WHITE PAPER

Artificial Intelligence as the Key to Cyber Resilience: Secure Integration, Protection Against Attacks and Intelligent Defence

Contributing authors/ companies

Ralf Benzmüller
Executive Speaker Security Labs G Data CyberDefense

Lisa Fröhlich
Corporate Communications Link11 GmbH

Christof Klaus
Director Global Network Defense Myra Security

Olaf Pursche
Pursche Tactical Consulting,
Head of the Security Competence Group at eco – Association of the Internet Industry

Cornelia Schildt
Senior Project Manager IT Security eco – Association of the Internet Industry

Dr. Or Sela
Channel Account Manager F5

Maurice Striek
Senior Consultant Cyber Strategy & Architecture NVISO Security

Marcel Rieger
Co-Founder Jamorie Consulting

Table of contents

Foreword 04

I. Introduction 05

II. Secure implementation of AI tools into corporate structures 06

III. Defence against AI-based attacks on companies 10

IV. AI-supported defence against cyber attacks 11

V. Conclusions and outlook 12

Foreword

Artificial intelligence has long been a business reality. Companies are no longer faced with the question of whether to use AI, but rather with the task of integrating it securely and effectively into their structures. However, the rapid technological advances in recent years have both produced innovative solutions for companies and created new attack vectors for cybercriminals.

Current research findings clearly confirm this development. According to recent studies, the quantity and quality of AI-generated misinformation has increased dramatically. Renowned security experts, including the independent research institute AV-TEST, have documented a concerning increase in deceptively realistic deepfakes and tailor-made phishing campaigns created using advanced AI models.

Also noteworthy is the emerging phenomenon of AI poisoning, in which actors attempt to influence training data or model parameters in order to manipulate the functioning of AI systems. While this type of attack on training data is currently less prevalent, occurring primarily in specific contexts, it nevertheless merits attention as a potential future challenge. AI-assisted deception and manipulation can target business processes and employees, resulting in both financial losses and reputational damage.

This white paper takes a practical approach and is aimed at decision-makers and IT security managers who fare tasked with using AI both as a tool for operational optimisation and as a defence mechanism against increasingly sophisticated threats. Based on the experience of leading experts and successful implementations in various industries, it offers concrete recommendations for action.

The aim is to provide practical knowledge to help companies harness the transformative potential of AI while minimising the associated risks, from the secure integration of AI solutions into existing corporate structures to the active defence against AI-supported cyberattacks. Particular attention is paid to regulatory requirements, such as the EU AI Act, and to ensuring compliance in a constantly evolving legal environment.

The insights gathered here underscore that the successful use of AI requires technological expertise and a deep understanding of security aspects, as well as strategic embedding in existing business processes. The case studies presented demonstrate how companies can use AI both as a driver of innovation and an effective shield against cyber threats.

As AI-powered attacks increase in complexity and frequency, this white paper provides valuable guidance for companies seeking to strengthen their digital resilience while capitalising on the opportunities presented by the AI revolution. The following pages serve as a resource for AI transformation – with the aim of promoting innovation while ensuring a high level of security.

Olaf Pursche
Pursche Tactical Consulting,
Head of the Security Competence Group
at eco – Association of the Internet Industry

I. Introduction

Artificial intelligence (AI) is now an integral part of many business processes. The key question is no longer whether companies should use AI, but how it can be integrated safely and responsibly. As its use increases, so do the threats: deepfakes, AI-supported phishing, automated malware and manipulation of training data are just a few examples.

This white paper provides an overview of the key challenges and demonstrates how organisations can implement AI successfully while establishing protective mechanisms against new forms of attack.

II. Secure implementation of AI tools into corporate structures

The introduction of AI changes processes, roles, and decision-making paths. A structured approach is needed to ensure that its use is secure and successful.

Learning models and trends

AI systems are based on various learning approaches – from supervised and unsupervised learning to reinforcement and transfer learning. The current focus is particularly on generative models that produce text, images or code, as well as highly specialised pattern recognition and expert systems.

Areas of application

AI offers potential in many areas: from marketing and customer service to medical diagnostics, cybersecurity and autonomous driving.

