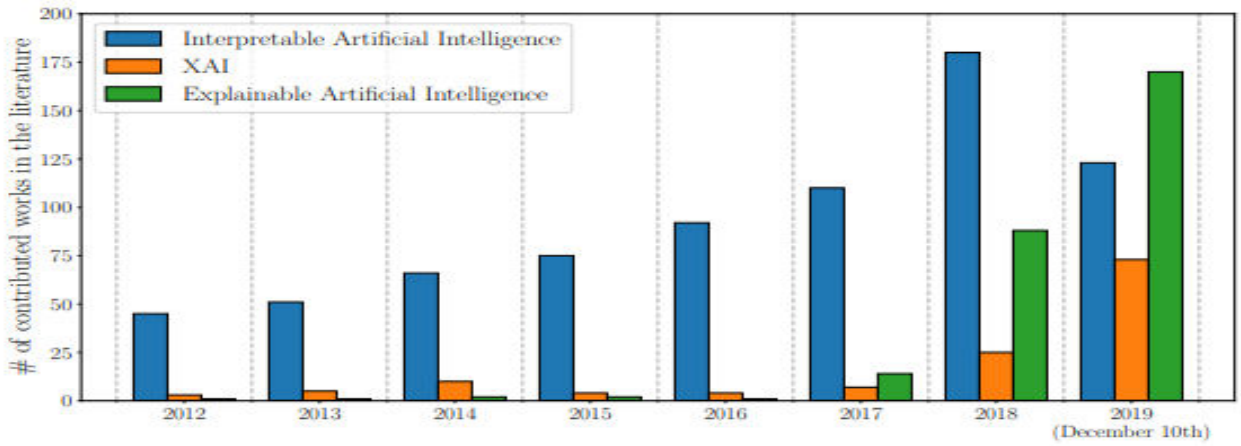

Figure 4: A horizontal graph showing the evolution of the number of total publications whose title, abstract and or keywords refer to the field of XAI during the last years.

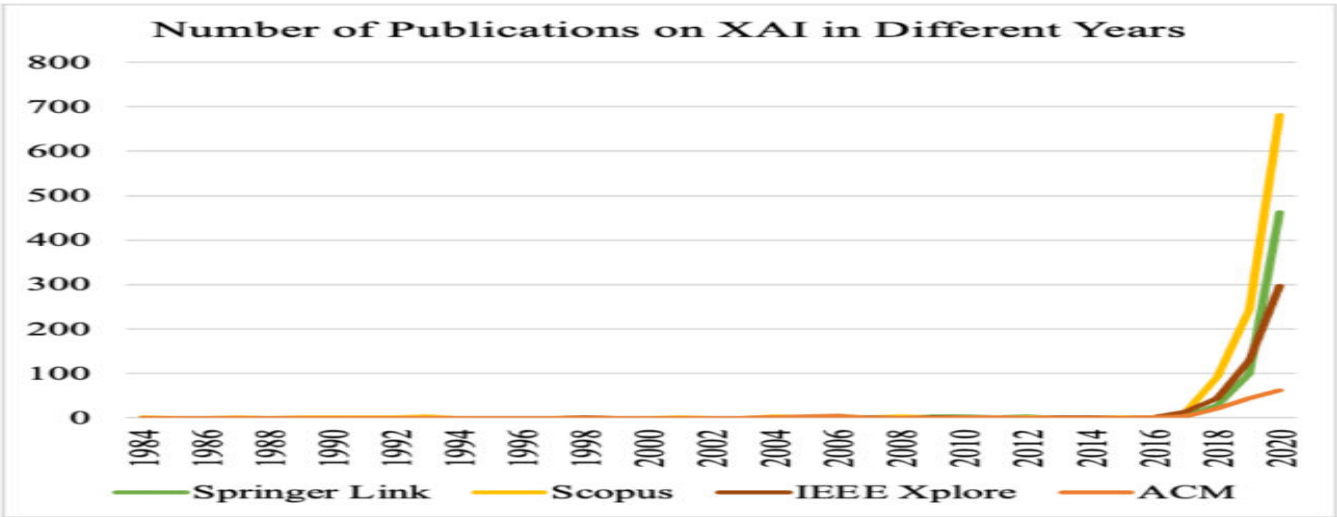

Figure 5: A horizontal graph showing the number of publications on Explainable Artificial intelligence in different years.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal) | Impact Factor: 7.512 |

