Novel Cybersecurity Challenges Within Artificial Intelligence

Anthony Caldwell
Letterkenny, Co.Donegal, Ireland

Abstract

Beyond the entertainment industry's dystopian rendering of AI, industry needs to understand how it might best practically integrate AI over the next few years. From the speed and accuracy of problem resolution to the discovery of new ideas, AI has been heralded as a panacea for many issues within many fields, but is it? There are certainly jobs that AI is best placed for, but others we are not ready for yet. Cybersecurity is a particular focus of this paper and in particular the challenges faced in the future. AI is regarded as a combination of studies including machine learning and data mining. Beyond technologies, the field also considers the ability of AI to perform tasks that would otherwise require human intelligence. Their integration into cybersecurity frameworks has opened new frontiers in digital defense while simultaneously presenting novel challenges. As AI technology evolves, so do the techniques employed by cybercriminals, resulting in a constant cat-and-mouse game between security experts and malicious actors. There is a lack of transparency in AI algorithms that raises concerns about the reliability and accountability of AI-driven security measures which begs the question, do we trust AI? For if AI algorithms make incorrect decisions or fall victim to adversarial attacks, the consequences could be severe. Ethical considerations also loom large in the realm of AI-powered cybersecurity. Questions about bias in AI algorithms, which can result in discriminatory outcomes, raise concerns. Ensuring that AI systems are fair and unbiased while maintaining their effectiveness is a delicate balance that the cybersecurity community must navigate. Insight As A Service (IAAS) may become a future offering beyond the current analytical capabilities of predictive analytics.

1. Introduction

As noted in a 2022 Forbes article, there are difficulties with defining what AI as a field is and while the promises (in the future) are impressive, for the moment, these aren't actionable [1]. However, if elements of AI are actionable, there are challenges. A statement from Chatham House's report on AI and its challenges suggests, "Probably the greatest challenge facing the AI industry is the need to reconcile AI's need for large amounts of structured or standardized data with the human right to privacy." [2]. This is most evident in the area of IoT where research from 2018 and 2020 suggests that AI integration is feasible for the security of IoT devices [3], [4]. Security concerns with this integration include issues with device authentication, DoS / DDoS attack defense, intrusion detection and malware detection [3], [4]. Moreover, in recent decades we have become comfortable having AI-like applications in areas like online shopping, advertising, navigation, language translation, smart home environmental regulation, health, transport, manufacturing and our personal security. There are a diminishingly few streets in the world where Google maps hasn't already explored for us, indeed the proliferation of security cameras in towns and cities go largely unnoticed. Perhaps the genesis of this fear lies within the novel Nineteen Eighty-Four, where George Orwell paints the picture of an oppressive, highly controlled society where such technology was also commonplace. Indeed, the entertainment industry's obsession with the fear of AI is centered around artificial general intelligence, or AGI, that is an AI which could meet and exceed the full capabilities of the human mind in the future leading to our demise. The integration of AI into many systems is most likely but what do we want these AI driven systems and machines to do? What do we prefer to do ourselves?

2. Literature Review

AI systems are autonomous, self-learning agents interacting with the environment. As AI applications grow in complexity and pervasiveness, new challenges in cybersecurity have emerged, demanding innovative solutions to safeguard sensitive data and critical systems. Traditional cybersecurity measures are struggling to keep pace with the evolving threat landscape introduced by AI. Cybercriminals are exploiting AI algorithms to create sophisticated attacks, making detection and mitigation increasingly difficult [2], [3], [4]. Adversarial attacks on AI systems involve manipulating input data to deceive the algorithms, leading to erroneous outcomes. Researchers have explored novel techniques to identify and counter these attacks, including robust AI model training and generative adversarial networks (GANs) for creating adversarial examples. Explainability remains a challenge in AI, particularly in cybersecurity applications where decisions have significant consequences. In this respect, interpretable AI models are crucial for understanding AI-driven security solutions, ensuring trust among users and decision-

