What is Artificial Intelligence and how is it changing cybersecurity?

Mahek Shah   
Baruch College  
New York, United States of Americamahekshah963@gmail.com

Abstract— Artificial Intelligence (AI)- has gained popularity in the last few years at an unmatched speed as none before. Since the introduction of OpenAI’s chatbot ChatGPT, many large companies, such as Google, Microsoft, and Grammarly, have released their own AI bot to compete with the market. AI is the transformation of technology to the next level. It is a revolutionary discovery that has already begun to start changing the way the world works, and this is just the beginning to the start of changing the world. There is a mind-boggling amount left to see what the future holds for AI. A concept that seemed unreal just a decade ago is in complete pace to take over the world in the next few years in all sectors. This paper discusses how AI is changing the cybersecurity field within technology and its pros and cons. Different domains such as vulnerability management, network security, and other specific applications will also be covered.

Keywords— Artificial Intelligence, network security, malware, machine learning, deep learning, neural networks, hacking, artificial neural networks, vulnerability, bug bounty, threat detection, Denial of Service (DoS) attack

1. Introduction (What is AI? When did it start?)

To fully understand and grasp the information in this paper, it is critical to understand what Artificial Intelligence or AI is. According to John McCarthy from Stanford University, it is “The science and engineering of making intelligent machines, brilliant computer programs” [1]. To put it in simple words, it is the science of making computers or robots that can perform tasks that traditionally require human intelligence. The definition explains how when asked open-ended questions to ChatGPT, it might try to give one a human-like response such as “What are some ideas for my nephew’s 4th birthday?”. One might think the term “Artificial Intelligence” is relatively modern, and it must be coined in the 21st century. However, this is not the case. AI research started as early as after WWII. According to McCarthy, “Alan Turing may have been the first. He gave a lecture on it in 1947” [1]. Alan Turing was an English mathematician who broke the Enigma code, the cryptographic algorithm that the Germans used in WWII to send messages across. It is essential to note that the development of a widely known concept was not made in minutes but rather decades of research which resulted in modern AI software today.

1. The Race of AI

From the release of Open AI’s ChatGPT in November 2022 to the present day, May 2023, the pace that large corporations are moving to catch up with the AI race. Since then, computers have come a long way and changed the world as humans know it. In the 1960s, the US Department of Defense became interested in this type of work and began training computers to mimic basic human reasoning [2]. Google released its AI chatbot known as Bard just a few weeks ago to encounter ChatGPT. The competition is up there, and the big companies are feeling the urgency to release their own AI bot to compete with the others out there. Google Bard did not get as much popularity and good word as ChatGPT due to lacking features and feeling rushed. Microsoft already has a massive stake of billions in startup Open AI since they realized the value of the lion hiding in the forest. Microsoft is even releasing CoPilot, an AI tool for different Microsoft applications such as Microsoft Teams, Microsoft Excel, Microsoft Word, and many more in which the capabilities are unimagined. Some billionaires, such as Elon Musk, founder of Tesla and Space X, seem to be wary about AI and the potential harm it can cause. They believe people are using a tool beyond their power whose impact will be negative if rolled out too quickly and therefore are calling for an AI pause. It is too far-fetched to bring the technology back, for sure. Shortly after ChatGPT’s popularity blew up, many companies and developers started making their own AI tools for different purposes, such as an AI to make videos with speaking bots with the text one input or an AI video generator that allows one to create and edit high-quality videos easily.

1. The different types of AI

AI is often confused with machine learning (ML) and several other types of AI. This section will clarify the difference between AI, ML, neural networks, and deep learning. All of them are related to data processing and learning due to the large volume of data available nowadays.A diagram of a machine learning

Description automatically generated with low confidence

Figure 1: AI in relation to its subsets

* 1. Artificial Intelligence

As mentioned, AI is when a computer is taught to perform human-like tasks or have cognitive thinking functions. This definition simply means that a machine is “learning” or “problem-solving,” just as a human brain does [3]. Applications of AI but not ML include chatbots and natural language processing (NLP). Chatbots are limited to human-defined rules which let them answer questions and help customers, but only a limited amount of knowledge is fed by human input. NLP allows a computer to understand text and speech just as humans do, which makes it a potent tool. Due to the advancement of graphics processing units (GPUs) by companies such as NVIDIA and AMD, data processing has been exponentially faster than a few years ago.[3] Parallel processing allows a computer to take an enormous task and divide it into smaller tasks while working on each of them smaller tasks at the same time, which makes the process more efficient and saves a colossal amount of time when a large amount of data is processed.