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

Year Wise Comparison Graph

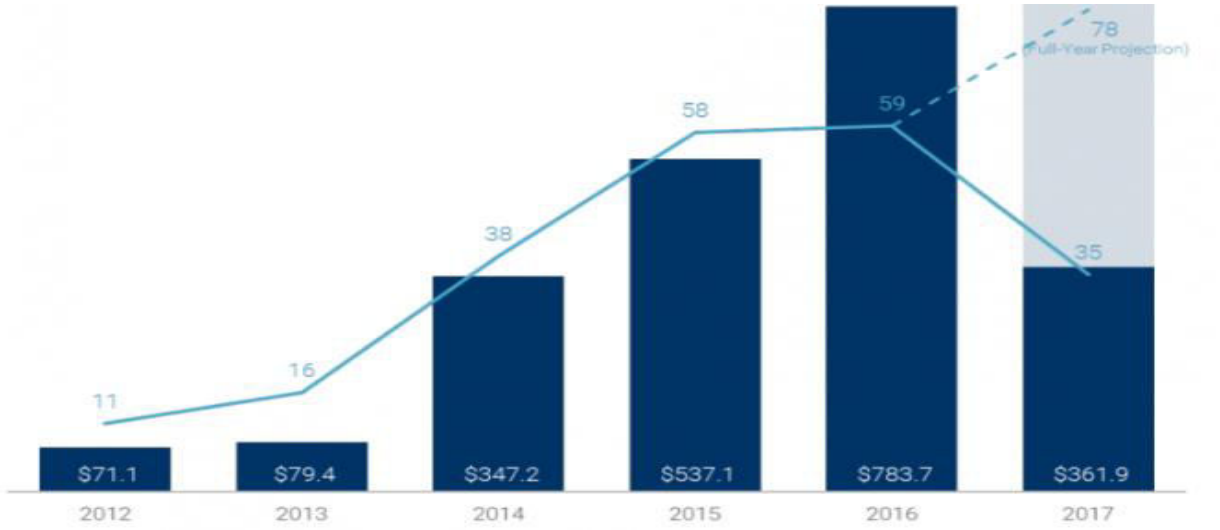

Figure 6: A year wise comparison graph showing the AI methods on cybersecurity

VII. MODEL COMPARISON

In context to AI for cybersecurity, there are few types of machine learning and few types of deep learning models that have been largely used which have their own pros and cons. Compared to each other, the respective models can reveal their efficiency in solving various cybersecurity problems including intrusion detection, malware categorization, and anomaly detection.

1. A particular subset of Bioinformatics, and amongst the most frequently employed algorithms in cyber security, particularly in the case of an intrusion and classification. Its major strength is its capability in the high dimensionality of data and very high classification performance where the problem involves regions in the space which are not linearly separable. However a major issue associated with SVM is that for large data sets, it can be time consuming and computationally expensive to train.
2. Convolution Neural Networks: The original CNNs have shown outstanding performance in image-specific security roles, including malware recognition in images or solving CAPTCHA. Their advantage is the capability to learn spatial hierarchies and relations between the data automatically without the need for feature extraction. However, CNNs are data intensive and need large amount of labeled data for training and over learning is a common problem if there is no strict regularization.
3. Random Forest; This type of learning is a combination of many decision trees to try to increase the classification rate and to fix the overfitting problems. It is useful in the analysis of large datasets which contain much noise, and therefore is ideal for intrusion and malware detection. Nonetheless, it identifies all the important features of the problem but can be less interpretable comparing to simpler models, and could require a lot of time to train.