Copyright © 2023, Infonomics Society | DOI: 10.20533/jitst.2046.3723.2023.0098

makers [5]. Regarding privacy concerns in AI, AI applications often require vast amounts of data, raising concerns about user privacy. Institutionalised learning and encryption are promising techniques that enable AI training without compromising individual privacy, addressing one of the core challenges in AI-driven cybersecurity[10], [11]. Regarding privacy concerns in AI, AI applications often require vast amounts of data, raising concerns about user privacy. Important ethical questions, especially concerning autonomous decision-making have been raised also where balancing the benefits of AI-driven threat detection with accountability and transparency is a crucial area of research [12]. But what do we mean by artificial intelligence? Moving beyond the entertainment industry's interpretation or indeed the general public, there are strict definitions as regards what might be classified as AI.

3. What Do We Mean by Artificial Intelligence?

AI systems are autonomous, self-learning agents interacting with the environment. Their robustness critically depends upon the inputs and interactions with other agents they ingest once deployed as well as their design and training [5]. As a field of study, AI is regarded as a combination of studies including machine learning and data mining. The field also considers the ability of AI to perform tasks that would otherwise require human intelligence [2] and has led to advances in medicine, astrophysics, education, mathematics and criminal prosecution [6]. The AI project, if we can call it that for a moment, is an ideological one in which we develop practices, processes, procedures that allow us to be less nervous in the face of this next revolution in technology. On this basis, we might summarize the main capabilities and limitations of AI as it is presented to date.

4. Capabilities and Limitations of AI

Advantages of artificial intelligence lie in the algorithms which are used to operate it. It is thought that these algorithms are based upon facts rather than emotions and overall have potential to add value to problem resolution from the perspectives of discovery of new ideas, calculation of long term and complex situations, speed and accuracy of tasks complete [7]. As noted by Katari (2022), "…AI is good at recognizing patterns, and humans are good at understanding when those patterns have meaning" [1]. Disadvantages include detrimental impacts upon jobs requiring human interaction similar to the advent of steam power versus horse power at the beginning of the industrial revolution, but this is unlikely to have an impact in the near future. Importantly, from a return-on-investment perspective, the time, resources and money needed to maintain the AI are prohibitive to most companies [7]. There is also a worrying lack of transparency of the algorithms the AI systems use, leading to a general mistrust of outputs of AI based upon the biases that may be present [3]. Table one below outlines some of the popular AI chat models with their associated organization, size and notable uses. Research into mitigating biases, ensuring transparency and accountability, establishing clear ethical guidelines, and involving diverse voices in the development and deployment of these technologies to minimize their negative impacts on society is ongoing. The following table shows the difference in large language model size today and importantly how the main companies that can get involved have dominance over a great many aspects of social media.

Table 1. Table highlighting a few key language models

Model	Organization	Size	Notable Uses
GPT-3	OpenAI	175 billion	Natural language generation, translation, Q&A
BERT	Google	340 million	Question answering, language understanding
T5 (Text-to-Text)	Google	11 billion	Various NLP tasks like translation, summarization
XLNet	Google	154 million	Text generation, language modeling
RoBERTa	Facebook AI	125 million	Text classification, language understanding
ELECTRA	Google	110 million	Pretraining for downstream tasks

5. AI Powered Tools

Unfortunately, cybercrime is a prevalent market for malicious actors, and the costs of risk mitigation and defense are increasing yearly [8]. The intersection of artificial intelligence (AI) and cybersecurity has given rise to powerful tools and techniques that both defenders and attackers employ. Among these, Generative Adversarial Networks (GANs) have emerged as a groundbreaking concept in machine learning, enabling the generation of synthetic data. Additionally, AI-powered exploitation tools and natural language processing (NLP) tools are being increasingly utilized in cybersecurity operations. Generative Adversarial Networks (GANs) consist of two neural networks, a generator, and a discriminator, that are trained simultaneously through adversarial training. The generator creates synthetic data, while the discriminator evaluates its authenticity. In cybersecurity, GANs find

