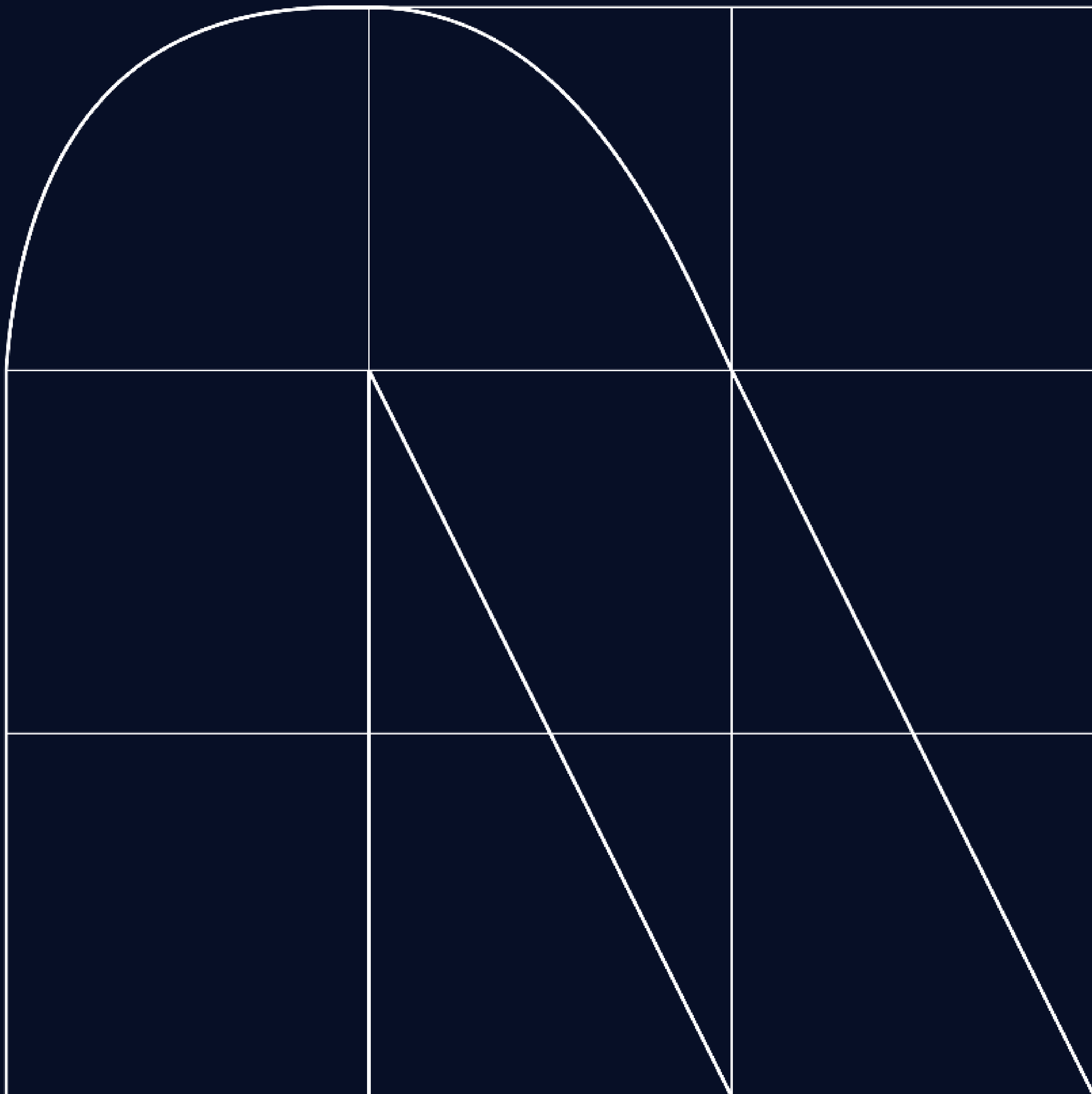
Issue 92 | July 2024

Radar

Cybersecurity magazine



The impact of artificial intelligence on cybersecurity

By David Sandoval Rodríguez-Bermejo

The impact of artificial intelligence on current society is undeniable and leaves behind an era of changes (internet, social networks, 5G...), opening the door to a change of era. While it is true that this technology is extremely powerful, its level of maturity is still quite premature. Until a few years ago, working with neural networks required a fairly technical level of knowledge. However, nowadays artificial intelligence has been democratised and it is no longer necessary to know what happens under the hood to be able to drive the car. This democratisation, combined with the digital capabilities already inherent in today's society, has contributed to its extremely rapid adoption.

