How Law Enforcement Utilizes AI to Deal with Crime

Sinchul Back, Ph.D.

President & CEO, the Royal Robotics & AI Security

Assistant Professor, the University of Scranton

ROYAL ROBOTIC & AI SECURITY LLC.

ABSTRACT:

Due to the advancement of Artificial Intelligence (AI), all sectors of society have been affected by AI. AI has been beneficial from public sector to private sector; however, criminals have also taken advantage of the new technology to conduct malicious activities, expand existing vulnerabilities, and introduce new threats. This article explores the malicious uses of AI capabilities. The purpose of this study is to articulate the types of activities and corresponding risks of using AI. Accordingly, this study will discuss the current use of AI in the criminal justice system to fight against the malicious use of AI. Finally, this study suggests a road map for government officials to effectively combat criminal use of AI by using strategic policies.

INTRODUCTION

Artificial Intelligence (AI) is dramatically changing the dynamics of the world around us. It is exhilarating, inspiring, and promising. Despite AI's benefits to society, there is also a grey area in the use of AI. Many law enforcement agencies know that criminals find ways to use AI in their malicious activities (McCarthy, 2023). "It is a similar situation in virtual kidnap for ransom cases," FBI Special Agent in Charge for Oregon, Kieran Ramsey stated. "Instead of using photos, AI generates a phone call with a fake voice of a family member. The voice will say something along the lines of being in danger and a payment would need to be made." Similarly, criminals can employ AI to assist with the scale and effectiveness of their social engineering attacks (Durbin, 2020). For example, AI can learn to spot patterns in behavior, understanding how to convince people that a video, phone call, or email is legitimate and then persuade them to compromise networks and hand over sensitive data. All the social techniques cybercriminals

currently employ could be improved immeasurably with the help of AI. Criminals can use AI as weapon to achieve their malicious goals.

Therefore, the purpose of this study is to articulate the types of activities and corresponding risks of using AI for malicious use. On the other side, this study will also discuss the current use of AI in the criminal justice system to fight against such malicious use of AI. Finally, this study suggests a road map for government officials to effectively deal with criminal use of AI by using strategic policies.

Overview of Artificial Intelligence, and Malicious Use/Abuse of AI

According to the European Union Artificial Intelligence Act (EU AI Act), the term AI system "means software that is developed with one or more techniques and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with." Blauth et al. (2022) claim that the use of AI can enable already existing forms of crime ('cyber-enabled crime') or establish new forms of crime ('cyber-dependent crime'). It means AI potentially enables attacks that are larger in scale and reach than previously possible with other technologies. In addition, they proposed the term "AI-Crime" to describe the situation in which AI technologies are re-oriented to

facilitate criminal activity. AI-Crime focuses on behavior already defined as criminal within the given legislation. Hibbard (2015) and Johnson and Verdicchio (2017) proposed that the term "harmful AI." This term was used in a context in which the AI program/application itself caused harm.

Given the concepts mentioned above, Ciancaglini et al. (2020) explains that "malicious use" refers to the use of AI to enhance, augment, or enable acts committed by individuals or organizations. He also illustrates that "malicious abuse" can be referred to the exploitation of AI with bad intentions, as well as attacks on AI systems themselves. The following section will discuss AI-enabled attacks (malicious use of AI) and the vulnerabilities of AI models (malicious abuse of AI).

Malicious Use of AI: AI-Enabled and AI-Enhanced Attacks

Social Engineering

Social engineering manipulates individuals to divulge confidential information through deceit. With the aid of AI, attackers can refine their deceptive methods, thereby enhancing the effectiveness and benefits of their attacks.

1) Deception and Phishing

AI can craft 'social bots' that mimic human behavior to deceive and manipulate individuals online. These bots can create content and communicate with users, potentially leading them to malicious websites (Freitas et al., 2015; Boshmaf et al., 2013; Ferrara et al., 2016). A prime example is 'CyberLover', a dating chatbot launched in 2007, which used natural language processing (NLP) to engage users in chat rooms and encourage them to reveal personal data or engage with deceitful links (Rossi, 2020).