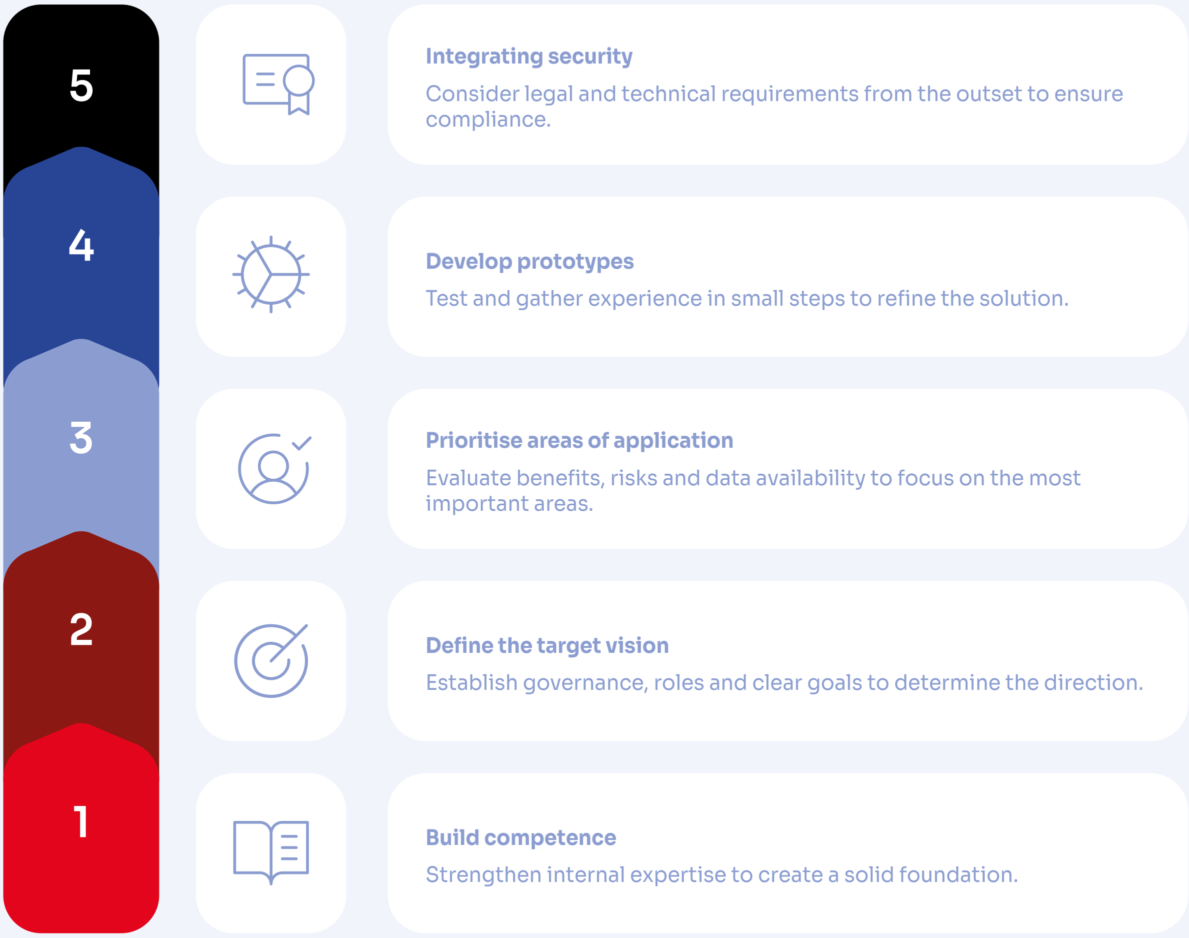
Challenges

Companies must take data protection, reliability, liability and acceptance into account. Striking a balance between the speed of innovation and security is crucial.

Procedural model

The following process model shows a structured and practical approach to introducing AI in companies. Five consecutive steps illustrate how organisations can systematically build up expertise, develop clear objectives and identify suitable areas of application. Complemented by iterative prototype development and the early integration of security and compliance aspects, the model offers compact guidance for responsible and successful AI implementation.