Copyright © 2023, Infonomics Society | DOI: 10.20533/jitst.2046.3723.2023.0098

applications in generating realistic-looking phishing websites, malware variants, and even fake biometric data for bypassing authentication systems. By creating convincing fake data, attackers can exploit vulnerabilities and breach security measures. GANs are also utilized for data augmentation in cybersecurity research, helping analysts understand attack patterns and enhance intrusion detection systems. Furthermore, they aid in generating realistic datasets for training security models, ensuring robustness against diverse cyber threats. AI-powered exploitation tools leverage machine learning algorithms to automate the process of identifying vulnerabilities and launching attacks. These tools use AI techniques to scan networks, identify weak points, and exploit them at scale. For instance, AI-powered penetration testing tools can simulate sophisticated attack scenarios, enabling organizations to identify and remediate vulnerabilities before malicious actors can exploit them. Additionally, AI-driven threat intelligence platforms analyze vast amounts of data to detect patterns and predict cyber threats. By processing large datasets, these tools can identify emerging threats and vulnerabilities, allowing organizations to proactively bolster their defenses. Natural Language Processing (NLP) tools in cybersecurity are instrumental in analyzing and understanding human language data, including emails, social media posts, and chat conversations. Sentiment analysis, entity recognition, and topic modeling are some NLP techniques used to detect phishing attempts, social engineering attacks, and insider threats. By automatically analyzing textual content, NLP tools help security professionals identify suspicious activities and respond swiftly. Moreover, NLP algorithms are employed in threat hunting and incident response. Security teams use NLP to parse and comprehend large volumes of security logs and reports, facilitating the identification of potential security incidents. By extracting relevant information from unstructured data, NLP tools enhance the speed and accuracy of incident investigations. One significant trend in the future of cybersecurity is the integration of AI-driven tools into Security Operations Centers (SOCs). These AI systems can autonomously detect and respond to threats, mitigating the need for human intervention in routine security tasks. While this can enhance efficiency, it also raises concerns about the potential for AI systems to make errors or be manipulated by attackers. Machine learning models within AI-driven SOCs continuously learn from new data and evolving threat landscapes. This adaptability is crucial in dealing with emerging and sophisticated cyber threats that may exhibit novel characteristics. As these systems become more adept at recognizing anomalies and potential security breaches, they contribute to a proactive defense posture, anticipating and thwarting threats before they can cause significant harm. Striking the right balance between automation and human oversight will be crucial in ensuring the effectiveness and reliability of AI-driven cybersecurity solutions.

6. AI in Critical Environments

AI in sensitive environments, such as critical infrastructure, military operations, and hazardous industrial settings, introduces a unique set of cybersecurity challenges. While AI technologies offer the potential to enhance safety, efficiency, and decision-making in these contexts, their integration also amplifies the consequences of security breaches. In dangerous environments, where the stakes are high, adversarial attacks can have severe consequences. An attacker could exploit vulnerabilities in AI algorithms to manipulate critical systems, compromise safety measures, or disrupt operations[2], [5]. Tampering with data inputs can lead to incorrect decisions by AI systems, potentially causing accidents or failures in critical infrastructure [2], [3], [5]. For example, new smart grid reliability is based on the integrity, confidence, and availability of control of communication systems. A source of vulnerability resulting from integrating ICTs through the public network is the known weaknesses of the internet protocol (IP), potentially leading to the risk of intrusions or interceptions of data. Malicious users could implement a man-in-the -middle attack where consumption patterns could be analysed to determine how much energy is used thus establishing if the consumer is present to execute attacks. In addition, criminals could use information from these patterns to harm specific consumers. AI technologies are already integrated into tools, such as malware and virus detection, endpoint detection and response, firewalls, data loss prevention and can automatically respond to attacks. Edge devices (e.g. sensors, cameras and controllers) are potential targets for cyber-attacks. Strengthening the security of these devices is crucial to prevent unauthorized access and tampering [5]. Typically, the edge networking devices in critical infrastructure face Denial of Service (DoS) attacks, service or resource manipulation, privacy data loss, and man-in-the-middle attacks. In the distributed infrastructure, where the user equipment is capable of performing all computations, one can argue that the security threat is negligible. Also, an adversary and they start to inject bogus information or manipulate services. Vulnerabilities in the supply chain, if exploited, can have cascading effects on the cybersecurity of AI systems deployed in dangerous environments [2], [5]. There is an increasing trend in using AI and machine learning in static and dynamic malware analysis to aid in and anomaly detection [3]. While helpful in discriminating malicious activities in the network traffic AI enhanced deep learning methods will be required to analyze AI-driven behavior over a series of actions. Cyber criminals are using AI to spread the threats faster and for the moment, AI solutions are less effective at deceiving people than humans.

