CyberGrant
SECURITY MADE EASY

WHITE PAPER

AI & Cybersecurity

Index

Introduction

Artificial Intelligence (AI) is increasingly used in cybersecurity to enhance the protection of IT systems, networks, and data from cyberattacks. It automates threat detection, analyzes large data volumes, identifies patterns, and responds to security incidents in real time. However, it can also serve as a powerful weapon for cybercriminals, enabling increasingly sophisticated attacks.

Evolution of Threats

AI-powered attacks can bypass traditional security measures, automate malicious activities, and exploit vulnerabilities at scale. Below are the main threat types and their impact.

Deepfake

Deepfakes are among the most visually and psychologically impactful tools available to cybercriminals. They rely on advanced machine learning models such as Generative Adversarial Networks (GANs) to create hyper-realistic videos and audio clips capable of deceiving most people into believing they are genuine.

Cybercriminals use deepfakes for various illicit purposes:

- **Corporate espionage –** Deepfakes can impersonate executives or decision-makers, authorize fraudulent transactions, trick employees into revealing sensitive business data, or spread false information to manipulate stock prices.
- **Blackmail and extortion –** Cybercriminals can fabricate compromising or damaging videos of an individual and threaten to release them unless a ransom is paid.
- **Disinformation campaigns –** Deepfakes can be used to produce fake interviews or speeches by public figures, such as politicians or health officials, spreading misinformation and undermining trust in institutions.

Detecting deepfakes presents growing challenges since traditional forensic methods struggle to keep pace with the realism of modern synthetic media.

Detection tools must analyze micro-expressions, inconsistencies in lighting, and audiovisual mismatches—tasks that often require advanced AI capabilities themselves.

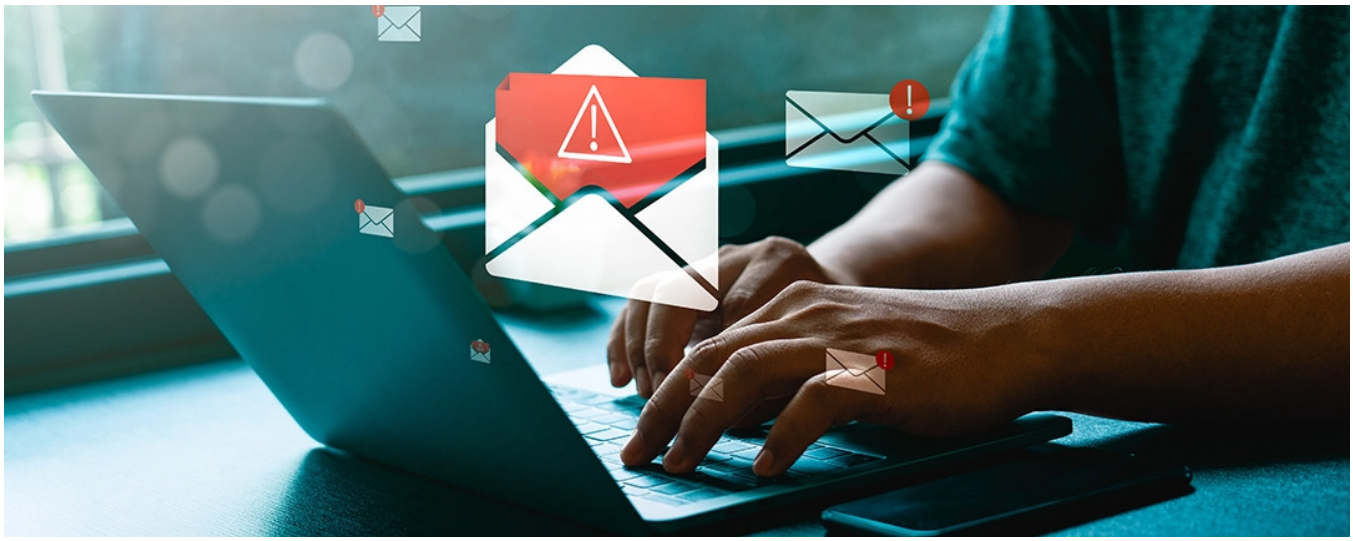
Phishing

> *"Phishing has evolved from simple email scams into highly sophisticated"*

Phishing has evolved from simple email scams into highly sophisticated, **AI-powered attacks that are difficult to detect and easy to fall for**. AI has made phishing scams far more effective due to the following factors:

