The Potential Emerging Threat of Artificial Intelligence

Eric Mung'aũ Preston

Old Dominion University

CYSE 526: Cyber War

Dr. Alex Korb

November 24, 2024

Abstract

Significant technological shifts have occurred in the past few years, changing daily life. Since the COVID-19 pandemic, arguably the most substantial change was the shift to emphasize online work and operations within academia and many other industries due to people being in lockdown. However, the most recent advancement has come from publicly available artificial intelligence (AI). AI chatbots such as ChatGPT and Gemini have quickly risen in popularity due to the number of questions that users can comprehensively answer and are substantially convenient. However, from a cybersecurity perspective, AI can be a significant threat in many ways. Results show that the local level has to manage polymorphic malware and human-like phishing attacks. The national level manages against adversarial AI and integration into the critical sectors of nations. Finally, the global level has to handle the legality of AI and the risk of perpetuating issues such as the ethics of privacy and bias. This paper will examine AI's potential security risks at the local, national, and global levels. Furthermore, due to the infancy of AI's commercialized use and the rapid evolution of cybersecurity within the law, more research should be conducted by scholars in the future on this topic.

Introduction

Artificial Intelligence (AI) has become a significant topic of discussion within our digitized society due to its quickly rising integration and popularity. This growing popularity can be seen through constant discussion within academia and many news outlets. Integrating AI into the devices we use daily is also prominent, as statistics from the National University and AIPRM show that 77% of devices currently utilize AI in some way. Two additional surveys concluded that 55% of Americans were willing to use AI in their daily lives and that from a sample of 11,000 Americans, 67% utilized ChatGPT more than Google. Furthermore, AI has become an

economic powerhouse, with the current global market for AI totaling $454.12 billion and the prediction for the worldwide economy from AI's contribution alone by 2030 being $15.7 trillion. (AIPRM, 2024; National University, 2024).

Background

These statistics show some of AI's current popularity and success gained. However, because AI is a broad term, there can be confusion when referring to different instances of AI's utility has been lumped into the same thing and described the same way. According to IBM, artificial intelligence is "technology that enables computers and machines to simulate human learning, comprehension, problem-solving, decision making, creativity and autonomy." Some of the subcategories that are under the umbrella term of AI include "machine learning (ML)," when a computer uses a large amount of data to learn about a topic, "deep learning (DL)" which is a multi-layered form of machine learning that mimics the human brain, and most recently, "generative AI" that uses deep learning models to generate original content (Stryker & Kavlakoglu, 2024).

AI may be a new phenomenon for some users due to recent popularity in the past few years. However, AI has a rich history that spans even before the 1900s. From 1914-1950, primary machines for chess, robotics, and computer science were developed, and the actual birth of AI began with Alan Turing's "Computing Machinery and Intelligence" and the Turing Test, laying the foundation for determining a machine's capability to mimic human intelligence in 1950. 1951-1973 housed more significant advancements, with the term "artificial intelligence" coined by John McCarthy in 1955 and the developing of the first neural networks and backpropagation for machine learning. There was also the creation of DENDRAL and ELIZA in 1965, the first expert system that mimic the human decision-making process, and the first

chatbot. However, due to the amount of time for research to develop and some failed expectations in this period, it led to an "AI Winter" from 1974-1980, in which there was reduced funding for research (Mucci, 2024).

From 1980 to 1998, Japan had innovations in improving humanoid robots and the creation of the Fifth Generation Computer Systems Project (FGCS), which was the start of 5G devices that are used today. Multiple studies on subjects such as the backpropagation algorithm, early chatbot designs such as A.L.I.C.E., and further developments of neural networks have been conducted (Mucci, 2024). However, according to Aggarwal (2023), there was a second AI Winter from 1987-1993 due to the field of expert systems needing expansion at the time. However, near the end of this period, IBM's Deep Blue expert system, developed in 1997, would beat the at-the-time world chess champion Garry Kasparov.