VIII. IMPACTS AND OBSERVATIONS

1. Enhanced Threat Detection and Response: First, their use of artificial intelligence in analyzing threat intelligence allows the identification of these threats in actual time given the large volumes of data involved. This greatly minimizes the time taken before the threats are detected hence allowing for timely intervention before most attacks occur.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal) | Impact Factor: 7.512 |

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

2. Proactive Cybersecurity Measures: Depending on the type of attack patterns that are studied through historical records, it is possible to provide a measure of prediction. This assists in predicting future threats in order to prepare organizations and take precautionary measures and prevent or reduce or risks of cyber-attacks.
3. Increased Operational Efficiency: Security tasks, for example, threat identification, examination, and counteraction, are repeatedly performed by AI. This automation partially relieves the human security teams to tackle a higher-order network while simultaneously decreasing the timeframe for the action and increasing the overall operational efficiency.
4. Adaptability to Evolving Threats: While unlike common, non-adaptive, rule-based RL systems, AI models are capable of learning and applying updates. On the contrary, AI-driven systems can easily adapt to them because they are designed to churn out cyber defense capability that counters the new strategies used by cyber criminals.
5. Reduction in Human Error: It was also found that one of the main causes of cyberattacks is human mistake. With the help of an AI system, threats can be monitored more closely all the time and formulated more accurately and quickly, which means that no mistakes that threaten security can be made.

Observations
1. Dominance of Machine Learning Models: The research also establishes that eight major machine learning algorithms are widely used in cybersecurity, with SVM, CNN, and ANN on top. These models are chosen due to their effectiveness in the search for patterns, and in pattern recognition and classification, and their ability to learn from Big Data; these algorithms are used in malicious software identification, intrusion prevention, and anomalies detection.
2. Geographical Distribution of Research: The study shows that Asia and especially countries in this region are leading in cybersecurity research with the help of AI. This is obvious, taking into account the growing rate of technological developments and the rising levels of cyber risks... While some regions such as South America and Africa contribute very few studies there is geographical imbalance in cybersecurity research.
3. Varied Application Domains: The paper discriminates AI applications to major areas, and it is noted that IDS&PS and malware detection are revealed to be the most popular areas of research. This is in line with current requirements for efficient and more importantly, automated methods and tools to combat modern sophisticated threats. The other fields like encryption and DoS attack seem to be less related to AI suggesting the fact that there is ample of opportunity to discover more related to these fields.
4. Challenge of Computational Complexity: One common theme throughout the studies is the problem of computational intensity, particularly for deep learning frameworks. Despite these advancements in detection and accuracy AI is still unable to fulfil the computational demand of models like CNN or deep neural networks. This could reduce the applicability of these methods particularly in areas of difficult economic realities.

Practical Implications
Multidisciplinary Focus: A final implication is that all the journals cited in this paper that publish AI and cybersecurity papers are interdisciplinary, rather than discipline-specific cybersecurity journals. However, based on the results, we have to admit that despite the importance of AI and cybersecurity as the research subject, it has not been explored enough in dedicated specialized cybersecurity journals. There is a possibility of seeing the appearance of journals that are exclusively dedicated to the topic of AI applications in cybersecurity since at the present moment this field does not have a sufficient number of specialized publications.

Increasing Focus on Intrusion Detection and Prevention Systems (IDPS): That is why the fact that the majority of researches are targeted at solutions to the problems in Intrusion Detection and Prevention Systems (IDPS) proves the significance of this field in the sphere of cyber defense. Since IDPS is in front-line-defense for physical as well as virtual attacks, handling challenges associated with intrusion detection could translate to appreciable enhancement of the general security infrastructure. This means that the IDPS should invest more funds in the future research of AI as it is connected to the minimization of other kinds of cybersecurity threats.

Malware and Phishing Threats: The emphasis made for malware, phishing and viruses corresponds with the increasing threat posed for such attacks, particularly on Android platform. The large number of tools for malware detection meant especially for Android shows that there is a need to tailor the solutions on mobile platforms. Therefore, conducting a

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal) | Impact Factor: 7.512 |

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

comprehensive overview of the existing AI approaches in securing mobile devices as more and more users resort to using this platform, and the number of threats on the same platform continues to rise.

Limited Research on Denial of Service (DoS) Attacks: The finding that the analysis of the papers considered for the review showed that Denial of Service attacks were a subject of comparatively short discussion is an important issue. However, sophistication in DoS attacks, and especially DDoS attacks that have evolved in recent years, this area of the AI cybersecurity sector has remained relatively untouched. The fact that there are comparatively fewer research works dealing with DoS attacks only may give rise to the presumption that the research may not have reached the stage where all important investigational facets must be explored. Since DoS attacks are becoming a significant threat for crucial infrastructures, researchers might have to extend their works in the corresponding AI-based protection solutions.