- **Linguistic proficiency -** Large Language Models (LLMs), including ChatGPT, can generate grammatically flawless, contextually relevant messages tailored to specific corporate communication styles.
- **Dynamic content generation -** AI can personalize phishing emails in real time based on a target's online behavior, job role, or recent activity.
- **Multilingual threat -** AI can translate phishing content into multiple languages while maintaining cultural nuances, expanding the reach of cybercriminals globally.

Common AI-powered phishing examples include fake HR department emails requesting employees to update login credentials, or fraudulent invoices and payment requests that perfectly mimic legitimate supplier communications.



AI-Based Malware and RaaS (Ransomware-as-a-Service)

AI-driven malware represents a major evolution. Unlike traditional variants with static behavior, these a**dvanced programs can dynamically adapt to target environments, analyze security measures in real time, and adjust tactics to bypass defenses**. They continuously refine their strategies during execution, making them progressively harder to detect.

At the same time, ransomware has evolved into the Ransomware-as-a-Service (RaaS) model, democratizing access to criminal techniques. The integration of AI has amplified this threat by enabling intelligent auto-targeting, automated security evasion, contextual adaptation, and continuous learning—making attacks far more impactful for organizations.

Social engineering

Social engineering exploits human emotions and cognitive biases. With generative AI, cybercriminals can automate and personalize these manipulations at scale. Specifically, they can:

- **Build trust** - AI can simulate long-term conversations, gradually establishing trust with a target before initiating fraud.
- **Trigger fear and urgency** - AI-generated messages can induce panic (e.g., "Your account has been compromised!"), prompting victims to act rashly.
- **Exploit authority** - AI-generated communications can impersonate a person's superior, pressuring them to bypass security protocols.

Cybercriminals can also create fake social media profiles with AI-generated photos and backstories or deploy AI-powered bots that engage in conversations to gather intelligence and influence opinions.

Public AI Tools and the Risk of Data Exfiltration

The growing use of public AI tools—such as chatbots, virtual assistants, and platforms like ChatGPT—has introduced new and significant data security risks. When employees use public AI systems for work tasks, they may inadvertently share sensitive or confidential information. Once entered into these AI systems, such data can:

- Be used for model training
- Become accessible to other users through generated responses
- Be stored on provider servers, resulting in loss of corporate control over critical information

Data at risk includes source code, business strategies, financial data, customer records, and intellectual property. A particularly concerning factor is employee unawareness—many staff members, drawn by the convenience of these tools, do not realize they are exposing critical corporate assets.

Moreover, privacy policies of public AI services vary widely and often fail to ensure the necessary confidentiality protections.

> ***"Sensitive information may become accessible to other users through generated responses."***

How Cybersecurity Players Are Responding

31,70%

The projected CAGR

between 2025 and 2032

Fortune Business Insight

According to Fortune Business Insight, the global market for AI-based cybersecurity solutions was valued at **$26.55 billion in 2024** and is projected to **grow from $34.10 billion in 2025 to $234.64 billion by 2032**, registering a **CAGR of 31.70% over the forecast period**.