7. Cybersecurity

"Darktrace found that the average linguistic complexity of phishing emails has risen by 17 percent since ChatGPT's November 2022 launch." [8], [9]. While the increase in linguistic complexity is a concern, it cannot be correlated with ChatGPT in any clear sense. With that in mind, the vulnerabilities of AI pose serious limitations to its great potential to improve cybersecurity in particular the lack of transparency of AI systems. Early research in IoT suggested that AI methods used in security led to potential threats such as poisoning, evasion, impersonation and reverse attacks [3]. In these cases, poisoning, evasion and impersonation are aimed at changing the training data used by the AI maliciously modifying samples, so that the model learns wrong and invalid knowledge. As a result, the ability to distinguish standard, normal data from the abnormal, led to the failure of the ability of the model to detect corrupted models. Using an application program interface (API), existing machine learning (ML) systems can collect basic information about the target model, engage in reverse analysis of this to obtain private data. These attack methods combined with IoT services suggest serious consequences [4] and have caused technologists to pause their enthusiasm for AI until we can be sure that these concerns are mitigated. Central to this hesitation is the learning strategies of AI which have implications for biases, ethical failings, and as indicated above and may lead to privacy breaches [2]. Adversarial AI or adversarial machine learning methods harness more sophisticated AI techniques to attack AI systems. Several classes of adversarial AI have been identified, including adversarial examples, the use of basic policies or more sophisticated machine learning methods to fool AI systems with inputs that cause the systems to fail to function properly. But what can be said of the power of AI tools in the fight against such cyber threats? AI can analyze vast amounts of data and identify patterns and anomalies that may indicate a potential cyber-attack. Machine learning algorithms, a subset of AI, can also learn from past incidents to identify new threats and predict future attacks. Additionally, AI-powered cybersecurity solutions can automate threat detection and response, allowing security professionals to focus on more critical tasks. One of the primary benefits of AI in cybersecurity is its ability to improve threat detection. Traditional security systems rely on rule-based detection mechanisms, which are often insufficient to detect sophisticated attacks. AI, on the other hand, can analyze vast amounts of data from multiple sources, including network traffic, logs, and user behavior, to identify potential threats. AI-powered systems can also continuously monitor the network and detect anomalies that may indicate a potential attack. AI also has the capacity to transform incident response by automating incident response processes thus enabling security professionals to respond quickly and effectively to cyber threats.

In terms of an organization's security posture, AI-powered cybersecurity solutions can analyze the organization's security infrastructure and identify areas that need improvement. This can include identifying vulnerabilities in the network, recommending security policy changes and providing guidance on how to improve security awareness among employees. A central aspect of this is the attention to regulatory guidance where many industries, such as healthcare and finance, are subject to strict data protection regulations. Helping organizations ensure compliance with these regulations by continuously monitoring the network for potential breaches and providing real-time alerts when necessary is a strength area of AI powered machines. But should we trust AI-powered systems to be integrated in the first place?