Phishing, where attackers disguise themselves as trustworthy entities to get victims to reveal information, can also be amplified by AI. An experiment on Twitter demonstrated that AI-generated text can be efficiently used for phishing because informal tones, typographical errors, and abbreviated links are commonly accepted on the platform (Seymour & Tully, 2016).

2) Big Nudging and Manipulation

Bots can potentially influence public perspectives and even election results. By amplifying specific content, such bots can falsely portray a political figure or movement as being more popular than it genuinely is (Bessi & Ferrara, 2016).

Astroturfing mimics grassroots support for an initiative or individual, even when little genuine backing exists. This strategy involves disseminating a multitude of posts from diverse accounts to manipulate public sentiment. Astroturfing can be discerned on platforms such as Twitter,

blogs, and news websites and can be wielded as disinformation tactics (Kovic et al., 2018; Keller et al., 2020; Zwitter, 2016).

The U.S. Federal Communications Commission's (FCC) net neutrality consultation exemplified potential bot interference. An analysis discovered that over 80% of the roughly 22 million comments received by the FCC during the consultation were bot-generated (Thuen, 2017).

AI's potential misuse also extends to online profiling and targeting. The scandal involving Cambridge Analytica highlights this. The GSRApp gathered user data, including psychological traits, which was then used to generate voter profiles. These profiles facilitated targeted advertising and message manipulation based on individual psychological profiles, influencing democratic processes (Bakir, 2020).

Misinformation and Fake News

Advancements in technology, the rise of blogging platforms, and the proliferation of social media have revolutionized how people access news and form opinions. This digital transformation has, however, facilitated the rapid spread of misinformation and fake news. Although the term "fake news" has faced criticism (Habgood-Coote, 2019; Sullivan, 2017), it remains an essential term to encourage debates about digital literacy and academic inquiry into the issue (Garcia, 2021). The spread of false news can have dire consequences, especially during

uncertain times like pandemics (Oyeyemi et al., 2014) or political events like elections (Bennett & Livingston, 2018; Wilder & Vorobeychik, 2019). AI can exacerbate the creation and dissemination of such content, posing a threat to society and potentially to democracy itself (Nemitz, 2018).

AI tools like GPT-3 can autonomously produce written content that mimics human writing, making it harder for readers to discern its authenticity (Floridi & Chiriatti, 2020). For instance, "NotRealNews.net" uses AI to craft artificial news articles, highlighting the tool's potential to generate convincing misinformation. Such AI-generated content, paired with advanced targeting techniques, can amplify the reach and impact of disinformation campaigns, influencing voter behavior or reinforcing pre-existing beliefs (Leyva & Beckett, 2020). As technology advances, AI can fine-tune content to fit audience preferences, fostering echo chambers and increasing polarization (Floridi & Chiriatti, 2020).

Hacking

1) Forgery: Deepfakes

Deepfakes are hyper-realistic videos and images crafted using AI, making it hard to differentiate between real and fabricated content (Thies et al., 2016; Westerlund, 2019). Although photo manipulation has existed since the creation of tools like Photoshop, AI has made these forgeries

more sophisticated. Prominent uses of deepfakes have mostly targeted celebrities and politicians (Guarnera et al., 2020). These forgeries can be used maliciously for propaganda, disinformation, bullying, or blackmail (Maras & Alexandrou, 2019). Awareness-raising initiatives, like satirical videos, help in educating the public about the potential risks of such technology. The "liar's dividend" complicates matters, where individuals discredit genuine videos by claiming they are deepfakes (Chesney & Citron, 2019).

