Number 106 | September 2025

NTT DATA

Radar

The Cybersecurity Magazine

AI Adoption: Are We Ready for Cybersecurity Challenges?

By David Sandoval Rodríguez-Bermejo

Artificial Intelligence (AI) is, without a doubt, the technology everyone is talking about right now. It doesn't matter what industry we work in or what use case we're trying to solve — if it doesn't involve AI, it's considered irrelevant. At least, that's the stance of the vast majority of companies today.

While it’s true that the adoption of generative artificial intelligence has become one of the fastest in human history, the reality is that this adoption has been carried out in a rather precarious way from a cybersecurity perspective. A new tech stack, capable of "reasoning" and easily automated, is every manager's dream, since its potential is virtually limitless.If we add to this the fact that humans now possess increasingly advanced digital skills (digital natives) and that generative AI has been democratized due to its ease of use, we have a perfect breeding ground where opportunities and risks coexist in equal measure.

I won’t dwell on the opportunities — they are widely known. You just need to open any newspaper and scroll through tech news to find them. The problem lies in the risks, and in the fact that most people implementing AI solutions (largely thanks to this democratization) lack proper cybersecurity knowledge.

To give a few quick examples: Integrations with AIs that exfiltrated company data to third parties. Poorly secured setups or mishandled information flows that allowed manipulation of LLMs (Large Language Models) to obtain products at ridiculous prices. Theft (exfiltration) of entire models, such as what happened with the first version of LLaMA, which can put an entire organization at risk by exposing its intellectual property.

In fact, from an unforeseen risk, a new paradigm emerged in the AI world: open-source LLMs. Due to the leak of LLaMA's first model, Meta decided to open its model to the world and publish the steps behind it — prompting many other companies to replicate this strategy.

This led to the rise of major communities like HuggingFace, which serves as both a centralized repository of AI models and a backbone resource for AI training — all for free. This shift has added an extra layer of competition for major players like Google, OpenAI, and Anthropic, who now face serious contenders among open model initiatives.On the other hand, there’s a new trend within AI: the definition of small operators (Agents) that coordinate with one another to carry out a specific task thanks to their specialization.

Just like what happened with the release of open-source models, the entire world has jumped into the use of agents without taking proper security measures to deploy safe and well-fortified systems. While many lessons learned from the first LLMs have been put into practice, there is still a long road ahead.

Given this new scenario, it's worth asking: what does the future hold for cybersecurity? We're entering a highly complex landscape with countless challenges: securing language models, managing and hardening AI agents, supply chain security (especially those partially managed by AI), and more.

Finally, from a business standpoint, I’d like to highlight a paradigm shift gradually emerging in today’s market. Product vendors, leveraging AI, are transforming into service providers by incorporating AI services into their infrastructure. This shift allows them to generate: Recurring revenue through new service subscriptions, and Passive income via token purchases required to execute those services.

In cybersecurity, a prime example of this shift is the Burp suite for dynamic analysis, which now includes Burp AI to speed up the execution of security audits.

To conclude, we are living in an era of significant challenges, especially in the intersection of artificial intelligence and cybersecurity. The rapid adoption of AI across the business sector is generating not only opportunities and innovation but also substantial risks. Therefore, it's crucial that organizations implement strong security policies and foster continuous education in both AI and cybersecurity.

David Sandoval Rodríguez-Bermejo
Cybersecurity evangelist

Reality Outpaces Code: AI, Frauds, Robots, and Quantum Warfare

Cyber chronicle by Rodrigo Rey

Artificial intelligence can now replicate your face, your voice, your gestures... and even your access credentials.In the words of Sam Altman himself, biometric barriers are being overcome by systems capable of deceiving the sensory signals used by our authentication interfaces.

We are standing at the threshold of a mass identity fraud crisis, where credentials will become meaningless if they can be forged by an AI that imitates you better than you can imitate yourself.

Cybersecurity is thus entering a new era: “Who you are” is no longer enough to prove that you are you. Digital trust, as we’ve known it, is compromised. Get ready.

On the brighter side, AI doesn’t just clone your identity — it could also save your life before your heart decides to shut down the system.

At Johns Hopkins University, researchers have developed MAARS, an AI model capable of analyzing cardiac images and predicting heart failure with 89% accuracy. This far surpasses traditional human and clinical capabilities, as it detects subtle signs in heart tissue that often go unnoticed. Science fiction five years ago. Precision medicine today.

Meanwhile, Japan takes a stand — and not with bonsai trees, but with qubits. Its ambition is crystal clear: to dethrone the U.S. and China in the race for quantum computing by 2030.

Their weapon: a quantum supercomputer developed by RIKEN and Fujitsu, already operating with 256 qubits and projected to be 25% more powerful than IBM’s current giant.

But this move isn't just geopolitical — it's strategic. The nation that dominates quantum computing will: Break modern cryptography Accelerate AI development Redesign global security Welcome to the next front in the digital battlefield.

So, if you thought ransomware was as bad as it gets, just wait until quantum-powered attacks become widespread.

Meanwhile, Shanghai has just become the runway for functional Terminators.

Over 150 humanoid robots — equipped with vision, language, and action capabilities — have been unveiled as part of an industrial initiative that moves beyond prototypes and into real-world deployment. These robots don’t just walk or lift objects — they understand commands, evaluate their surroundings, and act autonomously in real time.

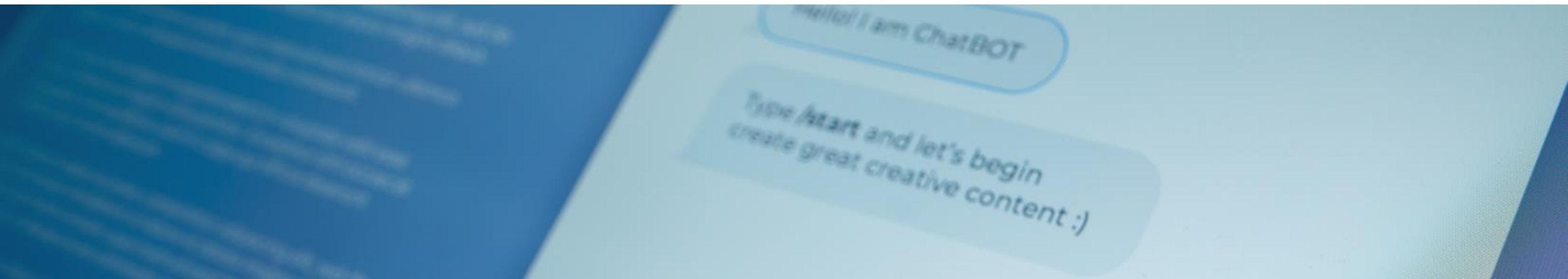
China is making one thing crystal clear: AI is no longer just an idea...It’s an industry.

