Politecnico di Torino

Cybersecurity and National Defence

AI in Cybersecurity

Research Report

Ali Fuat Sakaci S307607
Edip Shahinyoz S312087
Bruno Petutschnig S311940
Sendege Junior Urlich S297958
Divine Umuganwa S315758

Contents

1 Abstract

This study dives into the complex role of Artificial Intelligence (AI) in cybersecurity.The project focuses on identifying crucial areas where AI is transforming current security procedures. The research analyzes the use of AI community security ,its overall application in reinforcing cybersecurity infrastructures, and its efficiency in detecting and repelling cyberattacks. Additionally, The study further indicates increasing application of AI in cybersecurity simulations and training environments to increase readiness for real-world threats. Moreover,the research critically examine issues of privacy, bias, and accountability regarding AI deployment in cybersecurity. The findings reinforces the ability of AI to transform threat detection, response effectiveness and decision-making while highlighting the importance of careful and transparent application.

2 Introduction

As cyber threats become increasingly complex and prevalent, artificial intelligence (AI) has become a transformative tool in cybersecurity. AI's ability to process large amounts of data, detect patterns, and adapt to new challenges makes it a powerful ally in protecting digital infrastructures. From securing communications and real-time detection of cyberattacks to improving the training of cybersecurity professionals, AI is reshaping our defenses against malicious activity.

While artificial intelligence offers great benefits, it also presents new risks. Cybercriminals make use of it in order to automate and refine their attacks, posing greater risks and challenges to security professionals. Furthermore, the use of artificial intelligence in cybersecurity raises ethical challenges related to privacy, algorithmic fairness, and liability that need to be carefully considered.

This paper dives into AI in communication security, the application of AI in detecting and defending against cyberattacks, its use in training and simulations, and the ethical implications of its widespread adoption in cybersecurity. By exploring these areas, we aim to provide a comprehensive understanding of how AI is shaping the future of cybersecurity.

3 Background

History of Cyber Attacks

The history of cyberattacks is almost as old as the presence of computers. If we allow ourselves to stretch the meaning of cyberattacks we can see the first ever harmful acts throughout 1960 to 1980. These were called "Phone Phreaking", it was the act of manipulation of telephone routing in order to make free phone calls by demand. Phone companies at the time used various types of tones for instructing phone lines to be routed. Some people have discovered that mimicking these otherwise undisclosed frequencies could allow them to self route, and thereby having connections without fees.

Figure 1: Device Used for Phone Phreaking

By the end of the decade phone phreaking had forced the telecommunications industry to rethink its signaling protocols; early network viruses highlighted the critical need for antivirus tools. The issues that emerged made it apparent that new measures needed to be put into place by the authorities. As we entered the 2010's and the years that followed, the cyber threat landscape continued to rapidly evolve. Ransomware became paramount, encrypting corporate and municipal data. Today, whether it's a lone hacker probing a misconfigured cloud bucket or a clandestine intelligence outfit deploying zero-day exploits, the lessons of those early phreakers still resonate: in a world where every protocol and port represents both utility and vulnerability, stability in cybersecurity is an illusion; sustained protection requires ongoing adaptation and vigilance.[1][2]

4 Overview

4.1 AI in Communication Security

As digital communication becomes more integrated into everyday life, from emails and messaging to smart home networks and IoT infrastructure, the need for communication security has improved. Traditional rule-based systems and static filters are no longer sufficient to detect complex, adaptive threats. In response, Artificial Intelligence (AI) has risen as a transformative force in securing communication channels, offering adaptability, pattern recognition, and real time threat response capabilities.

AI contributes to communication security across multiple layers, including physical signal protection, email filtering, spoofing prevention, and IoT traffic authentication. One of the key advancements lies in using machine learning models and deep learning frameworks to analyze behavioral and network traffic patterns, which helps detect anomalies, prevent impersonation, and secure sensitive data transmissions.[3]

In the domain of email and messaging security, AI powered systems use Natural Language Processing (NLP) to detect phishing attempts and social engineering. For instance, platforms like Gmail use TensorFlow based spam filtering models that block over 100 million phishing emails daily. These models learn from large collections of communication data, identifying linguistic cues, urgency triggers, or suspicious attachments.

