MRIS.bd.1 Mythos-Resistant Hardening of the ISMS

MRIS.bd: User-Defined Layer — Mythos-Resistant Information Security

Version 1.2 | Date: 2026-08-14 | Author: Richard Peddi

1 Description

1.1 Introduction

Agentic frontier AI models ("Mythos class") can find vulnerabilities autonomously, produce exploits in hours instead of days, and execute complex attacks largely on their own. Publicly documented threat intelligence reports confirm real campaigns in which 80 to 90 % of the tactical attack steps were carried out autonomously. Since mid-2026, this capability class has been broadly available across vendors.

For an established ISMS, this does not break the methodology, but it destabilises its inputs: the probability, timing, and effectiveness assumptions of many implemented security safeguards are no longer reliable. In particular, safeguards whose protective effect relies on the limited patience and capacity of human attackers ("friction") lose effectiveness, while safeguards with a hard structural barrier (cryptography, least privilege, immutability, verifiable separation) remain robust.

This module operationalises the thirteen Mythos Hardening Controls (MHC) of the Mythos-Resistant Information Security (MRIS) framework as requirements in the IT-Grundschutz format. It is based on the systematic effectiveness assessment of all 93 controls of ISO/IEC 27002:2022 against the AI-accelerated threat landscape, as well as the convergence analysis with BSI C5:2026, NIST, DORA, the CRA, and NIS2 including its implementing regulation (see [MRIS]). Implementation is described in the accompanying implementation guidance.

1.2 Objective

The objective of this module is to harden an existing ISMS against AI-accelerated attacks: the effectiveness assumptions of the implemented security safeguards are realigned with the changed threat landscape, and targeted hardening requirements close the gaps that technology-agnostic standards such as ISO/IEC 27001 do not explicitly address. The module thereby establishes conformity with ISO/IEC 27001 and the IT-Grundschutz standards and complements them with effectiveness-oriented requirements.

1.3 Scope and Modelling

Module MRIS.bd.1 Mythos-Resistant Hardening of the ISMS must be applied once to the entire information domain.

To create an IT-Grundschutz model for a specific information domain, the entirety of all modules must be considered as a matter of principle. This module presupposes an established ISMS and complements in particular the modules of the ISMS, ORP, OPS, DER, and APP layers. It does not replace any of these modules.

This module covers

- the hardening of authentication, workload identities, data backup, and attack detection against AI-accelerated attack patterns,

- transparency and protection of the software supply chain, including automated security testing, and
- the governance of AI agents used in production, continuous effectiveness verification, and automated incident response.

The following aspects are also relevant, but are covered elsewhere:

- Establishment and operation of the security process and risk management (see ISMS.1 Security Management)
- Basic data backup, network architecture, patch and change management (see the CON, OPS, and NET layers)
- Basic handling of security incidents (see DER.2.1 Security Incident Handling)

This module does not cover

- the security of the organisation's own AI products towards end customers (product security),
- a complete risk management or management system pursuant to ISO/IEC 27001 clauses 4 to 10, and
- legal compliance evidence towards supervisory authorities or certification bodies.

2 Threat Landscape

Since IT-Grundschutz modules cannot address individual information domains, typical scenarios are used to describe the threat landscape. The following specific threats and vulnerabilities are of particular importance for module MRIS.bd.1 Mythos-Resistant Hardening of the ISMS.

2.1 Collapse of the Window Between Patch and Exploit

Under AI conditions, the time between the publication of a patch and the availability of a working exploit is measured in hours rather than days or weeks.

Agentic AI models master the reverse engineering of patches as a mechanical analysis task. If an organisation works with staggered patch cycles and manual approval processes, then internet-exposed systems may already be compromised before the regular patch process even begins. The effectiveness assumption of any vulnerability management that relies on a multi-day reaction window is thereby broken. This is aggravated by the disclosure waves of AI-assisted vulnerability research setting in since 2026, which significantly increase the volume of new vulnerabilities while central vulnerability databases are structurally overloaded.