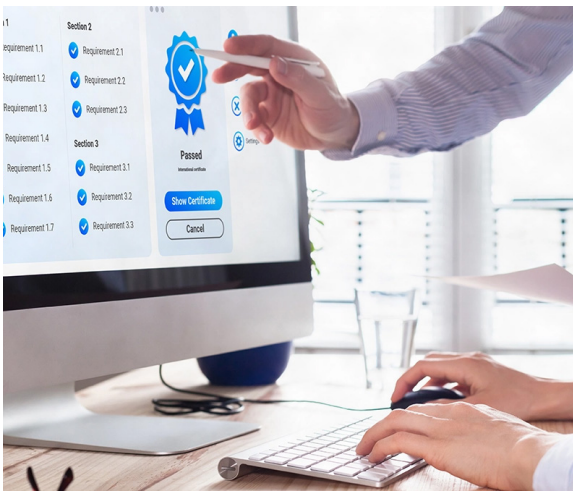
Leading companies in the field are heavily investing in the development of **AI-driven platforms** for threat detection, automated incident response, and predictive analytics. These platforms **integrate machine learning**, **natural language processing**, and **deep learning algorithms** to identify anomalous patterns, automate threat-hunting processes, and dramatically reduce response times to attacks.

However, the adoption of AI in cybersecurity presents several significant limitations. Machine learning models require massive amounts of training data, which can lead to **false positives** that overwhelm security teams.

Moreover, the reliance on historical datasets can make systems vulnerable to **zero-day or novel attacks**, while AI algorithms themselves can be **manipulated through adversarial machine learning** techniques. The “black box” nature of certain models also complicates **auditability and incident response**, as it is often difficult to explain how automated decisions are made.

From a compliance perspective, AI integration raises critical challenges in relation to existing regulatory frameworks, including:

- **GDPR** – Requires transparency in automated decision-making and guarantees the right to explainability in the processing of personal data.
- **NIS2** – Mandates proportional and verifiable risk management measures.
- **DORA** – In the financial sector, requires operational resilience testing, including validation of AI systems.



Given these considerations, organizations must balance technological innovation with obligations of **data minimization, privacy-by-design,** and **accountability**. They must also ensure accurate documentation of AI decision-making processes and maintain human oversight in critical operations.

Ultimately, the main challenge remains to **harmonize intelligent automation** with the principles of **transparency, responsibility,** and **control** required by the European regulatory framework.

Applications of AI in Cybersecurity

While generative AI is increasingly used as a weapon by attackers, it is equally leveraged by defenders to develop smarter, faster, and more adaptive cybersecurity systems.

These technologies deliver significant advantages by automating threat detection and enabling real-time responses to potential attacks. Below are the main areas of application.

Data Classification and Prevention of Exfiltration and Data Breaches

AI provides advanced tools for protecting corporate data through **automated and intelligent information classification**.

AI-driven systems can analyze and automatically categorize data based on sensitivity levels, assign appropriate access rights, and implement **dynamic controls** that adapt to context and user behavior.

These systems can detect anomalous access patterns, identify potential exfiltration attempts in real time, and **enforce granular security policies**. Additionally, continuous machine learning improves the precision of threat detection, significantly reducing the risk of data breaches and ensuring **regulatory compliance** through proactive and automated monitoring of information assets.

Threat Detection and Resolution

AI enables the following capabilities:

- **Code quality control** – AI can analyze source code to ensure security compliance and detect potential vulnerabilities early in the development cycle.
- **Vulnerability identification** – Machine learning algorithms can assess software and systems to identify weak points and prioritize remediation efforts.
- **Solution generation** – AI-based tools can suggest fixes or automatically apply patches to vulnerable systems.

Detection of Hacking and Malware Attacks

AI-powered systems can monitor network traffic and analyze behavioral patterns to detect hacking attempts and malware intrusions.

By leveraging machine learning, AI can identify anomalies and issue **real-time alerts**, enabling rapid intervention to mitigate potential threats.

Behavioral Monitoring and Anomaly Detection

AI excels at **behavioral analysis**, continuously learning normal user and system activity patterns to detect deviations that may indicate malicious behavior.

Intrusion Detection and Prevention

AI-based cybersecurity solutions include:

- **Network-based Intrusion Detection and Prevention (NIDP)** – Monitors network traffic to detect malicious activities.
- **Host-based Intrusion Detection and Prevention (HIDP)** – Analyzes activity on individual devices to identify potential threats.
- **Intrusion Detection and Prevention Systems (IDPS)** – Combine network- and host-based approaches to deliver comprehensive protection.