* 1. Machine Learning

Machine learning is a field of study within artificial intelligence focused on creating algorithms and models that can learn from data to make predictions or decisions without the need for explicit programming. The main objective of machine learning is to create models that can identify patterns and correlations within datasets and utilize this information to make predictions or decisions about the new data. The process of machine learning involves using statistical techniques to examine large datasets and identify any underlying relationships or patterns within the data. The machine learning algorithm is trained on a labeled dataset, where the expected output or answer is known. The algorithm adjusts its internal parameters through this training process to minimize the difference between its predicted output and the actual output. After the machine learning model has been trained, it can be used to make predictions or decisions regarding new data it has not previously encountered. Predictions can be accomplished by inputting the new data into the model, which will generate an output based on the patterns and relationships it has learned during the training process. There are different categories of machine learning algorithms, including supervised, unsupervised, and reinforcement learning. Supervised learning involves training a model on labeled data. Supervised learning maps an input to an output based on an example of an input-output pair [4]. Examples of supervised learning algorithms include decision trees, naïve bayes, and support vector machines. Unsupervised learning involves training a model on unlabeled data. Unlike supervised learning, there is no “right” answer or no “teacher” to tell the computer what should be the “right” answer. Examples of unsupervised learning include k-means clustering and principal component analysis. Reinforcement learning involves training a model to make decisions based on rewarding the model for the desired answer or punishing it for the undesired answer. Some examples of reinforcement learning include training robots to walk, teaching a game-playing agent to win in chess, and training self-driving cars to make decisions based on traffic.

A diagram of a machine learning

Description automatically generated with medium confidence

Figure 2: Three different categories of machine learning

* 1. Neural Networks and Deep learning

Neural networks are a type of computational model that imitates the behavior of the human brain. They consist of interconnected nodes or “neurons” organized into layers, each of which processes a distinct aspect of the input data [5]. During training, neural networks are provided with a vast dataset of input data and the corresponding output data. The network adjusts the weights of the connections between its neurons to learn how to map the input data to the correct output. The training process involves initializing the network weights, feeding the input data to the network, passing it on through the layers, computing the loss, backpropagating the error, and optimizing the network parameters to minimize the error [5]. Deep learning is a type of machine learning that utilizes neural networks with multiple layers to learn and recognize complex patterns within data. Unlike traditional machine learning algorithms, which may require specific features to be manually engineered, deep learning algorithms can learn features directly from the data itself, making them highly versatile and practical for processing large and complex datasets. To perform deep learning, a deep neural network is constructed with multiple layers, each responsible for learning different features or representations of the input data. [5] Deep learning has many applications, including image and speech recognition, natural language processing, autonomous driving, and robotics. One of the main advantages of deep learning is its ability to learn highly abstract and complex representations of data, which can be utilized for making accurate predictions or decisions.

1. Pros and Cons of AI

Every technology introduced has pros and cons; the same applies to artificial intelligence. Even though AI is all about automating tasks and computers being as intelligent as humans, they are made by humans. The amount of time AI saves us by automating tasks that usually take hours of thinking is mind-altering. Anything humans make is prone to human error, and there will always be errors that must be detected and then fixed. It is essential to note that AI is not always accurate or correct; therefore, one must only partially believe in its answer. It is essential to check through it for the credibility and trustworthiness of the answer one requested, and one must use common sense to validate the answer. An AI model is only as strong as the model it is trained on [6]. Therefore, there can be biases that the AI model can unintentionally have and produce biased results, which are unsuitable if one wants a detailed and proper response. Racism and sexism are examples of how AI models can be biased in institutions that use AI to filter out applicants for job screening [6]. Often in, movies and shows, they portray AI negatively as robots who have become too bright and started taking over the world, but so far, that is not true, and that stage is not still unattainable. AI has unleashed potential for the near future, but the world must be aware of its biases and skepticism.

