MRIS.bd.1 Mythos-resistente Härtung des ISMS

MRIS.bd: Benutzerdefinierte Schicht — Mythos-resistente Informationssicherheit

Version 1.2 | Stand: 14.08.2026 | Autor: Richard Peddi

1 Beschreibung

1.1 Einleitung

Agentische Frontier-KI-Modelle („Mythos-Klasse") können Schwachstellen autonom finden, Exploits in Stunden statt Tagen erzeugen und komplexe Angriffe weitgehend selbstständig ausführen. Öffentlich dokumentierte Threat-Intelligence-Berichte belegen reale Kampagnen, in denen 80 bis 90 % der taktischen Angriffsschritte autonom abliefen. Seit Mitte 2026 ist diese Fähigkeitsklasse anbieterübergreifend und breit verfügbar.

Für ein etabliertes ISMS bedeutet das keinen Bruch der Methodik, aber eine Destabilisierung ihrer Eingaben: Wahrscheinlichkeits-, Zeit- und Wirksamkeitsannahmen vieler implementierter Sicherheitsmaßnahmen sind nicht mehr belastbar. Insbesondere Maßnahmen, deren Schutzwirkung auf der begrenzten Geduld und Kapazität menschlicher Angreifer beruht („Reibung"), verlieren Wirksamkeit, während Maßnahmen mit harter struktureller Barriere (Kryptografie, Least Privilege, Unveränderlichkeit, nachweisbare Trennung) standhalten.

Dieser Baustein operationalisiert die dreizehn Mythos-Härtungs-Controls (MHC) des Frameworks Mythos-resistente Informationssicherheit (MRIS) als Anforderungen im IT-Grundschutz-Format. Grundlage sind die systematische Wirksamkeitsbewertung aller 93 Controls der ISO/IEC 27002:2022 gegen die KI-beschleunigte Bedrohungslage sowie die Konvergenzanalyse mit BSI C5:2026, NIST, DORA, CRA und NIS2 samt Durchführungsverordnung (siehe [MRIS]). Die Umsetzung wird in den zugehörigen Umsetzungshinweisen beschrieben.

1.2 Zielsetzung

Ziel dieses Bausteins ist es, ein bestehendes ISMS gegen KI-beschleunigte Angriffe zu härten: Die Wirksamkeitsannahmen der implementierten Sicherheitsmaßnahmen werden an der veränderten Bedrohungslage neu ausgerichtet, und gezielte Härtungs-Anforderungen schließen diejenigen Lücken, die technologie-agnostische Standards wie ISO/IEC 27001 nicht explizit adressieren. Der Baustein stellt damit Konformität zu ISO/IEC 27001 und den IT-Grundschutz-Standards her und ergänzt sie um wirksamkeitsorientierte Anforderungen.

1.3 Abgrenzung und Modellierung

Der Baustein MRIS.bd.1 Mythos-resistente Härtung des ISMS ist einmal auf den gesamten Informationsverbund anzuwenden.

Um ein IT-Grundschutz-Modell für einen konkreten Informationsverbund zu erstellen, muss grundsätzlich die Gesamtheit aller Bausteine betrachtet werden. Der Baustein setzt ein etabliertes ISMS voraus und ergänzt insbesondere die Bausteine der Schichten ISMS, ORP, OPS, DER und APP. Er ersetzt keinen dieser Bausteine.

Dieser Baustein behandelt

- die Härtung der Authentisierung, der Workload-Identitäten, der Datensicherung und der Angriffserkennung gegen KI-beschleunigte Angriffsmuster,

- die Transparenz und Absicherung der Software-Lieferkette einschließlich automatisierter Sicherheitstests sowie
- die Governance produktiv eingesetzter KI-Agenten, die kontinuierliche Wirksamkeitsprüfung und die automatisierte Vorfallsreaktion.

Folgende Inhalte sind ebenfalls von Bedeutung und werden an anderer Stelle behandelt:

- Aufbau und Betrieb des Sicherheitsprozesses und des Risikomanagements (siehe ISMS.1 Sicherheitsmanagement)
- Grundlegende Datensicherung, Netzarchitektur, Patch- und Änderungsmanagement (siehe Schichten CON, OPS und NET)
- Grundlegende Behandlung von Sicherheitsvorfällen (siehe DER.2.1 Behandlung von Sicherheitsvorfällen)

Dieser Baustein behandelt nicht

- die Sicherheit eigener KI-Produkte gegenüber Endkunden (Produktsicherheit),
- ein vollständiges Risikomanagement- oder Management-System nach ISO/IEC 27001 Clause 4 bis 10 sowie
- rechtliche Compliance-Nachweise gegenüber Aufsichtsbehörden oder Zertifizierungsstellen.