From 2000 to 2024, some of the current technologies were developed, such as the start of deep learning, the launch of Siri in 2011, and OpenAI's launch of GPT-3 in 2020 as one of the first large language models. Companies such as Google, Tesla, OpenAI, IBM, and NotebookLM then produced models like Multitask Unified Model, Full Self Driving, ChatGPT, Granite, and DeepDive (Stryker & Kavlakoglu, 2024; Mucci, 2024).

While AI has made significant strides within the technology community as individual technologies, there has been more excellent capability within cybersecurity. First, a traditional form of cybersecurity is "a collection of technologies, procedures, and practices designed to protect networks, computers, programs, and data from attack, disruption, or unauthorized access" (Sarker et al., 2021). According to an article discussing a new phenomenon called "AI in cybersecurity," the phrase is described as "applying AI technologies such as machine learning, deep learning, and data analytics to protect digital systems and networks from cyber threats."

With capabilities like machine learning algorithms to find potential threats, utilizing data analytics to sift through large amounts of data to find vulnerabilities and different activity, and pattern recognition to differentiate normal and abnormal traffic (Luna, 2024).

However, while AI and cybersecurity have promising capabilities, there is also potential for malicious actors to take advantage of their inherent vulnerabilities. The initial problem is false positives and negatives, where a system will either send an alert to a minimal or non-issue or not acknowledge a threat as one, adding unpredictability (Luna, 2024). Another issue to stem from this integration has been the start of AI-driven cyberattacks and the advent of "AI Weaponisation." According to an article on the concept of AI Weaponisation, the phrase refers to "the use of AI technology to launch cyber attacks" and that AI technology is utilized by being involved with the cyber-attack process, solely trusting the technology's output, and poisoning the model with malicious or false information. These types of attacks include using WormGPT, a version of ChatGPT without ethical restrictions, to generate phishing emails, generating scripts for malware, and using Deepfake technology to manipulate people into giving up information (Weaponisation of AI: The New Frontier in Cybersecurity, 2024). With this small handful of threats that can somewhat look at local threats, there is more significant curiosity in how AI technology integrated with cybersecurity affects other geographic levels of the world, such as other nations and the world.

As AI and cybersecurity have vulnerabilities simply from their integration, this prompts one research question: "How are the developing capabilities of artificial intelligence at the local, national, and global levels presenting potential cybersecurity risks?"

Artificial Intelligence as a technology is a possible and significant emerging cybersecurity danger due to its ability to manipulate individuals and businesses and damage

critical infrastructure. If not properly evaluated, the policies for AI will continue these dangers, and adequate time and collaboration between international organizations is required to prevent it.

Methodology

The methodology for this conceptual research paper will consist of a literature review. It will gather existing evidence from multiple resources to provide insight into increasing geographic scales and the potential threat AI presents. The review will also note conflicting information and research gaps that may inform future research.

Literature Review

The Local Level

This literature review focuses on how AI impacts systems in sectors and communities. Locally, this risk affects individuals, businesses, and local governments. As stated previously, phishing attacks, generated malware scripts, and deepfake technology are utilized to manipulate people into giving up data or compromising information. While these attacks are severe, AI can further their damage potential. According to Arif et al. (2024), additional attacks such as automated virus deployment and malicious use of natural language processing (NLP), the ability for a computer to understand and communicate human language, are also prevalent among the average person (Holdsworth & Stryker, 2024). For virus and malware deployment, AI not only significantly speeds up the creation of malware but also provides the protection of a process known as polymorphism, where, in this instance, malware can change around protective mechanisms. Malicious use of NLP coupled with algorithms to gather significant amounts of data can lead to significantly improving the quality and execution of phishing attacks.