2) Repetitive Tasks

AI excels at repetitive tasks, which can be harnessed maliciously. An instance is Ticketmaster's incident, where AI was used to bypass Captcha for buying and reselling tickets (Zetter, 2010). Advanced AI might elevate concerns about crimes like password-cracking, with AI-enhanced brute-force attacks having a higher success rate (Trieu & Yang, 2018).

3) Malware

Malware attacks have evolved over decades, becoming a significant cybersecurity threat (Gibert et al., 2020).

The AV-TEST Institute (2021) identifies over 350,000 new malware and PUAs daily.

There is rising apprehension about AI being used to create more potent malware (Ciancaglini et al., 2020; Zwitter, 2016). IBM's DeepLocker (Kirat et al., 2018), presented at Black Hat USA 2018,

showcases malware combined with AI to enhance its evasion capabilities. The potential risks associated with AI-enabled or enhanced malware are profound, and preemptive actions should be taken.

Malicious Abuse of AI: Vulnerabilities of AI Models

Integrity Attacks

Machine learning (ML) models can be vulnerable to integrity attacks where attackers manipulate the software or underlying data (Luo et al., 2018; Garcia, 2021). `Adversarial examples' are malicious inputs crafted to trick ML models, resulting in misclassifications. Some of these alterations are so subtle that humans might not notice, but they can still mislead AI systems (Kurakin et al., 2016; Goodfellow et al., 2014). In `poisoning attacks', attackers introduce corrupted data points into the training set, thereby manipulating the classifier's outcomes (Jagielski et al., 2018). Microsoft's AI chatbot, Tay, faced an adversarial attack when users exploited its features to make it produce inappropriate content. As a result, Microsoft had to suspend Tay within 16 hours of its release (Schwartz, 2019; Lee, 2016). Researchers (Gu et al., 2019) at NYU introduced the concept of BadNet, a network that behaves normally until triggered. For instance, a self-driving car's traffic sign detection could misinterpret a stop sign with a specific trigger (e.g., a yellow `Post-It' note). The EU AI Act has addressed training data concerns for 'high-risk systems' (Renda et al., 2021).

Unintended Outcomes of the Use of AI

AI training models might produce results different from developers' intentions. Specifically, models, especially those based on neural networks, might unintentionally memorize and disclose sensitive data (Carlini et al., 2019). Google's Smart Compose for Gmail underwent rigorous testing to ensure it did not unintentionally suggest private information based on memorized data. The objective is to ensure the model does not inadvertently expose personal details (Chen et al., 2019). This exemplifies the challenge where even without malicious intent, AI might behave unexpectedly and potentially disclose private information.

Algorithmic Trading/Stock Market Manipulation

AI systems enable financial decisions to be made at an unprecedented speed, which has both positive and negative implications (Scopino, 2020). This modern financial technology has led to reduced transactional costs and capital costs (Lin, 2017). Rapid algorithmic trading, however, introduces market instability, resulting in the risk of high-speed market crashes, also known as "flash crashes" (Wiener-Bronner, 2018). The 2010 flash crash, which resulted in a nearly $1 trillion loss, highlighted the challenges of automated decision-making in finance (Martin, 2020). Practices like spoofing and layering, used to manipulate high-frequency trading, have been prohibited (Zwitter, 2017). Regulatory discussions often revolve around the harm from malicious actors, but accidents or insufficient algorithm testing also pose threats (Scopino, 2020). To

counteract the risks of flash crashes, there are suggestions for the establishment of insurance systems like a "National Protection Fund" to ensure stability and compensate affected investors (Yadav, 2016). Strengthening cybersecurity and algorithm assessments can also be crucial preventive measures.

Membership Inference Attacks

Membership inference attacks aim to determine the samples used in training an ML model (Webster et al., 2021). These attacks are effective on various systems, including classification models, sequence-to-sequence models, and generative adversarial networks (GANs) (Hu et al., 2022). A study highlighted that the faces generated by the website "This person does not exist" resemble the individuals from the training data (Webster et al., 2021). This indicates that malicious actors might, through such attacks, identify the real identities of the data sources. Such inference attacks pose significant privacy concerns. For example, an attacker might link an illness to an actual individual if the model was trained on medical data (Hu et al., 2022). To mitigate the risks, it's essential to train models on diverse datasets, reduce dataset bias, and ensure rigorous pre-deployment testing.