2.2 Compression of the Response Timeline

AI-driven attacks progress faster than human decision chains can react.

In documented campaigns, 80 to 90 % of the tactical attack steps were executed autonomously, at request rates unattainable for human operators. If a human has to decide at every point of the immediate response chain, then the organisation's reaction time is no longer in the same order of magnitude as the attack speed. Detection and response safeguards thus lose effect not because of insufficient quality, but because their latency is structurally too high. An attack can be completed before the first triage decision has been made.

2.3 Fragmentation of Attacks into Inconspicuous Micro-Steps

Agentic attackers decompose an overall attack into individual actions, each of which appears to be a legitimate technical request.

Vulnerability scans, credential validation, lateral movement, and data extraction are executed as separate, individually inconspicuous micro-steps. If attack detection is based on the assessment of discrete individual events — signature-based detection without behavioural context, static thresholds, alerting on single actions — then the attack remains invisible even though every single safeguard works as designed. The risk only materialises in the aggregation of events. Organisations without behaviour-based, cross-event correlation detect such attacks only by the damage caused.

2.4 Decoupling of Attacker Capability and Actor Type

The capacity barrier between opportunistic attackers and highly professional attacker groups has largely disappeared.

Classic probability assessment rests on the assumption that individualised campaigns at nation-state level require scarce specialist capacity. Agentic AI models, however, make this capacity available even to actors with little technical competence of their own. If an organisation continues to base its risk analysis on the scarcity of capable attackers, then it systematically underestimates the likelihood of targeted attacks. The probability axis of every risk matrix shifts upwards without any new threat actors emerging.

2.5 Unregulated Use of Privileged AI Agents

AI agents used in production with far-reaching privileges form a new, often unmanaged attack surface within the organisation.

Coding agents, agentic workflows, and their components (MCP servers, IDE extensions, skills) frequently receive blanket privileges, personal user accounts, and no audit trail in practice. If such an agent is taken over through prompt injection, compromised components, or manipulated tool definitions, then the attacker acts with the agent's privileges inside the legitimate infrastructure. Without an inventory, shadow AI additionally emerges: agents operated outside any security consideration. Existing control frameworks do not explicitly capture this attack surface.

3 Requirements

The specific requirements of module MRIS.bd.1 Mythos-Resistant Hardening of the ISMS are listed below. The Chief Information Security Officer (CISO) is responsible for ensuring that all requirements are met and verified in accordance with the agreed security concept. The CISO must always be involved in strategic decisions.

Further roles are defined in the IT-Grundschutz Compendium. They should be filled insofar as this is reasonable and appropriate.

Responsibilities	Roles
Overall responsibility	Chief Information Security Officer (CISO)
Further responsibilities	IT Operation Department, Developer

Exactly one role should have overall responsibility. In addition, there may be further responsibilities. If one of these further roles has primary responsibility for meeting a requirement, this role is stated in square brackets after the heading of the requirement.

3.1 Basic Requirements

The following requirements MUST be met as a matter of priority for this module.

MRIS.bd.1.A1 Reassessing the Effectiveness Assumptions of Existing Security Safeguards (B)

The organisation MUST reassess the effectiveness and probability assumptions of the implemented security safeguards against the AI-accelerated threat landscape. Security safeguards whose protective effect relies predominantly on attacker friction rather than on a hard structural barrier MUST be identified in the process. The results of the reassessment MUST be incorporated into the risk analysis. If the protective effect of a safeguard is assessed as degraded, then a decision MUST be taken between hardening, compensation, and justified risk acceptance. The reassessment MUST be repeated upon significant changes to the threat landscape.

MRIS.bd.1.A2 Using Phishing-Resistant Multi-Factor Authentication (B)

All privileged access MUST be protected with phishing-resistant mechanisms based on FIDO2/WebAuthn (passkeys or hardware tokens). All externally reachable services MUST be protected with phishing-resistant mechanisms. Access to systems with increased protection needs MUST be protected with phishing-resistant mechanisms. SMS-based mechanisms and simple push confirmations without number matching MUST NOT be used for new access. For existing legacy MFA mechanisms, a phase-out plan with a deadline MUST be documented. The enrolment and recovery processes for authentication means MUST be protected against social engineering.