Businesses have become more reliant on the utilization of AI in their daily operations. In particular, in the manufacturing industry, AI has become increasingly integrated into developing and manufacturing products to packaging and shipping them to customers (Bécue et al., 2021). These integrations have helped expand and optimize systems, leading to efficiency and high performance. According to Bécue et al. (2021), the manufacturing sector significantly benefited from the integration of AI. For example, AI allows quality control, including monitoring products to minimize defective products reaching consumers. Another task includes detecting and isolating products that may cause harm, such as contaminated food or chemicals. By monitoring and isolating these products, standards are maintained, and unnecessary harm is avoided. Additionally, human error is decreased, and time is saved. Other benefits of AI in manufacturing are predictive maintenance, including decreasing incidents of non-operational machines and inventory monitoring (Shahrukh et al., 2024). For example, the company Lowe introduced the LoweBot in 2016, which meets customer service tasks and detects misplaced and out-of-stock inventory. Another benefit is the management of the supply chain. Most manufacturing industries supply products globally. AI plays an integral role in managing this task and mitigating potential disruptions that may occur. AI is utilized to detect possible supply shortages or delays that may be caused by external events such as political unrest or weather. Altogether, it is clear how critical AI has become in manufacturing (Bécue et al., 2021).

However, while AI has enhanced the overall functioning of business, it has also created potential security threats that could have significant implications that include disruptions of operations, systems and could cripple businesses. More specific examples come from malicious actors using deepfake technology to impose as important figures and utilizing finely tuned denial of service (DoS) or distributed denial of service (DDoS) attacks to halt operations completely.

With DoS and DDoS attacks, ML algorithms can be used to find the perfect time and method to attack the inside of a business, which makes it much more difficult for a business to retaliate. According to Arif et al. (2024), there was a case in 2019 in which deepfake technology was used to impersonate the CEO of a company, and it was convincing enough for an employee to send € 220,000 into a fake account.

Recommended strategies that may limit these threats include integrating AI into cybersecurity strategies to mitigate risks and threats (Bécue et al., 2021; Kasaraneni, 2019). These integrations of AI include AI monitoring, machine learning algorithms, analyzing data, detecting anomalies and responding to them in real time, and creating multi-layered defense strategies by complementing existing measures that include firewalls, intrusion detection systems, and inscription protocols. The goal is for manufacturing businesses to become proactive in threat detection and continuously update and upgrade these AI tools to get ahead of emerging threats and vulnerabilities (Kasaraneni, 2019). Engaging in research to increase knowledge related to how AI continuously influences human factors adds to the risks and vulnerabilities of most businesses (Kasaraneni, 2019).

Studies have shown that multiple local governments have deemed the integration of AI beneficial in their operations. For example., the United States has used AI in traffic management and measuring government performance (Distor et al., 2021). A study on administrators in a local government in the Philippines sought to explore their perceived acceptance and adoption of AI in the public sector. The study revealed that participants found previous use increased confidence in AI's use and perceived usefulness. While AI was helpful, there were concerns regarding how much power AI would have, anxiety about privacy, and overall concerns about self-efficacy in mastering AI. They also expressed concerns about older generations needing help

using AI as effectively as younger generations and the impact of implementation depends on whether leaders embraced AI. There was concern regarding the cost of implementation, which is a significant issue in developing and under-resourced countries. These findings could be applied to any local government in the world. Some of the risks and threats noted include risks that present significant potential for intrusion of citizens' privacy. Additionally, some risks of recent AI technology exploitation were the use of massive surveillance and deepfakes to spread fake news (Distor et al., 2021). These risks present significant potential to a local government by intruding on the privacy of citizens as well as spreading large amounts of disinformation, which is information that is intended to be deceitful (American Psychological Association, 2024).

The local level of AI-associated cybersecurity risks shows significant immediate prevalence. Individuals face attacks such as automated and polymorphic malware, phishing emails utilizing NLP, and algorithms maliciously used to gather people's data to make them susceptible. Businesses need to manage against upgraded DoS and DDoS attacks and deepfakes that can manipulate employees. Local governments also need to do their best to mitigate disinformation campaigns and stop malicious actors from being able to survey their general population.