APPLICATIONS FOR CRIMINAL JUSTICE AND PUBLIC SAFETY

AI's role in the criminal justice system is expanding across various domains to enhance accuracy and efficiency. In particular, AI is primarily employed in four key areas: analyzing public safety videos and images, DNA assessment, detecting gunshots, and predicting criminal activities.

Analyzing Public Safety Videos and Images

In the fields of criminal justice and law enforcement, video and image analysis serve as pivotal tools for gathering insights on individuals, objects, and activities, aiding criminal investigations. However, processing this visual data is an arduous task, necessitating the deployment of knowledgeable personnel and subject matter experts. The sheer amount of data, combined with rapidly evolving technology like smartphones and their operating systems, makes the analysis susceptible to human errors, especially given the scarcity of specialized experts in this domain (Brynjolfsson & McAfee, 2018). Advancements in AI technology offer a solution to these challenges by emulating expert-level proficiency. While conventional software algorithms guiding human analysts are restricted to predefined attributes, such as eye shape, color, or inter-eye distance for facial identification, or basic demographic data for pattern analysis, AI-driven algorithms for video and image analysis are more dynamic. They not only master intricate tasks but also autonomously formulate their distinctive facial recognition criteria and

other parameters, often surpassing human benchmarks. This capability enables these AI systems to accurately recognize faces, pinpoint weapons or various items, and identify intricate scenarios like accidents or criminal events. Recognizing these challenges and the transformative potential of AI, the National Institute of Justice (NIJ) has channeled resources into multiple domains. Their goal is to augment the efficiency, quality, and precision of data acquisition, image processing, and analysis, as well as to enrich the contextual relevance of the data (Marr, 2016).

DNA Analysis

From a scientific and evidence processing perspective, AI presents significant advantages to the law enforcement sector, especially in the realm of forensic DNA testing which has transformed the landscape of the criminal justice system in recent decades (Brynjolfsson & McAfee, 2018). In the course of a crime, various biological samples, including blood, saliva, semen, and skin cells, can be transferred via interactions between individuals and objects. Advancements in DNA technology have bolstered the precision of DNA analyses. This heightened sensitivity enables forensic experts to identify and analyze degraded or minuscule DNA samples that were previously deemed unfeasible for use. For instance, DNA samples from older crime cases, such as those of sexual assaults or unsolved homicides, are now being assessed in labs. Due to this enhanced sensitivity, even minute DNA quantities are detectable, raising the potential of

identifying DNA from multiple sources, even if they are present in minuscule amounts. This poses a novel set of challenges for forensic labs. One of these challenges, for example, involves the potential detection of DNA from various individuals, possibly unrelated to the crime, when using advanced analysis techniques. This introduces complications in DNA mixture interpretation, necessitating the disentanglement and identification of separate DNA profiles to provide valuable leads for criminal investigations (Marr, 2016).

In this context, AI could offer a solution. DNA analyses yield vast quantities of intricate electronic data. These datasets embody patterns, some of which might be imperceptible to human scrutiny but could become invaluable as analysis techniques grow more sensitive (Rigano, 2019). Syracuse University, in collaboration with the Onondaga County Center for Forensic Sciences and the New York City Office of Chief Medical Examiner's Department of Forensic Biology, is exploring a hybrid of human expertise and AI for DNA mixture deconvolution (Rigano, 2019).