And yes, if robots scare you, maybe this time it’s for good reason. As if that weren’t enough, Q2 of 2025 saw a boom in phishing attacks using real brand names. According to CheckPoint, the most impersonated companies were: Microsoft, Google, Apple and Spotify. But the most shocking case?Booking.com, which suffered a 1000% increase in fake domains.

The reason? Summer vacation season. The goal? Your banking credentials. So if you get an email saying,“Booking confirmed in Cancún for €3,452” —you might want to think twice before clicking.

We are witnessing a perfect storm, where AI is the protagonist, the antagonist, and the oracle. Cybersecurity is no longer just prevention —it's prediction, it's anticipation, it’s understanding that the future can no longer be hacked... it has to be trained.

Rodrigo Rey
Cybersecurity Lead Architect



Hello! I am ChatBOT

Type /start and let's begin create great creative content :)

The Impact of AI on Cybersecurity and Cyberdefense

Article by Carlos González Parrado

Cybersecurity is a constantly evolving field, driven by the rise of digital connectivity and the increasing sophistication of threats. Initially, security measures focused on protecting systems from basic attacks using antivirus software and firewalls. However, over time, new strategies have emerged — such as advanced encryption, multi-factor authentication, and intrusion detection — to address more complex threats.

This mutual evolution of security and attacks can be effectively understood through game theory, a branch of mathematics that studies strategic decision-making between different agents.

In this context, we find defensive agents (organizations, cybersecurity professionals) and attackers (hackers, cybercriminals). Every time an attacker develops a new tactic or tool to breach a system, defenders must adapt their strategies to counter these moves — creating a constant strategic game of cat and mouse.

Over the past decade, cyberattacks have become more organized and targeted, exploiting vulnerabilities in both software and human interaction. This has pushed organizations to adopt a proactive security posture, relying on data analysis and continuous monitoring to detect and mitigate risks before they escalate into incidents.

The emergence of technologies like the Internet of Things (IoT) has further expanded the cybersecurity landscape, making the protection of data and systems an even greater challenge.

In recent years, numerous computing techniques — particularly various forms of Artificial Intelligence (AI) — have emerged, enhancing the development of tools for both defenders and attackers. This has led to an acceleration in the evolution of both sides of the cybersecurity arms race..

Why AI Means Acceleration

Artificial Intelligence (AI) is revolutionizing cybersecurity by providing tools that enable faster and more efficient responses to threats. Thanks to its ability to analyze vast amounts of data and learn from behavioral patterns, AI can detect anomalies that could indicate a possible cyberattack, with a speed and precision far beyond human capabilities.

Traditional information analysis techniques require the knowledge of one or more subject matter experts, along with the implementation of specific code to perform analysis and management tasks — resulting in significantly longer development times.AI enables cybersecurity systems to adapt to new threats with minimal reaction time, identifying the tactics and techniques used by attackers. This not only accelerates threat detection, but also improves incident response, allowing organizations to mitigate risks before significant damage occurs.Moreover, AI can automate repetitive tasks, freeing up cybersecurity professionals to focus on more complex and strategic challenges. In fact, it doesn’t just automate — it also provides strategic guidance in prioritizing different vulnerabilities, thereby speeding up their mitigation.But it’s not all on the defensive front — AI can also be used by attackers to build more sophisticated tools. Even though LLMs (Large Language Models) have built-in restrictions, cybercriminals can circumvent these limits and get them to generate the necessary code. This opens the door to new types of attacks, such as vishing (voice phishing).

Use Case: Cyberdefense

We see this acceleration particularly clearly in the domain of cyberdefense, especially within Security Operations Centers (SOCs), where AI improves threat detection, response, and analysis.First and foremost, AI enables the automation of network monitoring, analyzing massive volumes of data in real time to identify anomalous behavior patterns that could indicate a cyberattack — something that was very difficult to automate in the past.

Moreover, AI systems can learn from past incidents, continuously improving their ability to recognize new threats and reduce false positives. This leads to a faster and more effective response to incidents, as AI can prioritize alerts and suggest corrective actions.

Finally, these techniques facilitate the integration of different information sources, providing a more complete view of the threat landscape and helping analysts make informed decisions.

Altogether, these capabilities make Security Operations Centers (SOCs) more proactive and efficient in defending against cyberattacks.

AI Adoption Among Professionals

As Artificial Intelligence (AI) becomes established as an essential tool in the field of cybersecurity, a recent survey conducted by CrowdStrike (CrowdStrike State of AI in Cybersecurity Survey) offers valuable insight into its adoption among industry professionals.

The data reveals that, while there is widespread recognition of the benefits AI can provide, there's a clear gap between the intention to use AI and the effective implementation of AI projects in production environments.

One of the main obstacles identified is the lack of adequate data and poor data management, which limits professionals' ability to develop truly effective solutions. This issue is especially pronounced in cybersecurity, where data itself is a highly valuable asset, and mishandling it can create high-risk profiles for AI applications.

According to an article published in VentureBeat, a staggering 87% of AI-based projects never make it to production. This highlights the fact that, despite the hype around AI, its real-world implementation often falls short.

What we see is that—apart from a small group of experts capable of using models in more rudimentary or self-developed ways—most professionals aim to use AI at a user level to perform specific tasks. This creates a strong demand for more accessible and less technical tools.

However, the lack of professionals with multidisciplinary training — not just data scientists or mathematicians, but those who also possess skills in software engineering and security — limits the full transformative potential that AI could offer to the cybersecurity sector.

Given the data and findings, it becomes clear that there is a strong need to promote a comprehensive approach: one that not only trains data scientists, but also prepares engineers and other cybersecurity specialists to effectively manage AI projects. Only then will we see broader and more effective adoption of AI in cybersecurity, transforming organizations' defensive capabilities and equipping them with an evolutionary edge comparable to that of cybercriminals.

Conclusion

Artificial Intelligence has enabled massive advances for both sides of the security equation — offering greater speed in executing attacks and enhanced adaptability in defense. Cybercriminals will continue to exploit these new techniques to carry out more sophisticated, large-scale, and previously unseen types of attacks.

That's why it's crucial to have cybersecurity experts capable of developing tools for monitoring, management, and remediation against emerging security gaps — ensuring we don't fall behind in this rapidly evolving landscape.