Generative AI is also being used to secure wireless communication at the physical layer. According to a 2025 IEEE survey, Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) can encrypt signals in a way that maintains low bit error rates for legitimate users while scrambling data for unauthorized receivers. In jamming-prone environments, Deep Reinforcement Learning (DRL) agents adapt transmission strategies in real time, maintaining communication integrity with up to 89% success even under attack.[4]

In the IoT ecosystem, AI techniques such as Random Forests, K-Nearest Neighbors (KNN), and Deep Neural Networks (DNNs) are widely used for device authentication and malware detection. A 2023 study demonstrated that RF fingerprinting, when paired with GAN-generated adversarial training data, could authenticate devices with 99.9% accuracy. This is crucial for securing communication between IoT devices such as smart locks, thermostats, and wearable health trackers, which often operate unattended and are vulnerable to spoofing and eavesdropping.

AI also plays a role in preventing chatbot impersonation attacks and email account takeovers by monitoring behavioral patterns such as login timing, typing rhythm, and semantic analysis of text. For example, Microsoft's Defender uses AI to correlate signal anomalies across endpoints and user behavior to flag compromised accounts before any data is leaked.[5]

Moreover, communication protocols such as ZigBee, Bluetooth, Wi-Fi, Z-Wave, and Sigfox—all integral to IoT networks—can be enhanced by integrating AI-based anomaly detection systems. These AI models monitor communication flows for inconsistencies and detect devices from their expected behavior, even under encrypted environments.[6]

Real world incidents, such as the Mirai botnet DDoS attack, highlight the need for smarter communication defense strategies. AI enables early identification of such threats by analyzing traffic for bot-like patterns. Supervised learning algorithms like SVMs and unsupervised anomaly detectors have been proven effective in this context.

4.2 Usage of AI in Cyber Attack

Among many things AI has proven itself to be capable of, it is also shown to be useful in malicious intentions in cyberspace.

Ever since AI tools became accessible to virtually anyone, it has not shown any slowness in its evolution. With the implemented self-learning algorithms, AI evolves towards faster, more practical and more bulletproof ways of executing its given tasks. This undoubtedly creates challenges in keeping up with its progression in some regards.[7]

Most powerful tropes of AI seem to be doing repetitive tasks optimally, learning and to mimic human-like behaviour. As the pattern recognition algorithms advanced, AI understood more on how to act like a human being. This property is used almost everywhere in cyberspace. With its ability to disguise itself in a space populated and used by humans, its influence became more impactful. As the example of the 2016 election case, the spread of AI accounts through social media is very much capable of causing damage to the society and possibly to the culture of it. A sudden and planned spread of ideas by a massive group of, seemingly human, AI controlled accounts could manifest a belief that will be adapted by the society. As we humans tend to look for other people's opinions of a matter before we come up with our own. Therefore we can see that the possibility of these other opinions were to be of a single group, using AI to manipulate real world matters like elections, stock market or war, the danger seems very prominent, since misinformation is one of the biggest causes of societal collapse.

AI's ability to mimic human behavior is not a miracle by any chance. AI driven systems usually process great amounts of data, analyzing, learning and improving themselves. This trope is the foundation of all that makes AI so fascinating to us.[8] As AI scours the internet, it collects all the data it needs to learn about human behavior. It learns how we act, how we don't act. Once a reasonable amount of information is collected on the matter, it can start filtering the data in order to use the more concentrated database for any given task. The presence and the amount of AI in social media is growing at an alarming rate, the day of distinguishing the authenticity of human actions being impractical is very much in us.