Gunshot Detection

The identification of specific patterns in gunshot sound analysis has highlighted a new potential application for AI algorithms. Under the sponsorship of NIJ, Cadre Research Labs, LLC, embarked on a project to examine gunshot audio recordings taken from smartphones and other advanced devices. This investigation was rooted in the understanding that various factors, such as the

type of firearm and ammunition, the environmental setting, and the nature of the recording equipment, influence the recordings' characteristics and fidelity (Rigano, 2019). Drawing from a comprehensive mathematical framework, the experts at Cadre are formulating algorithms that can pinpoint gunshots, discern muzzle explosions from shockwaves, evaluate the interval between consecutive shots, calculate the total number of guns involved, allocate distinct shots to their respective firearms, and assess the likelihood of the gun's type and caliber. These technological strides hold significant promise in bolstering law enforcement's investigative capabilities (Rigano, 2019).

Crime Forecasting

Predictive analytics is a multifaceted procedure that leverages extensive data to predict and infer potential future outcomes. In the criminal justice sector, this responsibility predominantly falls on police officers, probation officials, and other specialists, who acquire this proficiency over extensive periods. However, this process is labor-intensive and is vulnerable to biases and inaccuracies (Rigano, 2019). Utilizing AI, vast data sets comprising legal precedents, societal data, and media content can be harnessed to propose judicial decisions, pinpoint criminal networks, and anticipate and identify individuals vulnerable to criminal enterprises. Researchers backed by NIJ at the University of Pittsburgh are delving into computational methods for statutory interpretation. These techniques might significantly enhance the pace and precision of

such interpretations undertaken by judges, lawyers, prosecutors, administrative teams, and other professionals. The foundational theory is that a software solution might autonomously identify particular statement types pivotal for statutory interpretation. The overarching aim is to devise a prototype system that aids in interpretation and autonomously conducts it for cybercrime scenarios (Rigano, 2019).

Moreover, AI possesses the capability to scrutinize vast criminal justice records to foresee potential criminal re-offenses. A collaborative effort between the Research Triangle Institute, Durham Police Department, and the Anne Arundel County (Maryland) Sheriff's Office aims to develop an automated tool to prioritize warrant services for the North Carolina Statewide Warrant Repository. With NIJ's backing, the team employs algorithms to dissect datasets encompassing over 340,000 warrant records. These algorithms create decision trees and undertake survival analysis to ascertain the time until a subsequent event and assess the recidivism risk for absconding culprits (if a warrant remains unexecuted). This model aspires to guide practitioners in prioritizing warrant services amid backlogs and will also include geographical references to enable practitioners to target clusters of high-risk absconders and others with active warrants, thereby maximizing resource utilization (Rigano, 2019).

In addition, AI can be instrumental in identifying potential elderly victims of both physical and fiscal abuse. NIJ-sponsored researchers from the University of Texas Health Science Center in

Houston employed AI techniques to study elder victimization patterns. These algorithms can discern the victim, the offender, and situational elements that distinguish financial exploitation from other elder abuse forms. They can also segregate cases of exclusive financial abuse from instances where financial exploitation co-occurs with physical maltreatment or neglect. The ultimate vision is to convert these data-driven algorithms into accessible web apps, empowering practitioners to gauge the probability of financial exploitation and facilitate timely interventions (Rigano, 2019).

Lastly, AI's potential is being harnessed to anticipate potential victims of violent crimes based on relationships and behavioral patterns. A joint initiative between the Chicago Police Department and the Illinois Institute of Technology utilized algorithms to gather data and establish preliminary groupings emphasizing social network construction and analysis to pinpoint high-risk entities. This research, supported by NIJ, has subsequently been integrated into the Chicago Police Department's Violence Reduction Strategy (Rigano, 2019).

Discussion and Conclusion

The rise of AI and its growing accessibility indicates that its role in cybercrime will only intensify. Thus, it is imperative for law enforcement agencies to equip themselves with instruments and skills to detect and analyze AI-generated content like that from ChatGPT. While tools such as GPTZero, Huggin Face GPT2, and Writer AI detector are being utilized to identify AI-crafted

content, it is vital that they are thoroughly tested and validated before large-scale adoption. The global community must stay proactive, ensuring platforms like ChatGPT are utilized positively and criminal activities are monitored.