.

Carlos González Parrado
Cybersecurity Analyst

The Use of Artificial Intelligence in Offensive Activities

Article by Fernando Echevarria Gutierrez

Artificial Intelligence is no longer just a promise — it has become a fundamental force in the field of cybersecurity. This powerful tool has turned into a double-edged sword, forcing organizations to rethink their protection models.

While attackers use AI to automate, personalize, and scale their attacks, defenders must adapt their mechanisms to respond with equal speed and intelligence.

1. Types of AI-Based Attacks

In the past, risk was mostly centered on technical exploits. Today, new attack vectors emerge that leverage generative, adaptive, and predictive capabilities. These attacks not only increase in frequency but also in complexity — exploiting the very same technologies that many organizations use to innovate.

The integration of language models into digital products — from chatbots to support tools, analytics, or automation systems — has opened up a new surface of exposure. Unlike traditional vulnerabilities, these attacks do not necessarily rely on code flaws but rather on language logic, semantic interpretation, and overreliance on non-deterministic models.

The following attack vectors, also highlighted in the OWASP Top 10 for LLM Applications, are currently among the most dangerous and common:

1.1. Prompt Injection

In applications based on Large Language Models (LLMs), prompt injection attacks involve modifying the original instructions via malicious user input. An attacker can embed hidden commands within user prompts, causing the model to behave in unintended ways.

These types of attacks do not require advanced technical skills, and their detection is particularly difficult if explicit boundaries are not set in the model’s instruction handling.

A real-world example occurred with AI assistants that generated temporary passwords or access tokens. When they received a subtly camouflaged instruction hidden in a question, the model responded by providing the token directly, bypassing the security rules defined by the developers.

1.2. Deepfakes

The use of Generative Adversarial Networks (GANs) allows for the creation of highly realistic fake images, videos, or audio. While initially associated with entertainment or technical experimentation, their use has expanded to identity fraud, political manipulation, and impersonation in corporate or legal contexts.

They pose a growing threat in scenarios where biometric or visual validation is critical. In 2020, a bank in Hong Kong was scammed out of more than $35 million when a voice deepfake mimicked the CEO of a partner company, authorizing a fraudulent transfer. Unfortunately, this is not an isolated incident — such attacks are becoming increasingly common, both in business and personal spheres.

1.3. *Automated Phishing*

One of the most widespread attacks today is AI-powered phishing. Unlike traditional phishing — often generic and poorly written — AI enables the automation of phishing campaigns, generating highly personalized messages with natural language and specific references to the victim. This drastically increases the chances that a user will fall for the scam.

There have been documented cases where AI collects information from public profiles and crafts messages that mimic co-workers or banking institutions.

In some instances, the messages even include emotionally tailored language — urgency, trust, or a professional tone — to maximize click-through rates on malicious links.

1.4 Hallucinations (Generation of Incorrect or Nonexistent Content)

Trust in AI leads many users to assume that the content it produces is reliable, without verifying it. Generative models not only accidentally generate false information but can also be manipulated to produce misleading content for malicious purposes.

These so-called "AI hallucinations" become a risk when the generated content is used to make decisions without external validation.

A real-world case involved a generative model integrated into a legal department, offering assistance in drafting documents. In several instances, it generated references to non-existent laws and fictitious legal precedents. Because this wasn't detected in time, one of these documents was sent to a client, triggering a reputational crisis.

2. Adaptation and Mitigation Strategies

The increasing sophistication of AI-powered cyberattacks demands a deep transformation in how organizations design their security systems. It is no longer enough to react; it is essential to anticipate.The approach must shift from a static, defensive logic to a dynamic, integrated model, focused on intelligent detection and operational resilience.Below are the key priority measures:

2.1. Continuous Monitoring and AI Validation in Production

Many incidents don't stem from massive attacks but rather from unexpected behavior of models that were integrated without adequate oversight. Implementing continuous monitoring mechanisms allows organizations to detect deviations, anomalous outputs, and suspicious patterns. Key Actions: Implement alert systems for unexpected outputs. Regularly audit logs of model interactions. Use tools to detect prompt injections and manipulated outputs.

2.2. Segmentation of Responsibilities: Limiting Model Power

A critical strategy is to prevent AI models from having direct access to critical systems or functions that could cause irreversible impact (such as money transfers, case closures, sending sensitive data, etc.). Instead, an intermediate validation layer should be implemented to review or supervise any action suggested by the model.

Recommended Measures:

- Use gateway-based architectures or define explicit authorization rules.
- Apply the principle of least privilege: grant the model access only to the data or functions strictly necessary.
- Require human approval before executing any irreversible action.

2.3. Staff Training and a Culture of Verification

AI-driven cybersecurity cannot rely solely on automated systems. The ability of human personnel to interpret signals, identify manipulation, or question AI-generated results is key to reducing risk.

It is essential that development, product, and security teams understand the specific risks associated with the use of generative models. Internal users should also be trained on how to safely interact with AI-based tools.

Effective Measures:

- Conduct regular awareness programs on AI-based phishing, deepfakes, and semantic manipulation.
- Train staff to spot inconsistencies in emails, phone calls, or documents generated by AI.
- Run attack simulations to strengthen the response capacity of non-technical teams.

2.4. Using AI to Defend Against AI

Defense must also leverage Artificial Intelligence. Tasks such as detecting unusual patterns, identifying real-time threats, or analyzing generated content are areas where AI can provide significant advantages.

Notable Applications:

- MDR platforms (Managed Detection and Response) that combine automated analysis with human-led response.
- Dynamic authentication systems that adjust controls based on user behavior.
- Semantic analysis tools to detect prompt injection, conversational spam, or data leakage.

2.5. Redesigning the Secure Development Lifecycle

The use of AI requires that security principles be embedded from the design phase. This means evaluating which model to use, how to train it, how to limit its scope, and how to control it in real time. Security should not be treated as an add-on, but rather integrated from the very beginning of the model's lifecycle.ç

Recommended Best Practices:

- Define a specific threat model for each AI-based application.
- Include security reviews in every update of system prompts.
- Establish rollback processes and incident response protocols related to generative model failures.

Conclusions

Artificial Intelligence has introduced a new dimension to cybersecurity — one where language, context, and algorithmic autonomy become attack vectors. What were once predictable technical threats have evolved into emergent behaviors that are difficult to foresee.

As we’ve seen, attacks like prompt injection, automated phishing, and exploitation of hallucinations don’t require advanced technical skills — just a precise understanding of how AI systems operate.

