AI-BASED CYBER CRIMES AND CYBER REGULATIONS: ISSUES AND CHALLENGES

By Pranav Choudhary[*]

ABSTRACT

Artificial intelligence (AI) is the science and engineering of creating intelligent machines, particularly computer programs. It has the potential to play a role in criminal acts known as AI-crimes (AIC), which pose threats to cybersecurity, such as stealing sensitive data, hacking and weaponizing autonomous weaponry, social engineering attacks, and political disruption. Currently, there is no regulatory framework targeting AI crimes, and the legal texts do not address the contentious issue of criminal liability for AI. Traditionally, AI has been regulated through product licensing, R&D supervision, and tort responsibility. However, applying existing AI concepts to assess the harm that autonomous AI may pose in the future is insufficient. This research will be divided into four areas: defining AI crime and potential AI hostile acts, examining existing cybercrime frameworks, surveying international organizations that have made progress in AI crimes, and presenting legal notions focused specifically on AI crimes. The fourth half will discuss the challenges in determining the jurisdiction of AI crimes due to their global nature. The author will analyze how the law should deal with these concerns and provide recommendations on how the legal system should specifically address these obstacles.

Keywords- AI, AI-crimes, criminal liability, artificial intelligence, criminal liability of AI, Legislation, Human Rights, etc.

[*] Ph.D. Research Scholar, School of Law, Raffles University, Neemrana, Rajasthan

I. INTRODUCTION

The digital age has brought about tremendous advancements in technology, but it has also given rise to new challenges and issues related to cybercrime and the need for effective cyber regulations. Artificial Intelligence (AI) plays a significant role in both cybercrime and cybersecurity, presenting a double-edged sword. Here are some key issues and challenges in this domain:[1]

AI-Based Cyber Crimes:

AI-based cybercrimes refer to criminal activities that involve the use of artificial intelligence and machine learning techniques to carry out malicious actions in the digital space. These crimes often exploit the capabilities of AI to automate and enhance various aspects of cyber-attacks, making them more sophisticated, scalable, and challenging to detect. Here are some common AI-based cybercrimes:

1. **Automated Phishing Attacks:** Automated phishing attacks involve the use of artificial intelligence (AI) and machine learning (ML) techniques to streamline and enhance the effectiveness of phishing campaigns. Phishing is a type of social engineering attack where attackers attempt to trick individuals into revealing sensitive information, such as usernames, passwords, or financial details.
2. **Email Generation:** AI is used to generate phishing emails that mimic legitimate communication. These emails often imitate trusted entities like banks, government agencies, or well-known companies. Natural language processing (NLP) and other AI techniques are employed to make the content of the emails more convincing and contextually relevant.
3. **Personalization:** AI algorithms analyze publicly available information about the target, such as social media profiles, to personalize phishing messages. This makes the emails appear more authentic and increases the chances of success.
4. **Spoofing Techniques:** AI is used to mimic the look and feel of legitimate websites by creating realistic replicas. This can involve copying the design, layout, and content of

[1] Burgess, Matt, "Police built an AI to predict violent crime. It was seriously flawed", WIRED, August 6, 2020, available at: https://www.wired.co.uk/article/police-violence-prediction-ndas.

trusted sites to deceive users into providing login credentials or other sensitive information.[2]

5. **Dynamic Content:** AI-driven phishing campaigns may incorporate dynamic content, adjusting elements of the phishing site or email in real-time based on the user's behavior or responses. This adaptability makes detection more challenging.
6. **Automated Delivery:** AI can be utilized to automate the delivery of phishing emails at scale. By analyzing the best times to send messages and adjusting the content based on user behavior, attackers can optimize their campaigns for maximum impact.[3]
7. **Evasion Techniques:** AI is employed to evade email security filters and other detection mechanisms. This can involve the generation of polymorphic content that changes with each iteration to avoid pattern-based detection.
8. **Credential Harvesting:** AI-driven phishing attacks often aim to harvest usernames and passwords. Once users submit their credentials on a fake website, the attackers can collect and exploit this sensitive information.
9. **Continuous Learning:** AI models in phishing attacks can continuously learn and adapt based on the success or failure of previous campaigns. This iterative process allows attackers to refine their techniques over time.