The problem with adopting a technology of this calibre without providing training or raising awareness among the population is quite critical. From a productivity perspective, not understanding what goes on behind the scenes means it is not used to its fullest potential (e.g., by correctly creating prompts). However, this lack of knowledge has an even greater impact from a cybersecurity standpoint due to the associated risks it entails (and that go unnoticed).

Integrating immature solutions into production environments implies exponentially increasing the attack surface. Currently, the literature is investigating how to attack both the models and their integrations within a client's ecosystem. In this vein, various attacks have been carried out, such as extracting countless Windows licenses by pretending that my grandmother sang them to me at night to sleep; breaking the models' protections to respond to crimes against humanity; or buying a car for one euro through a company's chatbot.

The problem is that the risk does not lie only in the use (from a technology consumption perspective), but also in its use as a lever to drive and optimise our productive efforts. Many companies use LLMs to develop their products without properly checking and auditing the code. It is known that many of these code-generating models have been trained with GitHub, GitLab, and other public repositories (of which an estimated 70% of the code has some vulnerability). This means that, probabilistically speaking (and with the current level of maturity), the code they generate will be vulnerable.

Another interesting attack resulting from the lack of awareness about this technology occurs under the concept of hallucination. Many times (due to its functioning), AI "invents" the answers. At first glance, this invention is not a problem; the answer is incorrect and is discarded. However, many malicious actors have decided to take advantage of the hallucinations, generating these packages that the AI invents (and that theoretically should not exist) to bypass all security protections quickly and easily.

Another critical risk associated with AI relates to honour and reputation. Due to the generative capabilities of AI, voices and videos can be cloned, and people can be impersonated in various contexts. This fact, combined with the ease of obtaining data to carry out the attack (through social networks), breaks down the barrier between reality and the virtual world, leaving us defenceless against these attacks. While it is true that the attack is not real, the damage it causes is.

The aim of this article is not to instil fear about artificial intelligence but to raise awareness about its misuse and to warn that, due to its level of maturity, it should be used with great caution. To mitigate these risks, it is recommended to adopt existing methodologies such as OWASP LLMs, which provide basic recommendations for implementing generative AI solutions based on LLMs in production environments.

David Sandoval Rodríguez-Bermejo
Cybersecurity Expert Architect

Generative AI is revolutionising the market, and attackers are well aware of it

Cyberchronicles by Alejandro Bernal Almeyda

A recent study by the IBM Institute for Business Value revealed that 64% of CEOs face constant pressure from their investors, creditors, and lenders to accelerate the adoption of generative AI in their companies; in contrast, 84% expressed concern regarding the cybersecurity attacks that this adoption could lead to.

A fundamental part of the adoption of generative AI at a corporate level is understanding the potential risks associated with this emerging technology that need to be mitigated. Some of these risks are associated with attacks specifically designed for this type of AI, such as prompt injection, and others are linked to the exploitation of vulnerabilities within a supply chain attack scheme.

Prompt injection attacks manipulate Large Language Models (LLMs) by using malicious inputs to override the system prompts; prompts are the initial instructions provided to the AI by its developer, and their evasion can result in the AI generating misleading responses or revealing sensitive information.

In September 2023, a university student named Kevin Liu from Stanford University carried out a prompt injection on Bing's chat (Microsoft): by asking Bing to "ignore previous instructions" and to write what is at the "beginning of the previous document," Liu prompted the AI model to disclose its initial instructions, which were written by OpenAI or Microsoft and are generally hidden from the user.

Recently, as this article is being written, The Synopsys Cybersecurity Research Centre (CyCR) announced the presence of a new prompt injection that exploits a security vulnerability to steal data; this vulnerability, which has been assigned the code CVE-2024-5184, was found in the EmailGPT service, an API service and a Google Chrome extension that helps users write Gmail emails using OpenAI's GPT models.

