As per UGC guidelines an electronic bar code is provided to seure your paper

International Journal for Modern Trends in Science and Technology, 8(03): 25-33, 2022

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0803005
Available online at: http://www.ijmtst.com/vol8issue03.html

An Exploration into Artificial intelligence of Security threat, Crime and Forensics

Muhsiena K H [1] | Amrutha N [2]

Department of Computer Science, St. Alberts College (Autonomous) Banerjee Road Cochin, Kerela.
Corresponding Author Email: muhsienakh878@gmail.com

To Cite this Article
Muhsiena K H and Amrutha N. An Exploration into Artificial intelligence of Security threat, Crime and Forensics. International Journal for Modern Trends in Science and Technology 2022, 8(03), pp. 25-33. https://doi.org/10.46501/IJMTST0803005

Article Info
Received: 20 January 2022; Accepted: 22 February 2022; Published: 01 March 2022.

ABSTRACT

Artificial Intelligence (AI) advances have affected nearly every subject, including computer science, criminology, robotics etc Because of AI's outstanding ability to acquire and analyse vast amounts of data, its methodologies are suitable for tackling a range of crime-related issues. Despite the fact that AI has solved a variety of problems, AI experts have warned about the possible security risks associated with AI algorithms and training data. As AI systems inherit existing computer system security concerns, concern about unique cyberattacks aided by AI is growing. This work review literature assessment on security risks and AI-related criminality in this setting. Based on the literature analysis, this article defines AI crime and divides it into two categories: tool crime and target crime. In addition, forensic approaches are discussed. We also look at the features of AI crimes in the past and now. Traditional forensic approaches are unable to tackle problems that are tough to solve. Finally, there are some unresolved difficulties discussed, with a focus on the need to develop new AI forensics tools.

KEYWORDS: Artificial Intelligence, AI Crimes, Forensic

1. INTRODUCTION

AI has the ability to acquire and analyse vast amounts of data, its methodologies are suitable for tackling a range of crime-related issues. Many people are interested in AI cyber security because it is a popular and relevant scientific issue. The concerns we will encounter will include not just pure science and technology issues, but also political and social influence challenges that will cause issues in our society. Therefore AI is every importand subject in every field like computer science, criminology, robotics etc Deep Learning, which is inspired by the structure and function of the brain, has been a key accomplishment in the AI area, igniting AI research in a variety of fields. Deep learning research has been investigated to handle large amounts of data (e.g., photos, social media, crime data information, and so on) in order to do medical picture analysis, speech analysis, and so on. While AI's rapid progress has provided benefits in terms of innovation, it has also come with considerable concerns. Unexpected problems (e.g., terrorism, security threats, cybercrime, privacy violation, etc.) arose when ICT advanced at fast speeds in the past, resulting in significant social costs. Similarly, there are rising concerns about the potential for AI to generate a variety of issues. We look at AI security concerns, predictable crimes, and digital forensics for AI in this context. After defining AI crime, we offer a taxonomy

for new sorts of crime: the AI as tool crime and AI as target crime, inspired by an existing taxonomy [1] used in cybercrime: Computer as tool crime and computer as target crime.

2. LITERATURE REVIEW

Various academics has already analyzed AI. From many viewpoints, this section discusses studies on the AI security threat and AI-related criminality. We also look at cybercrime as defined by the cybersecurity and digital forensics communities.

2.1 AI security threads and crime

Because the term 'crime' is associated with law and ethics, the phrase 'AI crime' was originally coined by the humanities sector. Several research have highlighted security dangers and malevolent applications of AI that can cause numerous crimes, despite the fact that the phrase AI crime has not been explored in the computer science field. The concerns we will encounter will include not just pure scientific and technological issues, but also political and social impact challenges that will cause troubles in our society. Many contemporary issues and concerns about AI security have been documented, by some recent investigation and conversations. Adopting online identities, known as socialbots, that behave like humans is a great example of harmful AI use. Though the original goal of socialbot was to promote awareness and collaboration among people, it has been used negatively in the past for phishing, fraud, and political infiltration into online social networks campaigns [2]. Machine learning, according to Seymour and Tully [6], can be used for social engineering; for example, using AI, mass-produced tweets with phishing links could be broadcast on Twitter without causing any disruption. Because the harmful socialbot is based on a specific user's previous actions and public profiles, detecting it has become a computer security issue. When harmful socialbots are created to carry out a political attack, the tactic may affect or inflame public opinion, according to social science [3] [4]. According to some academics, hackers have already begun to weaponize AI in order to improve their hacking abilities and develop new sorts of cyber attacks [5]. Traditional cybercrimes such as financial fraud, cyber terrorism, and cyberextortion are all using AI to improve their strategies.