AI enhances intrusion detection and prevention by enabling:

- **Real-time monitoring** – Continuous surveillance of systems and networks.
- **Anomaly detection** – Identification of deviations from established behavioral baselines.
- **Automated response** – Immediate, automated countermeasures to contain threats.
- **Predictive analysis** – Forecasting and prioritization of potential threats based on historical data and current trends.

Real-Time Analysis and Incident Response

AI-powered tools deliver real-time security event analysis and automate incident triage processes. They play a crucial role in **incident response** by providing:

> *"AI is capable of analyzing security events in real time and automating incident triage processes."*

- **Early detection** – Identifying threats before they cause significant damage.
- **Rapid response** – Automating mitigation through predefined protocols.
- **Automated investigation** – Using threat intelligence to determine the scope and impact of an attack and generating detailed, customized forensic reports.
- **Behavioral analysis** – Evaluating user and system activity to identify potential risks.

Identifying the Source and Cause of Security Incident

Immediately after a security attack, AI can play a vital role in mitigating risk by ensuring rapid containment, analysis, and documentation across several key areas.

Protection of the Affected System

- **System isolation** – AI-driven systems can automatically detect compromised devices and isolate them from the network to prevent lateral movement of threats.
- **Automated threat containment** – Tools such as Endpoint Detection and Response (EDR) solutions powered by AI can automatically block suspicious activities.
- **Dynamic access control** – AI models can enforce adaptive authentication measures and lock affected accounts to prevent further compromise.

Incident Documentation

- **Automated logging** – AI can collect and organize logs from multiple sources in real time, ensuring that no critical information is lost.
- **Pattern recognition** – Machine learning algorithms can detect abnormal activity in logs, simplifying documentation and analysis.

Evidence Preservation

- **Data integrity checks –** AI tools can apply cryptographic hashing to files to ensure integrity and authenticity.
- **Automated backups** – During an incident, AI systems can create system and file snapshots for forensic analysis.

- **Chain-of-custody management** – AI helps maintain the authenticity and traceability of digital evidence.

Initial Incident Analysis

- **Incident triage** – AI systems can prioritize incidents based on severity, impact, and urgency.
- **Threat intelligence integration** – AI can correlate incidents with external threat databases to identify known attack patterns.

Data Recovery

- **Data reconstruction techniques** – AI can use predictive models to reconstruct lost or corrupted data.
- **File carving tools** – Machine learning algorithms can recognize and restore fragmented files from disk images.

Behavioral and Network Malware Analysis

- **AI-based sandboxes for controlled malware execution** – AI-powered sandboxes safely run malware in isolated virtual environments to observe behavior and assess impact. AI automates the analysis and identifies malicious activity patterns.
- **Monitoring malware-system interactions** – AI detects how malware interacts with the system, including file modifications, registry changes, network connections, and exploitation attempts.
- **Comparison with known threats and anomaly detection** – AI continuously compares malware behavior with known threats, using anomaly detection to identify new attack patterns.
- **AI analysis results** – Findings can be used to create malware signatures, develop countermeasures for future attacks, and enhance threat intelligence.
- **Identification of suspicious network patterns** – AI monitors and detects abnormal network activity that may indicate a breach.
- **Signature-based vs. behavioral analysis** – Signature-based analysis identifies known malware by matching hashes or code patterns, while AI-driven behavioral analysis detects new or unknown threats by monitoring process behavior.
- **Advantages of the combined approach** – Signature analysis ensures fast detection of known threats, while behavioral analysis provides proactive defense against emerging and fileless attacks.
- **Advanced correlation capabilities** – AI correlates behavior over time, improving detection of sophisticated threats such as fileless malware and Advanced Persistent Threats (APTs), reducing false positives and validating suspicious behavior against benign activities.

Timeline Reconstruction

- **Event correlation** – AI can automatically sequence events by analyzing logs, network traffic, and user activities.
- **Graphical representations** – AI tools can generate visual timelines illustrating the progression of a security breach.