A recent theory led itself to rise, called "Dead Internet Theory". The theory states that the internet is 99 percent if not all of it is used by AI controlled users, that there are barely any human interaction happening on the internet as it is near impossible for one to prove themselves if the billions of people seen on the internet is real outside of the devices we use.[9]

AI has proven useful in brute force attacks as well. Instead of trying many magnitudes of possible entry passwords for a security wall. The AI can eliminate large groups of previously thought as possible answers. This can be done with analyzing the source of the password creation. And assigning weights to certain groups of words of meanings, for example a human made password is more probable to be consisting of meaningful words instead of completely random text characters.[10] These kinds of information and more could be useful to AI in trying to crack a password. An AI attack concentrated on a single person or a group with similar traits can be prepared by researching the entire online presence of the victim to learn about hobbies, friends, ideologies and such. This will let AI understand the personal traits of the victim, eventually creating its own data set for what this person "might" have used as their personal password. This shows the bigger the footprint we leave behind on all accessible internet, the more information we give out that can be used against us or people close to us.[11]

4.3 AI in detecting and defending against Cyber Attacks

In the modern age of technology and its rapid evolution, Artificial intelligence (AI) is a big topic. It comes with revolutionizing benefits, but can also be seen as a threat to online data systems. Its implementation has become paramount to keeping with the times and to not only conquer but also prevent modern cyber attacks. AI can be applied in many cybersecurity fields, most importantly defense.

The Training of Artificial Intelligence Systems

Before an AI system is allowed to manage vital data for the systems it intended to protect, it needs to undergo a learning phase using machine learning algorithms. These phases can be supervised to confirm correct behaviour. Meanwhile it is fed with raw data with which it is able to distinguish between normal and malicious activity . This data can include network traffic logs, system event logs, and user activity records. [12]

The successful development of a threat detection system run by AI is complex. The productivity of the model is dependent on many factors, especially the data quality. The process to ensure a successful training includes; defining the problem, collecting and preparing data, choosing the correct AI model, training, optimization and testing. Once the training has been proven to be effective it is ready to be put to work.

Incorporation of AI in preexisting defence systems

As artificial intelligence is still relatively new and constantly being improved, it is important to understand the many ways it can be integrated into defense systems that have already been set up. AI is able to leverage its ability of pattern recognition and adaptive learning to detect constantly evolving threats which often go unnoticed by human experts. Considering it is capable of processing vast data sets, automating responses, and using predictive analytics allows for fast and accurate detection of possible threats.[13] Not only detecting anomalies but also reducing the amount of false positives, the overall time consuming threat assessment is reduced. AI systems are then able to flag uncertainties, which are then reviewed by experts for final judgement.

Specific Roles of AI in Threat Detection

Artificial intelligence models can have various functions in threat detection. This can include areas such as advanced anomaly detection, where instead of relying on predetermined “red flags” to detect threats, the AI model can analyze and outline any deviations or anomalies that may indicate an attack. This is beneficial as, using traditional methods, only known types of attacks would be identified. Zero-day exploits, especially ones created by AI would also be detectable.[14]

Using artificial intelligence also allows the automation of tedious and time consuming tasks. This would allow security analysts to transfer focus on more important and urgent threats. Models can also be taught to analyze previous threats and attacks by analyzing and predicting patterns to negate the possibility of any ensuing attacks. By doing so the overall security is improved and solidified.

Limitations of AI use in Threat Detection

Despite having many advantages, there are some drawbacks and limitations of using AI to predict and identify threats. AI models heavily rely on available data, which need to be high quality to allow for optimal efficiency. This means data which is inaccurate or biased can change outcomes in identification of threats. [15][16]

There are also concerns about privacy, especially considering the amount of data being fed to these models. Data is constantly being collected and analysed, which definitely calls for sufficient data security measures.

AI is able to detect the vast majority of threats, but as threats continue to evolve, AI may fall back in its efficiency in attack detection. These models can only adapt to a certain extent. This pushes for more training and modernization of AI systems.

As we are dealing with an artificial intelligence, which is constantly learning and susceptible to input, a new concern arises. There is the risk of attackers attempting to manipulate AI models, by feeding them corrupt data. This may lead to threats being overlooked.

Another major weak link in organizations is the supply chain. As many companies depend on third party data, this is an entry point and vulnerability, which hackers are targeting. Tainted data is being introduced through third party data. Some deep learning models lack transparency, when it comes to reasoning. They may still be efficient, but do not allow analysts to easily understand the nature of its decision.