MRIS.bd.1.A3 Maintaining Immutable Backups with Regular Restore Tests (B) [IT Operation Department]

For all business-critical data sets, at least one immutable or air-gapped backup copy MUST be maintained. The backup regime SHOULD be built according to the 3-2-1-1-0 principle. The backup infrastructure MUST be administered via dedicated access paths and privileges separated from the production environment. Restore tests MUST be performed at least quarterly. The restore tests MUST be documented. If a restore test fails, then the cause MUST be remediated without delay.

MRIS.bd.1.A4 Establishing Behaviour-Based Attack Detection with Event Correlation (B)

The organisation MUST have procedures in place for detecting attacks based on behavioural anomalies and the correlation of related events. The detection rules MUST be aligned with a recognised catalogue of attack techniques (MITRE ATT&CK). The coverage of the most relevant attack techniques MUST be measured. The measured coverage MUST be reviewed regularly. Threat hunting SHOULD be established as a permanent function with a documented methodology. If no in-house security operations centre is operated, then a managed detection and response service with a comparable scope of services SHOULD be commissioned.

MRIS.bd.1.A5 Introducing Workload Identities and Identity-Based Access Control (B) [IT Operation Department]

Privileged service accounts and workloads MUST be equipped with cryptographically verifiable, short-lived identities. Communication between critical services MUST be authorised on the basis of this workload identity. Network location alone MUST NOT serve as the basis for authorisation. Long-lived shared secrets stored in plain text MUST NOT be used on critical communication paths. Communication between production services SHOULD be protected end-to-end with mTLS. The migration SHOULD proceed in stages, starting with the privileged workloads.

MRIS.bd.1.A6 Creating a Cryptographic Inventory and a Post-Quantum Migration Strategy (B)

The organisation MUST maintain an inventory of all cryptographic mechanisms, key lengths, and implementations in use. If data with a long confidentiality lifetime is processed, then a documented migration strategy towards quantum-safe mechanisms MUST be in place. If data with a long confidentiality lifetime is transmitted, then hybrid mechanisms SHOULD be used during the transition. The inventory and the strategy MUST be reviewed at least annually or on an event-driven basis. For new procurements, support for quantum-safe or replaceable mechanisms SHOULD be included as a requirement.

MRIS.bd.1.A7 Regulated Use of AI Agents (B)

The organisation MUST maintain an inventory of all AI agents and agentic components used in production. If AI agents are used in production, then they MUST be treated as privileged systems. Every production AI agent MUST be given a technical identity limited to the privileges required. Actions of AI agents with write or network privileges MUST be logged completely and attributably to the initiating human identity. For every production AI agent, a named responsible person with the ability to shut it down MUST be designated. Irreversible actions of an AI agent MUST require human approval. Agentic components such as MCP servers, IDE extensions, and skills MUST ONLY be used via a maintained allow list.

3.2 Standard Requirements

Together with the basic requirements, the following requirements correspond to the state of the art for this module. They SHOULD be met as a matter of principle.

MRIS.bd.1.A8 Generating Software Bills of Materials and Build Provenance Evidence (S) [Developer]

For every piece of software created and distributed by the organisation, a software bill of materials (SBOM) in a standardised format SHOULD be generated automatically in the build pipeline. The SBOM SHOULD also capture transitive dependencies. The captured components SHOULD be correlated automatically against vulnerability databases. If a new vulnerability becomes known, then the organisation’s own exposure MUST be determinable within 24 hours on the basis of the SBOMs. For critical artefacts, tamper-proof build provenance SHOULD be maintained. The provision of an SBOM SHOULD be contractually required from critical software suppliers.