User and System Analysis

- **User Behavior Analytics (UBA)** – AI systems can detect unusual user activity, such as accessing files outside regular working hours.
- **System state analysis** – Machine learning can identify system modifications, including registry changes and file alterations.

Reporting Results

- **Automated report generation** – AI tools can consolidate technical findings into reports tailored to both technical and non-technical audiences.
- **Customizable dashboards** – AI-driven platforms can build dashboards showing key metrics and incident status for stakeholders.
- **Support for legal and compliance reporting** – AI can generate customized reports aligned with regulatory requirements such as GDPR, NIS2, DORA, or other data protection laws.

On-Premise AI for Data Security

On-premise AI is becoming an increasingly strategic choice for organizations operating in highly sensitive sectors such as **finance, healthcare**, and **public administration**.

Deploying AI infrastructures and models within the corporate perimeter - rather than relying on external cloud providers - offers substantial advantages in terms of **control, security, and regulatory compliance**.

Given that AI now represents both one of the **greatest threats** and the **most effective defense** in today's cybersecurity landscape, identifying the most appropriate **deployment model** for organizational needs has become a critical priority.

Opportunities and Benefits of On-Premise Solutions

On-premise AI solutions primarily ensure **data sovereignty**, keeping sensitive information entirely within an environment controlled by the organization.

This approach drastically reduces exposure to cloud-based threats and minimizes dependence on third-party infrastructures.

Direct management of the infrastructure provides full visibility into system activities, with **traceable logs** that enhance security oversight and facilitate **forensic investigations** in case of an incident.

Additionally, organizations retain complete control over **security protocols**, enabling them to adapt measures in real time to evolving threats and align them with internal data policies.

The on-premise model also supports the core principles of the **CIA triad** (confidentiality, integrity, availability) and simplifies compliance with stringent frameworks such as **GDPR, HIPAA, NIS2, and DORA**, while keeping data and infrastructure under **direct organizational control**.

Limitations and Challenges of the On-Premise Approach

> *"On-premises AI involves implementation and maintenance costs that are considerably high."*

Despite its advantages, adopting on-premise AI poses significant challenges, particularly in terms of **implementation and maintenance costs**, which are considerably high.

It requires continuous investment in **hardware, updates, and specialized personnel training**.

Moreover, **multi-agent AI systems**, while highly efficient for complex tasks, introduce potential vulnerabilities such as **data breaches, prompt injection, and privacy risks**.

Scalability represents another key limitation.

While cloud solutions offer instant flexibility, on-premise expansion requires meticulous planning and upfront infrastructure investments.

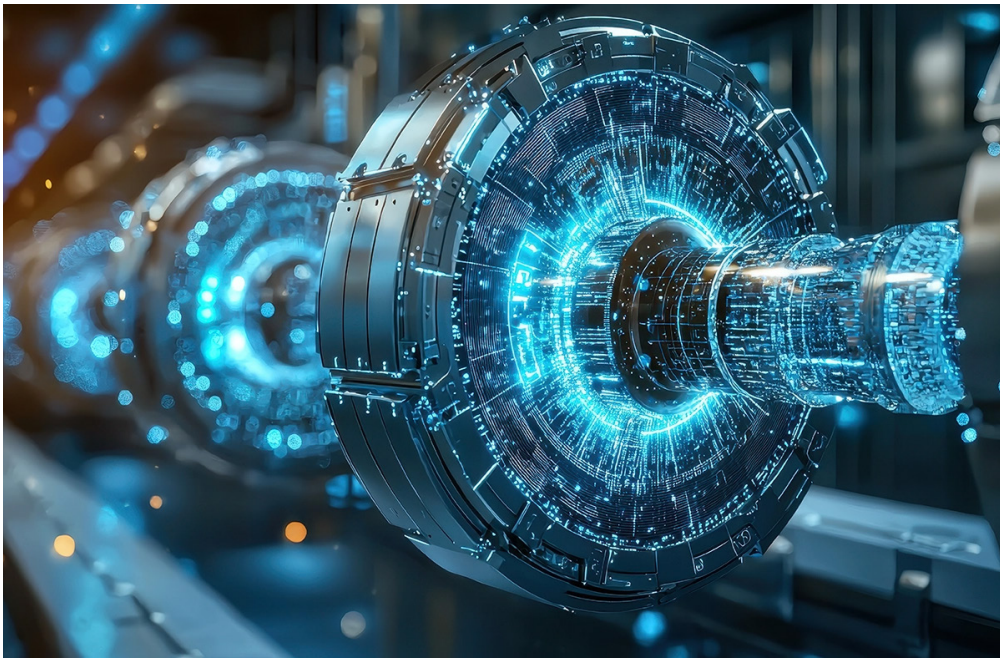
Organizations must also employ professionals capable of effectively managing AI implementation, security, and maintenance - skills that are often scarce and difficult to retain.