A series of complementary components appear around LLMs, allowing their capabilities to be extended. Among these are agents, chains, and plugins that exploit the power of LLMs, enabling users to build applications that search for information in a database or solve a problem. The risk arises when these LLM extensions are constructed without a security concept: given that the output of an LLM serves as input to these extensions, and the output of the LLM comes from a user’s input (a prompt), an attacker can alter the behaviour of one of these components if it has been designed incorrectly.

Some examples can be found in public documentation:

- Three LangChain vulnerabilities identified and verified by the NVIDIA AI Red Team:
    - CVE-2023-29374 - llm_match allows remote code execution (RCE)
    - CVE-2023-32786 - APIChain.from_llm_and_api_docs allows exploitation of SSRF (server-side request forgery)
    - CVE-2023-32785 - SQLDatabaseChain allows SQL injection attacks

- Three vulnerabilities were reported by Protect AI in May 2024 through their bug bounty platform called huntr, impacting open-source LLM applications:

    - CVE-2024-4078 – This vulnerability may allow an attacker to execute arbitrary remote code on the server (LoLLMs).
    - CVE-2024-3153 – This vulnerability allows an attacker to shut down the server through a file upload endpoint (AnythingLLM).
    - CVE-2024-3104 – This vulnerability may allow an attacker to execute arbitrary remote code on the server (AnythingLLM)).

- The Synopsys Cybersecurity Research Centre (CyCR) discovered a vulnerability in the EmbedAI application, which allows users to interact with documents through LLM capabilities. This vulnerability has been catalogued with the code CVE-2024-5185 and, if exploited, can lead to unauthorised access or data poisoning attacks.

The adoption of generative AI must be accompanied by a security strategy similar to any other conventional application, which includes end-to-end protection from conception, applying concepts of Zero Trust, the principle of least privilege, Security by Design, among others.

Alejandro Bernal Almeyda
Cybersecurity Lead Architect

Artificial Intelligence: Navigating the boundary between defense and attack

By Mafalda Maciel Querido

We have verified that Artificial Intelligence is not just a new buzzword or a "sexy" trend in the world of technology. In fact, Artificial Intelligence incorporates fields that we have known for a long time, such as Machine Learning and Deep Learning, with proven results, and now has a new focus and its use has been democratised. However, its rapid evolution and societal adoption, as several studies have already shown, concern us as cybersecurity professionals. Furthermore, I would say that society will inevitably have to face these challenges.

We can view the problem from two perspectives. On one hand, we know that this technology will change the way we work, accelerate slow processes, boost productivity, and address the shortage of professionals in this field. Organisations that do not get on the innovation train will inevitably be left behind, as we have seen historically. On the other hand, we know that the line between the positive aspects and the imminent dangers brought by Artificial Intelligence is thin, and attackers generally always try to stay one step ahead. Like any hero, Artificial Intelligence will also have its villain.

The positive impacts that Artificial Intelligence brings to cybersecurity are undeniable: greater and better automation in threat detection and response, with the ability to analyse massive volumes of data at unprecedented speeds and, therefore, identify anomalies more quickly, allowing security teams to anticipate risks and threats more effectively, and also help in this crisis of specialised human resources we are experiencing; faster pattern and behaviour analysis; adaptive systems that evolve to counter new threats; increased predictability and the capacity and speed of decision-making based on concrete data and information. In short, Artificial Intelligence can and should be used as an allied tool that helps us in terms of productivity, information analysis, and speed of response in this rapidly transforming environment we live in.

However, like any technology, it also brings new risks, and for cybersecurity, it represents a new factor of speed, sophistication, and reach of attacks. As defence barriers evolve, so do the tactics used by malicious actors. Automation leads to the large-scale exploitation of vulnerabilities that, also benefiting from the adaptability of systems, learn new ways to circumvent security barriers as they encounter them; deception and evasion tactics that mimic legitimate human behaviour, making detection more difficult; AI-guided reconnaissance that allows thorough and faster analysis of potential targets, identifying vulnerabilities and entry points in an organisation's infrastructure; the ability to create highly targeted and convincing phishing, smishing, and vishing messages, which, combined with the use of deepfake, takes the entire field of social engineering to a more sophisticated, unpredictable, and difficult-to-detect level, and brings a new disruption regarding the precautions and defence mechanisms we must equip our employees with.