* 1. Tesla Hack

Autonomous driving cars are the most prominent example of impactful AI, which people have seen in their daily lives, and that starts with no other company than Tesla. All Tesla cars are equipped with technologically advanced features, including AI, that allow them to self-drive while detecting cars around them and reading traffic signs and other sensor systems. Since the whole selling point of Tesla is based on technology and modernized vehicles, one would expect its software system to be solid and impenetrable. However, that is not the case. With Tesla’s features of self-driving and self-parking, if it were to get hacked by a remote user, the car could simply start moving on its own and possibly cause accidents based on the hacker’s intentions. Last year, Synacktiv’s white hackers penetrated through a Model 3 Tesla at a hackathon named “Pwn2own” [7]. The breach would have allowed them to open the car’s doors and front hood. Soon after, the hackers were able to gain access to the infotainment system, which would have theoretically given them the power to take over the car [7]. This is just of the many hacks that have happened so far as Tesla tries to patch up their system to make it better and not hackable.

* 1. ChatGPT Hack

Alex Polyakov was able to hack into GPT-4 just in a few hours by bypassing OpenAI’s safety systems. Polyakov made a “universal” jailbreak, not sure for GPT-4 but even Google Bard and other language models. [8] This vulnerability shows how easy it is for hackers to easily hack into AI systems and make them give the wrong output making it a security threat. Black hat hackers can use chatbots like GPT-4 to make it write scripts for hacking users, which OpenAI is trying to mitigate, but there are possible workarounds. There is no quick fix for this; in the long term, companies must find a way to automate finding safety and security risks. Large companies are willing to pay hundreds of thousands of dollars to hackers who can penetrate the companies’ systems and report the vulnerability instead of misusing the vulnerability. HackerOne and BugBountyHunter are a couple of websites where companies can post a prize for finding vulnerabilities. *Bounty bug hunts* are a new way for companies to patch their system and could save them millions of dollars by finding the bug(s) early on.

1. AI and cybersecurity

Cybersecurity can be described as the practice of identifying, detecting, and defending against cyber threats on any computer, network, or other electronic device. The field of cybersecurity within technology is upcoming at a rate faster than any other sector of technology. The role of cybersecurity has been more significant than ever in today’s age of more and more users going online with multiple devices. Large corporations that hold millions of user data have been heavily investing in cybersecurity to prevent hacks and data breaches from maintaining the integrity of their customer’s data and their company’s reputation. Not only do large corporations have to worry about acceleration in the world of digitalization, but so do small and medium businesses and companies. The projected spending on service providers for the cybersecurity market in 2025 is around $101.5 billion, while cybercrime is expected to increase by around 15% to 10.5 trillion in 2025 [9]. Just a few years ago, these numbers would seem unrealistic, but it is soon to be true. After the COVID-19 pandemic, ransomware attacks were up by 148%, while phishing attacks were up by 510% in a few months [9]. More and more hackers are using AI, machine learning, and other methodologies to launch increasingly sophisticated attacks. For hackers, the enormous reliance on the internet and cloud services targeting an organization’s cybersecurity is justified for making the most money quickly. In 2020, every person globally created 1.7 megabytes of data each second [9]. This number is shocking as it accounts for 8 billion + in the world and the data that can be processed and analyzed. AI does not only come with flaws, as it has its fair share of benefits that help the cybersecurity industry. AI can help with vulnerability management by utilizing sophisticated methodologies to scrutinize potential security risks and abnormal behavior, ultimately enabling prompt detection and containment of threats [10]. AI can also monitor transactional data in a company’s network and protect it from threats.

Artificial intelligence can detect real-time threats through behavioral analysis, enabling swift responses to potential security breaches [10]. Additionally, AI can establish a baseline for a company’s network traffic, which is then used to monitor and protect the network by identifying and mitigating potential threats. Even though automation sounds like a no-brainer, it might not always be a go-to solution for many reasons. One of the reasons is that AI and machine learning is expensive to implement, which many businesses might not be able to afford [10]. Another application of artificial intelligence in cybersecurity is intrusion detection, which involves scrutinizing traffic patterns for any anomalous behavior that may indicate an attack is underway [11]. Naïve Bayes is a commonly used machine learning algorithm in cybersecurity that applies the Bayesian theorem to classify data. This algorithm assumes anomalous activities stem from independent events rather than a single attack [11]. As mentioned previously, decision trees, a form of supervised learning, is applied in cybersecurity to detect Denial-of-Service (DoS) attacks by examining the traffic flow rate, size, and duration. The versatility of artificial intelligence in detecting misinformation lies in its ability to analyze vast amounts of data rapidly. In an experiment with 4.4 million Facebook messages, they were classified as fake and legitimate by employing Naïve Bayes, Decision Trees, AdaBoost, and RandomForest. Fake news was separated from legitimate messages with 86.9 percent accuracy [12]. Fake news should be detected early to avoid chaos and miscommunication on social media. This is where Artificial Neural Networks (ANN) come in place. Many news and social media companies are starting to use it to classify billions of messages and posts. It is impossible to go through each one manually by a human. AI techniques have been employed to detect malicious JavaScript websites and drive-by-download attacks [12]. This technique is a way to address human weakness in cybersecurity awareness, as it is difficult to train every single employee or individual on phishing links and malware files, especially for individuals who are not tech-savvy.