Furthermore, integrating I with **legacy systems** or **industry-specific software** can introduce additional complexity, often requiring extensive customization efforts.

The Threat of Quantum Computing

Quantum computing technology is advancing rapidly, and it has long been recognized that it could **break current encryption standards**, threatening the security and privacy of individuals, businesses, and entire nations.

This scenario makes it urgent to adopt **post-quantum cryptography strategies**, especially considering that the threat may already exist: cybercriminals could be **harvesting encrypted data today to decrypt it later** in what is known as a **"harvest now, decrypt later"** attack.



Consequently, organizations implementing on-premise AI must plan their transition toward **quantum-resistant algorithms**, integrating post-quantum standards into their security architectures to safeguard long-term data protection.

Open-Source vs. Proprietary Models in On-Premise AI for Cybersecurity

Choosing between **open-source** and **proprietary AI models** is a crucial strategic decision for organizations implementing on-premise AI for cybersecurity.

This choice has a deep impact on **security, regulatory compliance, operational costs**, and the organization's **ability to respond to emerging threats**.

Advantages of Open-Source Models

Open-source AI models offer complete **transparency**, allowing organizations to examine the source code, understand the algorithmic decision-making mechanisms, and identify potential vulnerabilities before deployment.

This transparency is especially valuable in contexts where **explainability** of automated decisions is a regulatory requirement - as mandated by **GDPR.**

The **collaborative nature** of open-source accelerates innovation: global developer communities continuously contribute improvements, identify bugs, and deliver security patches quickly.

Customization is another major advantage. Organizations can modify models to meet specific operational needs, integrating custom features and optimizing performance for their environment.

Additionally, the **absence of licensing costs** lowers barriers to entry, making AI accessible even to smaller organizations with limited budgets.

In an on-premise context, open source also eliminates vendor dependency, ensuring strategic autonomy and reducing the risk of technological lock-in.

Challenges and Risks of Open Source

However, open-source models also pose notable cybersecurity challenges.

Because their code is publicly available, **malicious actors** can study these models, identify weaknesses, and develop targeted **adversarial attacks.**

Data security may also be weaker compared to enterprise-grade proprietary solutions, as protection protocols might

not meet the strict standards required by regulations such as **NIS2, DORA**, or other sector-specific frameworks.

Security updates depend on community activity - less popular projects may receive **irregular maintenance**, leaving vulnerabilities unpatched for extended periods.

Furthermore, many organizations struggle with a **shortage of advanced technical expertise** required to manage the deployment, maintenance, and security of open-source models.

There is also the **supply chain risk** of compromised or malicious code being introduced into open-source components.

> *"In proprietary models, features are optimized for specific use cases, with user-friendly interfaces and comprehensive documentation."*

Advantages of Proprietary Models

Proprietary AI solutions, on the other hand, provide **enterprise-grade security guarantees**, with **tested and certified protocols** that meet international compliance standards.

Vendors typically offer **dedicated technical support, regular updates**, and **timely security patches**, reducing the operational burden on internal teams.

