Revista ARACÊ
ISSN: 2358-2472

ARTIFICIAL INTELLIGENCE AND CYBERSECURITY: A STUDY OF ARTIFICIAL INTELLIGENCE IN CYBERNETIC DEFENSE

https://doi.org/10.56238/arev7n5-133

Date of submission: 07/04/2025 **Date of publication:** 07/05/2025

Ricardo Marciano dos Santos[1], Alfredo Nazareno Pereira Boente[2], Vinícius Marques da Silva Ferreira[3], Renata Miranda Pires Boente[4], Danielle Oliveira da Luz[5], Lawrence Matheus Evangelista Duarte[6], Antonio Jorge Borges dos Santos[7] and George Coelho Kleinau Vasconcelos[8]

[1] PhD in History of Sciences and Techniques and Epistemology
Federal University of Rio de Janeiro, HCTE/UFRJ
E-mail: rms221070@gmail.com
ORCID: https://orcid.org/0000-0002-9031-1608
Lattes: http://lattes.cnpq.br/6329550960331880
[2] PhD in Production Engineering
Federal University of Rio de Janeiro, COPPE/UFRJ
E-mail: boente@nce.ufrj.br
ORCID: https://orcid.org/0000-0002-2718-4917
Lattes: http://lattes.cnpq.br/7741044822342404
[3] PhD in Production Engineering
Federal University of Rio de Janeiro, COPPE/UFRJ
E-mail: vinicius.ferreira@pep.ufrj.br
ORCID: https://orcid.org/0000-0003-3664-3510
Lattes: http://lattes.cnpq.br/6490780573139543
[4] PhD student in History of Science and Technology and Epistemology
Federal University of Rio de Janeiro, HCTE/UFRJ
E-mail: renata@hcte.ufrj.br
ORCID: https://orcid.org/0000-0001-7856-5691
Lattes: http://lattes.cnpq.br/8792693794416432
[5] Master in History of Science and Technology and Epistemology
Federal University of Rio de Janeiro, HCTE/UFRJ
E-mail: danielleluz@ufrj.br
ORCID: https://orcid.org/0009-0006-7235-5953
Lattes: http://lattes.cnpq.br/2584342363676913
[6] Master's student in History of Science and Technology and Epistemology
Federal University of Rio de Janeiro, HCTE/UFRJ
E-mail: lawrence@hcte.ufrj.br
ORCID: https://orcid.org/0009-0007-7779-2367
Lattes: http://lattes.cnpq.br/9437015403780414
[7] Specialist in Information Technologies Applied to Education
Federal University of Rio de Janeiro, HCTE/UFRJ
E-mail: anjsantos@firjan.com.br
ORCID: https://orcid.org/0009-0008-7584-6218
Lattes: http://lattes.cnpq.br/0700765800418877
[8] Graduated in Business Intelligence & Analytics
Pontifical Catholic University of Minas Gerais, PUC-Minas
E-mail: gcvasconcelos@firjan.com.br
ORCID: https://orcid.org/0009-0004-3301-502X
Lattes: http://lattes.cnpq.br/4175103920417277

ABSTRACT

The article presents the integration of Artificial intelligence (AI) aligned with cybersecurity, analyzing the opportunities and threats faced by companies, particularly in the year 2024. The technology behind Artificial intelligence, including machine learning, deep learning, and natural language processing, has proven to be crucial to strengthening cybersecurity and data protection in a scenario of constant digital threats. The use of Artificial intelligence allows for a faster and more efficient response to cyberattacks, identifying patterns and detecting anomalies in real time. In addition, the article discusses the importance of governance and regulation of Artificial intelligence in the context of cybersecurity. Implementing appropriate ethical and legal guidelines becomes essential to ensure that Artificial intelligence solutions are used responsibly, minimizing the risks of abuse and vulnerabilities. Appropriate regulation also contributes to transparency and user trust in automated systems. In an increasingly interconnected digital world, the balance between innovation and regulation is essential for the security and success of companies.

Keywords: Artificial intelligence. Cybersecurity. Cyber defense. Machine learning. Deep learning. Zero trust.

INTRODUCTION

Artificial intelligence (AI) has been consolidating itself as a growing trend in several areas, automating mechanical and electromechanical processes and allowing professionals to optimize their time to focus on strategic activities. A study by IBM indicates that 41% of Brazilian companies have already incorporated this technology into their operations, and this percentage is expected to increase significantly by 2024 (IBM, 2022).

Despite the benefits of Artificial intelligence for cybersecurity, it has also been exploited for malicious purposes, being used in cyberattacks. In this context, cyberattacks involve the use of malicious codes to modify computer systems and networks of various institutions.

According to Souza and Morais (2021), advances in the science and technology of Artificial intelligence are not restricted to legitimate applications, but are also appropriated by cybercriminals, who use them to develop increasingly sophisticated malware, capable of causing substantial damage to different organizations.

The complexity of artificial intelligence encompasses a wide range of tools and technical issues that require in-depth knowledge of Information Science for their full understanding. This specialized nature often directs debates toward technical, mathematical, and electronic aspects, leaving aside broader discussions about the impacts of these technologies (Farias, 2022, pp. 53).

When Artificial intelligence is used for malicious actions, the phenomenon is called adversarial artificial intelligence. In this case, cybercriminals use machine learning to develop sophisticated cyber threats, creating algorithms capable of operating autonomously and stealthily, making it difficult for security systems to detect them (Rodrigues, 2021).

According to Farias (2022, pp. 57), when analyzing Brazilian legislation and the governance of artificial intelligence in the country from a legal perspective, one can see the need to improve regulatory instruments to ensure a balance in the use of this technology.

Although Brazil is considered innovative in the implementation of Artificial intelligence in areas such as facial recognition, GPS geolocation, automation of the Judiciary and digitalization of public services, there are still regulatory gaps in addressing emerging threats caused by the misuse of technology by criminals.

The growth in the use of Artificial intelligence by both the private sector and the government has not yet been widely discussed with civil society. This lack of robust dialogue compromises the development of effective control and regulatory strategies, especially when it comes to the application of Artificial intelligence in population monitoring and public safety, including its use for criminal and penal policy purposes. This lack of discussion can expose institutions to vulnerabilities in the field of cybersecurity (Faria, 2022).