Unlike the above research, which focused on the difficulties that certain techniques could cause, Brundage et al. [6] provided a holistic view of AI's malevolent use. They focused on three shifts in the threat landscape: the growth of current risks, the introduction of new threats, and a shift in the threat's characteristic character. The cost of jobs that require human labour could be reduced thanks to the AI system's scalability. As a result of the cost-cutting strategies (e.g. mass spear phishing), perpetrators are able to attack more targets, resulting in the expansion of existing dangers. New dangers may also arise to fulfil jobs that are impossible for humans to complete (for example, impersonating individual voices or commanding many drones). The normal character of threats will be altered when highly effective AI strikes become increasingly widespread. Security domains were also divided into three categories by Brundage et al.: digital security, physical security, and political security. Cyberattacks that target human or AI systems are included in the digital security realm. Physical threats, such as forcing autonomous vehicles to crash and manipulating thousands of drones, are under the physical security realm. New dangers in profiling, repression, and targeted disinformation efforts are all part of the political security arena.

By introducing the phrase 'AI crime,' King et al. [7] gave a distinct perspective on AI security issues. They looked at the issue from a different angle. Commerce, financial markets, and insolvency (e.g. market manipulation, price fixing, collusion), harmful or dangerous drugs (e.g. trafficking, selling, buying, possessing banned drugs), offences against the person (e.g. harassment, torture), sexual offences (e.g. sexual assault, promotion of sexual offence), theft and fraud, and forgery and personation are all classified as AI crimes in the article (e.g. spear phishing, credit card fraud). They argued that the offences are classified as having one or more threats. They focused on human nature when identifying AI security threats: emergence, liability, monitoring, and psychology. The psychology threat, for example, implies that AI can influence a human's mental state to the point of aiding or instigating crime. This approach differs significantly from that of computer science; this diversity of viewpoints is due to AI's inherent interdisciplinarity.

Some research has focused on AI privacy concerns emerging from the processing of personal data.According to Li and Zhang [8], AI applications in healthcare, banking, and education may cause privacy issues. Because the quantity and quality of training data have a significant impact on AI performance, developers want to acquire as much data as possible. The acquisition of extensive data, according to Li et al., has inherent privacy risks. Mitrou [9] used the General Data Protection Regulation to tackle the issue of privacy (GDPR). Although GDPR does not expressly address AI, the author emphasised that it can be applied to AI when it manages personal data. The prior research has three consequences for AI stakeholders. First, due to AI's dual-use nature, researchers and engineers should be aware that the technology could be used to execute criminal acts, even if it was created for legal purposes. Second, completely new forms of security risks will develop that have never been considered before. Because AI can do activities that were previously thought to be impossible for people or traditional programmes to complete, the threats will be outside the core purview of known threats. To prevent AI security threats and respond to AI crime, AI researchers should interact closely with specialists from other industries. Finally, the AI security field should learn from the cybersecurity industry's mistakes. The anticipated AI crimes are very intimately involved in cybercrime, as reported in prior studies. The dual-use nature of ICT led to cybercrime; the current state of AI security mimics the early stages of cybersecurity.

2.2 Cybercrime

Cybercrime is seen as the evil side of the internet.It is divided into two types: computer as target crime and computer as tool crime[10][11]. New sorts of crimes, such as cyberterrorism, cyberextortion, and cyberwarfare, have evolved as information has become digitised and connected via network; these crimes are known as computer as target crime. The goal of computer-as-a-target crime is to disable or destroy computer systems. As a result, when criminals execute a computer-as-a-target crime, they employ tools or procedures that have been created to break into computer systems. In the meantime, every data in our daily lives has been digitised, from personal to professional. Offline crimes like as fraud, threats, child abuse, stalking, and so on now enter the online world as a result of this transformation.It is known as computer as a tool crime in the online world. The taxonomy of cybercrime has aided in the development of practical measures to combat the crime. When forensic investigators look into computer-as-a-tool crime, they focus on demonstrating the perpetrator's previous conduct to see if any illegal activity has place. Criminals that utilise computers as tools typically use well-known technologies and manipulate well-known infrastructures such as text messages, websites, social media, and so on. When investigating a computer as a target crime, however, detectives should concentrate on malicious applications, sometimes known as malware. To respond promptly to the crime and determine the amount of the damage, they must first locate the malware and then undertake reverse engineering to determine the virus's purpose and the source of the attack. [12] [13]