Mitigation Strategies:

1. **Employee Training:** Conduct regular training sessions to educate users about phishing risks and how to recognize suspicious emails.
2. **Email Filtering:** Implement advanced email filtering solutions that use AI to detect and block phishing emails.
3. **Multi-Factor Authentication (MFA):** Enable MFA to add an extra layer of security, even if credentials are compromised.
4. **Website Blacklisting:** Utilize threat intelligence services to identify and block known phishing websites.
5. **Behavioral Analysis:** Implement solutions that analyze user behavior to detect anomalies that may indicate a phishing attempt.

[2] Supply Chain, "Lessons Learned from the Vaccine Supply Chain Attack", January 16, 2021, available at: https://www.supplychaindigital.com/supply-chain-risk-management/lessons-learned-vaccine-supply-chain-attack.

[3] The New York Times, "Cyber Attack Suspected in German Woman's Death", September 18, 2020, available at: https://www.nytimes.com/2020/09/18/world/europe/cyber-attack-germany-ransomeware-death.html.

Combating automated phishing attacks requires a multi-layered approach that combines technological solutions, user education, and ongoing vigilance. As AI continues to advance, defenders must evolve their strategies to stay ahead of sophisticated phishing techniques.

(1) Credential Stuffing:

Credential stuffing is a type of cyberattack in which attackers use automated tools to systematically input large sets of stolen usernames and passwords (credentials) into online services, with the aim of gaining unauthorized access to user accounts. This attack is successful when individuals reuse passwords across multiple online platforms. Here's how credential stuffing typically works:[4]

1. **Compromised Credentials:** Attackers obtain username and password combinations through various means, such as data breaches, phishing attacks, or purchasing them on the dark web.
2. **Automated Tools:** Cybercriminals use automated tools, often powered by scripts or specialized software, to rapidly input these stolen credentials into multiple online platforms.
3. **Credential Database:** Attackers may use a database of known credentials or leverage automated methods to generate potential username and password combinations.
4. **Large-Scale Login Attempts:** The automated tools systematically attempt to log in to various websites or online services using the stolen or generated credentials.
5. **Account Takeover:** If the same credentials are reused by the account owner on multiple platforms, successful login attempts can lead to unauthorized access and account takeover.
6. **Monetary Gain:** Once unauthorized access is gained, attackers may engage in various malicious activities, such as stealing personal information, conducting fraudulent transactions, or using the compromised accounts for other illicit purposes.

[4] INTSIGHTS, "The Dark Side of Latin America: Cryptocurrency, Cartels, Carding and the Rise of Cybercrime", p.6, available at: https://wow.intsights.com/rs/071-ZWD-900/images/Dark%20Side%20of%20Latin%20America.pdf. See also, "The Next, El Chapo is Coming for your Smartphone", June 26, 2020, available at: https://www.ozy.com/the-new-and-the-next/the-next-el-chapo-might-strike-your-smartphone-and-bank/273903/.

Mitigation Strategies:

1. **Multi-Factor Authentication (MFA):** MFA adds an additional layer of security by requiring users to provide a second form of authentication (e.g., a temporary code sent to a mobile device) along with their passwords.
2. **Password Policies:** Encourage or enforce strong, unique passwords and educate users about the risks of password reuse.
3. **Account Lockout Policies:** Implement account lockout mechanisms to prevent multiple unsuccessful login attempts within a short period, making automated credential stuffing less effective.
4. **Monitoring and Anomaly Detection:** Employ systems that monitor user behavior and detect unusual patterns, such as a sudden increase in login attempts or logins from unusual locations.
5. **IP Blocking and Rate Limiting:** Implement measures to block or limit login attempts from specific IP addresses, especially if they show suspicious behavior.
6. **Regular Password Changes:** Encourage users to regularly change their passwords, reducing the window of opportunity for attackers to use stolen credentials.
7. **Dark Web Monitoring:** Monitor the dark web for the sale or distribution of stolen credentials associated with your organization.
8. **Educating Users:** Provide ongoing cybersecurity awareness training to educate users about the risks of password reuse and the importance of using strong, unique passwords.[5]

Credential stuffing attacks highlight the importance of implementing proactive security measures and promoting good password hygiene practices among users. As cyber threats continue to evolve, organizations must stay vigilant and adopt a multi-layered security approach to protect against unauthorized access and account takeovers.