4.4 AI in Cybersecurity Training and Simulation: Enhancing Preparedness through Innovation

In today's increasingly digital world, cybersecurity threats are growing more sophisticated and frequent. As a result, organizations need better ways to prepare their employees and security teams. Traditional training methods, like static presentations or theory-heavy lessons, are no longer enough. Artificial Intelligence (AI) is now at the heart of a major shift in cybersecurity training and simulation, making it more interactive, adaptive, and effective. Through realistic threat replication, personalized learning, and immersive training environments, AI is transforming how individuals and teams learn to defend against cyberattacks.

One of the most impactful uses of AI is in creating highly realistic cyberattack simulations. These simulations replicate actual threats—like ransomware or data breaches—allowing trainees to experience and respond to them in a safe environment. Some platforms, such as the SCORPION Cyber Range, use AI to dynamically adjust the behavior of these simulated threats based on the actions of the trainee. This means that the simulated "attacker" adapts and responds much like a real cybercriminal would, providing a more authentic and challenging learning experience. In advanced setups, even biometric data like heart rate is tracked to measure trainee stress and performance, giving more detailed feedback for skill improvement.[17]

AI also plays a crucial role in phishing awareness training. Phishing remains one of the most common and dangerous cybersecurity threats because it targets human error. AI-powered tools generate realistic phishing emails that are customized based on an individual's behavior and vulnerabilities. If a user falls for a simulated phishing email, the system provides instant feedback and education, helping them learn from the experience. Over time, the AI improves the training by adapting future emails to challenge users more effectively. These simulations don't just help users recognize typical phishing attempts—they expose them to more advanced techniques like spear phishing or smishing (SMS-based phishing), offering a more thorough training experience.[18][19]

To further enhance engagement and learning outcomes, many cybersecurity training programs are integrating gamification. AI makes it possible to create learning environments that function like strategy games, where users must protect virtual assets, make real-time decisions, and experience the consequences of their actions. For example, the game CyberCIEGE allows players to simulate the management of secure computer networks, balancing factors like cost, usability, and security. As users play, the system evaluates their performance, adjusts difficulty levels, and recommends areas

for improvement. This gamified, AI-driven approach makes learning more enjoyable and helps users retain information better through hands-on experience.[20][21]

Overall, the integration of AI in cybersecurity training and simulation is revolutionizing the way individuals and organizations prepare for cyber threats. By combining realistic simulations, personalized phishing tests, and engaging learning formats, AI ensures that training is not only more effective but also more accessible and tailored to each learner's needs. As cyberattacks continue to evolve, AI-driven training will be a critical tool in building resilient defenses and informed cyber professionals.[22]

4.5 Ethical relations regarding usage of AI in Cybersecurity

While there is much discussion around technical solutions to cybersecurity issues, there is far less focus on the ethical issues raised by cybersecurity. Cybersecurity is of critical ethical significance because cybersecurity technologies have an important impact on human well-being as they make possible many contemporary human organisations which rely on the accessibility and integrity of data and computer systems. Cybersecurity raises important ethical trade-offs and complex moral issues, such as whether to pay hackers to access data encrypted by ransomware or to intentionally deceive people through social engineering while undertaking penetration testing.

With the growing threat of cyber attacks in today's digital world, organizations and governments are turning to artificial intelligence (AI) as a tool to enhance their cybersecurity capabilities. AI has the potential to transform the way we defend against cyber threats, from detecting and preventing attacks to responding and recovering from them.

AI in cybersecurity is helpful in learning and adapting to new threats. As cybercriminals become more sophisticated in their attacks, traditional cybersecurity methods may no longer be sufficient. By using machine learning algorithms, AI can learn from past attacks and adapt its defenses to new threats, making it more difficult for cybercriminals to breach security defenses.

The following are 6 key ethical issues of using AI in cybersecurity and suggested resolutions:

Privacy Risks from Increased Surveillance

AI cybersecurity tools often rely on continuous data monitoring, which can intrude on personal privacy. For example, systems analyzing network traffic might unintentionally collect sensitive user information. Regulations like the General Data Protection Regulation (GDPR) in the European Union require organizations to collect only necessary data and use it transparently (Voigt & Von dem Bussche, 2017). Ethical practices, such as anonymizing data and obtaining user consent, are critical to balancing security needs with privacy rights (Zimmer, 2008). [23]

Unfair Bias in AI Decisions

AI systems trained on biased data may produce discriminatory outcomes. For instance, a cybersecurity tool might incorrectly flag users from certain regions or groups as threats due to flawed training data. Studies show that biased algorithms can worsen existing inequalities, such as in facial recognition systems (Buolamwini & Gebru, 2018). To address this, organizations must regularly test AI systems for fairness and inclusivity, as recommended by guidelines like IEEE's Ethically Aligned Design (IEEE, 2019).