2.3 Digital forensics

"The application of scientifically derived and validated methodologies to the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence" is how digital forensics is defined [14]. Many principles and guidelines have been proposed in the field of digital forensics because each country and organisation has its own set of laws and rules. Nonetheless, they all agree on the fact that a forensic method is only considered forensically sound if it adheres to five principles: Meaning,Errors,Transparency,Trustworthiness,Reproducibility, and Experience are all important factors to consider[15][16]. The evidence would be difficult to accept in court if the forensic method did not follow any of the five principles. As a result, investigators must collect and examine evidence while following the guidelines. In addition, forensic researchers established a proactive technique called Digital Forensic Readiness that is intended to handle events before they happen [17]. (DFR). During an incident response, DFR strives to collect digital evidence fast and accurately while minimising the cost of conducting forensic investigation [18]. Digital forensic experts, like those in other domains, have looked into the use of AI in investigations. To detect malware, Karbab and Debbabi [19] used a natural language processing system and supervised machine learning. Furthermore, numerous researchers have investigated forensic investigation

methods employing AI [20], [21], [22], but no research on AI as a forensic investigation subject has yet been published.

3. AI AS TOOL CRIME

Given the dual-use nature of AI, this section describes predicted AI as tool crime. Because AI systems are built on digital infrastructure, they are vulnerable to cybercrime, which includes both computer as tool and computer as target crimes. Furthermore, AI can be used to manipulate autonomous technologies such as smart cars, drones, and Internet of Things (IoT) devices to commit physical crimes. In this part, we'll look at how artificial intelligence (AI) can be utilised to sharpen cyberattacks. Then we turn our attention to physical crime, which is considered an unique attack.

3.1 Cyber crime with advanced technology

3.1.1 Computer as tool crime

Previous studies have shown that AI may be utilised for phishing, and its effectiveness has been demonstrated [23]. Scam email with profiling is one of the most prevalent phishing strategies. In the corporate world, AI-assisted profiling is a hot topic; targeted advertising is a good example. However, the targeted advertising method, which is based on the customers' previous purchasing history or interests, may be useful to the attacker. Previous studies referred to the AI programme as a chatbot. According to Kietzmann et al. [24] and Paschen et al. [25], AI will improve techniques for defrauding clients using malevolent chatbots. The chatbot may converse with customers continuously and collect large amounts of data about their behaviour and profile. In academics and industry, the chatbot has already been developed and is in use. However, as Natural Language Processing (NLP) technology has progressed, the chatbot has been designed to speak verbally with individuals [26], [27]. While some studies have suggested that AI-assisted voice could be used for the common good in areas such as social therapy [28], education [29], medical diagnostics , and health , there are also concerns that AI-assisted voice could increase theft and fraud. Because speech is one of the biometrics that is an indispensible measure in security mechanisms, it can be a powerful weapon for attackers (for example, voice phishing).

Another example of advanced crime is fake news. Although fake news has a long history in social engineering, with the development of social network services (e.g. Twitter, Facebook, and YouTube) , it has recently gained attention. Fake news, in particular, has a significant impact on political issues such as policymaking, propaganda, and elections . Fake news becomes more potent using the deepfake approach. AI anchors have been developed by news organisations to improve productivity and minimise coworkers. It means that fake news can be produced using virtual anchors who appear to be real individuals.

3.1.2 Computer as target crime

AI can execute previously unsolved jobs at a cheaper cost and with less manpower. It can have the same effect as hiring extra human analysts by making replicas of the AI system. As a result of this feature, attackers can acquire illegal access. For example, password authentication, the most basic method of identifying users, would be jeopardised. The dictionary attack is one of the most successful methods of obtaining the password because it includes well-known words or phrases that are likely to have been used in the password. The social engineering technique of obtaining victim's information from the internet (e.g. birthday, phone number, address, etc.) is frequently utilised when generating the dictionary [30]. Collecting information takes a lot of time and effort, but AI systems built to automate social engineering can do it quickly and easily. For criminals, automated detection techniques to find vulnerabilities might be a handy tool. Russell et al. [31] suggested that AI may be used to find vulnerabilities. They showed that using a convolutional neural network (CNN) and a tree ensemble over standard static analysis has some advantages. Grieco et al. [32] proposed an approach for detecting large-scale flaws. Without inspecting source code, the proposed method could be used to identify programmes having vulnerabilities. Aside from those investigations, a number of strategies for detecting susceptibility have been investigated [33] – [34] .Despite the fact that the approaches were created for the general good, offenders may utilise them to locate susceptible systems.