Implementing a practical approach



5 **Integrating security**
Consider legal and technical requirements from the outset to ensure compliance.

4 **Develop prototypes**
Test and gather experience in small steps to refine the solution.

3 **Prioritise areas of application**
Evaluate benefits, risks and data availability to focus on the most important areas.

2 **Define the target vision**
Establish governance, roles and clear goals to determine the direction.

1 **Build competence**
Strengthen internal expertise to create a solid foundation.

Best Practices

The use of AI tools opens up a wide range of opportunities for companies, but also brings with it new requirements in terms of security, transparency and governance. Clearly defined best practices are needed to integrate AI into existing corporate structures in a responsible and sustainable manner. The following graphic provides a concise overview of key areas of action that are essential for the secure implementation of AI tools.

KI-Best Practices

Orientation towards standards

Orientate yourself towards standards such as ISO/IEC 42001:2023. This ensures a basis for the responsible implementation of AI.

Risk classification

Classify risks based on purpose and criticality.
This helps to effectively prioritise mitigation measures.

Transparency

This builds trust and facilitates audits.
Ensure transparency through explainable AI and documentation.

Technical measures

Implement technical measures such as encryption and access control.
Monitoring is also important for security.

Organisational measures

Establish clear responsibilities and governance structures.
Training is also important for employees.

A look at practical applications

The benefits are particularly evident in data-intensive areas: in medical imaging, AI supports the detection of minute anomalies in MRI data, thereby improving diagnostic quality. Personalised content on digital platforms and fraud detection in large data streams also benefit from pattern recognition and, where appropriate, generative processes. In addition, the fields of application range from security analyses and autonomous driving to warehousing and gaming, where AI optimises processes and automates complex decisions. These examples illustrate how combined learning approaches (supervised/unsupervised/transfer) create added value in real-world environments.

Legal framework for the use of artificial intelligence: AI Act – European AI regulation

The EU AI Act, also known as the Artificial Intelligence Act, is an EU regulation that aims to regulate artificial intelligence (AI). It seeks to create a uniform legal framework for trustworthy and safe AI systems. The AI Act contains a number of key points, including the definition of prohibited AI systems, the classification of risks and the establishment of implementation rules.

Definition of AI according to the AI Act

An AI system is a machine-based system designed to operate with varying degrees of autonomy, capable of demonstrating adaptability after its initial deployment, and which derives explicit or implicit objectives from the inputs it receives, such as generating results such as predictions, content, recommendations or decisions that may influence the physical or virtual environment.

Transparency obligations

Providers of AI systems must inform users about their interaction with the AI. In particular, interaction with AI-generated content, such as deep fakes, must be clearly labelled. The EU is developing codes of conduct and testing procedures to facilitate the labelling of artificially generated content.

Risk classification of AI systems

The AI Act divides the risk classification of AI systems into four risk categories:

- Unacceptable risk: Prohibited AI systems (e.g., social scoring, real-time biometric identification with few exceptions)
- High risk: Strict testing is required, for example for systems used in law enforcement, critical infrastructure or medicine
- Low risk: Transparency requirements for users, e.g., for automated recommendations
- Minimal risk: No specific regulations, e.g. AI-supported video games

Requirements for High-risk AI

- Establishment of a risk management system
- Quality requirements for training data to avoid bias
- Technical documentation for assessing system conformity
- Transparent and comprehensible instructions for use
- Human surveillance with emergency shutdown

Conformity assessment prior to market access

- Internal control: manufacturers check compliance themselves
- Notified bodies: state-authorised testing centres for biometric applications and other high-risk systems

Areas of application

The Regulation applies only to areas covered by EU law (i.e. not to areas falling within the competence of Member States or areas of national security).

AI systems used esclusively for military or defence purposes are excluded from the Regulation. The same applies to AI systems that organizations develop and put into service exclusively for research and development purposes or that they use for research, testing or development activities before placing them on the market or putting them into service. Persons who use AI systems for non-commercial purposes are also exempt.