In response, organizations must move beyond traditional perimeter-based defense paradigms and adopt an adaptive, continuous approach centered on intelligent oversight. The key is no longer just to prevent breaches, but to design systems that tolerate failure, detect deviations early, and respond with agility..

AI is not just a threat — it's also part of the solution. When used wisely, it can enhance threat detection, automate responses, and strengthen digital resilience.

The challenge is not to resist its use, but rather to understand its risks, mitigate its weaknesses, and harness its potential through a responsible approach.

The cybersecurity of the future won’t depend solely on technology, but on how we choose to integrate it into our processes, decision-making, and organizational culture.

.

Fernando Echevarria Gutierrez
Cybersecurity Analyst

The SOC Revolution Driven by Generative AI: A Leadership Perspective

Article by Javier Portabales Campos

In today’s cybersecurity landscape, Security Operations Centers (SOCs) are under unprecedented pressure. The growing sophistication of threats, the shortage of specialized talent, and the overload of alerts have turned security management into a strategic challenge.From my position as SOC Director at NTT DATA, I see how Generative AI (Gen AI) is emerging as a catalyst for profound transformation in our operations.

A New Operational Paradigm

Generative AI is not just a technological evolution — it represents a true paradigm shift. Unlike traditional AI, which focuses on classification and prediction, Gen AI creates new content, detects complex patterns, and enables natural language interactions.

Within the context of the SOC, this translates into unprecedented capabilities to: Automate repetitive tasks, reduce false positives and enhance threat intelligence.

At NTT DATA, we’ve begun integrating Generative AI into our operations with promising results, developing use cases such as: SASTIA, for code analysis and AI-driven triage and investigation during incident response.

In addition, the automation of data collection, predictive analysis, and the generation of executive-level reports are just some of the other areas where this technology is making a clear impact.

Strategic Capabilities and Technology Partnerships

Our approach is supported by a network of partnerships with major players in the industry. Within the SOC, our strategy is to combine technologies from our tech partners — such as Crowdstrike, Microsoft, Google, Swimlane, Palo Alto, Splunk, and TrendMicro — with our own AI assets, ensuring 100% alignment with privacy standards and the contractual obligations we've committed to. This collaborative structure allows us to scale quickly and adapt to client needs while maintaining high standards of quality and compliance. Specialization in AI usage is now mandatory for the entire SOC team this fiscal year. We also emphasize to our teams that AI makes them better professionals, enabling higher performance.

SOC Transformation: From Reactive to Proactive

Historically, SOCs have operated under reactive models, primarily focused on log monitoring and incident response.

However, technological evolution and environmental pressure have driven a shift toward proactive, semi-autonomous, and in some cases, fully autonomous models. Generative AI is enabling this transition by taking on tasks such as: Incident classification, Data correlation, Automated playbook generation or Reporting on new developments in Threat Hunting or Cyber Threat Intelligence processes.

This not only improves operational efficiency, but also frees analysts to focus on strategic decision-making and advanced threat management. We often use a simple but powerful analogy to explain this shift to our team: We can keep working the field with hoes, or we can switch to tractors and drones. Generative AI is precisely that generational leap — one that multiplies our productivity and operational effectiveness.

The Human Factor: Collaboration and New Skills

One of the most significant aspects of this transformation is its impact on talent. Generative AI does not replace analysts — it empowers them. The human–machine collaboration becomes the backbone of a modern SOC. Junior analysts can lean on tools like Microsoft Copilot to interpret malicious scripts. Senior analysts can use Gen AI to simulate scenarios and optimize responses. Additionally, a new key skill is emerging: prompt engineering. Knowing how to interact with AI models to get accurate results will become an essential capability in the coming years. At NTT DATA, we are developing dedicated training programs to prepare our teams for these new skills, and we are making them a strategic goal for our SOC staff.

Risks and Challenges in Adopting Gen AI

Implementing Generative AI is not without risks. These include:

- Data privacy concerns
- Biases within models
- Adversarial attacks
- Integration challenges with legacy platforms

SOCs must address these issues with a strong sense of responsibility and ethical oversight.

In our strategy, we follow frameworks such as the NIST AI Risk Management Framework (AI RMF) and evolving regulations like the EU AI Act, among others.Governance, transparency, and continuous auditing are fundamental pillars to ensure an ethical and secure adoption of Generative AI.

Implementation Strategy: The SMART Model

To guide our adoption of Generative AI, we have structured our roadmap around the SMART model:

- S (*Strategize*): We define high-value, low-risk use cases, such as optimizing incident response.
- M (*Map*): We assess our infrastructure and available models, ensuring alignment with our security policies.
- A (*Assess*): We monitor model performance, continuously adjusting to stay relevant against evolving threats.
- R (*Refine*): We promote ongoing team training through simulations and real-world scenarios.
- T (*Train*): We encourage collaboration between cybersecurity and AI experts to share knowledge and boost operational efficiency.

Generative AI Maturity Model in the SOC

In the maturity assessments we currently perform for our clients, we evaluate not only the overall state of their SOCs, but also their level of Generative AI adoption.Most organizations today are in a transitional phase between levels 1.0 and 2.0 of the maturity model, where they are exploring early practical applications of Gen AI.At this stage, the primary focus is on justifying the return on investment (ROI) of the initiatives already underway. As a result, operational efficiency and cost reduction become key performance indicators for progressing toward more advanced stages — and for making the process measurable at the executive level.In general, we distinguish the following maturity levels:

- **Gen AI 1.0 – Initial Automation:** Language models are integrated to perform basic tasks such as incident triage, log analysis, and threat detection.At this stage, AI is incorporated into existing platforms (e.g., SIEM, SOAR) to enhance operational efficiency and reduce human error.
- **Gen AI 2.0 – Specialization and Context:** AI models are specifically developed for cybersecurity, capable of processing multimodal data and delivering contextualized results.This stage introduces lightweight models (SLMs) and training programs to empower teams in effectively using these tools.
- **Gen AI 3.0 – Adaptive and Modular Intelligence:** AI enables real-time analysis during active attacks, is delivered as a service (AI-as-a-Service), and becomes self-improving.It integrates seamlessly across all SOC workflows, significantly enhancing the analyst experience.
- **Gen AI 4.0 – Advanced Reasoning and Governance:** Near-human reasoning capabilities are achieved through breakthroughs in Artificial General Intelligence (AGI).Robust governance and compliance frameworks are established, along with sustainable, energy-efficient infrastructures for training and operating models

Conclusion: Toward a Resilient and Adaptive SOC

The integration of Generative AI in SOCs is not a passing trend, but rather a necessary evolution. At NTT DATA's SOC, we are fully committed to leading this transformation — combining cutting-edge technology with highly skilled human talent.

