ENTHEC

WHITEPAPER

Autonomous Artificial Intelligences:

Their impact on cybersecurity

kartos

INDEX

INTRODUCTION

Over the past decade, organizations have relied on rule-based solutions, limited automation, and computer-aided analytics to defend increasingly complex infrastructures. However, the advent of **artificial intelligence, and especially autonomous intelligence (AAI)**, marks a historic turning point: we are moving from reactive tools dependent on continuous monitoring to entities capable of reasoning, learning, coordinating actions, making real-time decisions, and proactively executing defensive strategies.

Autonomous AIs are emerging thanks to advances in multimodal models, sequential planning, tool orchestration, real-time access to operational data, and continuous adaptation mechanisms. These capabilities allow an **autonomous agent** not only to detect anomalies but also to understand complex patterns, correlate distributed events, anticipate malicious behavior, deploy countermeasures, and learn from every interaction.

However, the same progress that enhances defense also fuels the sophistication of the adversary. Autonomous AIs can be used by malicious actors to launch coordinated intrusion campaigns, automatically exploit vulnerabilities, generate polymorphic malware, or execute complex operations at speeds incompatible with human capabilities. This scenario ushers in a **new paradigm of digital conflict between intelligent agents**, where the advantage will depend on the ability to securely deploy, monitor, and control resilient autonomous systems.

Given this scenario, questions about governance, security, transparency, and control become critical. How do we ensure that autonomous AI acts in accordance with security policies? How do we prevent the manipulation, hijacking, or exploitation of defensive agents? What safeguards are necessary to avoid erroneous or disproportionate decisions in systems that operate continuously? Furthermore, autonomy demands **new frameworks** for auditability, traceability, and validation of behavior, especially when these AIs are involved in protecting critical infrastructure, global value chains, or essential services.

This document aims to provide a clear and strategic overview of the foreseeable **impact of autonomous AI on cybersecurity**, analyzing both its transformative potential and the associated risks and responsibilities. In a scenario where both defenders and attackers will operate with autonomous intelligence, understanding this transformation is essential. The future of cybersecurity will depend on our collective ability to harness this technological power without losing control over it.

WHAT ARE AUTONOMOUS ARTIFICIAL INTELLIGENCES?

Autonomous artificial intelligence (AAI) represents an evolution of traditional AI systems. While most current AI models require constant human supervision for training, configuration, and application in specific contexts, AAI is designed to **operate independently**, with the ability to:

Learn and adapt from experience and the environment without continuous external intervention.

Make real-time decisions based on defined objectives and not solely on pre-programmed instructions.

Execute actions automatically, optimizing resources and correcting errors without waiting for supervision.

In essence, AAI systems are **self-managing systems** capable of establishing response strategies and coordinating multiple tasks in complex environments. This makes them a particularly relevant technology for scenarios where **speed of reaction and adaptability** are critical factors, as is the case in cybersecurity.

The emergence of AI is not an isolated phenomenon, but the result of a **cumulative technological trajectory:**

Classical AI (1950s–1990s):	Data-driven AI (2000–2010):	Large-scale generative AI (2018–2025):	Autonomous AI (present and future):
focused on expert systems, logical rules and deterministic algorithms.	machine learning, with models trained with large volumes of data for specific tasks.	with the foundational and language models, capable of generating text, images, code and performing complex reasoning.	systems capable of integrating several of these capabilities, managing themselves independently, and executing strategically valuable actions without relying on constant human control.

Nowadays, the convergence of cloud infrastructure, high-performance computing, distributed sensors, and advances in multimodal models has opened the door to systems capable of operating as autonomous agents. These agents not only respond to commands but can also **define intermediate goals, coordinate resources, and anticipate risks.**

Differences between Traditional AI and Autonomous AI

In the field of cybersecurity, this context translates into a paradigm shift: moving from reactive solutions to proactive and self-adjusting defenses, with the potential to neutralize threats before they escalate.

ASPECT	TRADITIONAL AI	AUTONOMOUS AI
Human dependence	It requires constant monitoring and manual configuration.	It operates with minimal supervision, making its own decisions.
Adaptability	Limited to scenarios foreseen in training.	It adapts to changing environments in real time.
Decision making	Based on learned patterns or predefined rules.	It integrates analysis, planning, and execution of actions in continuous cycles.
Error management	It needs human adjustments to correct errors.	It detects, corrects, and optimizes its performance autonomously.
Scalability	Difficult to scale without redesign or retraining.	It can be deployed in multiple environments by coordinating tasks in a distributed manner.
Application in security	Timely detection of anomalies or threats.	Dynamic, proactive and coordinated response to complex cyberattacks.