8. In AI We Trust?

What do we want these AI driven machines to do? What do we prefer to do ourselves? Returning to the ideological hurdles we must surpass, trust is a key element of the future of AI when we look ahead, since either actor in any AI-Human system may be corruptible. In order to foster adoption of any new technology, trust is essential and with this comes the defining and developing of standards and certification procedures with the goal of developing trustworthy AI in cybersecurity. Trust, defined as the decision to delegate a task, without any form of control or supervision over the way the task is executed, may be an oxymoron in cybersecurity. For trust to be successful an appropriate assessment of the trustworthiness of the agent to which the task is delegated (the trustee) is needed. Both as a prediction about the probability that the trustee will behave as expected, given the trustee's past behaviour, and as a measure of the risk run by the trustor, should the trustee behave differently trust is a probability assessment. If the probability that the expected behaviour will occur is too low, then the risk is too high and trust is unjustifiable. With AI, the lack of transparency and the learning abilities of AI systems, combined with the nature and scale of the attacks possible to these systems, increase the difficulty in evaluating if the AI system will continue to behave as expected in any given context. The current records of past behaviour of AI systems don't predict the systems' strength to resist future attacks, nor is this an indication that the system has not been corrupted by a stealth, APT-styled attack. All of which impairs the assessment of trustworthiness and perhaps at this point, trust in AI applications for cybersecurity is unwarranted [5]. This leads to the ethics and regulation of a powerful new tool that potentially has dangerous capabilities.

9. Ethics, Bias, Regulations?

Multifaceted ethical challenges and legal complexities surrounding AI cybersecurity highlight the importance of responsible AI development and implementation. As a starting point we can take privacy and data protection. The collection and storage of user data raises significant ethical concerns regarding privacy and data protection. When data breaches occur, the sensitive information of individuals can be exposed, leading to potential identity theft, financial losses, and reputational damage. Protecting user privacy should be a top priority in AI cybersecurity [2], [10]. AI systems are not inherently unbiased, as they learn from historical data that may contain biases and prejudices, leading to discriminatory outcomes, particularly in areas such as hiring, criminal justice, and loan approvals [11]. The use of biased algorithms perpetuates societal inequalities and violates the principles of fairness and justice. Addressing bias in AI cybersecurity requires careful design, diverse datasets, and continuous monitoring. Coupled with this is the lack of transparency in the algorithm design and implementation, making it difficult to isolate responsibility in case of a cybersecurity breach. Ethical AI would require explainable AI models that can be audited and understood to ensure transparency and accountability. Importantly, there is a degree of legal lag in AI that creates uncertainty regarding the obligations and liabilities of AI developers and users. Governments and international organizations must collaborate to establish clear guidelines for AI cybersecurity practices to safeguard both individuals and organizations. To protect the public, the U.S. has long enacted regulatory instruments, such as rules against discrimination, equal employment opportunity, HIPAA Title II, Commercial Facial Recognition Privacy Act, and Algorithmic Accountability Act. All these instruments would be useful in guiding the development of legal and regulatory policies and frameworks for AI ethics [12] and in this respect, the significance of cross-border legal challenges raises issues. Given that cybersecurity threats know no borders, makes it challenging to enforce laws and pursue legal action against cybercriminals operating from jurisdictions with lenient regulations. Coordinating international efforts to combat cyber threats while respecting national sovereignty is essential to ensure effective cybercrime prosecution and deterrence [13]. Determining liability and responsibility in AI-related cybersecurity incidents can be complex, particularly when AI systems make autonomous decisions. As AI adoption increases, legal frameworks must clarify the roles and responsibilities of AI system creators, owners, and operators, ensuring that accountability is appropriately assigned in case of security breaches.

