MRIS.bd.1 Mythos-resistant hardening of the ISMS

MRIS.bd: user-defined layer — Mythos-resistant information security

Version 1.3 | Status: 12 September 2026 | Author: Richard Peddi

Contents

1 Description

1.1 Introduction

Agentic frontier AI models (the "Mythos class") can find vulnerabilities autonomously, generate exploits in hours rather than days, and carry out complex attacks largely on their own. Publicly documented threat intelligence reports evidence real campaigns in which 80 to 90 per cent of the tactical attack steps ran autonomously. Since mid-2026 this capability class has been broadly available across vendors.

For an established ISMS this does not mean a break in methodology, but a destabilization of its inputs: the likelihood, time and effectiveness assumptions of many implemented security measures are no longer sound. In particular, measures whose protective effect rests on the limited patience and capacity of human attackers ("friction") lose effectiveness, while measures with a hard structural barrier (cryptography, least privilege, immutability, demonstrable separation) hold.

This module operationalizes the thirteen Mythos Hardening Controls (MHC) of the Mythos-Resistant Information Security framework (MRIS) as requirements in IT-Grundschutz format. The basis is the systematic effectiveness assessment of all 93 controls of ISO/IEC 27002:2022 against the AI-accelerated threat landscape, together with 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 measures are realigned to the changed threat landscape, and targeted hardening requirements close those gaps that technology-agnostic standards such as ISO/IEC 27001 do not address explicitly. The module thereby establishes conformity with ISO/IEC 27001 and the IT-Grundschutz standards and supplements them with effectiveness-oriented requirements.

1.3 Scope and modelling

The module MRIS.bd.1 Mythos-resistant hardening of the ISMS is to be applied once to the entire information domain.

To produce an IT-Grundschutz model for a specific information domain, the totality of all modules must in principle be considered. The module presupposes an established ISMS and supplements in particular the modules of the ISMS, ORP, OPS, DER and APP layers. It replaces none of those modules.

This module addresses

- the hardening of authentication, workload identities, data backup and attack detection against AI-accelerated attack patterns,
- the transparency and securing of the software supply chain including automated security testing, and
- the governance of AI agents deployed in production, continuous effectiveness checking and automated incident response.

The following content is likewise significant and is addressed elsewhere:

- Establishment and operation of the security process and of 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 Handling of Security Incidents)

This module does not address

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

2 Threat landscape

Because IT-Grundschutz modules cannot address individual information domains, typical scenarios are taken as the basis for describing the threat landscape. The following specific threats and vulnerabilities are of particular significance for the module MRIS.bd.1 Mythos-resistant hardening of the ISMS.

2.1 Collapse of the time window between patch and exploit

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

Agentic AI models master the reverse engineering of patches as a mechanical analysis task. If an institution 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 response window of several days is thereby broken. This is aggravated by the disclosure waves of AI-supported vulnerability research beginning in 2026, which significantly increase the volume of new vulnerabilities while central vulnerability databases are structurally overloaded.

2.2 Compression of the response time axis

AI-driven attacks run faster than human response chains can decide.

In documented campaigns, 80 to 90 per cent of the tactical attack steps were executed autonomously, at request rates unattainable for human operators. If a human must decide at every point in the immediate response chain, then the institution's response time is no longer of the same order of magnitude as the speed of the attack. Detection and response measures thereby lose effect not because of inadequate quality but because their latency is structurally too high. An attack may be complete before the first triage decision has been taken.

2.3 Fragmentation of attacks into unremarkable micro-steps

Agentic attackers decompose an overall attack into individual actions that each appear as legitimate technical requests.

Vulnerability scans, credential validation, lateral movement and data extraction are executed as separate, individually unremarkable micro-steps. If attack detection is based on evaluating discrete individual events — signature-based detection without behavioural context, static thresholds, alerting on individual actions — then the attack remains invisible even though every individual measure functions technically. The risk materializes only in the aggregation of the events. Institutions without behaviour-based, cross-event correlation recognize such attacks only from the damage.

2.4 Decoupling of attacker capability and actor type

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

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

2.5 Ungoverned use of privileged AI agents

AI agents deployed in production with far-reaching entitlements form a new, often unmanaged attack surface within the institution.

Coding agents, agentic workflows and their components (MCP servers, IDE extensions, skills) in practice frequently receive blanket entitlements, personal user accounts and no audit trail. If such an agent is taken over through prompt injection, compromised components or manipulated tool definitions, then the attacker acts with the entitlements of the agent inside the legitimate infrastructure. Without an inventory, shadow AI

additionally arises: agents operated outside any security consideration. Existing control frameworks do not capture this attack surface explicitly.