1. Conlcusion

Today’s question commonly associated with AI is, “Will AI ever reach human-level intelligence?” It is nearly possible to reach human-level intelligence based on the connections and links in our brains and the vast amount of information stored in them. AI has a lot in stock for the future of humanity, especially in cybersecurity. Time will tell how AI shapes the future of technology overall. In a few decades, more and more tasks will be automated with discoveries performing jobs that humans would never think possible. AI is truly the key to a technological revolution that is already undertaking place.

##### References

1. V. L. McCarthy, "What is artificial intelligence?," Department of Computer Science, Dartmouth College, Hanover, New Hampshire, United States, 2007. [Online]. Available: https://www.diochnos.com/about/McCarthyWhatisAI.pdf. [Accessed: May 14, 2023].
2. SAS. (n.d.). What is artificial intelligence? [Online]. Available: https://www.sas.com/en\_us/insights/analytics/what-is-artificial-intelligence.html. [Accessed: May 14, 2023].
3. M. Batta and R. Singh, "Machine Learning Algorithms - A Review," International Journal of Computer Science and Mobile Computing, vol. 9, no. 10, pp. 104-111, Oct. 2020. [Online]. Available: https://www.researchgate.net/profile/Batta-Mahesh/publication/344717762\_Machine\_Learning\_Algorithms\_-A\_Review/links/5f8b2365299bf1b53e2d243a/Machine-Learning-Algorithms-A-Review.pdf?eid=5082902844932096. [Accessed: May 14, 2023].
4. P. Ongsulee, "Artificial intelligence, machine learning and deep learning," 2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE), Bangkok, Thailand, 2017, pp. 1-6, doi: 10.1109/ICTKE.2017.8259629.
5. Rene Y. Choi, Aaron S. Coyner, Jayashree Kalpathy-Cramer, Michael F. Chiang, J. Peter Campbell; Introduction to Machine Learning, Neural Networks, and Deep Learning. *Trans. Vis. Sci. Tech.* 2020;9(2):14. doi: https://doi.org/10.1167/tvst.9.2.14.
6. Ntoutsi, E, Fafalios, P, Gadiraju, U, et al. Bias in data-driven artificial intelligence systems—An introductory survey. *WIREs Data Mining Knowl Discov*. 2020; 10:e1356. https://doi.org/10.1002/widm.1356
7. B. Croucher, "Tesla Hackers Take Home a Model 3 After Cracking It in Minutes at Pwn2Own," Gizmodo, 22 March 2019. [Online]. Available: https://gizmodo.com/tesla-hack-hackers-model-3-elon-musk-hackathon-1850263319. [Accessed: May 14, 2023].
8. W. Knight, "The Chatbot Will See You Now," Wired, 17 April 2019. [Online]. Available: https://www.wired.com/story/chatgpt-jailbreak-generative-ai-hacking/. [Accessed: May 14, 2023].
9. M. Doepner, S. Fagard and T. Gall, "Cybersecurity trends: Looking over the horizon," McKinsey & Company, 29 May 2020. [Online]. Available: https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/cybersecurity/cybersecurity-trends-looking-over-the-horizon. [Accessed: May 14, 2023].
10. M. Ramakrishnan, “How AI is Changing Cybersecurity: New threats & opportunities,” Emeritus Online Courses, https://emeritus.org/blog/cybersecurity-how-ai-is-changing-cybersecurity/ (accessed May 14, 2023).
11. Kuzlu, M., Fair, C. & Guler, O. Role of Artificial Intelligence in the Internet of Things (IoT) cybersecurity. *Discov Internet Things* **1**, 7 (2021). https://doi.org/10.1007/s43926-020-00001-4
12. S. Zeadally, E. Adi, Z. Baig and I. A. Khan, "Harnessing Artificial Intelligence Capabilities to Improve Cybersecurity," in *IEEE Access*, vol. 8, pp. 23817-23837, 2020, doi: 10.1109/ACCESS.2020.2968045.