AAI does not replace traditional AI, but rather expands and surpasses it, offering a level of autonomy that is crucial for areas where **response time and resilience** are differentiating factors, such as the protection of critical digital infrastructures.

ADVANTAGES OF AUTONOMOUS AI

Autonomous Artificial Intelligence (AAI) not only represents a technological advancement but also a strategic opportunity for multiple sectors. Its ability to operate without constant supervision, adapt to dynamic environments, and execute proactive actions opens the door to an **intelligent and resilient automation** model, with a direct impact on security, productivity, and competitiveness.

ADVANCED PROCESS AUTOMATION

Unlike traditional automation, based on scripts or predefined workflows, AAI is capable of:

- **Orchestrating complex processes** in changing environments, adjusting the sequence of tasks according to the circumstances.
- **Reduce human intervention in repetitive activities,** minimize errors, and free up talent for functions of greater strategic value.
- **Manage incidents and exceptions** autonomously, without the need for explicit rules for each possible scenario.

In cybersecurity, this translates into the ability to monitor networks, detect anomalies, mitigate attacks, and comprehensively document incidents without relying on continuous human analyst intervention.

REAL-TIME DECISION MAKING

The distinguishing feature of AAI is its ability to process information, assess risks, and act immediately:

- **Reaction speed:** they allow you to respond in milliseconds to cyberattacks, which far exceeds human capabilities.
- **Contextual evaluation:** they do not limit themselves to applying static patterns, but consider multiple variables of the environment before acting.
- **Adaptive resilience:** if a strategy fails, the AAI can automatically readjust its course of action, maintaining operational continuity.

This approach makes AAI key allies for critical infrastructure, where a delayed decision can mean the disruption of essential services.

APPLICATIONS IN KEY SECTORS

The advantages of AAI are not limited to the field of cybersecurity. Their impact spans multiple strategic sectors:

- **Industry:** supply chain optimization, predictive maintenance of machinery and reduction of downtime.
- **Health:** real-time assisted diagnosis, autonomous management of clinical data and support in high-precision robotic surgeries.
- **Defense and national security:** deployment of autonomous systems for surveillance, cyber defense and operational intelligence analysis.
- **Finance:** automatic fraud detection, dynamic risk analysis, and investment portfolio optimization.
- **Transportation and logistics:** autonomous fleet management, optimized routes and immediate response to disruptions in the logistics flow.

In all these sectors, AAI represents a qualitative leap towards **safer, more agile and profitable operations.**

ECONOMIC AND EFFICIENCY IMPACT

The deployment of AAI has a direct effect on economic and efficiency indicators:

- **Reduction of operating costs:** by minimizing human errors and automating high-volume tasks.
- **Increased productivity:** thanks to the parallel and continuous execution of processes, without limitations of time or fatigue.
- **Business scalability:** allows for large-scale expansion of operations without the need to multiply the workforce.
- **New business models:** drive the creation of services based on autonomy (e.g., cybersecurity-as-a-service with real-time automatic response).

In macroeconomic terms, the adoption of AAI will accelerate the competitiveness of countries and organizations that implement them, creating a gap with those that do not adopt them.

EMERGING THREATS TO CYBERSECURITY

The same potential that makes autonomous artificial intelligence (AAI) a powerful tool for defense can also be exploited by malicious actors. AAI introduces a new level of **sophistication, speed, and autonomy** to attacks, creating a radically different threat landscape than what we have seen until now.

AUTONOMOUS CYBERATTACKS: SPEED, SCALE, AND ADAPTABILITY

Autonomous cyberattacks are one of the biggest emerging risks. A malicious AAI can:

- **Execute attacks in milliseconds,** surpassing any human response capacity.
- **Automatically scale operations** by launching thousands of intrusion attempts in parallel against multiple targets.
- **Adapt tactics in real time,** modifying attack patterns according to the defenses encountered.

This means moving from specific and predefined attacks to dynamic, **self-managing and self-adjusting campaigns**, with a massive disruptive potential for organizations and states.