The following recommendations are proposed to bolster this initiative:

- **Aligning Technology:** It is essential for all stakeholders in the global community, including investigators, forensic experts, and prosecutors, to have the required expertise to adeptly use AI technologies in criminal investigations.
- **Creating Universal Investigative Protocols:** Collaboration with the private sector is crucial for law enforcement to devise standard methods for handling international crimes. A mutual understanding between the two about the challenges and potential solutions is fundamental.
- **Standardized Training Modules:** Considering the growing intrigue around conversational AI and its advanced features, dedicating resources to formulate standardized training and educational content is paramount. Familiarity with these technologies is key to tackling emerging challenges.
- **Future-Oriented Vigilance:** We strive to keep our community updated about the latest breakthroughs in the rapidly advancing AI sector. Encouraging nations to share

intelligence with us will aid in early detection of potential threats or opportunities, especially considering their ramifications for law enforcement.

- **Establishing Defined Regulations:** As international crimes involving platforms like ChatGPT proliferate, it becomes imperative for member states of the global community to engage in dialogues to determine relevant regulatory measures that can be implemented to combat and prevent crime.
- **Ethical Application of AI by Law Enforcement:** While AI holds immense promise for augmenting law enforcement activities, it is crucial to use such technology ethically, given its intricate nature and the need for maintaining public trust. In collaboration with the United Nations Interregional Crime and Justice Research Institute (UNICRI), we have rolled out a Toolkit for Responsible AI Innovation in Law Enforcement as of June 2023. This guide is tailored to aid global law enforcement agencies in responsibly integrating AI tools and platforms into their operations.

Given the perpetual evolution of AI-chatbot systems like ChatGPT and their increasing inclusion in law enforcement practices, we remain committed to tracking these developments. By fostering strong collaborations with law enforcement bodies, industry leaders, and academic experts, we aim to ensure a unified and synergistic approach moving forward.

References

AV-TEST. (2021). Malware Statistics & Trends Report. AV-TEST: The Independent. IT-Security Inst. Accessed: June. 22, 2021. [Online]. Available: https://www.av-test.org/en/statistics/malware/

Bakir, V. (2020). Psychological operations in digital political campaigns: Assessing Cambridge Analytica's psychographic profiling and targeting. *Frontiers in Communication*, *5*, 67.

Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European journal of communication*, *33*(2), 122-139.

Bessi, A., & Ferrara, E. (2016). Social bots distort the 2016 US Presidential election online discussion. *First monday*, *21*(11-7).

Blauth, T. F., Gstrein, O. J., & Zwitter, A. (2022). Artificial intelligence crime: An overview of malicious use and abuse of AI. *IEEE Access*, *10*, 77110-77122.

Boshmaf, Y., Muslukhov, I., Beznosov, K., & Ripeanu, M. (2013). Design and analysis of a social botnet. *Computer Networks*, *57*(2), 556-578.

Brynjolfsson, E., & McAfee, A. N. D. R. E. W. (2017). Artificial intelligence, for real. *Harvard business review*, *1*, 1-31.

Carlini, N., Liu, C., Erlingsson, Ú., Kos, J., & Song, D. (2019). The secret sharer: Evaluating and testing unintended memorization in neural networks. In *28th USENIX Security Symposium (USENIX Security 19)* (pp. 267-284).

Chen, M. X., Lee, B. N., Bansal, G., Cao, Y., Zhang, S., Lu, J., ... & Wu, Y. (2019, July). Gmail smart compose: Real-time assisted writing. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2287-2295).

Chesney, B., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *Calif. L. Rev.*, *107*, 1753.

Ciancaglini, V., Gibson, C., Sancho, D., McCarthy, O., Eira, M., Amann, P., & Klayn, A. (2020). Malicious uses and abuses of artificial intelligence. *Trend Micro Research*.