3.2 Physical crime

The problem of AI security goes outside cyberspace, especially with the rapid adoption of IoT . A perpetrator can physically harm a target by controlling an AI system (e.g. human, pet, vehicle, house). In the

context of physical crime, the ethics of AI have been debated in the topic of scientific ethics. Lin et al. [35] discussed robot ethics by stating that AI robots can harm people intentionally or unintentionally. Scherer further emphasised that AI systems might do bodily injury, and that there are difficulty in assigning moral and legal culpability for such harm. Military AI, on the other hand, has been developed for military purposes and is fundamentally built to attack physical targets. Military AI is developed for the benefit of the public, but it can also be used to damage people outside of a military context ; the drone swarm is a notable example.

4. AI AS TARGET CRIME

In this article, an offence producing damage or impairment in the processing of data or the operation of an AI system is defined as an AI target crime. This definition is based on definition of computer crime as a target crime. There are many different AI systems, however most AI systems have the same fundamental notion as Fig. 1.

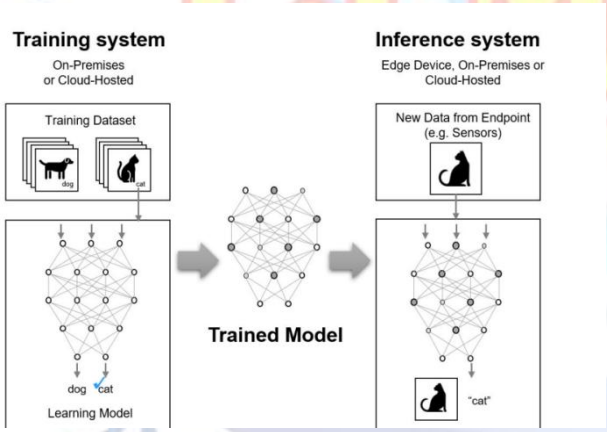

Fig1: The AI system's structure.

The AI system is made up of two parts: a training system and an inference system. Based on the training dataset, the training system develops a trained model, which is then used by the inference system to categorise new data from endpoints. The training system, for example, generates an algorithm that identifies dogs from cats in Fig. 1. The algorithm is loaded into the inference system, which then determines if the object image collected from sensors is a dog or a cat. AI as a target crime is primarily motivated by the security risks posed by AI systems. Several papers discussed threat classification, with white-box and black-box attacks being the most common threat models. The white-box assault is defined as an attack that uses knowledge of a target AI system's dataset, architecture, and parameters. The black-box assault, on the other hand, contains little or no information about the target system's structure. Threat models based on adversarial examples (AEs) were proposed in some studies. In order to misclassify the input and damage AI performance, the AEs are input data with imperceptible noise . This article, based on earlier work, portrays AI as a target crime from the victims' perspective. We concentrate on AI systems, such as the training and inference systems, which would be vulnerable to the above-mentioned assaults.

4.1 Training system as target crime

Direct access to the training system in a practical AI system appears difficult to acquire since the training system is protected with great confidentiality and not produced in a common computer system [36]. Insider spying, advanced persistent threat (APT), or malicious external storage could all help (e.g. USB, external hard drive). If the training system's security is breached, the AI system will suffer significant damage. The training system, in particular, comprises a training dataset that has a substantial impact on the learning model's effectiveness; it is for this reason that crimes against the training system would be disastrous. Because the examination of the breach is the domain of standard cyber forensics, the next section assumes that attackers have already intruded into the training system.

4.1.1 Training system attack

The goal of this crime is to undermine AI's trust worthiness.AI may misclassify fresh data from the inference system by inserting AEs or changing the existing dataset. If a perpetrator can influence the learning algorithm, which is known as logic corruption, the training system will be significantly harmed. The training system assault is divided into three areas in this article: data injection, data alteration, and logic corruption. By injecting AEs, data injection crime impairs the availability of AI systems. According to Goodfellow et al., AI incorrectly recognises a panda's image as a gibbon by adding a noise that people cannot perceive. The goal of AEs is to identify the smallest disruption that will fool AI. The data modification crime occurs when perpetrators have access to edit or remove some training data, allowing them to carry out

lethal attacks on AI systems. By modifying the labels of some training data, Zhao et al. [37] demonstrated that label contamination attack (LCA) can dramatically degrade AI performance. Hayes and Ohrimenko [38] demonstrated that supplying tainted assaults to the training system compromises the classifier's accuracy.