The key lies in adopting Gen AI responsibly, aligning its implementation with strategic objectives, and ensuring transparent collaboration between SOC analysts and AI systems. We believe the future of the SOC will be hybrid, collaborative, and adaptive. Organizations that embrace this vision will be better equipped to face tomorrow's threats and build robust, efficient, and trustworthy cybersecurity frameworks

Javier Portabales Campos
Cybersecurity Director

The Dark Side of AI

Article by Eduardo Alves

We’ve reached a point in history where Artificial Intelligence, once a sci-fi daydream, is now shaping the real world at astonishing speed. It’s in our hospitals, our factories, our banks, our armies and quietly woven into the daily routines of millions of people. For many, it’s a lifeline; for others, it’s the future knocking at the door. But progress, as we know, has a habit of bringing shadows along for the ride. The same algorithms that save lives and streamline industries can also be bent towards goals far less noble. The deeper our dependence on automated systems, the more we have to face awkward and sometimes uncomfortable questions about ethics, security, and the sheer scale of influence we’ve handed over to machines.

This isn’t just an academic debate. We’re already seeing AI used to launch complex cyberattacks, flood social feeds with crafted lies, and subtly steer public opinion. The danger isn’t on the horizon, it’s here. And the choices we make now will decide which way the story goes.

Automated Cyber Threats – the new battlefield

Criminals aren’t wasting any time. With machine learning at their fingertips, they can:

- Pinpoint weaknesses faster than human teams can react.
- Cook up shape-shifting malware that slips past traditional defences.
- Write phishing messages so convincing they’d fool even the wary.

The result? Old-school cybersecurity tactics feel increasingly like fighting a wildfire with a bucket of water. Defensive teams are scrambling to match pace, deploying AI of their own to sniff out abnormal activity before it spirals.

And then there are GANs (Generative Adversarial Networks), which can produce malicious code dressed up to look like the real deal, tricking even the most advanced scanners.

Deepfakes and the war on truth

Deepfakes aren’t just a parlour trick anymore; they’re a weapon. Videos, voices, and images that look so authentic you’d swear they were genuine. They’ve been used to:

- Sway public opinion with fabricated “evidence”.
- Pull off financial scams by imitating high-profile executives.
- Tarnish reputations, spread chaos, and deepen social divides.

The real threat isn’t the technology itself; it’s what it does to trust. Once people start questioning every image and every clip they see, confidence in media, politics, and even personal communication begins to fray. Add in armies of AI-driven bots, and the ripple effect becomes a wave.

Surveillance, privacy, and quiet control

AI also fuels surveillance on a scale that would make Orwell raise an eyebrow. Facial recognition cameras in public spaces. Software that claims to read emotions. Location tracking is baked into apps. Biometric scans for everything from banking to boarding a plane.

The concerns are obvious:

- Privacy is shrinking to a faint memory.
- The risk of governments pushing beyond legitimate security needs.
- The rise of “social credit” systems that judge and reward behaviour.

Left unchecked, these tools could shift the balance of power from citizens to systems, and once that line is crossed, history tells us it’s rarely reversed.

The black box problem

Ask a developer how their AI made a decision, and, in many cases, they’ll shrug. It’s not laziness; it’s complexity. Many systems are so intricate that even their creators can’t trace every step from input to output.

That’s fine if the AI is recommending a film. It’s far more troubling if it’s:

- Deciding someone’s eligibility for a loan.
- Filtering job applicants.
- Making recommendations in a court case.

Bias in, bias out, and without transparency, bad outcomes can’t be challenged. Explainable AI is a step in the right direction, but right now, it’s more an aspiration than a norm.

Autonomous weapons – machines with a trigger

In military labs, another uncomfortable reality is unfolding. Lethal Autonomous Weapon Systems (LAWS) don’t wait for human orders; they identify and strike targets on their own.

That raises blunt, sobering questions:

- Who takes the blame when a machine makes the wrong call?
- How do you stop such technology from falling into the wrong hands?

Despite calls from scientists and human rights groups to pause or ban them, the race continues without a clear set of global rules. It's a game where the stakes couldn't be higher.

The race with no brakes

Competition pushes innovation, but without guardrails, it can also push recklessness. Countries and companies alike are chasing AI dominance, often sidestepping ethics for the sake of speed. Without global agreements, we risk:

- A "winner-takes-all" environment where safety plays second fiddle.
- A widening gap between tech-rich and tech-poor nations.
- AI developing goals that don't match human values or even threaten them.

The scariest part? This isn't just sci-fi speculation. The groundwork is being laid right now.

Jobs, identity, and the human factor

AI isn't just changing how we work it's changing what work is. Roles based on repetition, analysis, or predictable tasks are being automated at a steady clip. For some, it means opportunity. For others, it means the fear of being left behind.

That fear isn't abstract. Job loss brings financial stress, yes, but also a hit to mental health, confidence, and purpose. Work isn't just about pay; it's tied to identity.

If AI is going to replace certain tasks, we need to replace more than just the income. Training, reskilling, honest conversations about the role of automation these aren't luxuries, they're necessities. And no machine can offer the creativity, empathy, and critical thinking that keep workplaces human.

Towards a safer, saner AI future

Slowing down the risks doesn't mean slamming the brakes on progress. It means steering it. Some steps we could take now:

- International laws with teeth – global agreements like those for chemical or nuclear weapons.
- Transparent AI – systems that can explain their decisions in plain language.
- Ethics from the start – bring in sociologists, ethicists, and human rights experts during design.
- Digital literacy for everyone – so people can spot manipulation and hold systems accountable.
- Global cooperation – because rogue AI doesn't stop at a border.

Conclusion

Artificial Intelligence might be the most powerful tool we've ever built, but a tool can be used for good or for harm. Whether it becomes a force for progress or a threat to stability depends on the choices we make now, not later.

The algorithms are getting smarter every day. The question is whether we can be wise enough to keep up.

Eduardo Alves
Cybersecurity Project Manager

Quantum Machine Learning and RNA

Quantum Space by María Gutiérrez

Quantum computing is beginning to transform the world of machine learning. This emerging field, known as quantum machine learning, is already being explored by researchers. And while AI and quantum computing are individually revolutionizing multiple sectors, their integration promises to multiply their capabilities and trigger a new wave of innovation.