3 Requirements

The specific requirements of the module MRIS.bd.1 Mythos-resistant hardening of the ISMS are set out below. The Information Security Officer (ISO) is responsible for ensuring that all requirements are met and reviewed in accordance with the defined security concept. The ISO must always be involved in strategic decisions.

Further roles are defined in the IT-Grundschutz Compendium. They should be filled where this is sensible and appropriate.

Responsibilities	Roles
Principally responsible	Information Security Officer (ISO)
Further responsibilities	IT Operations, Developers

Exactly one role should be principally responsible. Beyond that there may be further responsibilities. If one of those further roles is primarily responsible for meeting a requirement, then that role is stated in square brackets after the requirement heading.

3.1 Basic requirements

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

MRIS.bd.1.A1 Reassessment of the effectiveness assumptions of existing security measures (B)

The institution MUST reassess the effectiveness and likelihood assumptions of the implemented security measures against the AI-accelerated threat landscape. Security measures whose protective effect rests predominantly on attacker friction rather than on a hard structural barrier MUST be identified in doing so. The results of the reassessment MUST feed into the risk analysis. If the protective effect of a measure is assessed as degraded, then a decision MUST be taken on hardening, compensation or a justified acceptance of the risk. The reassessment MUST be repeated on material changes in the threat landscape.

MRIS.bd.1.A2 Use of phishing-resistant multi-factor authentication (B)

All privileged access MUST be secured with phishing-resistant methods to FIDO2/WebAuthn (passkeys or hardware tokens). All externally reachable services MUST be secured with phishing-resistant methods. Access to systems with elevated protection requirements SHOULD be secured with phishing-resistant methods. SMS-based methods and simple push confirmations without number matching MUST NOT be used for new access. For existing legacy MFA methods, a replacement plan with a target date MUST be documented. The registration and recovery processes for authentication means MUST be secured against social engineering.

MRIS.bd.1.A3 Maintaining immutable data backups with regular recovery tests (B) [IT Operations]

For all business-critical data holdings, at least one immutable or network-separated backup copy MUST be maintained. Data backup SHOULD be built on the 3-2-1-1-0 principle. The backup infrastructure MUST be administered through its own access paths and entitlements, separate from the production environment. Recovery tests MUST be carried out at least quarterly. The recovery tests MUST be documented. If a recovery test fails, then the cause MUST be remedied without delay.

MRIS.bd.1.A4 Establishing behaviour-based attack detection with event correlation (B)

The institution MUST have procedures for detecting attacks on the basis of behavioural anomalies and the correlation of connected events. The detection rules MUST be oriented to a recognized catalogue of attack techniques (MITRE ATT&CK). The coverage of the most important attack techniques MUST be measured. The measured coverage MUST be reviewed regularly. Threat hunting SHOULD be established as a fixed function with a documented methodology. If no security operations centre of its own is operated, then a managed detection and response service with a comparable scope of performance SHOULD be commissioned.

MRIS.bd.1.A5 Introduction of workload identities and identity-based access control (B) [IT Operations]

Privileged service accounts and workloads MUST be equipped with cryptographically verifiable, short-lived identities. Communication between critical services MUST be authorized on the basis of this workload identity. Network position alone MUST NOT serve as the basis for authorization. Long-lived shared secrets stored in cleartext MUST NOT be used on critical communication paths. Communication between production services SHOULD be secured throughout with mTLS. The migration SHOULD take place in stages, beginning with the privileged workloads.

MRIS.bd.1.A6 Producing a cryptographic inventory and a post-quantum migration strategy (B)

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

MRIS.bd.1.A7 Governed use of AI agents (B)

The institution MUST maintain an inventory of all AI agents and agentic workflows deployed in production. The inventory MUST cover self-developed agents as well as agents operated as software-as-a-service or through low-code and no-code platforms. If AI agents are deployed in production, then they MUST be treated like privileged systems. Every production AI agent MUST receive a technical identity limited to the entitlements required. If the platform in use does not support an agent-specific identity, then controlled delegation with minimal entitlements MUST take place through a dedicated service account, and the permission boundaries MUST be documented. Actions of AI agents with write or network entitlements MUST be logged completely and attributably to the triggering human identity. For every production AI agent, a named responsible person with the ability to shut it down MUST be established. Irreversible actions of an AI agent MUST require human approval. Agentic components such as MCP servers, IDE extensions and skills MUST ONLY be used through 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 generally be met.

MRIS.bd.1.A8 Generating software bills of materials and build provenance evidence (S) [Developers]

For every piece of software developed in-house and distributed onwards, a software bill of materials (SBOM) in a standardized format SHOULD be generated automatically in the build pipeline. The bill of materials SHOULD also capture the transitive dependencies. The components captured SHOULD be reconciled against vulnerability databases in automated form. If a new vulnerability becomes known, then the organization's own exposure MUST be determinable from the bills of materials within 24 hours. For critical artefacts, tamper-resistant build provenance SHOULD be maintained. The provision of a bill of materials SHOULD be required contractually from critical software suppliers.