The most serious crime in the training system is logic corruption. By tampering with the learning algorithm, thieves can change the architecture and parameters of the learned model . When the CNN system is hacked and then corrupted, for example, the attacker has control over the input layer, classification layer, and training parameters.

4.1.2 Training system theft

The training system consists of three components: a training dataset, a learning model, and a learned model. The training method is considered a trade secret by AI developers and makers of AI-related products because it is directly tied to the performance of AI. For AI stakeholders, the dataset is critical . They collect data from a variety of sources, including open-source data, to generate a dataset. Because creating a dataset takes a lot of time and effort, it's worth a lot of money. As a result, the dataset is a favourite target for criminals. If offenders steal sensitive data such as a medical image, a face image, or a voice, major privacy violations may ensue . Because they are developed with know-how, insight, and knowledge, the learning model and trained model are also valuable assets for AI developers. The algorithm, distribution of training data, and parameters of fully trained model architecture could be revealed to adversaries or the general public as a result of this crime. In particular, the information could be partially exploited for white-box or black-box attacks .

4.2 Inference system as target system

In addition, perpetrators may target the inference system. In comparison to the attack on the training system, criminals have a relatively easy time accessing the inference system because it is typically deployed at end devices. The criminal targeted inference system does not interact with the learning model, but it can cause the trained model to leak or classification to malfunction.

4.2.1 Inference system cracking

In the inference system, the parameters that were determined during the training phase play a significant role. Depending on where the parameters are located, there are two sorts of operation methods: centralised and distributed models (See Fig. 2).

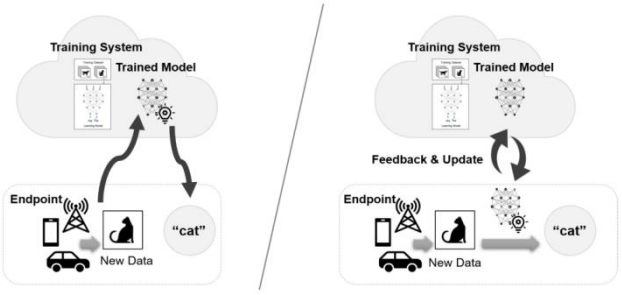

Fig 2: Comparison between the centralized model (left side) and distributed model (right side) in AI system.

In the centralised model, the inference procedure is performed by a central server provided by an AI provider. The job of the end device (e.g., smartphone, IoT device, in-vehicle infotainment) in a centralised face recognition system, for example, is to submit face images or extracted features to the central server, which then processes the image or feature. The results of the server's processing are used by the end devices. Although the centralised model is theoretically acceptable for AI service management and security, it may be less useful in practise due to the possibility of a bottleneck. As a result, the distributed approach is becoming more popular in real-world applications. . In a distributed model, the parameters defined in the training system are handled in the inference system, allowing end devices to perform image processing operations, such as face recognition. With the advent of the Internet of Things (IoT), it is assumed that the distributed model would be used for AI services]; the relationship between the central and distributed models is similar to that of cloud and fog computing. The trained model would be detected and then altered using traditional hacking techniques (e.g. reverse engineering, side channel attack) . Indeed, the perpetrators of this crime may be able to carry out a white-box attack.

4.2.2 Inference system abuse

Abuse of the inference system is a felony that results in AE misclassification. By breaking the end device or detecting that the target system uses common open-source project libraries, the criminals may be able to determine the learning model and its parameters. Abuse is defined as either a white-box attack or a