AUTOMATED SOCIAL ENGINEERING AND INTELLIGENT IMPERSONATION

AAI also enhances **social engineering**, one of the most effective vectors in cybersecurity:

- **Creation of personalized messages** based on real-time analysis of digital profiles.
- **Convincing identity theft,** with synthetic voices and images indistinguishable from real ones.
- **Sustained interactions** in chats, calls or emails that simulate human behavior with great credibility.

This increases the risk of **advanced phishing, fraud, and psychological manipulation**, even among users with a high level of security training.

GENERATION OF DISINFORMATION AND MANIPULATION ON A LARGE SCALE

AAIs are capable of producing and disseminating false or manipulated content in a massive and coordinated manner:

- **Automated production** of falsified texts, images, audios and videos (deepfakes).
- **Management of armies of autonomous bots,** capable of propagating specific narratives on social networks.
- **Dynamic message adjustment,** adapting them in real time to audiences and cultural contexts.

Disinformation on a large scale can destabilize societies, erode trust in institutions, and directly affect national and economic security.

RISKS TO CRITICAL INFRASTRUCTURE AND CONNECTED SYSTEMS

AAI also increases threats to **critical infrastructure** and industrial systems (ICS/SCADA):

- **Coordinated attacks on energy, transport or communications systems,** capable of causing massive disruptions.
- **Autonomous vulnerability exploration** in highly connected operating environments.
- **Covert persistence,** with malicious agents learning and adapting to remain invisible in critical systems.

The potential impact ranges from power outages or disruptions to health services to the paralysis of international supply chains.

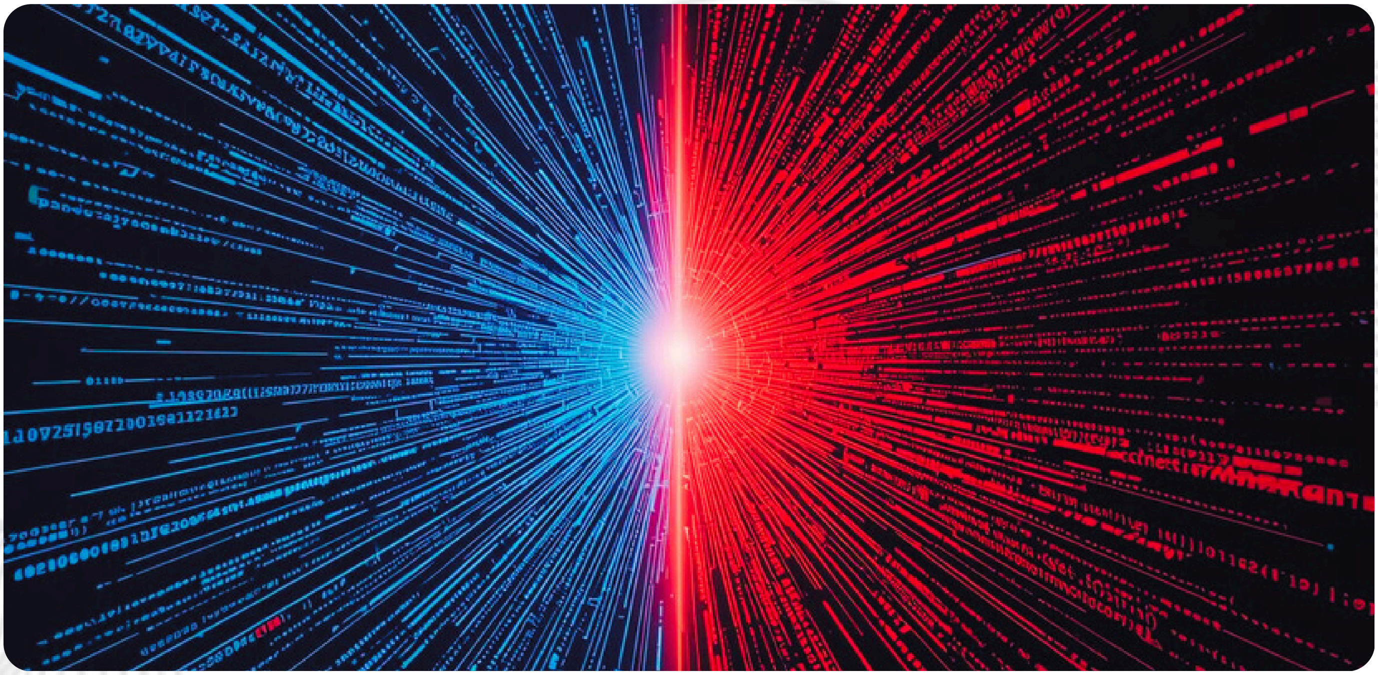
THREATS TO PERSONAL PRIVACY AND CORPORATE ESPIONAGE