This integration involves using quantum algorithms to enhance the efficiency, speed, and capacity of machine learning models. Instead of processing massive data volumes using classical computers, quantum properties are leveraged to perform parallel calculations and handle far more complex and expansive data spaces.

Although still in its early stages, this synergy offers potential benefits in several key areas, including:

- Faster model training.
- Improved analysis in high-dimensional spaces.
- Reduced energy consumption due to shorter compute times.

This synergy is already a reality in several fields and has shown particular promise in addressing one of the biggest current challenges in molecular biology: understanding how RNA (ribonucleic acid) folds — a molecule that is crucial to life, whose three-dimensional shape determines many of its functions.

RNA is essential in cellular processes, acting as a genetic messenger, regulator, and even as a catalyst. However, RNA must fold into a very specific 3D structure to perform its function and predicting how a sequence of nucleotides folds is extremely complex. The number of possible configurations grows exponentially with sequence length, and despite advances in computational biology, classical methods have failed to simulate these processes with both accuracy and efficiency.

This is where quantum machine learning comes into play — through algorithms such as Variational Quantum Circuits or the Quantum Boltzmann Machine, which can represent the energy state of RNA configurations and simultaneously explore multiple solutions using quantum superposition.

Instead of simulating all possible structures one by one, these approaches leverage phenomena like entanglement and quantum parallelism to more efficiently identify the optimal structure.

These models are not only capable of predicting how an RNA chain folds, but also of:

- Simulating its dynamic behavior
- Evaluating the stability of each conformation
- Identifying potential therapeutic intervention points

Quantum machine learning is still in the exploratory phase and must overcome several challenges before reaching commercial application, such as:

- The lack of clear quantum advantage
- The need for large volumes of quantum data
- The absence of a robust, consistent technological infrastructure.

However, it is likely that within the next five years, the integration of quantum computing and machine learning will become a core component of advanced R&D, particularly in:

- Personalized medicine
- Gene therapies
- Synthetic biology

Its application in this last area — for analyzing RNA sequences and automating vaccine design — may accelerate pandemic response and drive innovation in complex generics and biosimilars, ultimately paving the way for ultra-selective genetic treatments in a new era of precision medicine. The first steps in this direction are already underway.

The Impact of AI on the Evolution of Social Engineering

Article by Sergio Sánchez Encabo

For decades, social engineering has been one of the most effective tactics for exploiting human vulnerabilities for malicious purposes. Unlike purely technical attacks, social engineering is rooted in manipulating emotions such as trust, urgency, or fear.Initially, attackers relied on phone calls, rudimentary emails, or face-to-face interactions to achieve their goals. However, technological evolution has significantly expanded both the reach and sophistication of these practices.

Today, artificial intelligence has redefined the limits of deception. Language generation tools such as ChatGPT, Claude, Gemini, or LLaMA, along with voice and video synthesis technologies like ElevenLabs, Descript, or Synthesia, are transforming social engineering into a process that is automated, scalable, and more convincing than ever.AI enables attackers to carry out complex campaigns at a lower cost, exponentially increasing the risk for users and organizations of all kinds.

Phishing: From Mass Attacks to Precise Personalization

Phishing has evolved from being a generic mass attack to becoming a fully personalized threat. With the help of language models, attackers can customize the tone, style, and content of a message based on data gathered from social networks, forums, or even previous data breaches.

Using tools like ChatGPT, Claude, or Mistral, attackers can generate automated message templates in seconds. These tools allow messages to be tailored to specific professional roles. Combined with campaign platforms such as Gophish or Evilginx2, attackers can launch targeted phishing campaigns using advanced impersonation and credential harvesting techniques.

This level of personalization makes malicious messages more believable — and therefore more dangerous.

Moreover, AI can automate not only the message content but also the follow-up, generating dynamic responses that simulate a real conversation, making fraud even harder to detect.

Vishing: Voice as an Attack Vector

Vishing, or voice phishing, has also been profoundly transformed by artificial intelligence. Using voice cloning and speech synthesis techniques, attackers can replicate real voices with a high degree of fidelity, making impersonation far more convincing and harder to detect than ever before.

Among the most commonly used tools are ElevenLabs, Respeecher, and iSpeech, which can clone a person's voice with just a few seconds of audio. These tools generate realistic voice files that can be used in automated calls or recordings designed to increase the credibility of a scam.

This is particularly relevant in scenarios where voice conveys authority, such as impersonating a superior or trusted figure.In corporate environments where remote communication is common, verifying the authenticity of a voice becomes more difficult — making vishing an ideal tool for scams that rely on speed and trust.

Deepfakes: The Appearance of Power

The creation of synthetic audiovisual content using AI has opened a new dimension in social engineering attacks.Tools like Synthesia, DeepBrain, HeyGen, and even Runway ML allow the generation of hyper-realistic videos in which a seemingly real person makes statements that never actually happened.

Because these tools are highly accessible and do not require advanced technical knowledge, the ability to create realistic deepfakes has been democratized.

When integrated into disinformation campaigns or corporate contexts, their impact potential increases dramatically.

The ability to distribute deepfakes via email, collaboration platforms, or private groups makes them a particularly powerful and hard-to-detect attack vector.

Fraudulent Chatbots and the Risks Associated with LLMs

One of the less visible effects of implementing language models is the emergence of fraudulent chatbots.These bots, integrated into fake websites or simulated customer support channels, rely on platforms such as LangChain, Botpress, Rasa, or OpenChatKit.They can also be deployed on low-cost servers using lightweight solutions like Docker or FastAPI.

Unlike old rule-based systems, these bots can respond contextually, convincingly, and naturally, engaging in credible conversations for minutes or even hours. This capability makes them highly effective tools for stealing sensitive information, whether through social engineering or simply by imitating legitimate technical support. Given the increasing adoption of LLMs, the OWASP Foundation has identified specific risks in its preliminary list: "OWASP Top 10 for LLMs". In the context of chatbots, some of the most prominent threats include:

- Prompt injection (manipulating instructions to alter the bot's behavior).
- Unintentional disclosure of sensitive information.
- Generation of false or manipulative content.

These risks turn poorly configured models into unintentional attack platforms, as they can be exploited to facilitate fraud or deliver incorrect information.