From this scenario, it is essential to be aware of the use of Artificial intelligence by cybercriminals, as its malicious application can generate negative impacts. However, at the same time, Artificial intelligence has proven to be a powerful ally in protecting against these threats.

Artificial intelligence has been widely used in cybersecurity to detect financial fraud, conduct forensic investigations, mitigate denial of service (DDoS) attacks, and identify viruses and spam. According to Souza and Morais (2021), the technology stands out in applications such as analyzing the behavior patterns of users, devices, and systems, allowing the identification of anomalies and the anticipation of potential threats to public and private institutions.

Both Artificial intelligence and machine learning are essential tools for the proactive prevention and detection of cyber risks. Their ability to adapt to new types of attacks makes these resources indispensable, and investing in them throughout 2024 will be a fundamental strategic measure to ensure digital security.

METHODOLOGY

This article adopts a qualitative and exploratory approach, based on a bibliographic and documentary review, with the objective of understanding and critically analyzing the impact of Artificial intelligence on cybersecurity. The research was developed through consultation of academic sources, technical reports, scientific articles, institutional documents and specialized publications, selected based on their relevance and timeliness in the context of cybersecurity and emerging technologies.

The bibliographic survey included recognized authors and institutions in the areas of information technology, artificial intelligence, information security and digital law, aiming to identify the main cyber threats associated with the use of Artificial intelligence, as well as the mechanisms, strategies and practices, such as “Zero Trust”, used to mitigate them. The data analysis followed a critical and interpretative approach, seeking to highlight the interrelationships between technological advances and the risks arising from their improper application, in addition to exploring AI-based solutions in combating cyber threats.

The timeline prioritized publications from the last five years (2019-2024), given the rapid evolution of the topics covered, without disregarding classic works or important references for the theoretical basis. The methodology adopted allowed us to identify the main ethical, legal and technical challenges involved in the integration of Artificial intelligence into cybersecurity, as well as to propose reflections on possible paths for the development of safer policies and practices in the digital environment.

ARTIFICIAL INTELLIGENCE

Artificial intelligence is a scientific field dedicated to the study, development and application of systems capable of performing human activities autonomously. Its main objective is to create machines that can operate with cognitive capabilities equal to or even superior to those of human beings in certain situations.

IMPORTANT MILESTONES FOR AI

Over the past few years, Artificial intelligence, ilustred in Figure 1, established itself as a transformative force in several sectors, profoundly impacting the way companies operate and interact with their customers.

Figure 1 - The Artificial Intelligence.

Source: DALL-E, 2025.

The incorporation of Artificial intelligence into organizational processes has shown significant results, enabling the automation of complex tasks, the analysis of large volumes of data, and the improvement of strategic decisions. Through Artificial intelligence, companies can increase operational efficiency, optimize resources, and improve the customer experience by creating personalized and innovative solutions. In addition, Artificial intelligence has become essential in improving decision-making by providing valuable insights that enable more precise and assertive action in the market niche in which the company operates (Totvs, 2023).

The growing application of Artificial intelligence in companies is also reshaping business models, making them more agile and adaptable to constant market changes. However, the effective implementation of this technology requires a strategic approach, which considers both the technical aspects and the ethical issues related to the use of data and the automation of processes.

Artificial intelligence operates by analyzing large volumes of data to create intelligent models that simulate human activities, but there is no single approach to understanding how Artificial intelligence works, as there are several techniques, such as machine learning, deep learning and natural language processing, each with their own methodologies and applications (Ferreira et al., 2025).

Table 1 - Important Milestones for AI development.

Year	Important Milestones
1950	Alan Turing proposed the "Turing Test" to determine whether a machine could imitate a human in a written interaction.
1956	During a conference at Dartmouth College, John McCarthy coined the term "Artificial Intelligence" and laid the foundation for the field of research.
1957	Frank Rosenblatt introduced the perceptron, a type of single-layer artificial neural network used to classify data.
2022	ChatGPT has been launched, attracting attention for its ability to generate intelligent responses, marking a new phase in the everyday use of AI.
2023	The expansion of AI tools has accelerated, generating new creations such as AI-synthesized images and speech.

Source: TOTVS, 2023.

TECHNOLOGIES BEHIND AI

As Artificial intelligence advances, several technologies play a central role in its development. In fact, Artificial intelligence has been evolving significantly, supported by concepts such as Machine Learning, Deep Learning and Natural Language Processing.

In this context, Machine Learning is the basis of modern Artificial intelligence, and allows systems to learn from data and improve over time. Russell and Norvig (2016) describe this technology as fundamental to recommendation algorithms used by platforms such as Amazon and Netflix, which personalize experiences based on user preferences. According to IBM:

> "Generative AI can enhance developers' capabilities and reduce the growing skills gap in the domains of application modernization and IT automation. Generative AI for coding is made possible by recent advances in large language modeling (LLM) and natural language processing (NLP) technologies. It uses deep learning algorithms and large neural networks trained on vast datasets of existing source code" (IBM, 2024).

Neural networks and deep learning techniques, initially inspired by the structure of the human brain, enable machines to recognize complex patterns, such as images, sounds and voice commands. The advancement of deep learning has revolutionized several areas, especially speech recognition, driving the development of generative Artificial intelligence (IBM, 2024). These networks are composed of multiple processing layers that learn to extract relevant features from data autonomously and progressively, making systems more accurate and adaptable. In turn, Natural Language Processing (NLP) allows machines to understand, interpret and interact with human language in an increasingly natural and efficient way. Widely used technologies, such as virtual assistants - Siri, Alexa and Google Assistant -, as well as automatic translators, are based on NLP algorithms combined with machine learning (Russell and Norvig, 2016). These tools represent a milestone in communication between humans and machines, promoting more dynamic, personalized and accessible interactions.