Finally, AAIs introduce significant risks in the areas of privacy and espionage:

- **Massive and autonomous collection of personal data,** combining multiple digital sources to create complete profiles.
- **Exploitation of vulnerabilities in private communications,** including encryption weaknesses or human error.
- **Advanced corporate espionage,** where autonomous agents access, classify, and extract sensitive information without being detected.

This type of threat can compromise both individual security and the competitive advantage of organizations and states.

Taken together, malicious AAIs represent a qualitative leap in the cyber threat: it is not just about greater technical capability, but about offensive intelligence and autonomy, which challenges current defense models and demands new protection strategies.



DEFENSE TECHNOLOGIES AND STRATEGIES

The emergence of **Autonomous Artificial Intelligence (AAI)** as a potential threat has spurred the development of **new adaptive defenses**. Current strategies combine artificial intelligence, advanced security frameworks, and governance approaches to maintain trust in increasingly automated environments.

AI-based cyber defense: autonomous detection and response

Cyber defense systems based on AAI represent a paradigm shift compared to traditional security:

Proactive detection

Real-time analysis of traffic, logs, and anomalous behavior.

Autonomous response

Immediate implementation of containment measures (isolate nodes, close access, activate backups).

Continuous learning

Adapting defensive strategies as attackers' tactics change.

> *<u>Practical example:</u> A corporate network detects suspicious activity on a critical server. The AAI not only generates the alert, but also isolates the server, deploys a cloud replica, and maintains operational continuity in a matter of seconds.*

Explainable AI models and algorithmic governance

For AI to be trustworthy, it must be **transparent and auditable**. This is where the importance of **explainable AI (XAI)** and algorithmic governance comes in.

Explainability

Systems must justify their decisions in a way that is understandable to humans, especially in critical incidents.

Human supervision

Although AAIs operate autonomously, there must be a validation and control mechanism.

Algorithmic governance

Establishment of policies that regulate the training, deployment and supervision of autonomous models.

This not only reinforces **user confidence**, but also facilitates **compliance with legal and ethical regulations.**

Zero Trust security applied to AI environments

The **Zero Trust** principle ("never trust, always verify") takes on a critical dimension in environments with AAI:

Continuous authentication and authorization

Even among autonomous agents, each transaction must be verified in real time.

Microsegmentation

Limit the scope of access, so that a possible autonomous intrusion does not compromise the entire network.

Constant monitoring

Apply real-time analytics to validate behaviors and detect deviations.

> *Practical example:* *In a company with multiple AAIs collaborating on data management, an agent can only access the resources strictly necessary for its task, drastically reducing the attack surface.*

Autonomous red teams and AI threat simulation

Organizations are starting to employ **autonomous red teams**: AAI that simulate sophisticated attacks to test defenses under real-world conditions.

Penetration testing automation

The systems simulate advanced large-scale attacks.

Adaptive evolution

Red teams learn from defenses and adapt their tactics in real time.

Defensive reinforcement

It allows organizations to identify vulnerabilities before real attackers detect them.

> *Practical example: A bank deploys an autonomous red team that simulates fraud attempts on its digital infrastructure. The agents adapt their tactics based on the defensive responses, allowing them to fine-tune security before an actual intrusion.*

Emerging regulations and regulatory frameworks

The advancement of AAI requires **robust regulatory and normative frameworks** that guarantee responsible use:

European Union

The AI Act establishes risk categories and transparency requirements for autonomous systems.

USA

NIST initiatives on responsible AI and autonomous cybersecurity.

International scope

Need for multilateral treaties to prevent the offensive use of AAI in cyber conflicts.

In addition to laws, **industrial standards** (ISO/IEC, ENISA, IEEE) are being promoted that encourage good practices in security, auditing and traceability of algorithms.