In the event of a breach of the sanction provisions, the authorities may impose a fine, the amount of which is based on the global annual turnover of the company concerned and may not exceed a maximum of 7 %. A graduated assessment basis is permitted for SMEs and start-ups.

III. Defence against AI-based attacks on companies

Threat scenario

Generative models lower the barriers to entry for attackers and accelerate known tactics: highly personalised phishing, deceptively real deepfakes in image and sound, automatically generated or disguised malware, and rapidly adapted exploits following the publication of new vulnerabilities. This increases the speed, precision and scale of campaigns, while putting pressure on traditional defence mechanisms. Documented incidents involving deepfake video conferences and fake approvals show how easily established communication and approval processes can be exploited.

Detection and analysis

Technically, the identification of AI-supported attacks differs only partially from classic detection, but requires more refined telemetry and stronger context evaluation. In human-centred scenarios (phishing, deepfakes), awareness and tools for detecting synthetic content are key. In applications and web infrastructures, continuous, granular monitoring of protocols and parameters is needed to separate bot behaviour from legitimate sessions; for encrypted traffic, reputation, geo-signals and heuristics provide additional help.

Signatures reach their limits when it comes to new types of malware – signatureless, behaviour-based methods and ML-supported classifiers are gaining in importance.

Defence principles

A purely signature-based defence is no longer sufficient. Baselines for normal system and user behaviour, anomaly detection (e.g. UBA/UEBA approaches) and telemetry that makes deviations visible at an early stage are necessary. In addition, response processes should take AI-specific vectors into account (prompt manipulation, model abuse) and combine technical and organisational measures.

Protective measures and counterstrategies

Priority should be given to strong identities (MFA, including password-less methods where appropriate), end-to-end encryption with controlled inspection, strict access controls, and logging and analysis of all interactions. AI-specific controls address prompt injection, model manipulation and risky outputs; traffic management helps mitigate volumetric attacks without disrupting legitimate traffic. Exercises (tabletop), penetration tests including AI-supported tools and realistic phishing simulations harden processes and teams.

Role of humans and AI

AI-assisted detection (IDS/IPS with ML) identifies deviations in real time and reduces the workload on teams – yet experienced analysts remain indispensable: LLMs and ML accelerate routine tasks, but can hallucinate or misjudge context; critical decisions require human review and clear responsibilities.

Lessons learned from incidents

Deepfake-based instructions and approved payments show that technical controls must work in conjunction with process guidelines – e.g. binding callback channels, dual control principle, limit policies for transactions and identity verification outside the communication environment currently in use. This allows deceptively genuine but implausible requests to be reliably filtered out.

A look at practice

Several documented incidents show how deepfakes in video calls or audio messages imitate executives in a deceptively authentic manner in order to obtain approvals or payments. Such attacks circumvent familiar communication channels and underscore the need for combined measures: technical detection of synthetic content, strong identity checks and procedural guidelines such as callbacks via known channels and the dual control principle. At the same time, signature-based methods quickly reach their limits when it comes to novel, automatically modified malware; behaviour-based, signature-less approaches and ML-supported classification increase detection quality in this area.

IV. AI-supported defence against cyberattacks

Role of AI in defence.

AI is not only a tool for attackers, but also a central building block of modern cyber defence. It processes enormous amounts of data, recognises patterns that would remain hidden to human analysts, and can identify threats in real time.

Detection and prediction

AI models analyse log data, network flows and user behaviour to detect deviations from established normal values at an early stage. On this basis, predictive methods enable anticipatory protection by deriving probable attack paths and vulnerabilities from historical patterns. Threat hunting is also gaining momentum because the system automatically evaluates, contextualises and prioritises information from external sources (e.g., forums, social media or underground marketplaces).