Malware Generation and Modification:

Malware generation and modification using artificial intelligence (AI) represent a sophisticated approach that cybercriminals use to create malicious software capable of evading traditional security measures. AI techniques, particularly machine learning, enable attackers to develop

[5] Malwarebytes Lab, "When Artificial Intelligence goes awry: separating science fiction from fact", without publication date, available at: https://resources.malwarebytes.com/files/2019/06/Labs-Report-AI-gone-awry.pdf.

malware that can adapt, learn, and change its behavior to avoid detection. Here's an overview of how malware generation and modification with AI typically work:[6]

1. **Training the Malware Model:** Cybercriminals use AI algorithms, such as machine learning models, to train malware on various features, behaviors, and evasion techniques. This training involves exposing the model to different security environments and responses.
2. **Evasion of Signature-Based Detection:** Traditional antivirus solutions rely on signature-based detection, which involves identifying known patterns of malware. AI-powered malware is designed to continuously modify its code and behavior to avoid being recognized by signature-based methods.
3. **Polymorphic Malware:** AI is used to create polymorphic malware, which can change its code structure and appearance while retaining its malicious functionality. This variability makes it challenging for signature-based security tools to keep up.
4. **Behavioral Analysis Evasion:** AI-driven malware can adapt its behavior based on the analysis performed by security solutions. This includes modifying its execution flow, obfuscating its activities, or delaying malicious actions to avoid detection.
5. **Dynamic C2 Communication:** AI-enhanced malware can dynamically adjust its command and control (C2) communication patterns. This adaptability makes it more challenging for security solutions to identify and block malicious network traffic.
6. **Context-Aware Attacks:** AI algorithms analyze contextual information, such as the target environment, user behavior, or security measures in place. The malware can then adjust its tactics, techniques, and procedures (TTPs) accordingly for a more targeted and effective attack.
7. **Learning from Security Responses:** AI-powered malware can learn from its interactions with security defenses. If a particular evasion technique is successful, the malware may prioritize or enhance that technique in subsequent attacks.
8. **AI-Generated Malware Variants:** Automated AI systems can generate multiple variants of malware with unique characteristics. This variability further complicates the task of security analysts and tools in detecting and mitigating the threats.

[6] SIEMENS Energy, "Managed Detection and Response Service", 2020, available at: https://assets.siemens-energy.com/siemens/assets/api/uuid:a95b9cd3-9f4d-4a54-8c43-77fbdb6f418f/mdr-white-paper-double-sided-200930.pdf. SIEMENS Energy, "Managed Detection and Response Service", 2020, available at: https://assets.siemens-energy.com/siemens/assets/api/uuid:a95b9cd3-9f4d-4a54-8c43-77fbdb6f418f/mdr-white-paper-double-sided-200930.pdf.

Mitigation Strategies:

1. **Behavioral Analysis:** Employ advanced security solutions that focus on behavioral analysis rather than relying solely on signature-based detection.
2. **Heuristic Detection:** Implement heuristic approaches that identify suspicious behaviors or patterns, even if specific signatures are not known.
3. **Network Traffic Monitoring:** Regularly monitor network traffic for anomalous patterns, especially those indicative of command-and-control activities.
4. **User Education:** Educate users about the risks of downloading files from untrusted sources or clicking on suspicious links, as these are common entry points for malware.
5. **Patch and Update Systems:** Keep software and systems up to date to patch known vulnerabilities, reducing the potential attack surface for malware.
6. **AI-Enhanced Security Tools:** Invest in AI-powered security solutions that can adapt to evolving threats and employ machine learning to detect anomalous activities.
7. **Threat Intelligence Sharing:** Collaborate with threat intelligence sharing platforms to stay informed about emerging threats and vulnerabilities.
8. **Network Segmentation:** Implement network segmentation to limit the spread of malware within a network, minimizing the potential impact of an infection.

As AI-driven malware becomes more prevalent, organizations need to continuously enhance their cybersecurity strategies and adopt advanced technologies to detect and mitigate evolving threats effectively.[7]

(2) Adversarial Attacks:

Adversarial attacks in the context of artificial intelligence (AI) refer to deliberate and targeted efforts to manipulate the behavior of machine learning models. These attacks aim to exploit vulnerabilities in the model's decision-making process, leading to incorrect or undesirable outcomes. Adversarial attacks can occur in various domains, including image recognition, natural language processing, and other AI applications. Here's an overview of adversarial attacks:

1. **Input Perturbation Attacks:** Modifying input data to mislead the model without changing the overall perception of a human observer. Adding subtle, imperceptible noise to an image to cause misclassification by an image recognition system.

[7] POLITICO, "Automated racism: How tech can entrench bias", March 2, 2021, available at: https://www.politico.eu/article/automated-racism-how-tech-can-entrench-bias/.