These models are often optimized for **specific use cases**, providing **user-friendly interfaces** and **comprehensive documentation** that streamline deployment.

Legal accountability is also clearly defined through **Service Level Agreements (SLAs)**, which specify protection clauses in case of malfunctions or data breaches.

In on-premise deployments, proprietary models ensure **tested reliability**, rigorous **quality assurance**, and **certified compatibility** with enterprise infrastructures.

Limitations of Proprietary Solutions

Despite these benefits, proprietary solutions come with **high costs**, as licensing fees can represent a significant expenditure—especially for large-scale on-premise implementations.

The **lack of transparency** makes it difficult to verify how decisions are made, complicating compliance with **explainability requirements** under European law.

Vendor lock-in is another strategic risk: dependence on a single provider limits flexibility and can expose organizations to issues if the vendor changes its business policies or discontinues support.

Additionally, **customization options** are often limited by predefined functionalities, making it difficult to adapt the system to highly specific operational or security needs.

Hybrid Strategies and Recommendations

Many organizations are adopting **hybrid approaches**, combining open-source models for non-critical components with proprietary solutions for **core security functions**.

This strategy balances **innovation and control** with **reliability and vendor support.**

For on-premise implementations, it is essential to evaluate:

- The **criticality** of the data being processed
- **Regulatory compliance requirements**
- **Available budget** for licensing and maintenance
- **Internal capacity** to manage and secure AI infrastructure

Organizations should also:

- Establish dedicated **security review teams** for open-source code
- Implement robust **vulnerability management** processes
- Maintain a continuous **update and patching roadmap** regardless of technology choice

Ultimately, the decision should align with the organization's **risk management strategy**, ensuring that **benefits outweigh risks** in an increasingly complex and evolving cyber threat landscape.

Best Practices for Implementing AI in the Enterprise: Data and Security by Design

The effective implementation of AI within an organization requires a **methodological approach** that places **security at the core** of the strategy from the very beginning of system design.

The **security-by-design** principle dictates that security controls must be integrated directly into the architecture of AI systems, not added later as an external layer.

Data Quality and Governance

Data is the foundation of any AI system. Its **quality, accuracy,** and **representativeness** directly determine model performance.

Organizations must establish **rigorous data governance processes** to ensure the verified provenance of datasets, minimize data collection in compliance with **GDPR principles,** and classify data appropriately based on sensitivity.

Data lineage must also be traceable to ensure auditability and regulatory compliance throughout the entire lifecycle of the AI

Principles of Security-by-Design and by-Default

A security-by-design approach requires:

- **Pre-deployment risk assessment** – Conducting risk analysis before implementation to identify potential vulnerabilities.
- **Data encryption** – Ensuring protection of data both at rest and in transit.
- **Granular access controls** – Implementing least-privilege policies to minimize exposure.
- **Environment segregation** – Maintaining strict separation between development, testing, and production environments.
- **Continuous monitoring** – Detecting anomalies and suspicious behavior in real time.

Additionally, **privacy-by-default** must ensure that only strictly necessary data is processed, reducing exposure and improving compliance posture.

Continuous Testing and Validation

Before release, AI systems must undergo:

- **Penetration testing** to identify exploitable vulnerabilities.
- **Adversarial robustness assessments** to evaluate resistance to AI-specific attacks.
- **Regulatory compliance verification** to ensure alignment with standards such as GDPR, NIS2, and DORA.

Furthermore, **staff training** on AI-related risks and the establishment of **ethical oversight committees** are essential to guarantee responsible and secure AI deployment.

Conclusion

AI is transforming the cybersecurity landscape by enabling **advanced threat detection, automated incident response**, and **cost-efficient risk management**.

As cyber threats continue to evolve, integrating AI into cybersecurity strategies will be **essential** for organizations seeking to protect their data, maintain operational continuity, and ensure compliance with increasingly stringent regulatory frameworks.

A cura di **Federica Maria Rita Livelli**

www.cybergrant.net