10. Conclusion

The novel challenges in cybersecurity within the context of AI stem from the intricate nature of AI systems, the security of training data, the need for real-time threat detection, ethical considerations, and the necessity for international collaboration. Addressing these challenges is essential to harness the full potential of AI in enhancing digital security and safeguarding the ever-expanding digital landscape. AI will potentially enable new capabilities to address numerous needs such as providing not just information but deep intelligence. In this respect, Insight As A Service (IAAS) may become a future offering beyond the current analytical capabilities of predictive analytics. Transparency is the key to gaining trust from all stakeholders regarding AI integration. This trust must be built through explanatory power, which is a challenge with AI. The deepfake task force act [10] is one of the many legislative methods to prosecute the malicious use of powerful AI technologies that have strong potential to misinform the public at a global scale. Many corporate products are defined simultaneously by their security in data transfer and the efficiency with which this is done. But rather than had over control to an AI we still need experts to review the work carried out. To help with this, it is essential that governments provide legal frameworks for the private sector. But this is not yet in place. Protecting user privacy, addressing bias, ensuring accountability, and establishing clear legal frameworks are essential steps to create a more secure and ethically responsible AI ecosystem. Striking a delicate balance between innovation and regulation will be key to leveraging AI's potential while mitigating its risks. AI is transforming the field of cybersecurity by providing advanced analytical capabilities, proactive threat detection, and automated response mechanisms. The application of AI in cybersecurity has the potential to enhance the effectiveness and efficiency of cybersecurity operations, improve threat detection, automate incident response, improve overall security posture, and ensure compliance with regulatory requirements. As cyber threats continue to evolve, AI-powered cybersecurity solutions will become increasingly critical for organizations looking to protect their valuable assets and sensitive data.

11. References

[1] Katari, G. (2022). AI Trends For 2023: Industry Experts (And ChatGPT AI) Make Their Predictions.

Forbes. https://www.forbes.com/sites/ganeskesari/20 22/12/22/ai-trends-for-2023-industry-experts-and-chatopai -make-their-predictions/. (Access date: 22 December 2023).

[2] Chatham House. (2022). Challenges of AI. https://www.chathamhouse.org/2022/03/challenges-ai. (Access Date: 22 December 2023).

[3] Liu, Q. L. (2018). A survey on security threats and defensive techniques of machine learning: A data driven view. IEEE Access, 6, 12103-12117. DOI:10.1109/ACCESS.2018.2805680.

[4] Wu, H. H. (2020). Research on Artificial Intelligence Enhancing Internet of Things Security: A Survey. IEEE Access, 8, 153826-158848. DOI: 10.1109/ACCESS.2020.3018170.

[5] Taddeo, M. M. (2019). Trusting artificial intelligence in cybersecurity is a double-edged sword. Nature Machine Intelligence, 1, 557-560.

[6] Baidoo-Anu, D. O. (2023). Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4337484. (Access Date: 2 Februa -ry 2023).

[7] Khanzode, K. C. (2020). Advantages and Disadvantages of Artificial Intelligence and Machine Learning: A Literature Review. International Journal of Library and Information Science (IJLIS). Volume 9, Issue 1, January-April 2020, pp. 30-36, DOI: 10.17605/OSF.IO/GV5T4.

[8] Darktrace. (2023). Generative AI: Impact on Email Cyber-Attacks. https://darktrace.com/resources/generative-ai-impact-on-email-cyber-attacks. (Access Date: 12 March 2023).

[9] Hayes, J. (2023). AI defenders ready to foil AI-armed attackers. The Register. https://www.theregister.com/2023/04/20/ai_defenders_ready_to_foil/. (Access Date: 20 March 2023).

[10] C2PA. (2023). Coalition for Content Provenance and Authenticity (C2PA). (C2PA, Editor) Retrieved from Coalition for Content Provenance and Authenticity (C2PA): https://c2pa.org/. (Access Date: 26 January 2024).

[11] Akter, S., McCarthy, G., Sajib, S., Michael, K., Divedi, Y. K., D'Ambra, J., and Shen, K. N. (2021). Algorithmic bias in data-driven innovation in the age of AI. International Journal of Information Management, 60. DOI: 10.1016/j.ijinfomgt.2021.102387.

[12] Siau, K., and Wang, W. (2020). Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI. Journal of Database Management, 31(2), 74-78. DOI: 10.4018/JDM.2020040105.

[13] Messerschmidt, J. (2013). Hackback: Permitting Retaliatory Hacking by Non-State Actors as Proportionate Countermeasures to Transboundary Cyberharm. Columbia Journal of Transnational Law. http://ssrn.com/abstract=2309518. (Access Date: 1 August 2023).