2. **Model Evasion Attacks:** Crafting inputs specifically designed to exploit weaknesses in the model, causing it to misclassify or provide incorrect outputs. Creating a slightly modified version of an input image to force a misclassification.
3. **Model Inversion Attacks:** Attempting to reconstruct sensitive information about the training data or the model's parameters by exploiting its outputs. Reconstructing an image that was part of the training dataset based on the model's predictions.
4. **Membership Inference Attacks:** Determining whether a specific data point was part of the training dataset by leveraging the model's outputs. Inferring whether a given image was used during the training of a machine learning model.
5. **Transfer Attacks:** Creating adversarial examples on one model and successfully transferring those examples to another model, even if the models have different architectures. Crafting an adversarial example for one image classification model and successfully fooling another model with a different architecture.[8]

Challenges and Mitigation Strategies:

1. **Lack of Robustness:** Many machine learning models are susceptible to small changes in input data, making them vulnerable to adversarial attacks. Employ robust architectures, such as adversarially trained models, and augment training datasets with adversarial examples.
2. **Model Opacity:** Lack of transparency in complex models can make it difficult to understand and defend against adversarial attacks. Use explainable AI techniques to enhance model interpretability and identify vulnerabilities.
3. **Continuous Adaptation:** Adversarial attacks evolve over time, requiring continuous adaptation of defenses. Regularly update and retrain models with new data, and employ dynamic defenses that can adapt to emerging attack strategies.
4. **Transferability:** Adversarial examples crafted for one model can be effective on others, leading to transfer attacks. Implement model-specific defenses and ensemble approaches to increase resistance against transfer attacks.[9]

[8] In October 2021, the European Parliament adopted a resolution to ban the use facial recognition technologies in public spaces by law enforcement authorities to ensure the protection of fundamental rights. See European Parliament, "Use of Artificial Intelligence by the police: MEPs oppose mass surveillance". LIBE Plenary Session press release, October 6, 2021, available at: https://www.europarl.europa.eu/news/en/press-room/20210930IPR13925/use-of-artificial-intelligence-by-the-police-meps-oppose-mass-surveillance.

[9] See the Budapest Convention Chart of Signatures and Ratifications at: https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/185/signatures?p_auth=yUQgCmNc.

5. **Adversarial Training:** Adversarial training can be computationally expensive and may not cover all possible attack scenarios. Balance computational costs with the need for robustness and consider hybrid approaches that combine adversarial training with other defense mechanisms.
6. **Incorporating Security in Model Development:** Security considerations may not be adequately addressed during the development of machine learning models. Integrate adversarial robustness as a key component of the model development lifecycle, including during design, training, and testing phases.

Addressing adversarial attacks requires a holistic approach that combines robust model architectures, ongoing research in adversarial defense, and a deep understanding of the potential vulnerabilities in AI systems. As AI technology continues to advance, the development of more resilient and secure models remains a critical area of research and development.[10]

III. AI-ENHANCED RANSOMWARE

AI-enhanced ransomware refers to a type of malicious software that incorporates artificial intelligence (AI) techniques to enhance its capabilities, making it more sophisticated, evasive, and potentially more damaging. Ransomware is a type of malware that encrypts a victim's files or entire system, rendering them inaccessible, and then demands a ransom, usually in cryptocurrency, for the decryption key. When AI is integrated into ransomware, it can be used to optimize various aspects of the attack. Here are some characteristics and considerations regarding AI-enhanced ransomware:

1. **Automated Target Selection:** AI can be employed to analyze potential targets, identifying high-value systems or networks. This automated selection process allows attackers to focus their efforts on targets that are more likely to pay a ransom.
2. **Dynamic Encryption Techniques:** AI can be used to dynamically adjust encryption algorithms and keys, making it more challenging for cybersecurity experts to develop universal decryption solutions. This adaptability can increase the ransomware's effectiveness.
3. **Behavioral Analysis Evasion:** AI-enhanced ransomware may incorporate machine learning algorithms to analyze the behavior of security systems. This information can

[10] Trend Micro Research, EUROPOL EC3 and UN Interregional Crime and Justice Research Institute (UNICRI), Malicious Uses and Abuses of Artificial Intelligence, 19 November 2020, available at: https://www.europol.europa.eu/publications-documents/malicious-uses-and-abuses-of-artificial-intelligence.

be used to adjust the ransomware's behavior to evade detection and response mechanisms.