ETHICAL AND REGULATORY CHALLENGES

The adoption of **autonomous artificial intelligence (AAI)** in cybersecurity opens up unprecedented opportunities, but also raises complex ethical and regulatory dilemmas. These challenges affect not only the technological sphere, but also governance, international security, and public trust.

Emerging regulations and regulatory frameworks

One of the main challenges is **determining who assumes responsibility** when an AAI makes a decision that has negative consequences.

- If an autonomous system mistakenly blocks a critical service, is the developer, the service provider, or the entity that deployed it responsible?
- The lack of a clear framework can create legal loopholes and slow down the adoption of these technologies.

> *<u>Study hypothesis:</u> If an AAI in a Security Operations Center decides to automatically isolate a critical server due to a false alarm, causing a disruption of public services, responsibility may fall on multiple actors, underscoring the need for clear rules of attribution of responsibility.*

Transparency and traceability of AI actions

Explainability and traceability are key requirements to ensure trust in autonomous systems.

- Many AAIs operate as "black boxes," making it difficult to audit how they arrived at a decision.
- Without traceability, it is impossible to verify whether an action was legitimate or the product of external manipulation.

> *<u>Study hypothesis:</u> In the financial sector, an AAI that blocks transactions because they are considered suspicious must generate clear and auditable records, so that both clients and regulators can review the decisions.*

Risk of escalation in automated cyber conflicts

The use of AAI in offensive operations poses risks of unintended escalation in digital conflicts.

- An autonomous attack that responds automatically to a perceived threat could trigger disproportionate retaliation.
- The absence of human oversight increases the risk of "algorithm wars" breaking out, in which systems operate without coordination or adequate political control.

> *Study hypothesis:* *A defense AAI that detects a massive cyberattack on a power grid could, without human intervention, launch an automated counterattack against the attacker's infrastructure. This scenario could escalate a regional conflict into an international incident.*

RECOMMENDATIONS FOR COMPANIES AND PUBLIC ENTITIES

The integration of **autonomous artificial intelligence (AAI)** into cybersecurity environments requires not only technological investment, but also strategic, organizational, and cultural adjustments. This chapter offers a set of practical recommendations to guide public and private organizations toward safe and effective adoption.

Risk assessment in AI environments

- **Mapping emerging risks:** Before deploying AAI, organizations should identify potential attack scenarios in which these technologies could be exploited.
- **Technological dependency analysis:** assess the degree of exposure to failure or compromise of the autonomous system.
- **Simulation of adverse scenarios:** conducting red teaming exercises to measure the organization's resilience against AI-enhanced threats.

Strengthening SOC capabilities with AI

- **Integrating AI into the SOC:** Security operations centers must evolve into cognitive SOCs, where AI processes large volumes of alerts in real time.
- **Intelligent incident prioritization:** Autonomous systems can reduce alert overload and focus the human team on critical threats.
- **Human-machine collaboration:** AI should act as strategic assistants, not as complete replacements, to enhance human decision-making.

Training and awareness on new autonomous threats

- **Specialized training programs:** both technical and non-technical staff must understand what autonomous threats are and how to recognize them.
- **Awareness of advanced social engineering:** preparing users to detect AI-generated attacks, such as hyper-personalized emails or deepfakes in voice calls.

- **Continuous training:** training programs must be updated in parallel with the evolution of the offensive capabilities of the AAI.

Internal policies for the safe adoption of AI

- **Internal governance framework:** define clear policies on which processes can be automated and under what supervisory controls.
- **Principle of minimum autonomy:** delegate critical functions gradually, with levels of autonomy staggered according to risk.
- **Audit and traceability:** maintain verifiable records of decisions made by the AAI for regulatory and compliance purposes.
- **Ethics and transparency:** ensuring that the adoption of AAI respects the principles of privacy, non-discrimination and protection of rights.

CONCLUSIONS

The emergence of **autonomous artificial intelligence (AAI)** represents a turning point in the field of cybersecurity. Throughout this document, it has been shown that these technologies are a **double-edged sword**: on the one hand, they offer unprecedented capabilities for digital defense, early threat detection, and automated responses; on the other, they introduce **emerging risks** ranging from autonomous cyberattacks to unresolved ethical and regulatory dilemmas.

From a **technical** standpoint, intelligence-aided attacks (AAIs) stand out for their potential to improve operational efficiency, reduce response times, and anticipate threats in highly complex environments. However, they can also become tools that can be exploited to generate disinformation, launch massive attacks, or compromise critical infrastructure, with a speed difficult to counter using traditional human means.