black-box attack depending on the level of understanding of the target system. The attackers that are attempting a white-box attack are aware of the AI model and its parameters. Because the offenders already know the distribution of training data, they may imitate the targeted AI model by impersonating the AI system and creating a false training dataset based on their information. The black-box attack is carried out with limited information or without the AI model's knowledge. Non-adaptive black-box attacks, adaptive black-box attacks, and stringent black-box attacks are the three types of black-box attacks. Non-adaptive black-box attackers can collect alternative datasets with the distribution if they know the distribution of training data, even if they can't figure out the architecture or structure of the target AI model. They can really make AEs using their local AI model that has been trained using the alternative dataset.By interrogating the targeted AI model, the adaptive black-box attackers use input-output pairs. The oracle attack is frequently compared to this attack. The attackers can identify labels of queried data by collecting large amounts of query data, and then recreate the model using the queried data corpus. The strict black-box attack is similar to the adaptive black-box attack in that it is built on collecting input-output pairs, but it is more limited because the attackers are unable to make queries to the inference system. As a result, they should go after AI systems without using the oracle. Nonetheless, if the attackers receive a large number of input-output pairs and establish a pattern or distribution of them, it may be effective .

5. AI FORENSICS

To identify 5W1H, forensic investigators needs to collect and examine evidence (when and where the crime is committed ,who is criminal, what is targeted, why the criminal commit, and how the crime occurs). Forensic researchers have given issues and answers for forensic sub-fields such as smartphone forensics, cloud forensics, and IoT forensics, as well as the manner of gathering evidence and the type of digital data dependent on the device or platform. We discuss four important components of AI forensics that have not been covered in the forensic community, based on characteristics of AI systems and strategies utilised for perpetrating AI crimes: AI exploration, similarity analysis, detection of adversarial attacks, and damage assesment.

5.1 AI Exploration

When looking into AI as a tool crime, it's important to figure out how AI is employed. The dataset, learning model, trained model, inference model, and application of the AI system used to commit a crime should all be collected and analysed by the investigators. Investigators should be able to understand the AI's objective based on the assessment. It is critical for investigators to be able to distinguish between the developer's purpose and the AI product in this setting. Data and programme are processed on the computer to produce output in traditional programming; otherwise, data and output are utilised to create a programme in AI. As a result, even if the same dataset and learning model are provided, it may generate programmes with different parameters and outputs. With limited evidence, it is impossible to replicate the AI system. Indeed, because many AI systems use transfer learning, which starts with pre-trained models, getting origin data will grow increasingly difficult.

5.2 Similarity analysis

A traditional field of digital forensics involves investigating infractions such as copyright infringement, leaks of confidential documents, and invasions of privacy. Similarity analysis is one of the most essential ways for detecting illegal activity in these circumstances. When two models are built on the same dataset, determining their similarity is more difficult. Verifying or testing the models is more difficult if the investigators were unable to obtain the specific dataset. Similarity analysis for AI, with or without a training dataset, should be investigated in response to the theft. One of the most important research areas in AI forensics is the development of a file format for storing trained models.

5.3 Detection of adversarial attack

It is critical to proactively prevent or detect a hostile attack. Many researchers presented defence strategies aimed at correctly classifying AEs, however the methods have been thwarted by newer developed attacks. Because it is difficult to defend against

adversarial attacks, current research has focused on detecting AEs. Two-sample hypothesis testing, principle component analysis (PCA), and Bayesian uncertainty estimates are examples of statistical approaches to the problem. Several approaches detect adversarial attacks that were known at the time, but state-of-the-art AEs attacks have been created to neutralise the detection strategies. In the current situation, as the technique for creating AEs becomes more advanced, forensic experts are still working on improving detection techniques

5.4 damage assessment

Forensic investigators should determine the amount of the damage produced by AI crime. In the case of an attack using AEs, investigators must determine which data are AEs, how many AEs were actually injected, and how much the confidence was influenced. Finding AEs is the process of identifying data that increases the prediction error. We use the deep neural network (DNN) as an example to demonstrate the procedure.

6. FUTURE SCOPE AND CONCLUSION

Clearly, a more acceptable trade-off between science and technology and citizen's need is urgently needed and should be taken seriously. A more suitable way to working out a better trade-off between AI and cyber security, as well as the state's option, is a proposed future development. To achieve a better balance of compensation between citizens, government, and business enterprises. In the above summaries, we've looked at the overall balance between AI and security technological solutions, as well as government consideration of society's strategic viewpoints. Clearly, a more acceptable trade-off between 'science and technology and citizens' needs is urgently needed and should be taken seriously. The study work presented here is still in its early stages and requires more in-depth analysis based on worldwide examination

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] S. Gordon and R. Ford, ''Cyberterrorism?'' Computer &. Security, vol. 21, no. 7, pp. 636–647, 2002.