In addition to the moral and ethical conflict arising from the use of Generative Artificial Intelligence—on which more and more institutions, both governmental and non-governmental, are investigating—and the dangers related to the unintentional sharing of personal data and sensitive information, whether due to ignorance, lack of technological measures for its prevention, or even carelessness, there is an increased exposure of what many consider the weakest link, and for others, the first line of defence in organisations: the human element.

The massive use of this new technology has just begun, and it already has a greater reach than any other technology or platform seen before, and the consequences are already being felt. Although there are still no extensive studies on the impact that Artificial Intelligence will have on information security and cybersecurity from the perspective of human risk, nor very concrete statistical analyses, the first cases of attacks based on Generative Artificial Intelligence technologies are already emerging.

Awareness among employees and society at large regarding Information Security remains one of the least obvious and most difficult points to implement. We still face the challenge of preparing and alerting organisation employees about the risks and importance of security, doing so effectively and yielding results—results that are difficult to measure because there are many variables that are hard to quantify and qualify.

So, how should we proceed in the face of these new and improved threats? How do we teach people to detect increasingly credible attacks at first glance? How do we detect anomalous behaviours when they increasingly resemble our own? Will we need to reinvent ourselves and the way we raise awareness among our employees? At this moment, questions arise for which we currently have few concrete answers.

Security teams must rethink their approach, adopting a proactive stance and adapting to the new reality generated by the implementation of advanced defensive technologies, with a central focus on maximising automation, threat detection, operational agility, and improved decision-making. The urgent need to overcome resource limitations is, without a doubt, an area where Artificial Intelligence emerges as an essential ally.

Dependence on AI not only as a solution for the lack of resources but as a strategic approach to facing constantly evolving risks and threats is imperative. In this sense, the reorganisation of security teams must incorporate not only the implementation of advanced technologies but also the continuous exploration of new methodologies aligned with emerging challenges.

Building a strong security culture is crucial for long-term effectiveness, involving not only training employees with up-to-date knowledge but also promoting a vigilant mindset in daily activities, both professional and personal. We must foster critical analysis, constructive distrust, and the application of good practices in all aspects of daily life, thus establishing a solid line of defence.

Ultimately, the convergence of technology and cybersecurity is a challenging area that requires the strategic union of Artificial Intelligence with human capabilities. Recognising the inevitability of this technological battle of titans and embracing Artificial Intelligence as an indispensable ally is key to strengthening organisations against emerging threats.

Mafalda Maciel Querido
Cybersecurity Project Manager

Building a secure and effective artificial intelligence: The key to AI TRiSM in modern cybersecurity

By Melanie Brenis Valencia

In recent years, artificial intelligence (AI) has experienced exponential growth, transforming key sectors and enhancing the operational efficiency of countless organizations. However, it has also brought with it a series of significant risks, which can be evidenced by the following cases:

- May 2022: The United States Department of Justice publicly announces that the PATTERN algorithm, used to determine the eligibility for early release of certain individuals in federal custody, is undergoing review due to its inherent racial bias. According to reports from local media, the PATTERN algorithm displayed significant disparities by overestimating the risk of criminal recidivism in racial minorities.
- February 2024: A judicial ruling is made against Air Canada after its chatbot provided false information about certain airline policies to a customer. According to reports from local media, the chatbot made serious errors by "hallucinating" non-existent policies, leading the customer to miss their flight. This case highlighted the importance of oversight and verification of AI

As we can see, the current risks and challenges surrounding AI underscore the need for robust approaches that can ensure a safe and effective implementation of it. With this in mind, it is crucial to bring up the concept of **AI TRiSM**.