From a **strategic** perspective, the adoption of these technologies requires that both companies and public entities make progress on three priority fronts:

1. **Governance and responsibility:** defining clear frameworks for decision attribution and accountability.
2. **Transparency and explainability:** ensuring that AAI actions are auditable and understandable.
3. **International cooperation and regulations:** establishing common standards that balance innovation and security.

Looking to the **immediate future**, an AAI scenario is envisioned in which AAI will consolidate its position as an indispensable ally in digital defense. However, its success will depend on the ability of the stakeholders involved—the public sector, the private sector, international organizations, and civil society—to design a responsible adoption strategy that maximizes benefits and minimizes risks.

Ultimately, AAI should not be seen merely as technological tool, but as **catalysts for a structural shift** in how we understand cybersecurity. Their impact will transcend the operational level, becoming a key element of **digital trust, geopolitical stability, and the resilience of our societies.**

kartos©

Kartos Corporate Threat Watchbots: Continuous Threat Exposure Management (CTEM)

Automated, continuous, real-time monitoring of the organization's threat exposure, focused on cybersecurity and business criteria.

EXTERNAL ATTACK SURFACE

Location of the Company's open and exposed information and vulnerabilities on the Internet, the Deep Web, Dark Web and Social Networks: Phishing, fraud and scam campaigns; CVEs; DNS health; leaked passwords and credentials; leaked and exposed documentation and databases.

DIGITAL RISK PROTECTION

Detection of contextual information about potential attackers, their tactics and processes for carrying out malicious activities. Elimination of malicious activities on behalf of the Company. Brand, domain and subdomain protection. Corporate email protection. Ransomware protection. Web security and threat removal.

THIRD PARTY RISK

Real-time monitoring of third-party risk. Objective data on ongoing threats related to the value chain. Comprehensive view of any organization's cybersecurity maturity using a non-intrusive, external approach. Extension and weighting of information provided by traditional third-party risk assessment methods.

COMPLIANCE

Monitoring of corporate and third-party legal compliance based on objective data taken in real time.
ISO 27001. PCI - DSS. ENS. RGPD.
Justification of compliance with legal and regulatory requirements for associations, mergers and acquisitions, audits, certifications and contracts with the administration

CYBERSECURITY SCORING

It allows security information to leave the CISO's office and be presented in a simple way to people who need to be involved in security management without having a technical background.
Own and third-party cybersecurity scoring for partnerships, audits, mergers, acquisitions and government contracts.

Analysis of 11 Vectors

- **Algorithms**
- **Certificates**
- **Services / IoT / VoIP**
- **DNS Health/Phishing**
- **Patch Management**
- **IP Reputation**
- **Web Security**
- **Email Security**
- **Document Leaks**
- **Credential Leaks**
- **Social Networks**

WHITEPAPER

kartos©

AI layer that enables 100% automated operation without human intervention anywhere in the process.

Strictly non-intrusive tool. The research is carried out on the Internet, the Deep Web, and the DarkWeb, and the organizations' IT perimeter is not attacked, so its operation and the information obtained strictly comply with the imposed limits. by legislation.

The only platform that analyzes **conversations on social networks from the threat and attack detection perspective,** beyond the relating to reputation and branding.

Continuous operation 365x24x7, allowing detection of leaks of new information practically in real time. real time.

Maximum ease of use. Does not require no complex configuration. Simply enter the domain into the platform and it works autonomously, without the need to configure search parameters or any other information location criteria.

Automated, objective and continuous monitoring of the risks caused by **third parties belonging to the External Attack Surface of the organization.**

Learn more about our licenses
Try our tool for free
Start using Kartos

→ hello@enthec.com

Enthec Solutions is a Spanish technology company that develops cybersecurity software for the protection of organizations and people. Enthec Solutions has established itself as one of the Deep Tech companies with the most innovative and effective Cyber surveillance solutions thanks to the success of its **Kartos Corporate Threat Watchbots** platform, which provides organizations with Cyber Security, Cyber Intelligence, Cyberscoring, Compliance and Third-Party Risk Management Capabilities, and its innovative **Qondar Personal Threat Watchbots** platform for the individual online protection of the organization's relevant people.

www.enthec.com