CHALLENGES AND RISKS OF AI

The implementation of Artificial Intelligence in organizations brings a series of benefits, but it also presents significant challenges that cannot be ignored. One of the main obstacles is the high demand for specialized intellectual capital. Companies need to have qualified professionals to design, manage, train and improve Artificial intelligence systems, which requires continuous investment in training and hiring talent with advanced technical knowledge. Another relevant challenge, highlighted by Russell and Norvig (2016), is the urgent need to reskill the workforce. With the automation of routine and repetitive tasks through Artificial intelligence, many jobs are transformed or disappearing, requiring employees to acquire new skills to perform more analytical, creative or technical functions. This transition requires effective corporate training and education policies, promoting the continuous adaptation of workers to the new technological scenario. In addition, issues related to privacy and regulation are gaining prominence. The intensive use of data to train Artificial intelligence models raises concerns about the security of personal information, demanding stricter legislation to regulate the collection, storage and processing of this data. Data governance therefore becomes a fundamental pillar for the ethical use of Artificial intelligence.

Another critical issue is cybersecurity. As Artificial intelligence expands, so does the complexity of digital threats. AI-based tools can be targeted by sophisticated attacks or, in a more worrying scenario, used by malicious actors to develop cyber scams, fraud and system intrusions. Therefore, ensuring the protection of Artificial intelligence solutions and mitigating vulnerabilities becomes a strategic priority.

In light of this, the responsible implementation of Artificial intelligence requires not only technological innovation, but also a strong commitment to ethics, security and social sustainability. The future of Artificial intelligence will depend on the balance between technical progress and the ability of organizations and governments to deal with the challenges it poses.

CYBERSECURITY

Cybersecurity consists of a set of strategies, tools and procedures that aim to protect systems, networks and digital data against unauthorized access, malicious attacks and damage of various kinds. Its main objective is to ensure that information remains intact, confidential and available, both for individual users and organizations. It is a comprehensive area that involves everything from defense against malware and access control to network protection and prevention of threats such as phishing and ransomware, for example.

According to Gartner (2024), cybersecurity involves the adoption of defense mechanisms and strategies aimed at information security, protecting information technology environments against internal and external risks. It is defined as “the practice of protecting computers, servers, mobile devices, electronic systems and networks against malicious attacks”.

This protection involves not only preventive actions, such as the use of encryption, antivirus and firewalls, but also reactive measures, such as incident analysis, fault correction and recovery of compromised data. In this way, cybersecurity, ilustred in Figure 2, becomes essential for business continuity and digital trust in the connected society.

Figure 2 - The Cybersecurity.

Source: DALL-E, 2025.

THE IMPORTANCE OF CYBERSECURITY IN THE DIGITAL TRANSFORMATION SCENARIO

With the exponential growth of digital transformation, cybersecurity has become an essential component for companies across all industries. The integration of new technologies such as the cloud, Internet of Things (IoT) and Artificial intelligence expands attack surfaces, making networks and data more vulnerable to cybercriminals. Gartner (2024) reports indicate that digital transformation has not only increased connectivity and operational efficiency, but has also opened the door to new risks, requiring companies to rethink their security approaches.

In 2023, there was a significant increase in the use of emerging technologies, such as Artificial intelligence, by cybercriminals. These malicious actors began to employ automated techniques to exploit vulnerabilities in systems and gain access to sensitive information in an increasingly sophisticated and efficient manner. This evolution in attack methods represents a new challenge for digital security, requiring faster and more intelligent responses from organizations.

According to the report entitled "Predicts 2024: AI & Cybersecurity - Turning Disruption Into an Opportunity", published by Gartner (2024), the trend is that the use of Artificial intelligence in the area of cybersecurity will continue to develop rapidly until 2028. The expectation is that more than 70% of companies will adopt AI-based solutions as a fundamental part of their cyber defense strategies. This integration of Artificial intelligence promises to transform the way risks are identified, analyzed and mitigated, reinforcing the ability of organizations to respond to threats in real time and proactively.

TOP CYBER THREATS FACED BY ORGANIZATIONS

Cyber threats are a constant challenge for modern organizations, especially in an environment where digitalization is becoming increasingly present. According to a study presented by Gartner (2024), we can see below some examples of the most constant cyber threats in our daily lives:

Ransomware

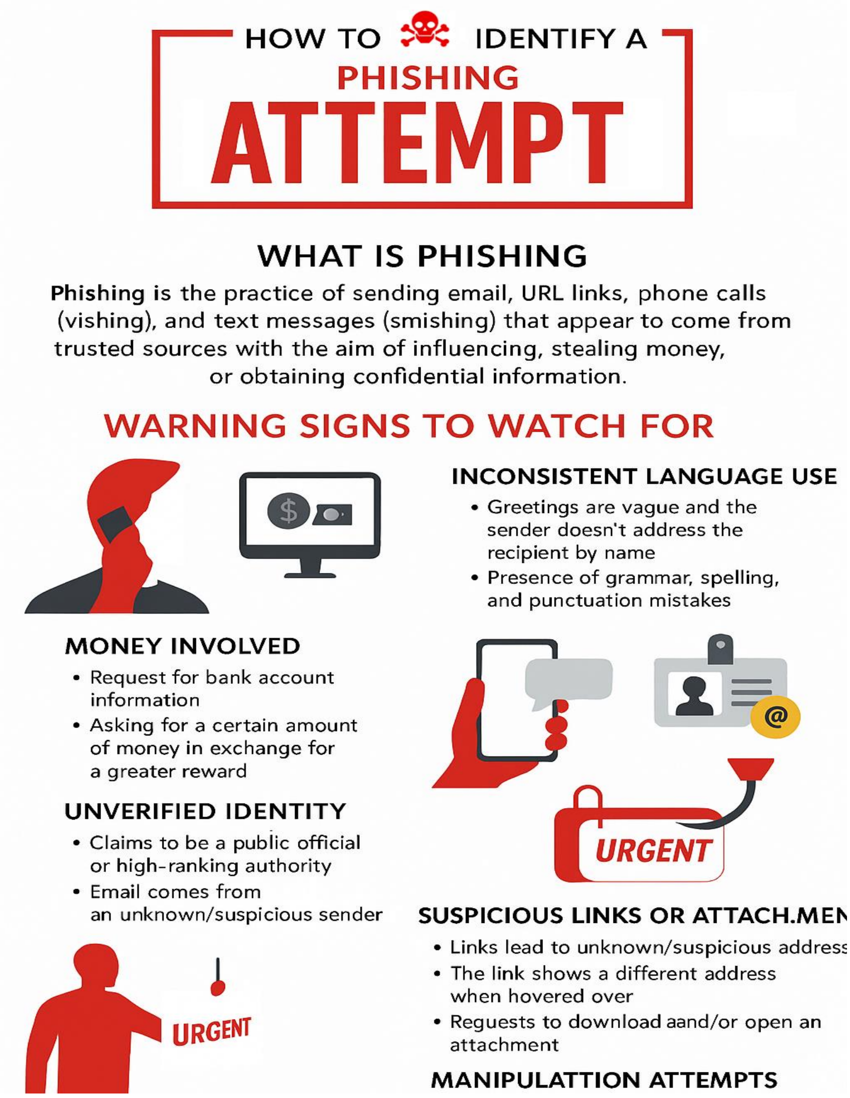
Ransomware is a form of malware that encrypts a victim’s data and demands a ransom to restore access. Ransomware attacks have increased dramatically in recent years and have caused significant losses to businesses around the world. According to Gartner (2024) estimates, the global financial impact of ransomware attacks is expected to continue to grow, with projected losses in the billions of dollars over the next few years.