MRIS.bd.1.A9 Integration of automated security testing into the development pipeline (S) [Developers]

Security checks of the source code, the dependencies and the running application SHOULD be executed automatically in the development pipeline with every change. Serious findings SHOULD block the merge into the main branch. AI-generated code SHOULD be taken into account as a risk factor in its own right in the secure development standard and in the risk assessment, and checked additionally. For suitable applications, AI-supported test methods SHOULD additionally be deployed. If an actively exploited vulnerability becomes known in a component in use, then an effective remediation SHOULD be implemented for internet-exposed systems within 24 hours. Patch, workaround, compensating measure, disconnection from the network or replacement of the component count as effective remediation.

MRIS.bd.1.A10 Automation of incident response with secured playbooks (S)

For unambiguous incident types, predefined, automated response playbooks SHOULD be in place. Containment of unambiguous incidents SHOULD be initiated automatically within minutes. Ambiguous or

serious cases SHOULD be escalated deliberately to the persons responsible for them. If an automatic action can have a large damage impact, then human approval MUST take place before execution. The automation layer itself SHOULD be hardened, in particular through authenticated triggers, a complete audit trail and rate limits. Exercises for several parallel incidents SHOULD be carried out at least half-yearly. If no security team of its own is available, then a managed detection and response service with a contractually agreed containment time MAY be used.

MRIS.bd.1.A11 Establishing continuous automated effectiveness checking (S)

Observance of the security measures carried in the Statement of Applicability SHOULD be checked on an ongoing, automated basis 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 MUST be alerted and tracked; an automated response to deviations SHOULD be sought. The institution SHOULD document which security measures are technically capable of automated checking. The share of measures checked in automated form SHOULD be measured in relation to those measures classified as automatable and increased step by step.

3.3 Requirements for elevated protection requirements

Set out below for this module are exemplary proposals for requirements that go beyond the protection level corresponding to the state of the art. The proposals SHOULD be considered where protection requirements are elevated. The concrete determination is made in the course of an individual risk analysis.

MRIS.bd.1.A12 Use of signed container images and protected execution environments (H) [IT Operations]

Production containers SHOULD be started exclusively from signed images, verified before start-up, drawn from controlled registries. The verification SHOULD be enforced technically through the platform, for example with admission controller policies. If workloads have a particularly high protection requirement, then protected execution environments (confidential computing) with remote attestation SHOULD be deployed.

MRIS.bd.1.A13 Implementation of demonstrable tenant separation (H) [IT Operations]

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 evidenced through regular, ideally automated tests. If cross-tenant access is identified, then it MUST be prevented without delay.

MRIS.bd.1.A14 Carrying out threat-led penetration tests with AI scenarios (H)

The effectiveness of the protective measures SHOULD be tested regularly through threat-led penetration testing. The test scenarios SHOULD include AI-typical approaches, in particular fragmented attack chains, parallel multi-vector attacks and AI-supported spear phishing. Between the test cycles, joint exercises of the attacking and defending sides (purple team) SHOULD take place. Critical findings identified SHOULD be remedied within an agreed time window.

4 Further information

The framework "Mythos-Resistant Information Security (MRIS)" (Version 2.0) contains the full effectiveness assessment of all 93 controls of ISO/IEC 27002:2022 against the AI-accelerated threat landscape, the gap analysis relative to 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 2.0) describes, for each Mythos Hardening Control, the organizational and technical implementation, the maturity path, measurement and audit evidence, typical mistakes and delimitation.

It is the substantive basis of the implementation guidance accompanying this module.

With the criteria catalogue C5:2026, the BSI provides requirements for cloud services that support several requirements of this module on the framework side, in particular on cryptography, development and tenant separation. The NIST standards FIPS 203, FIPS 204 and FIPS 205 specify the finalized quantum-safe methods; HQC was selected for standardization as a supplementary method. NIS2 Implementing Regulation (EU)

2024/2690, the Cyber Resilience Act and DORA contain binding requirements that overlap with individual requirements of this module.

The AI Audit and Assurance Assessment Architecture (A5) — published by the BSI as a Community Draft on 6 July 2026 — assesses the trustworthiness of AI systems themselves across their full life cycle and connects to the C5 catalogue through an operational module. A5 and this module pursue different directions: A5 examines AI systems as the object of assessment; this module, by contrast, hardens the effectiveness of existing ISMS measures against attackers who themselves deploy AI capabilities. The two works do not compete — they complement each other.