4. **Spear Phishing Customization:** AI can personalize spear-phishing messages by analyzing vast amounts of data about potential victims. This customization increases the likelihood of successful social engineering attacks, leading to the initial infection of the target system.
5. **Adaptive Ransom Demands:** AI algorithms can analyze the target's financial capabilities and adjust ransom demands accordingly. This adaptability enables attackers to set ransom amounts that are more likely to be paid while avoiding demands that may be deemed excessive.
6. **Automated Communication:** AI can be used to automate communication between the attackers and victims, facilitating negotiations and providing decryption keys upon payment. This automation allows cybercriminals to manage multiple attacks simultaneously.
7. **Predictive Analytics for Payments:** AI can analyze historical data to predict the likelihood of victims paying the ransom based on various factors. This information can inform attackers' strategies for setting ransom amounts and negotiating with victims.
8. **Evolution of Attack Techniques:** AI can be utilized to continuously evolve the ransomware's attack techniques, learning from previous successes and failures. This adaptability makes it challenging for security solutions to keep up with emerging threats.

Mitigation Strategies:

1. **Regular Backups:** Maintain regular and secure backups of critical data to ensure that even if ransomware strikes, data can be restored without paying the ransom.
2. **User Education:** Educate users about the risks of opening suspicious emails, clicking on unknown links, or downloading files from untrusted sources.
3. **Multi-Factor Authentication (MFA):** Implement MFA to add an extra layer of security, making it more difficult for attackers to gain unauthorized access.
4. **Network Segmentation:** Segment networks to limit the lateral movement of ransomware within an organization, minimizing potential damage.

5. **Security Software:** Utilize advanced cybersecurity solutions that employ behavioral analysis, anomaly detection, and AI-driven threat intelligence to detect and prevent ransomware attacks.
6. **Incident Response Plan:** Develop and regularly test an incident response plan to ensure a swift and coordinated response in the event of a ransomware attack.
7. **Collaboration and Information Sharing:** Engage in information sharing and collaboration within the cybersecurity community to stay informed about emerging threats and mitigation strategies.

As AI technologies continue to advance, both attackers and defenders are likely to leverage AI in their strategies. Staying ahead of AI-enhanced ransomware requires a combination of proactive security measures, user awareness, and collaboration within the cybersecurity community.

(3) Deepfake Threats:

Deepfake threats refer to the use of deep learning and artificial intelligence (AI) techniques to create highly realistic but fake audio, video, or other digital content, often for malicious purposes. Deepfakes have raised concerns due to their potential to deceive individuals, manipulate public opinion, and facilitate various forms of fraud or misinformation. Here are key aspects of deepfake threats:

1. **Generation of Realistic Content:** Deepfake technology uses AI algorithms, particularly deep neural networks, to analyze and synthesize realistic-looking and sounding content. Deepfake videos can manipulate the facial expressions and lip movements of individuals, creating seemingly authentic videos of people saying or doing things they never did.
2. **Manipulation of Audio:** Deepfake techniques extend beyond visual content to manipulate audio, enabling the creation of convincing voice recordings or altering the spoken words of individuals. Fake voice recordings of public figures making false statements or fraudulent phone calls impersonating someone.
3. **Impersonation and Identity Theft:** Deepfakes can be used to impersonate individuals, making it appear as if they are engaging in actions or making statements they never did. Impersonating a company executive to authorize financial transactions or creating fake videos of celebrities endorsing products.
4. **Misinformation and Disinformation:** Deepfakes can contribute to the spread of misinformation by creating false narratives, making it difficult for individuals to discern

between genuine and manipulated content. Fabricating political speeches, news reports, or public announcements to influence public opinion.

5. **Cybersecurity Threats:** Deepfakes may be used in cybersecurity attacks, such as creating fake audio or video messages to trick individuals into divulging sensitive information. Phishing attacks using deepfake voicemails from trusted contacts, urging recipients to take urgent actions.
6. **Political and Social Manipulation:** Deepfakes can be employed to manipulate public perceptions, influence elections, or create social unrest by generating misleading content. Creating fake videos of political figures making controversial statements or spreading false information during election campaigns.