CEO Fraud: Simulated Authority, Real Damage

The so-called "CEO fraud" is one of the most sophisticated forms of modern social engineering. These attacks aim to trick trusted employees into performing critical actions under the false authority of an executive. Today, this type of fraud can combine multiple AI tools to increase credibility and impact:

- Text generators like ChatGPT or Claude simulate emails in the executive's tone and style.
- Voice generators such as Descript or Murf.AI produce convincing phone calls.
- Video tools like HeyGen create realistic recordings of urgent requests.

All of this can be automated and launched from controlled environments, often using C2 (Command & Control) frameworks or infected devices acting as intermediary nodes. The multi-channel combination — email, voice, and video — makes these attacks extremely difficult to detect without strong internal verification procedures in place.

Conclusion

The integration of artificial intelligence into social engineering tactics is not just a technological evolution — it's a radical shift in the very nature of attacks.Now, the credibility of deception no longer depends on individual human ingenuity, but on systems capable of generating synthetic text, voices, and images that simulate authenticity with astonishing precision.

In this new landscape, traditional trust signals — a superior's writing style, a familiar voice on a call, or even a personalized video — are no longer reliable indicators.

Cybersecurity must evolve, not only through technical solutions but also by revisiting internal processes, reshaping organizational culture, and prioritizing digital education.

Key measures to face this growing threat include:

- Two-step verification protocols
- Training based on realistic simulations
- The ethical implementation of AI tools.

Artificial intelligence has opened the door to countless opportunities, but it has also introduced unprecedented ethical and operational challenges.

Only through clear awareness of the problem and collective action can we prevent trust — that invisible yet essential component of digital life — from becoming its greatest vulnerability.

Sergio Sánchez Encabo
Cybersecurity Analyst

Current AI Trends: Challenges and Opportunities from a Cybersecurity Perspective

Trends by David Sandoval Rodríguez-Bermejo

When analyzing current trends in artificial intelligence, they can be grouped into three main categories: AI agents, model specialization, and the disruption brought by open models. To this analysis, we can add a fourth category focused on the applicability of AI within the context of cybersecurity.

Artificial intelligence agents represent a new stage in the evolution of this fast-paced and disruptive sector. These agents operate independently (and usually with a high degree of specialization), and are capable of coordinating with one another to solve much more complex problems.

If we draw a parallel with software development, the emergence of agents is quite similar to the advent of containers and orchestrators.

Thanks to this new way of consuming AI, it is now possible to unify automation, task atomization, and artificial intelligence to build a large ecosystem out of small components (agents).Many companies are taking the concept of agents a step further by incorporating low-code / no-code platforms that allow people with no programming experience to deploy AI-powered workflows with minimal effort. In the case of NTT DATA, the company has developed its own proprietary asset known as Axet Flows.

In the cybersecurity context, the use of agents is highly relevant. They can be leveraged from: an offensive standpoint (to orchestrate an audit or a penetration test), and a defensive standpoint (to automate numerous daily tasks such as playbook generation or ticket classification).

Regarding the development of language models, the current trend is toward model simplification through Mixture of Experts (MoE) architectures. In these solutions: Language models share a common base of weights. Specific regions are activated depending on the task to be performed. The idea behind this type of model is similar to the agent concept, as they involve predefined task-specific workflows. However, they differ in that MoE models are not modular — they are a single, unified weight package rather than a network of orchestrated modules.

Thanks to this new approach, more specialized models are being developed, with fewer parameters, making them lighter, more resource-efficient, and still able to deliver responses as effective as their much larger “parent” models. Additionally, in the case of quantized models — where model size is reduced at the expense of some accuracy — the speed gains are even more noticeable.

At the same time, the current landscape features a growing competition between proprietary commercial models (like OpenAI, Google, and Anthropic) and open-source models (such as DeepSeek, LLaMA, and Mixtral).The presence of so many influential players with diverse business models is fueling competition and, to a large extent, accelerating technological evolution and the progress of generative AI.A final point worth highlighting in the realm of model specialization concerns their capabilities: On one hand, there is a wave of multimodal models — designed to understand and respond across multiple output formats (text, audio, images, and video).On the other hand, some developers are focusing on models tailored to specific output types, such as voice cloning or the creation of avatars that mimic human behavior.

While these capabilities are extremely useful in fields like education and training, misuse can easily lead to fraud, such as vishing (voice phishing).To conclude, within cybersecurity, one of the most direct and impactful areas of synergy with artificial intelligence is in the auditing and code review process, particularly in SAST (Static Application Security Testing).

At NTT DATA Spain – IBIOL, the cybersecurity team has developed a specialized tool called SASTIA, designed to accelerate and optimize this type of work.

With the evolution of AI models, their ability to understand source code has steadily increased, reaching levels that, in some cases, rival those of expert human auditors.

A clear example of this capability is seen in CVE-2025-37899, a zero-day vulnerability in the Linux kernel that had remained hidden for several years — and was discovered thanks to the use of artificial intelligence.

Conclusion

We are living in an exciting time in the world of technology, characterized by the rapid integration and deep penetration of AI across the various services in companies' portfolios.AI adoption is no longer optional — it's a non-negotiable necessity for organizations aiming to remain competitive in such a challenging environment.

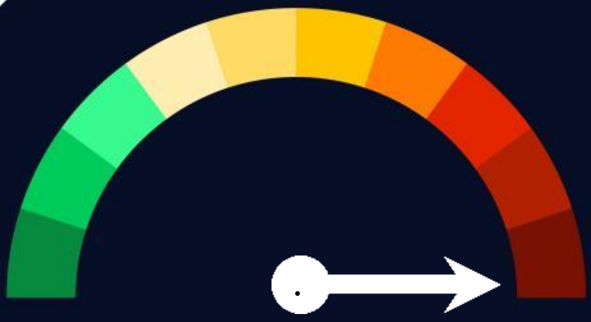
It is essential to approach this transformation with a responsible and fearless mindset, implementing the right controls and safeguards to ensure the protection of our digital assets.

David Sandoval Rodríguez-Bermejo
Cybersecurity Analyst

Vulnerabilities

Multiple Vulnerabilities in IBM Cloud Pak System

Date: July 28, 2025
CVE: CVE-2025-30065 and 3 more



CVSS: 10

CRITICAL

Description

IBM Cloud Pak System, with 2 critical vulnerabilities, 1 high-severity, and 1 medium-severity issue identified.
One of the critical vulnerabilities, CVE-2025-30065, is caused by the deserialization of untrusted data in the parquet-avro module of Apache Parquet (versions 1.15.0 and earlier), which could allow arbitrary code execution.

The other critical vulnerability, CVE-2025-3357, results from improper index validation in a dynamically allocated array, which could also enable an attacker to execute malicious code on the system.