MRIS.bd.1.A9 Integrating Automated Security Testing into the Development Pipeline (S) [Developer]

Security testing of the source code, the dependencies, and the running application SHOULD be executed automatically in the development pipeline on every change. Severe findings SHOULD block the merge into the main branch. AI-generated code SHOULD be additionally tested as a distinct risk category. If an actively exploited vulnerability in a deployed component becomes known, then the patch for internet-exposed systems SHOULD be rolled out within 24 hours.

MRIS.bd.1.A10 Automating Incident Response with Hardened Playbooks (S)

For unambiguous incident types, predefined automated response playbooks SHOULD be in place. The containment of unambiguous incidents SHOULD be initiated automatically within minutes. Ambiguous or severe cases SHOULD be escalated in a targeted manner to the persons responsible for them. If an automatic action can cause major damage, then human approval MUST be obtained before execution. The automation layer itself SHOULD be hardened, in particular through authenticated triggers, a complete audit trail, and rate limits. Exercises for multiple parallel incidents SHOULD be conducted at least every six months. If no in-house security team exists, then a managed detection and response service with a contractually agreed containment time MAY be used.

MRIS.bd.1.A11 Establishing Continuous Automated Effectiveness Verification (S)

Compliance with the security safeguards listed in the Statement of Applicability SHOULD be verified continuously and automatically against defined target states. Security requirements SHOULD, where possible, be formulated as executable rules (policy as code). These rules SHOULD be integrated into the deployment and operations processes. Deviations SHOULD automatically trigger a defined response. The share of security safeguards verified automatically SHOULD be measured. This share SHOULD be increased step by step.

3.3 Requirements in Case of Increased Protection Needs

Exemplary proposals for requirements that go beyond the level of protection corresponding to the state of the art are listed below for this module. They SHOULD be considered in case of increased protection needs. The specific requirements are determined in an individual risk analysis.

MRIS.bd.1.A12 Using Signed Container Images and Protected Execution Environments (H) [IT Operation Department]

Production containers SHOULD be started exclusively from signed images from controlled registries that are verified before start-up. Verification SHOULD be enforced technically by the platform, for example with admission controller policies. If workloads have particularly high protection needs, then protected execution environments (confidential computing) with remote attestation SHOULD be used.

MRIS.bd.1.A13 Implementing Verifiable Tenant Separation (H) [IT Operation Department]

If multi-tenant services are operated, then data, networks, and compute resources MUST be separated per tenant. Data separation SHOULD be implemented with tenant-specific keys. The effectiveness of the separation SHOULD be demonstrated through regular, preferably automated tests. If cross-tenant access is detected, then it MUST be stopped without delay.

MRIS.bd.1.A14 Conducting Threat-Led Penetration Tests with AI Scenarios (H)

The effectiveness of the protective safeguards SHOULD be verified regularly through threat-led penetration testing. The test scenarios SHOULD include AI-typical attack approaches, in particular fragmented attack chains, parallel multi-vector attacks, and AI-assisted spear phishing. Between the test cycles, joint exercises of the offensive and defensive sides (purple teaming) SHOULD take place. Critical findings SHOULD be remediated within an agreed time frame.

4 Additional Information

The "Mythos-Resistant Information Security (MRIS)" framework (Version 1.9) contains the complete effectiveness assessment of all 93 controls of ISO/IEC 27002:2022 against the AI-accelerated threat landscape, the gap analysis against BSI C5:2026, NIST, DORA, the CRA, and NIS2, and the catalogue of the thirteen Mythos Hardening Controls with maturity levels (CC BY-SA 4.0).

The "MRIS Implementation Guide" (Version 1.5) describes for each Mythos Hardening Control the organisational and technical implementation, implementation examples, maturity paths, key figures, and typical mistakes. It is the substantive basis of the implementation guidance for this module.

With the C5:2026 criteria catalogue, the BSI provides requirements for cloud services that support several requirements of this module on the framework side, in particular regarding cryptography,

development, and tenant separation. The NIST standards FIPS 203, FIPS 204, and FIPS 205 specify the finalised quantum-safe mechanisms. The NIS2 implementing regulation (EU) 2024/2690, the Cyber Resilience Act, and DORA contain binding requirements that overlap with individual requirements of this module.