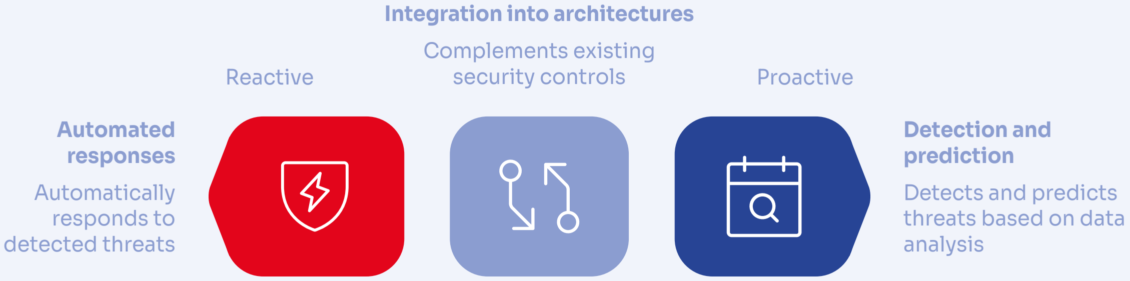
Automated responses

In SOAR environments, AI links detection events to predefined playbooks: suspicious activities are automatically throttled or blocked, compromised systems are isolated, and policies are dynamically adjusted. Continuous feedback loops with human-in-the-loop ensure quality, reduce false alarms and improve models during operation.

Integration into existing security architectures.

AI-based functions complement established controls such as firewalls, EDR/XDR, API security or DDoS protection. Clear responsibilities, standardised telemetry interfaces and complete logging are prerequisites. In large-scale infrastructures, GPU/DPU-supported pipelines ensure the necessary performance for analysis and inference without slowing down core processes.

AI security features range from reactive to proactive defence


Integration into architectures
Complements existing security controls
Reactive
Proactive
Automated responses
Automatically responds to detected threats
Detection and prediction
Detects and predicts threats based on data analysis

A look at practical applications

In productive environments, AI-supported systems analyse network flows in real time, detect DDoS patterns early on and initiate automated countermeasures – such as dynamic throttling or isolating suspicious sources. For API and application protection, AI-enabled WAF/IDS/IPS functions are combined with user/entity behaviour analyses to identify deviant behaviour more quickly. Where encrypted traffic limits visibility, controlled TLS inspection and orchestration mechanisms ensure that anomalies are visible despite end-to-end encryption. To ensure performance and latency, modern architectures also rely on DPU/GPU-accelerated paths that offload cryptography, packet processing and telemetry, further reducing detection and response times.