IX. FUTURE DEVELOPMENT TRENDS IN CI IDENTIFICATION OF AI TECHNIQUES

Since 2008, papers about integration of the AI techniques in the cybersecurity field can be classified as belonging to a number of broad and wide-ranging categories because the methods in question can be used in a number of fields. This has emerged as an indication that with new cyber threats emerging, it is hard for a study to concentrate on a particular cyber threat domain. Much of the specific related work from within these fields is actively testing a number of different AI approaches and machine learning methodologies. Table 4 shows that algorithms that included here as “other” comprised 36% of reviewed studies. These are algorithms that were used in fewer than two applications and may be an indication that more researchers are testing out new and unique methods. In the future, future research may again pay attention to new methods in real related domains to solve new cybersecurity problems. However, it is necessary to stress once more that there is a demand for applications which are more computationally efficient. It has been clearly identified that the cumulative computational complexity has to be minimized and work performance efficiency increased while searching for new solutions. However, increasing the algorithm in areas like Intrusion Detection and Prevention Systems (IDPS) is strategic since the researchers intend to optimize the performance of the jihadist of cybersecurity products.

X. CONCLUSIONS

AI enhances the capacity to identify threats as they occur, predict attacks and also automates security tasks that cyber security was traditionally associated with. However, there are some obstacles which are related to the computation and the required flexible and robust artificial intelligence models. The field is increasing the research work done in the areas of IDPS and malware detection but lacks sufficient research in other areas such as DoS attack defence. More future research effort should be directed towards applying new AI paradigms, improving computational performance and diverse applications using, for instance, quantum computers. There is a clear requirement for the establishment of more niche-oriented journals for AI in the cybersecurity field to improve concept and idea advancements in this important area.

REFERENCES

1. Apiecionek, Ł., Makowski, W., Biernat, D., & Łukasik, M. (2015, June). Practical implementation of AI for military airplane battlefield support system.
2. Applications of Artificial Intelligence Techniques to Combating Cyber Crimes: A Review Selma Dilek, Hüseyin Çakır and Mustafa Aydın (2015)
3. Bass, L., Weber, I., & Zhu, L. (2015). DevOps: A Software Architect’s Perspective. Addison-Wesley.
4. Chandrasekaran, K. C., & Meghanathan, N. (2017). Big Data Analytics: A Hands-On Approach. CRC Press.
5. Das, A., & Rad, P. (2020). Opportunities and challenges in explainable artificial intelligence (xai): A survey. arXiv preprint arXiv:2006.11371.
6. Dhiman, V. (2019). Dynamic analysis techniques for web application vulnerability detection. Journal of Basic Science and Engineering, 16(1).
7. Dhiman, V. (2020). Proactive security compliance: leveraging predictive analytics in web applications. Journal of Basic Science and Engineering, 17(1).

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal) | Impact Factor: 7.512 |