The National Level

Next, at the national level, are risks that threaten the governments of nations or other federal organizations. A significant cybersecurity risk that has sprouted from AI on the national level is Adversarial AI. Adversarial AI stems from ML, which "involves malicious actors deliberately attempting to subvert the functionality of AI systems" (What Is Adversarial AI in Machine Learning?, 2015). According to Fordham et al. (2023), a prominent vulnerability within the governance sector is a lack of standardization or framework for governance. Because there is

no established standard for adversarial AI, the risk of ML models being tampered with little to counteract it is still prominent. Furthermore, some of the dangers in the ML life cycle were data poisoning during the training phase and model privacy and evasion attacks in the inference stage. When a malicious actor puts incorrect or malicious data into an AI model, data poisoning can be severely harmful, such as tampering with facial recognition in potential federal court cases. Evasion attacks are another iteration of this attack, where malicious data is input into the completed model to trick it.

Another sector that is worth looking into is the integration of AI into the healthcare system. Policymakers around the world have determined and acted upon the idea of integrating AI into healthcare systems to increase access and quality of care for its citizens while at the same time mitigating risks. According to (Castonguay et al., 2024), most countries are in the emerging integration and implementation phase. Developed countries such as the United States and the United Kingdom have made significant progress towards integrating AI into their healthcare systems. While it is critical to focus on healthcare, it is imperative to focus on its risks and challenges. For example, it is the capacity to enhance access to services, empowering patients to communicate effectively with healthcare providers and achieve self-efficacy in their healthcare journeys. Additionally, it reduces waste and, most importantly, mitigates risks related to bias and breach of privacy regarding health care services. Some challenges associated with integrating AI into healthcare include the requirement of innovative approaches that ensure regulations and collaboration models, as healthcare tends to interact with multiple systems that pertain to wellness. Additionally, managing large data sets can be challenging as continuous monitoring is required to avoid threats tied to privacy breaches. Another challenge is the task of producing a workforce that is trained in AI technologies. While these challenges may be easier to navigate in

countries such as the USA and UK, they are a hardship in developing nations that are forced to focus on multiple needs all at the same time while burdened with a lack of resources. Other challenges include inadequate research focusing on hidden risks and challenges of integrating AI into healthcare in developing nations.

Challenges with AI technology also exist with other nations' security strategies, such as China. The country has made significant progress toward expanding big data, AI, and an extensive technology network that allows its systems to function optimally. According to Zeng (2022), the application of big data AI has effectively strengthened their computer network security systems. However, despite this advancement, major security incidents have occurred frequently. These attacks include physical infrastructure, major network systems, and social media security information. Because of these attacks, governments have applied strategies to empower AI to respond to cyber-attacks and security, which includes positioning themselves in a defense mode. Specifically, China has adopted lines of defense that include guarding against new threats and actively responding to challenges to resolve the application of AI in network defense. Additionally, it has fostered the ability to deal with network incidents and cyber-attack disasters (Zeng, 2022).

The AI-associated risks at the national level focus more on affecting nations' critical sectors, such as data poisoning affecting federal court cases, poor integration in healthcare creating more vulnerabilities, and attacks on physical infrastructure.

The Global Level

The global level includes interconnected organizations meant to connect the different nations. AI has shown the potential to offer multiple benefits that include nations having the

ability to share information, technology, goods, and services and advance global health. While these advances attributed to AI or machine learning continue transforming how the world functions, possible threats to global stability have become evident. For example, AI has been known to demonstrate bias from insufficient or tampered data, the data collection for AI models being an invasion of privacy, and possible effects of its implementation on current and future paradigms (Al-Mansoori & Salem, 2022; Erdélyi & Goldsmith, 2018). Because of these risks, the world has grappled with the legal and ethical issues of AI use, and a conclusion has been made that global problems require global solutions. Risk comes from a solely national approach to the topic, as AI has categories that "transcend national boundaries" and will conflict with legal issues within the domestic sphere. Research has indicated a need for an AI regulatory agency that offers a unified framework for the regulation of AI technologies and informs the development of AI policies around the world. The recommendations indicate an urgent need to create this body and framework to mitigate personalized and political ad hacking, preventing the possibility of autonomous weaponizing agents that may destroy nations, leading to the devastating destruction of international trade, politics, and war.