Solution

The manufacturer recommends the following:

- For Intel: Update to IBM Cloud Pak System v2.3.6.0 with Foundation 2.1.28.1 and ITM 1.0.29.1 pTypes, available on IBM Fix Central.

- For Db2 pType: Download the fix for IBM Db2 11.5.9 Special Build 58840.

Affected products

Some of the affected products include:

- IBM Cloud Pak System 2.3.3.6, 2.3.3.6 iFix1, and 2.3.3.6 iFix2

- IBM Cloud Pak System 2.3.4.0, 2.3.4.1, and 2.3.4.1 iFix1

References

- www.ibm.com
- www.incibe.es

Vulnerabilities

Critical Vulnerability in a WordPress Plugin

Date: August 4, 2025
CVE: CVE-2025-5394

CVSS: 9.8

CRITICAL

Description

The CVE-2025-5394 vulnerability affecting the WordPress theme "Alone – Charity Multipurpose Non-profit" is currently being actively exploited.

The issue stems from the alone_import_pack_install_plugin() function, which lacks proper security validations and is exposed via the wp_ajax_nopriv hook.

This allows an unauthenticated attacker to send malicious AJAX requests to upload files and install plugins from external sources — potentially leading to remote code execution and full takeover of the site.

Solution

Developers recommend:

- Immediately update to version 7.8.5, released on June 16, 2025.

Affected products

This critical vulnerability affects:

- Alone – Charity Multipurpose Non-profit WordPress Theme (versions prior to 7.8.3).

References

- unaaldia.hispasec.com
- thehackernews.com

Android Fixes 6 Vulnerabilities in Its August Security Patch

Date: August 4, 2025
CVE: CVE-2025-48530 and 5 more

Critical

Description

Android has released its August security patch, addressing a total of 6 vulnerabilities. Among them are two critical severity vulnerabilities and four rated as high severity.

The critical vulnerability CVE-2025-48530 is located in the Android system and could allow an attacker to execute code remotely without requiring additional privileges or user interaction.

The second critical vulnerability, CVE-2025-21479, affects a closed-source Qualcomm component. At this time, none of the vulnerabilities included in the update are known to be actively exploited.

Solution

It is recommended to apply the security patches released by the manufacturer as soon as possible.

Affected products

The affected products in this update include:

- Android Open Source Project (AOSP): Versions 13, 14, 15, and 16
- Components from: Arm and Qualcomm

References

- source.android.com
- incibe.es

Dell Releases Security Patch for PowerProtect Data Domain Platform

Date: August 5, 2025
CVE: CVE-2025-36594 and 4 more

Critical

Description

Tech giant Dell has recently issued a security advisory for its PowerProtect Data Domain platform, informing customers of a critical vulnerability classified as Authentication Bypass.

This vulnerability, identified as CVE-2025-36594, allows a remote, unauthenticated attacker to bypass security mechanisms, potentially enabling them to: Create unauthorized accounts, expose sensitive data, compromise the integrity and availability of the system.

In addition to this critical issue, Dell also disclosed other vulnerabilities related to local privilege escalation, though these were classified with lower severity.

Solution

Dell recommends that customers update their software to the latest available versions, which are currently: 8.3.1.0 and 8.4.0.0.

Additionally, Dell advises customers to limit access to administrative portals to prevent exposure to untrusted or unsecured networks.

Affected products

The affected versions of Dell PowerProtect Data Domain are as follows:

- Data Domain OS Versions: 7.7.1.0 to 8.3.0.15
- DD OS LTS 2024: Versions 7.13.1.0 to 7.13.1.25
- DD OS LTS 2023: Versions 7.10.1.0 to 7.10.1.60

References

- dell.com
- secure-iss.com
- securityvulnerability.io

|

Eventos

ISACA Latin American Conference 2025

9 - 12 september

Organized by the ISACA Bogotá Chapter, this conference focuses on IT auditing, cybersecurity, and governance. Under the theme "How to Generate Value in the Age of Innovation and Digital Trust," it offers: Hands-on workshops, Keynote sessions and interactive panels with regional and international experts.

Link

DragonJAR Security Conference 2025

10 – 11 september

Recognized as Colombia's most important cybersecurity conference — and one of the most prominent Spanish-language events. It gathers experts, professionals, and enthusiasts to: Share knowledge, explore the latest trends and build strategic connections in the field of cybersecurity.

Link

Mind The Sec Brasil 2025

16 - 18 september

A leading cybersecurity event in Latin America, it brings together over 16,000 professionals and offers more than 200 hours of content. This event serves as a powerful platform to: Stay up to date with the latest cybersecurity trends and threats, connect with industry leaders and discover innovative solutions for protecting digital assets.

Link

Cyber Security & Cloud Expo Europe 2025

24 - 25 september

This event focuses on cybersecurity and cloud computing, covering key topics such as: Zero-day threat monitoring, Threat detection, Generative AI, Quantum computing, and much more. It brings together industry leaders to explore essential strategies and build valuable connections in the digital security landscape.

Link

Recursos

➢ DefAgent.io – Automated Pentesting for AI

DefAgent.io is a specialized platform for conducting automated penetration testing on artificial intelligence models and large language models (LLMs).It combines military-grade cybersecurity techniques with AI technologies to detect vulnerabilities such as: Adversarial logic manipulation, prompt-based data exfiltration. Its DefAgent Shield™ system provides military-grade monitoring and ensures compliance with frameworks like NIST AI RMF 1.0.

Link

➢ TU Latch – Dynamic Digital Access Control

TU Latch enables users and organizations to manage real-time authorizations for accounts and services. It acts as a "digital latch", enhancing protection against unauthorized access or cyberattacks — making it a key tool for identity and access management (IAM).

Link

➢ Aqua Security – Cloud-Native Security

Aqua Security provides a comprehensive platform to protect applications deployed in containers and cloud-native environments.It includes powerful tools such as:Trivy – for vulnerability scanningKube-hunter – for Kubernetes security testingTracee – for real-time monitoringThese tools help ensure regulatory compliance and enable efficient threat detection.

Link

➢ Olvid – Encrypted and Private Messaging

Olvid is an open-source encrypted instant messaging app that does not require a phone number or collect any personal data.Certified by France’s National Agency for the Security of Information Systems (ANSSI), it is recommended for secure communications at both personal and institutional levels.

Link

Subscribe to RADAR

Powered by the
cybersecurity
NTT DATA team

es.nttdata.com