In most cases, ransomware infection occurs as follows. Malware first gains access to the device. Depending on the type of ransomware, the entire operating system or just individual files are encrypted. A ransom is then demanded from the victims in question. If you want to minimize the risk of a ransomware attack, you should rely on high-quality software. Ransomware is therefore part of the malware family, which is a combination of the English words “malicious” and “software”.

Phishing

Phishing is a tactic used to trick users into providing personal information, such as login credentials and banking details, through fraudulent emails or messages that appear to be from trusted sources. This threat continues to be one of the most prevalent due to its simplicity and high success rate. For Artificial intelligence Applications, the use of AI is making phishing attacks even more sophisticated and personalized, increasing the risk to victims.

Figure 3 - Phishing Attempt.



HOW TO ☠ IDENTIFY A PHISHING ATTEMPT

WHAT IS PHISHING

Phishing is the practice of sending email, URL links, phone calls (vishing), and text messages (smishing) that appear to come from trusted sources with the aim of influencing, stealing money, or obtaining confidential information.

WARNING SIGNS TO WATCH FOR

INCONSISTENT LANGUAGE USE
- Greetings are vague and the sender doesn't address the recipient by name
- Presence of grammar, spelling, and punctuation mistakes

MONEY INVOLVED
- Request for bank account information
- Asking for a certain amount of money in exchange for a greater reward

UNVERIFIED IDENTITY
- Claims to be a public official or high-ranking authority
- Email comes from an unknown/suspicious sender

SUSPICIOUS LINKS OR ATTACH.MEN
- Links lead to unknown/suspicious address
- The link shows a different address when hovered over
- Reguests to download aand/or open an attachment

MANIPULATTION ATTEMPTS

Source: Adapted to Global Sign, 2024.

Phishing is considered an effective method of tricking employees into revealing confidential information or performing actions that result in unauthorized access to company or industry systems.

Distributed Denial of Service (DDoS) Attacks

Distributed Denial of Service attacks, known as DDoS, have the main objective of making the normal functioning of systems, servers or networks unfeasible by overloading them with an excessive volume of malicious or unwanted traffic. This type of attack can cause temporary shutdowns or even complete outages of essential services, generating significant operational and financial losses for companies and institutions (Ferreira et al., 2025).

With the expansion of the Internet of Things (IoT) and the significant increase in the number of connected devices, the attack surface has become even larger, which favors the spread and sophistication of these threats.

IoT devices, often with poor security, are frequently recruited into botnets that carry out large-scale attacks. Given this scenario, it has become essential to develop effective mitigation strategies, such as the use of real-time detection systems, load balancing, use of advanced firewalls and network segmentation. Furthermore, constant monitoring and the adoption of proactive security policies are essential to ensure the resilience of technological infrastructures and the continuity of services in increasingly complex digital environments.

Deepfakes

The use of artificial intelligence to generate fake content, such as so-called deepfakes - videos, audios or images manipulated to accurately replicate the appearance and voice of real people - has become a significant threat in the field of digital security. This technology, despite its impressive advances, raises serious concerns when applied maliciously, especially in processes that rely on biometric facial recognition for authentication.

Identity verification systems, which were previously considered highly secure, now face the challenge of distinguishing real faces from synthetic representations generated by AI. According to Gartner's (2024) forecasts, by 2026, approximately 30% of organizations that currently use facial recognition as their main form of authentication will have to abandon this technology, precisely for fear of attacks based on deepfakes. This scenario requires the adoption of complementary verification methods, such as behavior pattern analysis, multifactor authentication and liveness detection, capable of offering additional layers of protection against digital fraud. Therefore, the need to balance technological innovation with solid cybersecurity strategies becomes evident.

Shadow IT

The term Shadow IT describes the practice of using technological tools, software, and services within an organization without the authorization or supervision of the Information Technology (IT) department. This parallel use, although often motivated by the search for agility or practicality on the part of employees, can generate serious risks to information security. Among the main consequences are the non-compliance with compliance standards and the unintentional exposure of sensitive company data.

With the advancement of artificial intelligence, this problem intensifies. The adoption of unregulated AI solutions, such as content generation platforms, virtual assistants, and chatbots based on generative models, has expanded rapidly in corporate environments, often without clear usage policies. This scenario opens the door to the accidental transfer of confidential information to external systems, increasing the vulnerability of organizations to leaks and cyberattacks. In view of this, it is essential that companies implement strict technology governance guidelines, promote internal awareness, and invest in monitoring tools capable of identifying and controlling the misuse of digital resources.

CRIME AND CYBERWAR

Cybercriminals carry out various types of cyberattacks not only against citizens, but also against companies and government agencies. This criminal practice was accentuated in 2019, during the Covid-19 pandemic.

Although it may seem like a technical issue and of interest only to information technology professionals, cyber risk should be closely examined by board members (Ferreira et al. 2025).