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

8. Eom, J. -H., Kim, N. -U., Kim, S. -H., & Chung, T. -M. (2012). Cyber military strategy for cyberspace superiority in cyber warfare, Proceedings Title: 2012 International Conference on Cyber Security, Cyber Warfare and Digital Forensic (CyberSec).
9. Fitzgerald, B., Stol, K. J., & O'Sullivan, P. (2014). Continuous software engineering and beyond: Trends and challenges. Information and Software Technology, 56(5), 365-386.
10. Forsgren, N., Humble, J., & Kim, G. (2018). Accelerate: The Science of Lean Software and DevOps: Building and Scaling High Performing Technology Organizations. IT Revolution Press.
11. Haines, M., & Righter, R. (2016). Securing DevOps: Security in the Cloud. O'Reilly Media.
12. Karaman, M., Hayrettin, Ã., & Aybar, C. (2016). Institutional cybersecurity from military perspective.
13. Karvonen, H., Heikkilä, E., & Wahlström, M. (2020). Safety Challenges of AI in Autonomous Systems Design – Solutions from Human Factors Perspective Emphasizing AI Awareness. In Lecture Notes in Computer Science (pp. 147–160). https://doi.org/10.1007/978-3-030-49183-3_12
14. Le, V. H., & Chua, T. S. (2017). A survey on data fusion in the era of big data. ACM Computing Surveys (CSUR), 49(1), 1-42.
15. Lewis, L. (2017). Insights for the Third Offset: Addressing challenges of autonomy and artificial intelligence in military operations.
16. Liu, S., Yu, S., & Guo, Y. (2019). A survey on security threats and defensive techniques of machine learning: A data driven view. Journal of Network and Computer Applications, 131, 36-57.
17. Luiijf, E. A., & Buijs, J. C. (2017). Securing Smart Cities. Springer.
18. M. Karaman , H. Ã‡atalkaya and C. Aybar , "Institutional Cybersecurity from Military Perspective".
19. Mettikolla, P., Calander, N., Luchowski, R., Gryczynski, I., Gryczynski, Z., & Borejdo, J. (2010). Kinetics of a single cross-bridge in familial hypertrophic cardiomyopathy heart muscle measured by reverse Kretschmann fluorescence. Journal of Biomedical Optics, 15(1), 017011-017011.
20. Mettikolla, P., Calander, N., Luchowski, R., Gryczynski, I., Gryczynski, Z., & Borejdo, J. (2010). Observing cycling of a few cross-bridges during isometric contraction of skeletal muscle. Cytoskeleton, 67(6), 400-411.
21. Mettikolla, P., Luchowski, R., Gryczynski, I., Gryczynski, Z., Szczesna-Cordary, D., & Borejdo, J. (2009). Fluorescence lifetime of actin in the familial hypertrophic cardiomyopathy transgenic heart. Biochemistry, 48(6), 1264-1271.
22. Masuhr, N. (2019). Ai in military enabling applications.
23. O'Connell, M. E. (2012). Cyber security without cyberwar.
24. O'Reilly, T., & Battelle, J. (2009). Web Squared: Web 2.0 Five Years On. O'Reilly Media.
25. Parnin, C., & Bird, C. (2016). Usage, costs, and benefits of continuous integration in open-source projects. Empirical Software Engineering, 21(3), 1-35.
26. Pires, M., & Duboc, L. (2017). Towards a DevSecOps process model: Organizational patterns of integration of security in DevOps. Journal of Systems and Software, 130, 141-159.
27. Pombinho, J., & Silva, A. R. (2018). DevSecOps: Shifting Security Left with Continuous Delivery. Proceedings of the 1st International Workshop on Secure Development Lifecycle.
28. Rasch, R., Kott, A., & Forbus, K. D. (2003). Incorporating AI Intelligent Systems, 18(4), 18-26.
29. Ransford, B., Clarke, D., & Duquennoy, S. (2019). Security for the Internet of Things: A Survey of Existing Protocols and Open Research Issues. IEEE Access, 7, 12950-12988.
30. Rubinoff, S. (2018). Web and Network Data Science: Modeling Techniques in Predictive Analytics. CRC Press.
31. Rubinoff, S., & Rajkumar, T. (2016). Applied Data Science: Lessons Learned for the Data-Driven Business. O'Reilly Media.
32. Svenmarck, P., Luotsinen, L., Nilsson, M., & Schubert, J. (2018, May). Possibilities and challenges for artificial intelligence in military applications.
33. Tikk-Ringas, E., Kerttunen, M., & Spirito, C. (2014). Cyber security as a field of military education and study.
34. Van Den Bosch, K., & Bronkhorst, A. (2018, July). Human-AI cooperation to benefit military decision making. NATO.
35. Vacca, W. A. (2011). Military culture and cyber security.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal) | Impact Factor: 7.512 |

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

36. ZHANG Zhi-min, SHI Fei-fei, WAN Yue-liang, XU Yang, ZHANG Fan, NING Huan-sheng. Application progress of artificial intelligence in military confrontation.
37. Council of Insurance Agents & Brokers. (2019, June 11). Artificial Intelligence in Cybersecurity | The Council of Insurance Agents & Brokers. The Council of Insurance Agents & Brokers. https://www.ciab.com/resources/artificial-intelligence-cybersecurity/
38. Wiafe, I., Koranteng, F. N., Obeng, E. N., Assyne, N., Wiafe, A., & Gulliver, S. R. (2020). Artificial intelligence for cybersecurity: a systematic mapping of literature. IEEE Access, 8, 146598-146612.
39. KUNUNGO, S., RAMABHOTLA, S., & BHOYAR, M. (2018). The Integration of Data Engineering and Cloud Computing in the Age of Machine Learning and Artificial Intelligence.