In general terms, AI TRiSM (AI Trust, Risk, and Security Management) is a set of solutions and/or frameworks focused on **ensuring trust, minimizing risks, and guaranteeing security management** in the use of AI systems by organizations. Its aim is to ensure cross-cutting aspects of AI systems such as governance, reliability, effectiveness, and data protection

Of the mentioned cases, we can note that AI systems can be vulnerable to malicious attacks, inherent biases, and even operational failures that can compromise the integrity of data and the trust of those who use and/or consume them. Considering this, the concept of AI TRiSM aims to address these challenges by providing a set of solutions and/or frameworks to manage and mitigate AI risks, ensuring that implementations are not only effective in operational and economic terms for organizations but also safe and ethical.

Now, to meet this objective, what does AI TRiSM actually encompass? It is comprised of four fundamental pillars, which, succinctly summarized, are as follows:

- **Explainability and Model Monitoring**: Focused on making AI systems more transparent. On one hand, explainability refers to AI systems being able to clarify their decisions and internal processes in a clear and concise manner to those who use/consume them. On the other hand, model monitoring refers to the continuous supervision of the performance and behaviour of AI systems to detect potential biases, poisoned data, information leaks, among other issues.

- **ModelOps**: Focused on end-to-end governance and management of the AI systems lifecycle, which includes operational practices and processes for the implementation, management, and maintenance of AI systems in production environments. ModelOps encompasses activities such as automated model deployment, real-time performance monitoring, version management, and continuous model optimisation.

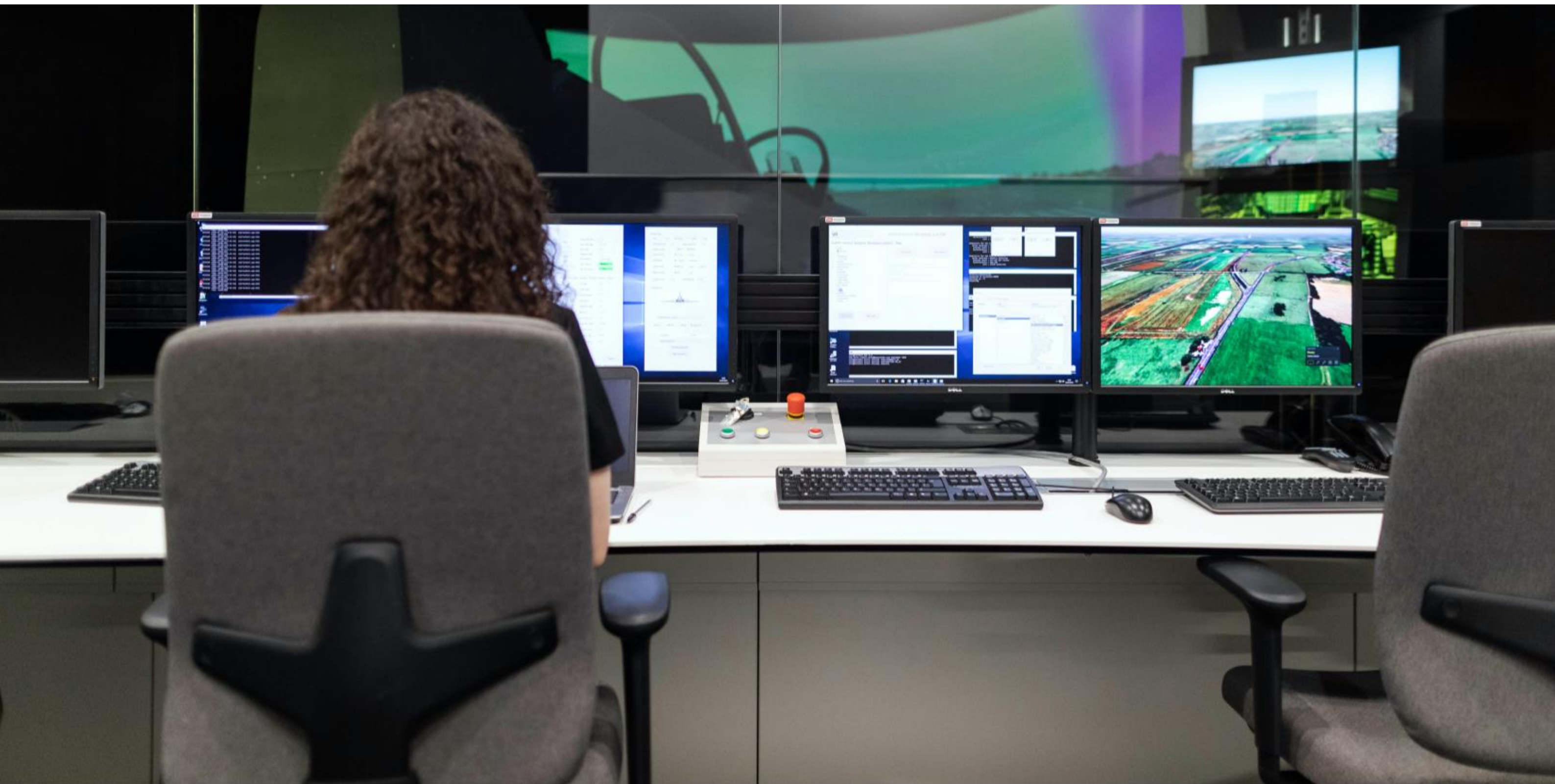
- **AI Application Security**: Focused on detecting and blocking attacks on AI systems. It includes measures to protect AI systems against threats and vulnerabilities, thereby ensuring the integrity, confidentiality, and availability of the data they utilize. This pillar is of utmost relevance because malicious attacks on AI entail losses and damages not only economically but also reputationally.