Those who think that this type of risk only affects the technology side of the company are mistaken: it has direct implications for the business and may even compromise its sustainability. It is up to the board to inform itself and lead the company's awareness and commitment to the importance of cybersecurity, as well as to monitor the implementation of the culture and concrete initiatives aimed at information security.

When we talk about Information Security, Cyber Security, we are obviously dealing with a subject directly related to Information Technology, which is the area responsible for producing, storing, transmitting and providing access, security and use to the information of individuals, both natural and/or legal entities. Cognitive warfare is also imminent from the cybersecurity aspect. According to Ferreira:

Revista ARACÊ
ISSN: 2358-2472

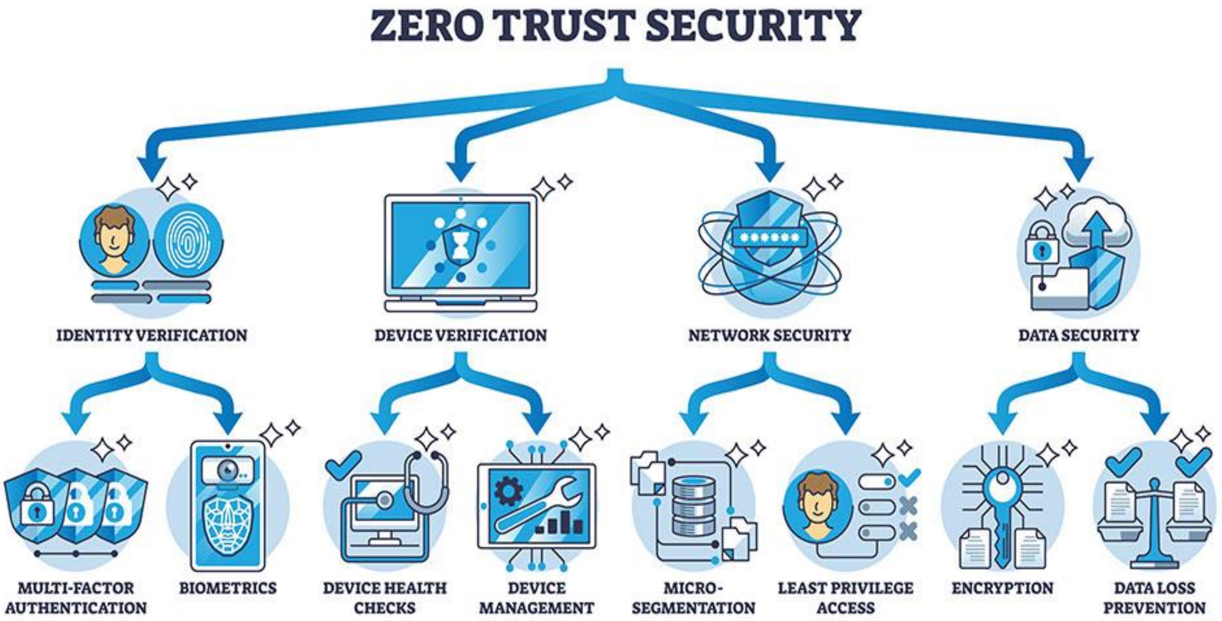
> "Cognitive warfare on social media represents a virtual battlefield where information is used strategically to influence perceptions and behaviors, shaping the digital landscape in favor of certain actors, be they states, organizations or individuals. This phenomenon has expanded with technological advances and the ubiquity of digital platforms, causing significant implications for society, politics and security (Ferreira et al., pp. 14291, 2025)".

Information Technology (IT) offers a set of activities and solutions processed by computational resources to take care of a company's data and information (Pinto et al. 2021).

In this Cyberwar, the Zero Trust policy must be adopted. Zero Trust presents a strategic cybersecurity approach that protects an organization by eliminating implicit trust and continually validating each stage of a digital interaction.

According to Pinto et al. (2021), the Zero Trust security model, ilustred in Figure 4, assumes that a breach is inevitable or has probably already occurred, therefore, it constantly limits access to only what is necessary and looks for anomalous or malicious activities.

Figure 4 - The Zero Trust Security Model.



Source: iStock, 2024.

The Zero Trust model is based on five basic principles:

1. All users on a network are always considered hostile;
2. External and internal threats exist at all times on the network;

3. Network locality is not enough to decide the trustworthiness of a network;
4. Every device, user, and network flow is authenticated and authorized;
5. Policies must be dynamic and calculated from as many data sources as possible.

Today’s modern workforce is becoming increasingly mobile, accessing applications from multiple devices outside the corporate walls. Therefore, Zero Trust employs the model of “Don’t trust, always verify,” i.e. “Zero Trust” (Pinto et al., 2021).

HOW DOES AI IMPACT CYBERSECURITY?

Artificial intelligence is transforming the world, and cyberspace is no exception. According to Leite and Ribeiro (2023), although Artificial intelligence has the potential to improve cybersecurity, it is also being used by malicious actors to enhance attacks, making them more sophisticated, automated, and difficult to detect. By exploring how Artificial intelligence is being used to fuel cyberattacks, ilustred to Figure 5, analyzing specific use cases and discussing the emerging challenges related to the growing attack surface driven by AI.

Figure 5 - Cybersecurity impact by AI.

Source: DALL-E, 2025.

Artificial intelligence can be used to generate more sophisticated and customized malware code, making it more difficult to detect by traditional antivirus software. In addition, it can optimize malicious code to avoid detection, making it more difficult to analyze and eliminate. An example of this is the use of advanced algorithms to develop malware variants that quickly adapt to security measures.

The set of Artificial Intelligence technologies significantly increases the cyberattack surface. Thus, it can write convincing phishing messages and create realistic deepfakes.

IBM reports that 67% of organizations have seen their attack surfaces grow over the past two years. Industry analyst Gartner has named expanding attack surfaces as the top security and risk management trend in 2022 (IBM, 2022).

The "Ryuk" ransomware, known for its targeted attacks on companies, uses machine learning algorithms to identify vulnerable systems and maximize its profits. According to Constantin (2020), FBI agents identified payments of $144.35 million in bitcoins to ransomware groups between 2013 and 2019.