Lack of Transparency and Accountability

Many AI systems operate as "black boxes," making it difficult to understand how decisions are made. This lack of transparency can erode trust, especially if an AI tool wrongly blocks a legitimate user. Clear accountability frameworks are needed to determine who is responsible for errors—developers, users, or organizations. The National Institute of Standards and Technology (NIST) emphasizes creating explainable AI systems to improve transparency (NIST, 2020). [24]

Risks of Misusing AI Tools

AI technologies designed for cybersecurity, such as tools that detect software vulnerabilities, can also be exploited by attackers to launch more sophisticated cyberattacks. Researchers warn that malicious actors could repurpose defensive AI for harmful purposes, creating a "dual-use" dilemma (Brundage et al., 2018). Global cooperation, as promoted by agreements like the Paris Call for Trust and Security in Cyberspace, is vital to reduce these risks (French Government, 2018). [25]

Over-Reliance on Automated Systems

Excessive dependence on AI may reduce human oversight in critical decisions. For example, fully automated responses to cyber threats could accidentally disrupt legitimate services. Ethical frameworks, such as the ACM Code of Ethics, stress the importance of keeping humans involved in AI decision-making processes (ACM, 2018). [26]

Inequality in Access to AI Tools

Advanced AI cybersecurity tools are often expensive, leaving smaller organizations or developing nations at greater risk of cyberattacks. This imbalance raises ethical concerns about global security disparities. Initiatives like the Cybersecurity Tech Accord advocate for affordable solutions to ensure equitable access (Microsoft, 2018). [27]

5 Conclusion

AI has become an unquestionable tool in reshaping cybersecurity, offering both revolutionary advantages and complex ethical considerations. Its ability to detect and defend against cyber threats, optimize communication security, and enhance training has strengthened defenses across multiple sectors. However, as AI continues to evolve, it introduces new challenges such as privacy concerns, biases in decision-making. These challenges remind us that while AI offers great potential, it also requires careful application, ethical overview, and a continuous commitment to ensuring transparency and fairness.

Moving forward, the balance between leveraging AI for improved security and addressing its ethical implications will be crucial in shaping a safer, more equitable digital future. As the digital world becomes increasingly intertwined with AI technologies, the importance of fostering responsible innovation and collaboration will only grow, making it essential for stakeholders to work together in creating a secure, inclusive, and ethically sound cyberspace.

Bibliography

[1] Heidi Marie Brush. *Phreaking*. Encyclopedia Britannica. Available at: https://www.britannica.com/topic/phreaking

[2] The Henry Ford. *Blue Box Designed and Built by Steve Wozniak (2017)*. Google Arts Culture. Available at: https://artsandculture.google.com/asset/blue-box-designed-and-built-by-steve-wozniak-and-marketed-by-steve-jobs-circa-1972-wozniak-steve-1
OQGxyXq2DFvKaw?hl=en

[3] Ma, C. *Trusted AI in Multiagent Systems: An Overview of Privacy and Security for Distributed Learning*, 2023. Available at: https://ieeexplore.ieee.org/document/10251703

[4] Zhao, C. *Generative AI for Secure Physical Layer Communications: A Survey*, 2026. Available at: https://ieeexplore.ieee.org/document/10623395

[5] Xiao, L. *IoT Security Techniques Based on Machine Learning: How Do IoT Devices Use AI to Enhance Security?*, 2018. Available at: https://ieeexplore.ieee.org/document/8454402

[6] Waqas, M. *The Role of Artificial Intelligence and Machine Learning in Wireless Networks Security: Principle, Practice and Challenges*, 2022. Available at: https://link.springer.com/article/10.1007/s10462-022-10143-2