- **Privacy:** Focused on the protection of personal and sensitive information used in the development and deployment of AI systems. This includes designing AI systems that minimize the collection and use of personal data, as well as implementing anonymization, encryption, and access control measures to protect the privacy of data subjects.

In that sense, we can conclude that AI TRiSM, with its four fundamental pillars, is key for effective, secure, and ethical AI management, which can lead to a competitive advantage for organizations. Aligned with this, according to Gartner, by 2026, organizations whose AI systems operationalize transparency, trust, and security will achieve a 50% improvement in terms of adoption, business objectives, and user acceptance.

Adopting AI TRiSM is not merely the implementation of a set of technical measures, but rather a commitment towards a trustworthy and ethical digital future

Melanie Brenis Valencia
Cybersecurity Consultant



Challenges of cybersecurity in the era of artificial intelligence: A path towards responsible democratisation

Trends by Mauro Pereira Almeida

In an increasingly digital world, the proliferation of Artificial Intelligence (AI) has propelled innovation and operational efficiency of organizations to unprecedented levels. However, this wave of progress is not without its challenges, especially concerning cybersecurity. The democratization of access to AI, while a vector for digital inclusion, amplifies the need for robust information security structures.

One of the most pressing challenges on this journey is managing and controlling access. In a context where AI is capable of processing and analysing vast volumes of data at a staggering speed, it is crucial to ensure that only authorized users have access to sensitive and/or confidential information. Implementing least privilege access and properly managing who has access to what is a premise that cannot be overlooked. Strict mechanisms for identification, authentication, and access privilege definition are essential to avoid the unwanted exposure of data and ensure compliance with strict regulations such as the General Data Protection Regulation (GDPR).

The classification and protection of information are also crucial pillars in this debate. It is essential for organizations to implement robust systems capable of identifying, classifying, and protecting information, considering its level of sensitivity. This process must be continuous, adaptable, and able to promptly respond to the volatile dynamics of cyberspace.

Furthermore, transparency and education are key elements in managing the risks associated with AI. Organizations must invest in training their employees, not only in terms of basic cybersecurity principles but also regarding the ethical and legal implications of AI use. Creating a security-conscious culture is vital for mitigating risks and promoting responsible AI use.

Collaboration among various stakeholders - regulators, academia, industry, and civil society - is another crucial element in building a safe and responsible AI ecosystem. Establishing regulations, security standards, and exchanging best practices are essential measures for addressing the inherent cybersecurity challenges in the context of AI.

In summary, the democratization of access to AI, while an encouraging step towards digital inclusion, requires a thoughtful and diligent approach to cybersecurity. Organizations, as protagonists in this field, have the responsibility to incorporate strong security practices, promote education, and actively collaborate with the community at large to ensure that the AI revolution unfolds safely and beneficially for all. Awareness, education, and multidisciplinary collaboration are therefore cornerstone principles to ensure that we navigate the turbulent waters of technological innovation with safety and confidence.