According to Warren (2012), the North Korean group Thallium has been using LLMs to research vulnerabilities and write content for phishing campaigns. The Iranian group Curium is also using LLMs to generate phishing emails and evasive codes to avoid detection by antivirus software. Microsoft also expresses concern about future use cases of Artificial intelligence in cyberattacks, such as fraud involving the use of voice and video.

According to a report by CNN BRASIL (2024), an employee of a multinational company was the victim of a scam involving deepfake, where criminals simulated a video conference with the company's CFO, resulting in the transfer of US$25 million. The worker, initially suspicious of a possible phishing attempt, was convinced of the authenticity of the meeting after recognizing simulated voices and faces of his colleagues. The fraud was only discovered after a subsequent inquiry at the company's headquarters. Hong Kong police arrested six suspects and identified the use of deepfakes to fool facial recognition systems.

The integration of Artificial intelligence and cybersecurity presents both unprecedented opportunities and challenges. While Artificial intelligence is being used to strengthen cybersecurity, malicious actors are also using it to enhance attacks, as demonstrated by deepfake scams and highly sophisticated phishing campaigns. This integration points to a future where cybersecurity will depend heavily on our ability to leverage technologies.

THE ROLE OF AI IN CYBER DEFENSE

Artificial intelligence has emerged as an indispensable tool in cyber defense, proactively identifying and mitigating cyber threats in real time. Its application in traffic monitoring and anomaly detection systems has enabled an unprecedented level of precision and agility, offering an automated response to attacks that would be difficult to identify with human intervention alone.

As discussed in the previous section, Artificial intelligence is capable of identifying anomalous behavior patterns, such as hacking attempts or unauthorized access, through machine learning, which continuously adapts to the cyber environment in which it is inserted. This enables organizations to detect suspicious activities before they become concrete threats, reducing response time and limiting potential damage.

AI-based cybersecurity tools, such as intrusion detection and response (IDR) systems and predictive analytics solutions, are examples of how technology has contributed to making cyber defense more robust. These tools use advanced algorithms to analyze large volumes of data, detecting patterns that indicate malicious activity, such as the use of malware, phishing attempts, and denial-of-service (DDoS) attacks. For example, Artificial intelligence solutions are being used to monitor networks and identify threats in real time, enabling fast and accurate responses that minimize the impacts of an attack.

Predictive analytics, in turn, can predict possible future threats, helping organizations create effective defense strategies and implement stricter security protocols.

As Schneier (2015) points out, “security is not a product, but a process.” This means that defending against cyber threats requires a continuous approach to protection, in which Artificial intelligence plays a key role by providing dynamic, real-time analysis, adapting to new threats and evolving to address emerging challenges.

However, it is important to note that the use of Artificial intelligence in cybersecurity does not eliminate the need for human intervention. Instead, Artificial intelligence acts as a complement, allowing security experts to focus their efforts on more complex activities and strategies, while Artificial intelligence takes care of repetitive tasks and the analysis of large volumes of data.

Therefore, the integration of Artificial intelligence in cybersecurity represents a powerful alliance between technology and the human factor, making cyber defense more effective, intelligent and adaptive. Given the challenges presented by cybercriminals who also use artificial intelligence to enhance their attacks, it is crucial that institutions invest in

Artificial intelligence technologies as a way to stay ahead of cyber threats and protect their systems and data in an efficient and innovative way.

CHALLENGES OF AI INTEGRATION IN CYBERSECURITY

The year 2023 was marked by the exponential growth of cyberattacks, driven by Artificial intelligence (Ferreira, 2024). In 2024, the challenges regarding cybersecurity will be greater, especially regarding the protection of personal data and corporate operations.

According to the consultancy IDC (2024), although the forecast shows a 12% growth in digital crimes in the IT market in 2024, the percentage will likely be much higher, due to the advancement of Artificial intelligence.

Although Artificial intelligence can be used as an ally in combating cyberattacks, it has also proven to be a point of vulnerability for companies that refuse to believe in this imminent possibility. Identifying scams with Artificial intelligence is a complex matter, which depends on pattern recognition, understanding the methods used by algorithms to exploit vulnerabilities, and detecting suspicious activities and deviations from usual patterns in a network.

According to Cibersecurity Fórum (2024), the dynamic cyber landscape demands robust, ethical solutions that can face ever-changing challenges. The synergy between creative minds in the cybersecurity field and Artificial intelligence innovators is essential to shape a secure digital future.

The successful integration of Artificial intelligence into cybersecurity represents a delicate balance between strengthening defenses and mitigating the risks associated with the use of this technology, aiming at data privacy in an ethically correct manner.

Shadow AI occurs when employees use public AI tools, such as ChatGPT, for corporate tasks (INFOMACH, 2024). Although it may seem harmless at first glance, this practice presents serious risks:

- Sensitive data leaks: When employees enter corporate information into public Artificial intelligence platforms, they inadvertently expose confidential data.
- Intellectual property infringement: Trade secrets and strategies can be compromised when shared with open Artificial intelligence models.
- Regulatory non-compliance: In highly regulated industries, the use of Shadow AI can result in violations of data privacy and security regulations.

- Loss of control over information: Once shared, data can be used to train Artificial intelligence models and is beyond the company’s control.
- Specific AI vulnerabilities: Hallucinations, exploitation of flaws, data manipulation.

A variety of incident detection tools and intelligent cyber threat solutions, coupled with constant vigilance through 24 per 7 monitoring, are vital components of maintaining cybersecurity integrity in the age of Artificial intelligence.

AI GOVERNANCE AND REGULATION IN CYBERSECURITY

Artificial intelligence has become a transformative force in many areas of modern society. Therefore, Artificial intelligence needs to be considered in any company's cybersecurity strategies, as its impacts range from social networks to the automation of industrial tasks, generating benefits and many challenges.

The central challenge in regulating Artificial intelligence is to balance public protection with fostering innovation. Rules must exist. However, excessively narrow rules can inhibit innovation, while the absence of standards can lead to the development of technologies that ignore ethics, privacy and individual rights. In this context, it is crucial that regulatory bodies and the industry build a legal framework that combines responsible innovation with the protection of citizens' rights.