Future risks exist globally, such as issues regarding people's personal data and bias. The legal approach towards AI technology also poses significant risks due to its potentially harmful effect on the future.

Results

This study's results show multiple threats occur within all three chosen geographical levels. At the local level, people are susceptible to improved versions of malware and phishing attacks that can change their protective measures if there are any, and messages are significantly more challenging to discern. Businesses have the risk of improved and well-timed DoS and

DDoS attacks, as well as deepfake software that can impersonate high-position figures. Local governments are at risk from disinformation campaigns and malicious actors using AI technology to gather personal data and spy on the public. At the national level, risk stems from harm within critical sectors, such as data poisoning essential information, increased vulnerabilities from poor integration of AI technology into the healthcare system, and damage to physical infrastructure from attacks. Finally, at the global level, there are risks with how AI technology gathers data as a potential invasion of privacy and how that data can also be skewed to have bias. Furthermore, the need for a standardized framework for AI technology has potential risks due to the harmful ramifications of all areas of AI not being included.

I would recommend future research that focuses on the risks of artificial intelligence and takes an interdisciplinary approach to the technological, ethical, and legal aspects. This would not only help us better understand the issue but also utilize that perspective to better address artificial intelligence in the future.

Conclusion

In conclusion, AI has been able to set a precedent for sharing information and be a solid pillar for day-to-day operations. Technology has also shown to be significantly profitable and famous in the past few years. Artificial intelligence has also had an impactful history, and its modern integration with cybersecurity has been beneficial for many. However, while that integration has been beneficial, there are also many cybersecurity risks that this emerging technology has brought alongside those benefits. Locally, AI has been integrated into many common cybercrimes, significantly improving their capabilities to gather data, impersonate others, and cause more damage. Nationally, the sectors of nations have taken damage from physical attacks and data poisoning. Globally, the risk to people’s privacy and the possibility of

bias are significant concerns, but the lack of a standardized understanding or framework is also a great concern. While AI as a tool has many established positives, it is important to highlight the potential threats that can create significant harm. Furthermore, as AI technology is still infantile in its potential capabilities, more research needs to be done on this topic with an interdisciplinary approach for a more holistic view in the future.

References

Aggarwal, A. (2023, December 18). *The Second AI Winter and the AI Resurgence Between 1980 and 2010 | Scry AI*. Scry Ai. https://scryai.com/blog/the-second-ai-winter-and-resurgence-of-ai-during-1980-2010/#the-era-of-ai

AIPRM. (2024, January 11). *AI Statistics 2024 · AIPRM*. Www.aiprm.com. https://www.aiprm.com/ai-statistics/#top-10-ai-statistics-2024

Al-Mansoori, S., & Salem, M. B. (2023). The Role of Artificial Intelligence and Machine Learning in Shaping the Future of Cybersecurity: Trends, Applications, and Ethical Considerations. *International Journal of Social Analytics*, *8*(9), 1–16. Retrieved from https://norislab.com/index.php/ijsa/article/view/36

American Psychological Association. (2024). *Misinformation and disinformation*. American Psychological Association. https://www.apa.org/topics/journalism-facts/misinformation-disinformation

Arif, A., Khan, M. I., & Khan, A. (2024). An overview of cyber threats generated by AI. *International Journal of Multidisciplinary Sciences and Arts*, *3*(4), 67–76. https://jurnal.itscience.org/index.php/ijmdsa/article/view/4753