Mitigation Strategies:

1. **Deepfake Detection Tools:** Develop and deploy advanced tools and algorithms designed to detect deepfake content. This may involve using AI-based techniques to identify anomalies in videos or audio recordings.
2. **Media Authentication Standards:** Establish standards for authenticating and verifying the origin of media content, making it more difficult for deepfakes to be accepted as genuine.
3. **Blockchain Technology:** Explore the use of blockchain to create secure and immutable records of original media content, ensuring that tampering is easily detectable.
4. **User Education:** Raise awareness among the general public and organizations about the existence of deepfakes, emphasizing the importance of critical thinking and verifying information sources.
5. **Watermarking and Digital Signatures:** Implement digital watermarking or digital signatures on media content to verify its authenticity. These can be used as indicators of tampering.
6. **Legislation and Regulation:** Enact and enforce laws and regulations that address the creation and malicious use of deepfake technology, imposing legal consequences for those who engage in deceptive practices.[11]
7. **Collaboration Between Tech Companies:** Foster collaboration between technology companies, researchers, and governments to share information and collectively develop effective solutions for detecting and mitigating deepfake threats.

[11] Cybercrime Convention Committee, "T-CY Rules of Procedure. As revised by T-CY on 16 October 2020", Strasbourg, 16 October 2020, available at: https://rm.coe.int/t-cy-rules-of-procedure/1680a00f34.

8. **Transparent AI Algorithms:** Promote transparency in AI algorithms and deep learning models, enabling researchers and the public to understand how these technologies operate and identify potential biases.

Addressing deepfake threats requires a multidimensional approach that combines technological solutions, regulatory measures, and public awareness. As the technology continues to advance, ongoing efforts are crucial to stay ahead of emerging deepfake challenges.

(4) Automated Social Engineering:

Automated social engineering refers to the use of automated tools, scripts, or artificial intelligence (AI) to conduct social engineering attacks. Social engineering involves manipulating individuals to divulge confidential information, such as passwords or sensitive data, or to perform actions that may compromise security. When automation is employed in social engineering, it allows attackers to scale their efforts and target a larger number of individuals more efficiently. Here are key aspects of automated social engineering:

1. **Automated Phishing Campaigns:** Automated tools are used to generate and distribute phishing emails or messages on a large scale. These messages often impersonate trusted entities to trick recipients into revealing sensitive information or clicking on malicious links. Mass phishing emails claiming to be from a bank, requesting users to click on a link and enter their login credentials.[12]
2. **Chatbots and Conversational Agents:** AI-powered chatbots or conversational agents are used to engage with individuals, attempting to extract sensitive information through natural language interactions. Chatbots mimicking customer support representatives to trick users into sharing account details.
3. **Credential Stuffing with Social Engineering:** Automated tools are used to systematically test large sets of stolen or guessed username-password combinations across various online platforms to gain unauthorized access. Automated scripts attempting to log in to multiple accounts using common passwords obtained from previous data breaches.
4. **Automated Vishing (Voice Phishing):** AI-driven voice synthesis technology is employed to create automated voice calls that impersonate legitimate entities, tricking individuals

[12] Council of Europe, "Second Additional Protocol to the Budapest Convention adopted by the Committee of Ministers of the Council of Europe", Strasbourg, 17 November 2021, available at: https://www.coe.int/en/web/cybercrime/-/second-additional-protocol-to-the-cybercrime-convention-adopted-by-the-committee-of-ministers-of-the-council-of-europe.

into providing sensitive information over the phone. Automated calls claiming to be from a government agency, requesting personal information for supposed verification purposes.

5. **Social Media Automation:** Automated tools are used to scrape information from social media platforms and craft personalized social engineering attacks based on the collected data. Automated messages on social media platforms pretending to be a friend or colleague, asking for sensitive information.
6. **SMS and Messaging Automation:** Automated tools send fraudulent text messages or messages through messaging apps, attempting to trick recipients into taking specific actions or revealing information. Automated SMS claiming to be from a service provider, asking users to click on a link to resolve an issue.

Mitigation Strategies:

1. **User Education:** Educate users about the risks of social engineering attacks, including phishing emails, phone calls, and messages. Encourage skepticism and critical thinking.
2. **Multi-Factor Authentication (MFA):** Implement MFA to add an extra layer of security, even if credentials are compromised.
3. **Security Awareness Training:** Conduct regular security awareness training sessions to keep users informed about evolving social engineering tactics.
4. **Email Filtering:** Use advanced email filtering solutions that can detect and block phishing emails before they reach users' inboxes.
5. **Endpoint Security:** Deploy endpoint security solutions that can detect and block malicious scripts and activities on users' devices.
6. **Phone Call Verification:** Establish protocols for verifying the authenticity of phone calls, especially those requesting sensitive information. Encourage users to call back using official contact details.
7. **Behavioral Analytics:** Implement behavioral analytics tools to monitor user behavior and detect anomalies that may indicate social engineering attempts.
8. **Regular Updates and Patching:** Keep software and systems up to date to patch known vulnerabilities that could be exploited by automated social engineering tools.