Durbin. (2020). How criminals use Artificial Intelligence to fuel cyberattacks. Forbes. https://www.forbes.com/sites/forbesbusinesscouncil/2020/10/13/how-criminals-use-artificial-intelligence-to-fuel-cyber-attacks/?sh=4ba35e7f5012

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, *59*(7), 96-104.

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, *30*, 681-694.

Freitas, C., Benevenuto, F., Ghosh, S., & Veloso, A. (2015, August). Reverse engineering socialbot infiltration strategies in twitter. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015* (pp. 25-32).

Garcia, R. T. (2021). Anonymous Twitter accounts in Brazil are pressuring advertisers to drop conservative media campaigns.

Gibert, D., Mateu, C., & Planes, J. (2020). The rise of machine learning for detection and classification of malware: Research developments, trends and challenges. *Journal of Network and Computer Applications*, *153*, 102526.

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*.

Gu, T., Dolan-Gavitt, B., & Garg, S. (2019). BadNets: Identifying Vulnerabilities in the Machine Learning Model Supply Chain.(2017). *arXiv preprint arXiv:1708.06733*.

Guarnera, L., Giudice, O., Nastasi, C., & Battiato, S. (2020, September). Preliminary forensics analysis of deepfake images. In *2020 AEIT international annual conference (AEIT)* (pp. 1-6). IEEE.

Habgood-Coote, J. (2019). Stop talking about fake news!. *Inquiry*, *62*(9-10), 1033-1065.

Hibbard, B. (2014). Ethical artificial intelligence. *arXiv preprint arXiv:1411.1373*.

Hu, H., Salcic, Z., Sun, L., Dobbie, G., Yu, P. S., & Zhang, X. (2022). Membership inference attacks on machine learning: A survey. *ACM Computing Surveys (CSUR)*, *54*(11s), 1-37.

Jagielski, M., Oprea, A., Biggio, B., Liu, C., Nita-Rotaru, C., & Li, B. (2018, May). Manipulating machine learning: Poisoning attacks and countermeasures for regression learning. In *2018 IEEE symposium on security and privacy (SP)* (pp. 19-35). IEEE.

Johnson, D. G., & Verdicchio, M. (2017). Reframing AI discourse. *Minds and Machines*, *27*, 575-590.

Keller, F. B., Schoch, D., Stier, S., & Yang, J. (2020). Political astroturfing on twitter: How to coordinate a disinformation campaign. *Political communication*, *37*(2), 256-280.

Kirat, D., Jang, J., & Stoecklin, M. (2018). Deeplocker–concealing targeted attacks with ai locksmithing. *Blackhat USA*, *1*, 1-29.

Kovic, M., Rauchfleisch, A., Sele, M., & Caspar, C. (2018). Digital astroturfing in politics: Definition, typology, and countermeasures. *Studies in Communication Sciences*, *18*(1), 69-85.

Kurakin, A., Goodfellow, I., & Bengio, S. (2016). Adversarial machine learning at scale. *arXiv preprint arXiv:1611.01236*.

Lai, A. (2021). Artificial Intelligence, LLC: Corporate Personhood as Tort Reform. *Mich. St. L. Rev.*, 597.

Lee, P. (2016). Learning from Tay's introduction. Official Microsoft Blog. *Geraadpleegd op*, *8*.

Leyva, R., & Beckett, C. (2020). Testing and unpacking the effects of digital fake news: on presidential candidate evaluations and voter support. *Ai & Society*, *35*, 969-980.

Lin, T. C. (2017). The New Market Manipulation Emory Law Journal, Vol. 66.

Luo, J., Hong, T., & Fang, S. C. (2018). Benchmarking robustness of load forecasting models under data integrity attacks. *International Journal of Forecasting*, *34*(1), 89-104.

Mahbub, S., Pardede, E., Kayes, A. S. M., & Rahayu, W. (2019). Controlling astroturfing on the internet: a survey on detection techniques and research challenges. *International journal of web and grid services*, *15*(2), 139-158.