Bécue, A., Praça, I. & Gama, J. Artificial intelligence, cyber-threats and Industry 4.0: challenges and opportunities. *Artif Intell Rev* 54, 3849–3886 (2021). https://doi.org/10.1007/s10462-020-09942-2

Castonguay, A., Wagner, G., Motulsky, A., & Paré, G. (2024). AI maturity in health care: An overview of 10 OECD countries. Health Policy, 140, 104938. https://doi.org/10.1016/j.healthpol.2023.104938

Distor, C., Odkhuu Khaltar, & M. Jae Moon. (2021). Adoption of Artificial Intelligence (AI) in Local Governments: An Exploratory Study on the Attitudes and Perceptions of Officials in a Municipal Government in the Philippines. *Journal of Public Affairs and Development*, *8*, 33–65. https://ovcre.uplb.edu.ph/journals-uplb/index.php/JPAD/article/view/798

Fordham, V., Caswell, D., & Diehl, A. (2023, April 11). *Securing government against adversarial AI*. Deloitte Insights. https://www2.deloitte.com/us/en/insights/industry/public-sector/adversarial-ai.html

Holdsworth, J., & Stryker, C. (2024, June 6). *What Is Natural Language Processing?* IBM. https://www.ibm.com/topics/natural-language-processing

Luna, C. dela. (2024, September 16). *AI and Cyber Security: Innovations and Challenges*. ESecurity Planet. https://www.esecurityplanet.com/trends/ai-and-cybersecurity-innovations-and-challenges/#Bottom-Line-AI-Driven-Solutions-for-Robust-Cybersecurity

Mucci, T. (2024, October 21). *History of artificial intelligence*. IBM; IBM. https://www.ibm.com/think/topics/history-of-artificial-intelligence

National University. (2024, March 1). *131 AI Statistics and Trends (2024)*. National University. https://www.nu.edu/blog/ai-statistics-trends/

Olivia J. Erdélyi & Judy Goldsmith. (2018). Regulating Artificial Intelligence: Proposal for a Global Solution. In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES '18). Association for Computing Machinery, New York, NY, USA, pp. 95–101. https://doi.org/10.1145/3278721.3278731

Ramana Kumar Kasaraneni. (2019). AI-Enhanced Cybersecurity in Smart Manufacturing: Protecting Industrial Control Systems from Cyber Threats. *Distributed Learning and Broad Applications in Scientific Research*, *5*, 747-784. https://dlabi.org/index.php/journal/article/view/128

Sarker, I.H., Furhad, M.H. & Nowrozy, R. AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions. *SN COMPUT. SCI.* 2, 173 (2021). https://doi.org/10.1007/s42979-021-00557-0

Shahrukh Khan Lodhi, Ibrar Hussain, & Ahmad Yousaf Gill. (2024). Artificial Intelligence: Pioneering the Future of Sustainable Cutting Tools in Smart Manufacturing. *BIN : Bulletin Of Informatics*, *2*(1), 147–162. Retrieved from https://ojs.jurnalmahasiswa.com/ojs/index.php/bin/article/view/355

Stryker, C., & Kavlakoglu, E. (2024, August 16). *What is Artificial Intelligence (AI)?* IBM; IBM. https://www.ibm.com/topics/artificial-intelligence

Weaponisation of AI: The New Frontier in Cybersecurity. (2024). Hkcert.org. https://www.hkcert.org/blog/weaponisation-of-ai-the-new-frontier-in-cybersecurity

What Is Adversarial AI in Machine Learning? (2015). Palo Alto Networks. https://www.paloaltonetworks.com/cyberpedia/what-are-adversarial-attacks-on-AI-Machine-Learning#types

Zeng, Y. (2022). AI Empowers Security Threats and Strategies for Cyber Attacks. *Procedia Computer Science*, *208*, 170–175. https://doi.org/10.1016/j.procs.2022.10.025