Combating automated social engineering requires a combination of technological solutions, user education, and ongoing vigilance. Organizations should stay informed about emerging tactics and continually refine their cybersecurity strategies to stay ahead of evolving threats.

(5) AI-Driven Fraud:

AI-driven fraud refers to fraudulent activities in which artificial intelligence (AI) technologies are used to perpetrate or enhance the effectiveness of fraudulent schemes. Fraudsters leverage AI algorithms and techniques to automate, optimize, and scale their illicit activities, making detection and prevention more challenging. Here are key aspects of AI-driven fraud:

1. **Automated Fraudulent Transactions:** AI algorithms can be employed to automate the execution of fraudulent transactions, such as unauthorized credit card purchases or fund transfers. Automated bots systematically testing stolen credit card information on e-commerce websites to make small transactions that may go unnoticed.
2. **Account Takeover (ATO):** AI can be used to analyze large datasets to identify patterns and behaviors associated with legitimate user accounts. Once compromised, these accounts can be exploited for various fraudulent activities. Automated attacks targeting weak passwords or leveraging stolen credentials obtained from data breaches to gain unauthorized access to user accounts.
3. **Identity Theft:** AI-driven techniques can assist in the creation of synthetic identities or the impersonation of real individuals to commit identity theft. Using AI to generate fake identification documents or manipulate facial images for fraudulent account creation.[13]
4. **Phishing and Social Engineering:** AI can be employed to enhance phishing attacks by customizing messages based on the analysis of individual profiles, increasing the likelihood of success. AI-generated phishing emails personalized with information about the target to trick individuals into revealing sensitive information.
5. **Fraudulent Credit Scoring:** AI algorithms can manipulate credit scoring models by generating false financial data to secure loans or credit lines. Generating fake financial transactions and credit history to artificially improve credit scores for fraudulent loan applications.
6. **Evasion of Anti-Fraud Systems:** Fraudsters use AI to analyze and adapt to anti-fraud measures, making their activities more difficult to detect and mitigate. Developing AI-driven evasion techniques that dynamically adjust fraudulent behaviors based on the responses of anti-fraud systems.

[13] The Conference program of the 2018 Octopus conference on cooperation against cybercrime is available at: https://rm.coe.int/3021-90-octo18-prog/16808c2b04.

7. **Ad Fraud:** AI is utilized to generate automated clicks, impressions, or views to manipulate online advertising systems for financial gain. Automated bots generating fake clicks on online ads to fraudulently inflate advertising revenues.
8. **Insider Threats:** AI can be used to analyze patterns of employee behavior to identify potential insider threats for financial or data theft. Automated analysis of employee activities to detect unusual patterns indicative of fraudulent activities.

Mitigation Strategies:

1. **Behavioral Analytics:** Implement behavioral analytics solutions to monitor and detect anomalies in user behavior, helping identify unusual patterns indicative of fraud.[14]
2. **Machine Learning for Fraud Detection:** Utilize machine learning models to analyze patterns in transaction data and detect anomalies associated with fraudulent activities.
3. **Biometric Authentication:** Implement biometric authentication methods to enhance identity verification and reduce the risk of account takeover.
4. **Transaction Monitoring:** Regularly monitor transactions for unusual patterns, such as large transfers, rapid transactions, or transactions from unfamiliar locations.
5. **User Education:** Educate users about the risks of phishing attacks, the importance of strong passwords, and the need to verify the authenticity of requests for sensitive information.[15]
6. **AI-Powered Fraud Detection Systems:** Deploy AI-powered fraud detection systems that can adapt to evolving fraud tactics and patterns.
7. **Multi-Factor Authentication (MFA):** Implement MFA to add an extra layer of security, making it more difficult for fraudsters to gain unauthorized access.
8. **Regular Audits and Reviews:** Conduct regular audits and reviews of security measures and fraud prevention strategies to identify and address potential vulnerabilities.
9. **Collaboration and Information Sharing:** Collaborate with industry partners, share threat intelligence, and participate in information-sharing initiatives to stay informed about emerging fraud trends.

[14] The Rules of procedure, adopted documents, activity reports and the Meetings of the 'Lanzarote Committee' are available at: https://www.coe.int/en/web/children/lanzarote-committee#{%2212441908%22:[].
[15] The presentation and materials of this panel are available at: https://www.coe.int/en/web/cybercrime/workshop-cybercrime-e-evidence-and-artificial-intelligence.