[2] R.W. Gehl and M. Bakardjieva, ''Socialbots and their friends,'' in Socialbots and Their Friends. Evanston, IL, USA: Routledge, 2016, pp. 17–32

[3] J. Seymour and P. Tully, ''Weaponizing data science for social engineering: Automated E2E spear phishing on Twitter,'' Black Hat USA, vol. 37, pp. 1–39, Aug. 2016

[4] N. Abokhodair, D. Yoo, and D. W. McDonald, ''Dissecting a social botnet: Growth, content and influence in Twitter,'' in Proc. 18th ACM Conf. Comput. Supported Cooperat. Work Social Comput., 2015, pp. 839–851.

[5] M.-A. Rizoiu, T. Graham, R. Zhang, Y. Zhang, R. Ackland, and L. Xie, ''#DebateNight: The role and influence of socialbots on Twitter during the 1st 2016 us presidential debate,'' in Proc. 12th Int. AAAI Conf. Web Social Media, 2018

[6] G. Dvorsky. Hackers Have Already Started to Weaponize Artificial Intelligence. Gizmodo.com. [Online]. Available: https://gizmodo.com/hackershave-already-started-to-weaponize-artificial-in-179768

[7] M. Brundage et al., ''The malicious use of artificial intelligence: Forecasting, prevention, and mitigation,'' 2018, arXiv:1802.07228. [Online]. Available: http://arxiv.org/abs/1802.07228

[8] T. King, N. Aggarwal, M. Taddeo, and L. Floridi, ''Artificial intelligence crime: An interdisciplinary analysis of foreseeable threats and solutions,'' SSRN Electron. J., vol. 26, no. 1, pp. 1–32, 2019

[9] X. Li and T. Zhang, ''An exploration on artificial intelligence application: From security, privacy and ethic perspective,'' in Proc. IEEE 2nd Int. Conf. Cloud Comput. Big Data Anal. (ICCCBDA), Apr. 2017

[10] L. Mitrou, Data Protection, Artificial Intelligence and Cognitive Services: Is the General Data Protection Regulation (GDPR) 'Artificial Intelligence-Proof '?. SSRN, 2018. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3386914

[11] S. Gordon and R. Ford, ''Cyberterrorism?'' Comput. Secur., vol. 21, no. 7, pp. 636–647, 2002.

[12] K. Dashora, ''Cyber crime in the society: Problems and preventions,'' J. Alternative Perspect. Social Sci., vol. 3, no. 1, pp. 240–259, 2011.

[13] M. Brand, C. Valli, and A. Woodward, ''Malware forensics: Discovery of the intent of deception,'' J. Digit. Forensics, Secur. Law, vol. 5, no. 4, p. 2, 2010.

[14] C. H. Malin, E. Casey, and J. M. Aquilina, Malware Forensics Field Guide for Windows Systems: Digital Forensics Field Guides. Amsterdam, The Netherlands: Elsevier, 2012

[15] B. Carrier and E. H. Spafford, ''An event-based digital forensic investigation framework,'' in Proc. Digit. Forensic Research Workshop, 2004, pp. 11–13.

[16] L. Pan and L. Batten, ''Reproducibility of digital evidence in forensic investigations,'' in Proc. 5th Annu. Digit. Forensic Res. Workshop (DFRWS), 2005, pp. 1–8.

[17] V. R. Kebande and H. S. Venter, ''Novel digital forensic readiness technique in the cloud environment,'' Austral. J. Forensic Sci., vol. 50, no. 5, pp. 552–591, Sep. 2018.

[18] R. Rowlingson, ''A ten step process for forensic readiness,'' Int. J. Digit. Evidence, vol. 2, no. 3, pp. 1–28, 2004.

[19] E. B. Karbab and M. Debbabi, ''MalDy: Portable, data-driven malware detection using natural language processing and machine learning techniques on behavioral analysis reports,'' Digit. Invest., vol. 28, pp. S77–S87, Apr. 2019.

[20] F. Amato, A. Castiglione, G. Cozzolino, and F. Narducci, ''A semanticbased methodology for digital forensics analysis,'' J. Parallel Distrib. Comput., vol. 138, pp. 172–177, Apr. 2020.

[21] W. Anwar, I. S. Bajwa, M. A. Choudhary, and S. Ramzan, ''An empirical study on forensic analysis of urdu text using LDA-based authorship attribution,'' IEEE Access, vol. 7, pp. 3224–3234, 2019.