Originally published in CNN Portugal in November 2023

Mauro Pereira Almeida
Cybersecurity Director

Vulnerabilities

Critical vulnerability in the WordPress wpDataTables plugin

Date: May 31, 2024
CVE: CVE-2024-3820

CVSS: 10

CRITICAL

Description

A critical vulnerability has been discovered in the wpDataTables – WordPress Data Table, Dynamic Tables & Table Charts Plugin for WordPress.

This vulnerability, identified as CVE-2024-3820, is due to insufficient escaping in the user-provided 'id_key' parameter of the AJAX action wdt_delete_table_row.This flaw could allow unauthenticated attackers to inject additional SQL queries into existing ones, which can be used to extract sensitive information from the database.

This vulnerability only affects the premium version of the plugin.

Affected Products

The vulnerability affects the following products:

- wpDataTables: versions up to 6.3.1 (included).

Solution

Se recomienda a los usuarios afectados que actualicen a la versión 6.3.2, que ya ha sido lanzada por los desarrolladores del plugin con los parches de seguridad correspondientes.

- wpDataTables: update to version 6.3.2 and later.

In addition, it is recommended to implement additional security measures, such as the use of web application firewalls (WAF), and regular monitoring of websites for suspicious activity, by conducting regular security audits.

References

- nvd.nist.gov
- wpdatatables.com
- www.wordfence.com

Critical Vulnerability in ThinkPHP Applications

Date: June 5, 2024
CVEs: CVE-2018-20062 and 1 more

CVSS: 9.8

CRITICAL

Description

An active exploitation campaign has been detected targeting ThinkPHP applications vulnerable to CVE-2018-20062 and CVE-2019-9082 (vulnerabilities disclosed several years ago), orchestrated by a Chinese cyber threat group since October 2023. The attacks originate from IPs associated with servers of the provider "Zenlayer" in Hong Kong.

The attackers download an obfuscated file from another compromised ThinkPHP server to gain initial access and install a web shell called "Dama" to maintain persistent access to the server, enabling actions such as privilege escalation and network port scanning.

Affected Products

The versions of the affected products are as follows:

- ThinkPHP: versions prior to 5.0.23
- NoneCMS: versions prior to 1.3.0
- Open Source BMS: versions prior to 1.1.1

Solution

Affected users are advised to update to the latest versions that do not include the vulnerable code:

- ThinkPHP 5.0.23 and later
- NoneCMS 1.3.0 and later
- Open Source BMS 1.1.1 and later

Users are advised to check and clean their systems from these affected versions.

References

- nvd.nist.gov (CVE-2018-20062)
- nvd.nist.gov (CVE-2019-9082)
- www.akamai.com

Patches

HIGH

Check Point Gateway VPN Zero-Day Vulnerability Patched

Date: May 26, 2024
CVE: CVE-2024-24919

Description

Check Point has released a hotfix for the critical zero-day vulnerability in its VPN Security Gateway product with IPsec, which allowed unauthorised remote access, potentially enabling an attacker to move laterally and gain domain administrator privileges.

The manufacturer has indicated that, in a small number of customers, numerous attempts to gain unauthorised access to VPN products have been detected, attempting to exploit this vulnerability.

Additionally, Check Point has created a remote access validation script that can be loaded into 'SmartConsole' and executed to review the results.

Affected Products

The vulnerability affects the following products:

- CloudGuard Network
- Quantum Maestro
- Quantum Security Gateways
- Quantum Spark Appliances

The specific versions affected are: R80.20.x, R80.20SP (EOL), R80.40 (EOL), R81, R81.10, R81.10.x, and R81.20.