AI-driven fraud presents a dynamic and evolving threat landscape. Organizations must continually enhance their cybersecurity strategies, leveraging advanced technologies and collaborative efforts to detect and prevent fraudulent activities effectively.

Law relating to Artificial Intelligence crime

As of January 2024, specific laws explicitly addressing artificial intelligence (AI) crimes may be limited, and legal frameworks often encompass broader categories of cybercrime, data protection, and technology-related offenses. However, the legal landscape is evolving to address the challenges posed by AI-based activities. Here are some aspects to consider:

General Legal Framework:

1. **Computer Crime Laws:** Laws like the Computer Fraud and Abuse Act (CFAA) in the United States and the Computer Misuse Act in the United Kingdom address unauthorized access, computer-related fraud, and offenses related to computer systems.
2. **Data Protection and Privacy Laws:** Regulations like the General Data Protection Regulation (GDPR) in the European Union and various national data protection laws govern the processing and protection of personal data, including data processed by AI systems.
3. **Anti-Fraud and Identity Theft Laws:** Laws related to fraud, identity theft, and financial crimes may apply to offenses facilitated by AI, such as automated phishing or fraudulent transactions.

Emerging Trends and Considerations:[16]

1. **AI Ethics and Responsible AI:** While not legal frameworks, ethical guidelines and principles for AI development and use are gaining attention. Organizations and researchers are encouraged to adhere to responsible AI practices.
2. **Algorithmic Accountability:** Some jurisdictions are exploring the concept of algorithmic accountability, emphasizing transparency, fairness, and accountability in the use of AI algorithms.

[16] The Final Virtual Plenary Meeting of CAHAI from 30.11.2021 to 02.12.2021 will facilitate meaningful discussions towards the adoption of a document outlining the possible elements of a legal framework on AI, which may include binding and non-binding standards based on the Council of Europe's standards on human rights, democracy and rule of law. See Council of Europe, "The CAHAI to hold its final meeting", Strasbourg, 24 November 2021, available at: https://www.coe.int/en/web/artificial-intelligence/-/cahai-to-hold-its-final-meeting.

3. **Bias and Discrimination:** Laws addressing discrimination and bias may indirectly impact AI systems, especially if algorithms lead to discriminatory outcomes.

International Collaboration:

1. **International Treaties and Agreements:** International cooperation on cybercrime is facilitated through treaties and agreements, such as the Budapest Convention on Cybercrime. These agreements may indirectly apply to AI-related offenses.
2. **UN Guidance on Cybersecurity:** The United Nations provides guidance on issues related to cybersecurity, which may include discussions on the implications of AI in cyberspace.

Future Developments:

1. **Legislation Addressing AI-Specific Offenses:** Some jurisdictions are considering or may develop legislation specifically addressing crimes facilitated by AI technologies, such as the malicious use of deepfakes or AI-driven fraud.
2. **Industry-Specific Regulations:** Certain industries, like finance and healthcare, may introduce regulations addressing the use of AI in specific contexts, considering the potential risks and challenges.
3. **Application of Existing Laws:** AI developers and users must comply with existing laws related to data protection, privacy, cybersecurity, and intellectual property, ensuring that AI activities adhere to legal standards.

Summary

It is important to note that the legal landscape is dynamic, and new laws and regulations may have been introduced since my last update. Additionally, legal interpretations and enforcement practices can vary across jurisdictions. Organizations and individuals involved in AI activities should stay informed about developments in the legal and regulatory environment, ensuring compliance with applicable laws and ethical guidelines.[17]

Addressing these challenges requires a multi-faceted approach involving technological innovation, international cooperation, and a dynamic regulatory environment that can adapt to the evolving landscape of AI-based cyber threats. Additionally, raising awareness and promoting cybersecurity education are crucial components in building a resilient digital

[17] UNICRI and INTERPOL, "Artificial Intelligence and Robotics for Law Enforcement", 2019, available at: https://issuu.com/unicri/docs/artificial_intelligence_robotics_la/4?ff.

society. To combat AI-based cybercrimes, cybersecurity professionals need to continually update and enhance their defence mechanisms, incorporating AI-driven solutions for threat detection, anomaly detection, and behavioural analysis. Additionally, collaboration between the public and private sectors, as well as international cooperation, is crucial in developing strategies to address these evolving cyber threats effectively.

****