Maras, M. H., & Alexandrou, A. (2019). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *The International Journal of Evidence & Proof*, *23*(3), 255-262.

Marr, B. (2016). What is the difference between deep learning, machine learning and AI. *Revista Forbes Recuperado de: https://www. forbes. com/sites/bernardmarr/2016/12/08/what-is-thedifference-between-deep-learning-machine-learning-and-ai*.

Martin, K. (2020). Flash crash—the trading savant who crashed the us stock market. *Financial Times*.

McCarthy. (2023). Criminal AI use provides unique challenges for law enforcement. KPTV. https://www.kptv.com/2023/06/10/criminal-ai-use-provides-unique-challenges-law-enforcement/

Nemitz, P. (2018). Constitutional democracy and technology in the age of artificial intelligence. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2133), 20180089.

Oyeyemi, S. O., Gabarron, E., & Wynn, R. (2014). Ebola, Twitter, and misinformation: a dangerous combination?. *Bmj*, *349*.

Renda, A., Arroyo, J., Fanni, R., Laurer, M., Sipiczki, A., Yeung, T., ... & de Pierrefeu, G. (2021). Study to support an impact assessment of regulatory requirements for artificial intelligence in Europe. *European Commission: Brussels, Belgium*.

Rigano, C. (2019). Using artificial intelligence to address criminal justice needs. *National Institute of Justice Journal*, *280*(1-10), 17.

Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L., Recchia, G., ... & Van Der Linden, S. (2020). Susceptibility to misinformation about COVID-19 around the world. *Royal Society open science*, *7*(10), 201199.

Rossi, S. (2020). Beware the CyberLover that Steals Personal Data.

Schwartz, O. (2019). In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation-IEEE Spectrum. *IEEE Spectrum: Technology, Engineering, and Science News*.

Scopino, G. (2020). *Algo bots and the law: technology, automation, and the regulation of futures and other derivatives*. Cambridge University Press.

Seymour, J., & Tully, P. (2016). Weaponizing data science for social engineering: Automated E2E spear phishing on Twitter. *Black Hat USA*, *37*, 1-39.

Sullivan, M. (2017). It's Time to Retire the Tainted Term" Fake News," WASH. POST (Jan. 8, 2017).

Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., & Nießner, M. (2016). Face2face: Real-time face capture and reenactment of rgb videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2387-2395).

Thuen, C. (2017). Discovering truth through lies on the Internet—FCC comments analyzed.

Trieu, K., & Yang, Y. (2018). Artificial intelligence-based password brute force attacks.

Webster, R., Rabin, J., Simon, L., & Jurie, F. (2021). This person (probably) exists. identity membership attacks against gan generated faces. *arXiv preprint arXiv:2107.06018*.

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology innovation management review*, *9*(11).Wilder, B., & Vorobeychik, Y. (2019, July). Defending elections against malicious spread of misinformation. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 2213-2220).

Wiener-Bronner, D. (Feb. 5, 2018). How the Dow Fell 800 Points in 10 Minutes. CNN Money. Accessed: June. 24, 2021. [Online]. Available: https://money.cnn.com/2018/02/05/news/companies/dow-800-points- 10-minutes/index.html

Yadav, Y. (2016). The failure of liability in modern markets. *Virginia Law Review*, 1031-1100.

Zetter, K. (2010). Wiseguys Plead Guilty in Ticketmaster Captcha Case. *Wired. url: https://www. wired. com/2010/11/wiseguysplead-guilty*.

Zwitter, A. (2016). The impact of big data of international affairs. *Clingendael Spectator*.

Zwitter, A. (July. 27, 2017). The Artificial Intelligence Arms Race. Policy Forum. Accessed: Apr. 12, 2021. [Online]. Available: https://www. policyforum.net/arti?cial-intelligence-arms-race/