[22] O. Mayer and M. C. Stamm, ''Forensic similarity for digital images,'' IEEE Trans. Inf. Forensics Security, vol. 15, pp. 1331–1346, 2020.

[23] T. King, N. Aggarwal, M. Taddeo, and L. Floridi, ''Artificial intelligence crime: An interdisciplinary analysis of foreseeable threats and solutions,'' SSRN Electron. J., vol. 26, no. 1, pp. 1–32, 2019

[24] J. Kietzmann, J. Paschen, and E. Treen, ''Artificial intelligence in advertising: How marketers can leverage artificial intelligence along the consumer journey,'' J. Advertising Res., vol. 58, no. 3, pp. 263–267, 2018.

[25] J. Paschen, M. Wilson, and J. J. Ferreira, ''Collaborative intelligence: How human and artificial intelligence create value along the B2B sales funnel,'' Bus. Horizons, vol. 63, no. 3, pp. 403–414, May 2020.

[26] S. A. Abdul-Kader and D. John, ''Survey on chatbot design techniques in speech conversation systems,'' Int. J. Adv. Comput. Sci. Appl., vol. 6, no. 7, 2015.

[27] L. Ciechanowski, A. Przegalinska, M. Magnuski, and P. Gloor, ''In the shades of the uncanny valley: An experimental study of human– chatbot interaction,'' Future Gener. Comput. Syst., vol. 92, pp. 539–548, Mar. 2019.

[28] S. D'Alfonso, O. Santesteban-Echarri, S. Rice, G. Wadley, R. Lederman, C. Miles, J. Gleeson, and M. Alvarez-Jimenez, ''Artificial intelligenceassisted online social therapy for youth mental health,'' Frontiers Psychol., vol. 8, p. 796, Jun. 2017.

[29] F. Clarizia, F. Colace, M. Lombardi, F. Pascale, and D. Santaniello, ''Chatbot: An education support system for student,'' in Proc. Int. Symp. Cyberspace Saf. Secur. Cham, Switzerland: Springer, 2018, pp. 291–302.

[30] S. Gupta, A. Singhal, and A. Kapoor, ''A literature survey on social engineering attacks: Phishing attack,'' in Proc. Int. Conf. Comput., Commun. Autom. (ICCCA), Apr. 2016, pp. 537–540.

[31] R. Russell, L. Kim, L. Hamilton, T. Lazovich, J. Harer, O. Ozdemir, P. Ellingwood, and M. McConley, ''Automated vulnerability detection in source code using deep representation learning,'' in Proc. 17th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA), Dec. 2018, pp. 757–762.

[32] G. Grieco, G. L. Grinblat, L. Uzal, S. Rawat, J. Feist, and L. Mounier, ''Toward large-scale vulnerability discovery using machine learning,'' in Proc. 6th ACM Conf. Data Appl. Secur. Privacy, 2016, pp. 85–96.

[33] S. M. Ghaffarian and H. R. Shahriari, ''Software vulnerability analysis and discovery using machine-learning and data-mining techniques: A survey,'' ACM Comput. Surv., vol. 50, no. 4, pp. 1–36, 2017.

[34] Z. Li, D. Zou, S. Xu, X. Ou, H. Jin, S. Wang, Z. Deng, and Y. Zhong, ''VulDeePecker: A deep learning-based system for vulnerability detection,'' 2018, arXiv:1801.01681. [Online]. Available: http://arxiv.org/abs/1801.01681

[35] H. Xue, S. Sun, G. Venkataramani, and T. Lan, ''Machine learning-based analysis of program binaries: A comprehensive study,'' IEEE Access, vol. 7, pp. 65889–65912, 2019.

[36] P. Lin, K. Abney, and R. Jenkins, Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence. London, U.K.: Oxford Univ. Press, 2017

[37] Q. Liu, P. Li, W. Zhao, W. Cai, S. Yu, and V. C. M. Leung, ''A survey on security threats and defensive techniques of machine learning: A data driven view,'' IEEE Access, vol. 6, pp. 12103–12117, 2018

[38] M. Zhao, B. An, W. Gao, and T. Zhang, ''Efficient label contamination attacks against black-box learning models,'' in Proc. IJCAI, 2017, pp. 3945–3951.

[39] J. Hayes and O. Ohrimenko, ''Contamination attacks and mitigation in multi-party machine learning,'' in Proc. Adv. Neural Inf. Process. Syst., 2018, pp. 6604–6615.