Solution

Upgrading to the following versions is strongly recommended:

- Quantum Security Gateway and CloudGuard Network Security: R81.20, R81.10, R81, R80.40
- Quantum Maestro and Quantum Scalable Chassis: R81.20, R81.10, R80.40, R80.30SP, R80.20SP
- Quantum Spark Gateways: R81.10.x, R80.20.x, R77.20.x

References

- support.checkpoint.com
- blog.checkpoint.com

HIGH

SolarWinds Releases Patches for Multiple Vulnerabilities

Date: June 4, 2024
CVE: CVE-2024-28996 and 2 more

Description

SolarWinds has announced in its security bulletin new patches to address 3 new security vulnerabilities affecting the SolarWinds Platform 2024 and FTP Serv-U MFT. The update fixes multiple high-severity vulnerabilities.

Below are details of some of these vulnerabilities:

- Firstly, the vulnerability CVE-2024-28996, discovered by a NATO member, could allow an attacker to perform an SQL injection, thus enabling them to query the SolarWinds database to obtain network information.

- The vulnerability CVE-2024-28995, a Path Traversal flaw, allows attackers to access directories and files outside the server's root directory. Its severity is based on the low complexity of the attack, as it can be exploited remotely without any user interaction.

Affected Products

The products affected by this vulnerability are the following:

- SolarWinds Serv-U 15.4.2 HF 1 and earlier versions (CVE-2024-28995).
- SolarWinds Platform 2024.1 SR 1 and earlier (CVE-2024-28996, CVE-2024-28999 and CVE-2024-29004).

Solution

Apply the latest updates available in the SolarWinds 2024.2 version released by the manufacturer.

References

- solarwinds.com
- documentation.solarwinds.com

Events

SANS London (1 and 6 July)

The SANS London July 2024 is an event that will be held in London, United Kingdom, over 5 days. This event, which is also accessible online, aims to offer practical training in various areas of cybersecurity such as incident management, threat intelligence, digital forensics, and others.
Link

2024 DataConnect Conference (11-12 July)

The DataConnect Conference 2024 will take place from 11 to 12 July in Columbus, Ohio, and is organised by Women in Analytics. This event includes keynote speeches, panels, and workshops on topics such as data analytics, machine learning, and artificial intelligence. It is an inclusive forum that promotes learning, collaboration, and networking among professionals from various sectors. Additionally, there will be a recruitment session and a space for startups.
Link

AI Summit 2024 (17 July)

The AI Summit 2024 will be held on 17 July in San Diego, as part of the Esri User Conference. This event, which can also be attended virtually, will explore the latest advancements in GeoAI and generative AI, and their application in ArcGIS. Attendees will learn about new tools and techniques for data extraction and analysis, and hear success stories in the use of AI. Additionally, there will be opportunities to expand professional networks and collaborate with industry experts.
Link

Gartner Security & Risk Management Summit Tokyo (24 – 26 July)

The Gartner Security & Risk Management Summit Tokyo 2024 is an event that will be held in Tokyo, Japan, from 24 to 26 July. This event will cover various key topics such as generative AI, risk management, cloud security, and more. Attendees will also be able to participate in sessions with experts on threat intelligence, incident response, and the critical role of human factors in building resilient security systems.
Link

Resources

Foresight Cybersecurity Threats For 2030 - Update 2024: Extended report
ENISA (European Union Agency for Cybersecurity) has published the second edition of the study "ENISA Foresight Cybersecurity Threats for 2030," which represents a comprehensive analysis and assessment of the new cybersecurity threats anticipated for the year 2030. The report re-evaluates the top ten previously identified threats and their respective trends, while also exploring their evolution over the past year.
Link

AI RMF Generative AI Profile
NIST (National Institute of Standards and Technology) has published the document "Generative AI Profile," developed over the past year and based on contributions from the public working group on generative AI, which consists of over 2,500 members. This document aims to help organisations identify the unique risks posed by generative AI and proposes actions for managing these risks in a way that best aligns with their objectives and priorities.
Link

ChatGPT-4o
OpenAI, the company responsible for ChatGPT, has announced the launch of ChatGPT-4o, a free version of ChatGPT 4.0, the most advanced version of the conversational chatbot that initiated the generative AI race. GPT-4o accepts any combination of text, audio, image, and video as input and generates any combination of text, audio, and image as output. It can respond to audio inputs in as little as 232 milliseconds, with an average response time of 320 milliseconds, which is similar to human response time in a conversation.
Link

Powered by the
cybersecurity
NTT DATA team

es.nttdata.com