[7] Anonymous. *Russian interference in the 2016 United States elections*, 2016. Available at: https://en.wikipedia.org/wiki/Russian_interference_in_the_2016_United_States_elections

[8] First Line Software. *AI Fundamentals: How do we teach machines to act like humans?* Available at: https://firstlinesoftware.com/blog/fundamentals-of-ai-how-do-we-teach-machines-to-act-like-humans/

[9] Jake Renzella. *The Dead Internet theory makes eerie claims about AI-run web. the truth is more sinister*, 2024. Available at: https://theconversation.com/the-dead-internet-theory-makes-eerie-claims-about-an-ai-run-web-the-truth-is-more-sinister-229609

[10] Yi Yang. *Artificial intelligence based password brute force attacks*, 2018. Available at: https://aisel.aisnet.org/mwais2018/39/

[11] Gcore. *How AI is making brute force attacks more dangerous*, 2025. Available at: https://gcore.com/blog/ai-brute-force

[12] Palo Alto Networks, *AI in Threat Detection*, Available at: https://www.paloaltonetworks.com/cyberpedia/ai-in-threat-detection

[13] Sangfor Technologies, *The Role of Artificial Intelligence (AI) in Threat Detection*, February 20, 2024. Available at: https://www.sangfor.com/blog/cybersecurity/role-of-artificial-intelligence-ai-in-threat-detection

[14] ISACA, *The Need for AI-Powered Cybersecurity to Tackle AI-Driven Cyberattacks*, 2024. Available at: https://www.isaca.org/resources/news-and-trends/isaca-now-blog/2024/the-need-for-ai-powered-cybersecurity-to-tackle-ai-driven-cyberattacks

[15] Sangfor Technologies. *The Role of Artificial Intelligence (AI) in Threat Detection*, February 20, 2024. Available at: https://www.sangfor.com/blog/cybersecurity/role-of-artificial-intelligence-ai-in-threat-detection (Accessed: 2025-05-28)

[16] Radware, *Fortifying Defenses Against AI-Driven Cyber Threats*, February 28, 2024. Available at: https://www.radware.com/blog/application-protection/fortifying-defenses-against-ai-driven-cyber-threats

[17] Nespoli, P., Albaladejo-González, M., Ruipérez-Valiente, J. A., & Garcia-Alfaro, J. (2024). *SCORPION Cyber Range: Fully customizable cyberexercises, gamification, and learning analytics to train cybersecurity competencies.*

[18] MetaCompliance. (2024). *MetaCompliance phishing simulation.* Available at: https://www.metacompliance.com

[19] Symantec Enterprise Blogs. (2021). *Phishing simulation and AI tools for employee training.* Available at: https://symantec-enterprise-blogs.security.com

[20] Rahartomo, A., Ghaleb, A. T. A., & Ghafari, M. (2025). *Phishing awareness via game-based learning.*

[21] Naval Postgraduate School. *CyberCIEGE: An Interactive Tool for Information Assurance Training and Education.* Available at: https://nps.edu/web/c3o/cyberciege (Accessed: 2025-05-28)

[22] Almukaynizi, M., Oprea, A., & Nita-Rotaru, C. (2021). *Adversarial machine learning in cybersecurity: Current advances and challenges.*

[23] Mark Coeckelbergh. *Ethics of AI and Cybersecurity When Sovereignty is at Stake.* Minds and Machines, 29, 635–645, 2019. Available at: https://link.springer.com/article/10.1007/S11023-019-09508-4

[24] P. Jonathon Phillips et al. *Four Principles of Explainable Artificial Intelligence.* NIST Interagency Report 8312, 2021. Available at: https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8312.pdf

[25] Paris Call Secretariat. *Paris Call for Trust and Security in Cyberspace.* 12 November 2018. Available at: https://pariscall.international/en/

[26] European Commission. *Ethics Guidelines for Trustworthy AI.* 2019. Available at: https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1.html

[27] Cybersecurity Tech Accord. *Cybersecurity Tech Accord Year Seven Annual Report.* 30 April 2025. Available at: https://cybertechaccord.org/cybersecurity-tech-accord-launches-year-seven-annual-report/