The involvement of leadership is vital to emphasize governance related to the topic, encouraging open communication about vulnerabilities, incidents and the functioning of the digital universe as a whole.

The responsible use of Artificial intelligence in organizations is an important factor in the discussion about AI governance principles. With the advancement of Artificial intelligence and its increasing application in companies, it is essential to establish robust Artificial intelligence governance policies. Therefore, it is important to create Artificial intelligence governance policies to be implemented in organizations.

In this context, Artificial intelligence governance establishes policies, processes, and structures to ensure the ethical, safe, and responsible use of AI in organizations. In addition to promoting transparency, it aims to ensure legal compliance and risk mitigation, contributing to the sustainability and competitiveness of companies.

According to BCANLAW (2024), creating Artificial intelligence governance involves the following aspects: Legal certainty, Risk management, Reputation and competitiveness, and Transparency and Trust. To create an effective Artificial intelligence governance policy, the following structured steps are necessary: Defining objectives, Team formation, Initial assessment, Risk analysis, Defining policies, Defining responsibilities, Monitoring and auditing, and Training and education.

In this way, it is expected that companies can implement standards and policies for the responsible use of Artificial intelligence, disseminating them to all their employees.

THE FUTURE OF AI AND CYBERSECURITY

The future scenario of Artificial intelligence applied to cybersecurity points to a dynamic and complex development, characterized by technological innovations that will strengthen digital defense mechanisms, but also by increasingly sophisticated threats that will challenge traditional protection strategies. Artificial intelligence must remain a key player in the modernization of security systems, automating risk identification, pattern analysis and rapid decision-making in the face of cyber incidents.

In recent times, generative Artificial intelligence (GenAI) has emerged as one of the most promising technologies in this field. This approach allows the generation of synthetic content and data from extensive sets of information, considering the analytical capabilities of security solutions. This enables, for example, simulations of attack scenarios, development of predictive responses and constant refinement of defenses based on continuous learning.

The report "Predicts 2024: AI & Cybersecurity - Turning Disruption Into an Opportunity", published by Gartner (2024), points out that GenAI is at the peak of the so-called "Hype Cycle", driven by optimistic predictions regarding its ability to transform digital security operations. However, the consultancy warns that current expectations may exceed what technology can actually offer in its initial phase, recommending a strategic and realistic approach to its adoption.

Among the most obvious benefits of Artificial intelligence in cybersecurity is its ability to reduce response times to critical incidents. Algorithms trained on massive volumes of data can identify anomalous patterns with high accuracy, often before an attack is initiated. In addition, Artificial intelligence can execute corrective actions in an automated manner, reducing the operational burden on security teams and allowing them to focus on more complex decisions.

Gartner projects that by 2028, the use of architectures based on intelligent agents for threat detection and incident response should grow exponentially, rising from 5% to 70% of AI applications in the sector. This growth reflects not a replacement of human professionals, but an expansion of their capabilities. The perspective is that Artificial intelligence will act as a strategic reinforcement, operating in synergy with the human factor, rather than eliminating it.

In this sense, organizations will need to invest in platforms that integrate automation with conscious supervision, avoiding misdirected automated decisions or actions that do not consider specific contexts. Technologies such as multi-agent systems and so-called "action transformers", highlighted by Gartner, have been gaining ground due to their ability to learn from human choices and collaborate in real time with security analysts (Warren, 2012).

In short, the future of Artificial intelligence in cybersecurity will require a balanced approach between technological innovation, responsible governance and continuous training of teams. The challenge lies in exploring the full potential of Artificial intelligence without neglecting the risks associated with its autonomy and misuse, always maintaining digital security as a strategic priority for companies and institutions.

CONCLUSION

Artificial intelligence has played an increasingly important role in the field of cybersecurity, significantly impacting the way digital threats are addressed and how systems are protected. Its use has proven to be a powerful ally in identifying and mitigating cyberattacks, offering automated mechanisms capable of detecting anomalous behavior, responding to incidents in real time, and intelligently adapting to new attack vectors. AI's ability to analyze large volumes of data quickly provides a significant advance in cyber defense.

However, the same potential that strengthens organizations' defenses has also been exploited by malicious agents. Cybercriminals have used Artificial intelligence algorithms to enhance their offensive actions, making attacks more precise, difficult to track, and highly personalized. Techniques such as deepfakes, automated phishing, and social engineering attacks generated by Artificial intelligence have increased the degree of complexity of threats, requiring more sophisticated responses from security teams.

In this context, the integration of Artificial intelligence and cybersecurity requires more than just technology. It is necessary to adopt a holistic approach that encompasses ethical, legal and regulatory aspects. Artificial intelligence governance becomes a strategic priority, as it establishes clear guidelines for its development, use and monitoring, ensuring that the solutions adopted respect principles such as transparency, fairness, privacy and data security.

The definition of governance policies and solid regulatory frameworks is essential to mitigate risks associated with the indiscriminate application of AI. This includes establishing audit mechanisms, monitoring automated decisions and strengthening accountability for the use of algorithms. Digital ethics, therefore, must be a fundamental pillar in the construction of Artificial intelligence solutions focused on cybersecurity.

Collaboration between Artificial intelligence experts, information security professionals and regulatory bodies is another key element for the success of this integration. Only through dialogue and multidisciplinary cooperation will it be possible to face the challenges posed by the misuse of technology and, at the same time, explore its enormous potential in a safe and sustainable way.

As the digital environment becomes more complex and interconnected, continued investment in AI-based solutions will be essential to ensure the resilience of critical infrastructures and the protection of sensitive information. Building a culture of digital security, combined with the development of responsible technologies, can transform Artificial intelligence into a fundamental ally in defending against increasingly advanced cyber threats. Remember, “Zero Trust”.

Therefore, it is essential that companies, governments and academic institutions join forces to develop proactive strategies capable of anticipating risks and responding effectively to future challenges. Artificial intelligence, if used responsibly and consciously, can be a central piece in consolidating a safer, more ethical and trustworthy digital ecosystem for all users in the information age.

REFERENCES

1. BCANLAW. **Política de governança de IA**. Disponível em: <https://www.bcan.law/2024/08/24/politica-de-governanca-de-ia/>. Acesso em: 25/09/2024.