V. Conclusions and outlook

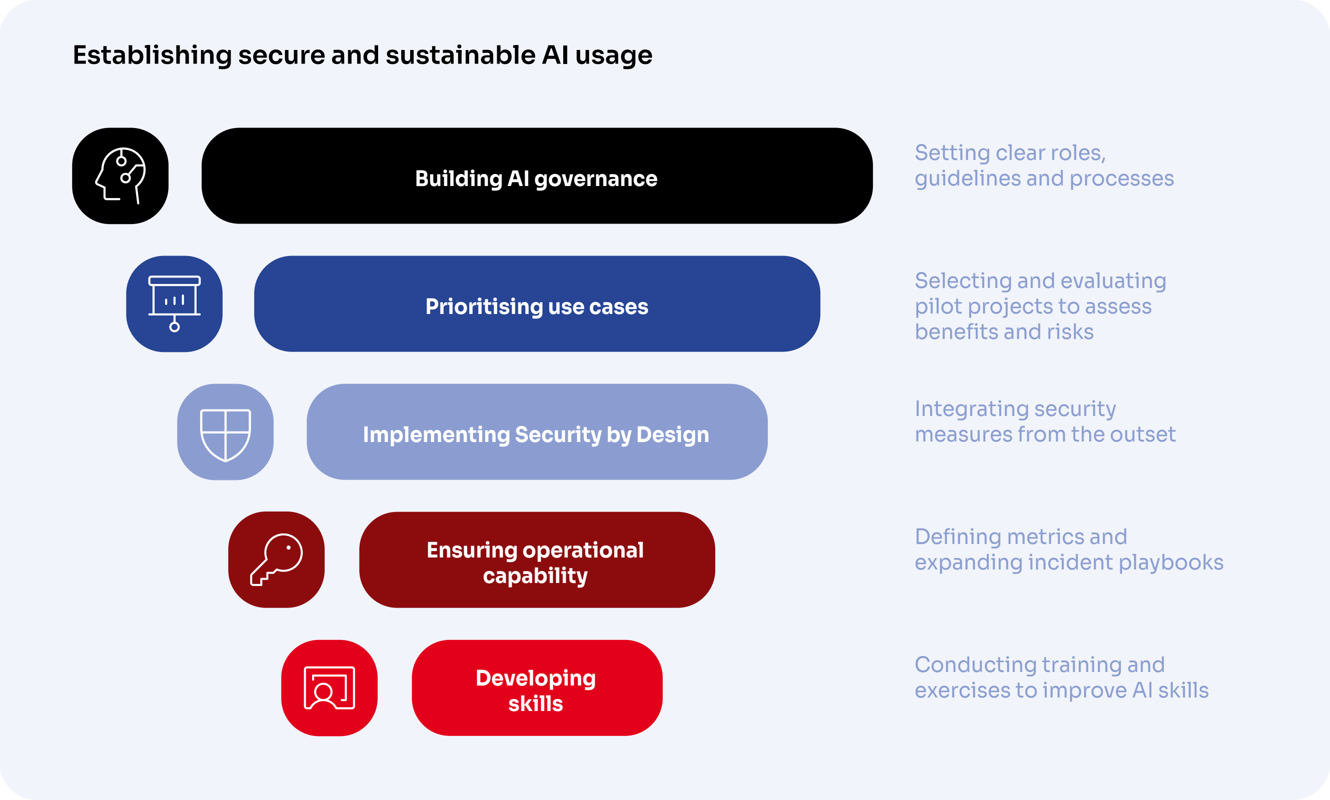
AI is both a value driver and a target for attack. Success comes to those who integrate governance, security, transparency and compliance from the outset and make operations measurable (metrics, SLOs, drift controls, incident playbooks) – in short, those who consistently document and improve small, low-risk steps.

Next steps

To ensure the secure and sustainable use of AI, companies should first establish robust AI governance. This includes clear roles, guidelines, approval processes and a sound risk and impact assessment. Building on this, it is important to prioritise use cases: pilot projects help to transparently assess the benefits, risks and data availability, as well as gathering initial experience in a controlled manner.

Security by design should be embedded from the outset, incorporating end-to-end access controls, encryption, seamless logging, clear output and prompt security mechanisms, and drift monitoring for the models. Operability must also be ensured by defining relevant metrics and service level objectives, expanding incident playbooks to include AI-specific scenarios and consistently recording telemetry data.

Finally, companies must develop competencies, including raising awareness among all employees, providing in-depth training for the development, operations and legal departments, and conducting regular exercises that incorporate AI-based attack scenarios.


Establishing secure and sustainable AI usage
Building AI governance
Setting clear roles, guidelines and processes
Prioritising use cases
Selecting and evaluating pilot projects to assess benefits and risks
Implementing Security by Design
Integrating security measures from the outset
Ensuring operational capability
Defining metrics and expanding incident playbooks
Developing skills
Conducting training and exercises to improve AI skills

Legal notice

eco – Association of the Internet Industry
(eco – Verband der Internetwirtschaft e. V.)

Lichtstr. 43h
50825 Cologne
Germany

Tel.: +49 221 70 00 48-0
Email: info@eco.de

international.eco.de

Contact persons

Cornelia Schildt
Senior Project Manager IT Security

Email: cornelia.schildt@eco.de

Olaf Pursche
Head of the Security Competence Group

Email: sicherheit@eco.de

Authors of the study

Ralf Benzmüller, Lisa Fröhlich, Christof Klaus,
Olaf Pursche, Cornelia Schildt, Dr. Or Sela,
Maurice Striek, Marcel Rieger

Disclaimer

This publication is for general information purposes only and does not constitute legal or other professional advice. It cannot and should not replace individual advice that takes into account the specific circumstances of each case. The contents have been compiled with due care; however, no guarantee is given for the accuracy, completeness and timeliness of the information.

Liability for damages arising from the use of this publication is excluded, regardless of the legal basis, to the extent permitted by law.