2. Cibersecurity Fórum. **2024 será conhecido como o ano da inteligência artificial em ataques cibernéticos, aponta, Apura**. Disponível em: <https://tiinside.com.br/18/03/2024/2024-sera-conhecido-como-o-ano-da-inteligencia-artificial-em-ataques-ciberneticos-aponta-apura/>. Acesso em: 20/09/2024.

3. CNN BRASIL. **Funcionário de multinacional paga US$ 25 mi a golpista que usou deepfake para simular reunião**. CNN Brasil, 2024. Disponível em: <https://www.cnnbrasil.com.br/internacional/funcionario-de-multinacional-paga-us-25-mi-a-golpista-que-usou-deepfake-para-simular-reuniao/>. Acesso em: 21 set. 2024.

4. Constantin, L. "**Ryuk ransomware explicou**: Um ataque direcionado, devastadorame nte eficaz". CSO Online. Grupo Internacional Dados. Acesso em: 27 set. 2024.

5. Dall-E. **Creating images from text**. Disponível em: <https://openai.com/index/dall-e/>. Acesso em: 01/04/2025.

6. Farias, K.H. **Impactos dos crimes cibernéticos e os riscos da inteligência artificial**: os pilares do direito na proteção dos dados sensíveis. Dissertação de Mestrado. Faculdade de Direito. Universidade Federal da Bahia. Bahia: Salvador, 2022.

7. Ferreira, T. **Cibersegurança 2024**: com IA, crimes cibernéticos serão os desafios do ano. Olhar Digital. Disponível em: <https://olhardigital.com.br/2024/01/01/seguranca/ciberseguranca-2024-com-ia-crimes-ciberneticos-serao-os-desafios-do-ano/#:~:text=Com%20o%20fim%20oficial%20de,dados%20pessoais%20e%20opera%C3%A7%C3%B5es%20corporativas.>. Acesso em: 20/09/2024.

8. Ferreira, V.M.S.; Cosenza, C.A.N.; Boente, A.N.P.; Boente, K.P.; Boente, R.M.P.; Vianna, A.M.S.; Pareto, E.L.; Bianchi, J.M.B. GUERRA COGNITIVA NAS REDES SOCIAIS: AMEAÇAS, DESAFIOS E IMPLICAÇÕES PARA A SOCIEDADE. **ARACÊ** , [S. l.], v. 7, n. 3, p. 14287-14303, 2025. DOI: 10.56238/arev7n3-240. Disponível em: https://periodicos.newsciencepubl.com/arace/article/view/4023. Acesso em: 17 apr. 2025.

9. Gartnet. **Predicts 2024**: AI & Cybersecurity - Turning Disruption Into an Opportunity. Gartner, 2024. Disponível em: <https://www.gartner.com>. Acesso em: 21 set. 2024.

10. Global Sign. **Como identificar uma tentativa de phishing?** Disponível em: <https://www.globalsign.com/pt-br>. Acesso em: 09/09/2024.

11. IBM. **Estudo IBM**: 41% das empresas no Brasil já implementaram ativamente inteligência artificial em seus negócios. IBM Comunica, 2022. Disponível em: <https://www.ibm.com/blogs/ibm-comunica/estudo-ibm-41-das-empresas-no-brasil-ja-

implementaram-ativamente-inteligencia-artificial-em-seus-negocios/>. Acesso em: 10-Set-2024.

12. IBM. **O que é deep learning?** Disponível em: <https://www.ibm.com/br-pt/topics/deep-learning>. Acesso em: 27 set. 2024.

13. IDC. **ICD Consultoria em Gestão Empresarial**. Recife: IDC, 2024.

14. INFOMACH. **Shadow AI**: O desafio invisível da segurança no uso da inteligência artificial. Disponível em: <https://www.infomach.com.br/shadow-ai-o-desafio-invisivel-da-seguranca-no-uso-da-inteligencia-artificial/>. Acesso em: 27/09/2024.

15. Istock. **iStock-Photo**. Disponível em: <https://www.istockphoto.com/br>. Acesso em: 01/12/2024.

16. Leite, E.H.; Ribeiro, D.F. O papel transformador da inteligência artificial na segurança. **Revista Interface Tecnológica**, [S. l.], v. 20, n. 1, p. 181-190, 2023. DOI: 10.31510/infa.v20i1.1669. Disponível em: <https://revista.fatectq.edu.br/interfacetecnologica/article/view/1669>. Acesso em: 21 set. 2024.

17. Pinto, D.J.A. et al. (orgs). **A geopolítica das estratégias em defesa cibernética**: Como EUA, China, Rússia e Israel protege seu ciberespaço. São Paulo: Editora Alpheratez, 2021.

18. Rodrigues, A.R.D. **Inteligência Artificial, Transformação Digital e Cibersegurança no Sector Bancário**: Uma Estrutura Multi-Stakeholder. Dissertação de Mestrado. Instituto Universitário de Lisboa. Mestrado em Gestão. Faculdade de Marketing, Operações e Gestão Geral. Portugal: Lisboa, 2021.

19. Russel, S.; Norving, P. **Artificial Intelligence**: A Modern Approach. 3. ed. Upper Saddle River: Pearson, 2016.

20. Schneier, B. (2015). **Liars and Outliers**: Enabling the Trust that Society Needs to Thrive. Wiley.

21. Souza, J.P.A.; Morais, M.J. Fortalezas e fragilidades no uso da inteligência artificial na cibersegurança. **Revista Tecnológica da FATEC Americana**, vol. 09, n. 02, julho/dezembro de 2021.

22. TOTVS. **O que é Inteligência Artificial?** 2023. Disponível em: <https://www.totvs.com/blog/inovacoes/o-que-e-inteligencia-artificial/>. Acesso em: 28 set. 2024.

23. Warren, T. **Microsoft and OpenAI develop new cyberattack tools using AI and ChatGPT**. 2012. Disponível em:<https://www.theverge.com/2024/2/14/24072706/microsoft-openai-cyberattack-tools-ai-chatgpt> Acesso em: 22 set. 2024.