2 Gefährdungslage

Da IT-Grundschutz-Bausteine nicht auf individuelle Informationsverbünde eingehen können, werden zur Darstellung der Gefährdungslage typische Szenarien zugrunde gelegt. Die folgenden spezifischen Bedrohungen und Schwachstellen sind für den Baustein MRIS.bd.1 Mythos-resistente Härtung des ISMS von besonderer Bedeutung.

2.1 Kollaps des Zeitfensters zwischen Patch und Exploit

Zwischen der Veröffentlichung eines Patches und der Verfügbarkeit eines funktionierenden Exploits liegen unter KI-Bedingungen Stunden statt Tage oder Wochen.

Agentische KI-Modelle beherrschen das Reverse Engineering von Patches als mechanische Analyseaufgabe. Wenn eine Institution mit gestaffelten Patch-Zyklen und manuellen Freigabeprozessen arbeitet, dann können internet-exponierte Systeme bereits kompromittiert sein, bevor der reguläre Patch-Prozess überhaupt beginnt. Die Wirksamkeitsannahme jedes Schwachstellenmanagements, das auf ein mehrtägiges Reaktionsfenster baut, ist damit gebrochen. Verschärfend wirken die seit 2026 einsetzenden Offenlegungswellen KI-gestützter Schwachstellenforschung, die das Volumen neuer Schwachstellen deutlich erhöhen, während zentrale Schwachstellendatenbanken strukturell überlastet sind.

2.2 Kompression der Reaktionszeitachse

KI-gesteuerte Angriffe laufen schneller ab, als menschliche Reaktionsketten entscheiden können.

In dokumentierten Kampagnen wurden 80 bis 90 % der taktischen Angriffsschritte autonom ausgeführt, mit Anfrageraten, die für menschliche Operatoren unerreichbar sind. Wenn in der unmittelbaren Reaktionskette an jeder Stelle ein Mensch entscheiden muss, dann liegt die Reaktionszeit der Institution nicht mehr in derselben Größenordnung wie die Angriffsgeschwindigkeit. Erkennungs- und Reaktionsmaßnahmen verlieren dadurch nicht wegen mangelnder Qualität an Wirkung, sondern weil ihre Latenz strukturell zu hoch ist. Ein Angriff kann abgeschlossen sein, bevor die erste Triage-Entscheidung getroffen wurde.

2.3 Fragmentierung von Angriffen in unauffällige Mikroschritte

Agentische Angreifer zerlegen einen Gesamtangriff in Einzelaktionen, die jeweils wie legitime technische Anfragen erscheinen.

Schwachstellen-Scans, Credential-Validierung, Lateral Movement und Datenextraktion werden als getrennte, einzeln unauffällige Mikroschritte ausgeführt. Wenn die Angriffserkennung auf der Bewertung diskreter Einzelereignisse basiert — signaturbasierte Erkennung ohne Verhaltenskontext, statische Schwellwerte, Alarmierung auf Einzelaktionen —, dann bleibt der Angriff unsichtbar, obwohl jede einzelne Maßnahme technisch funktioniert. Das Risiko materialisiert sich erst in der Aggregation der Ereignisse. Institutionen ohne verhaltensbasierte, ereignisübergreifende Korrelation erkennen solche Angriffe erst am Schaden.

2.4 Entkopplung von Angreiferfähigkeit und Akteurstyp

Die Kapazitätsbarriere zwischen Gelegenheitstätern und hochprofessionellen Angreifergruppen ist weitgehend entfallen.

Die klassische Wahrscheinlichkeitsbewertung stützt sich auf die Annahme, dass individualisierte Kampagnen auf Nation-State-Niveau knappe Spezialkapazität erfordern. Agentische KI-Modelle stellen diese Kapazität jedoch auch Akteuren mit geringer eigener technischer Kompetenz zur Verfügung. Wenn eine Institution ihre Risikoanalyse weiterhin auf die Seltenheit fähiger Angreifer stützt, dann unterschätzt sie systematisch die Eintrittswahrscheinlichkeit gezielter Angriffe. Die Wahrscheinlichkeitsachse jeder Risikomatrix verschiebt sich nach oben, ohne dass neue Bedrohungsakteure hinzukommen.

2.5 Ungeregelter Einsatz privilegierter KI-Agenten

Produktiv eingesetzte KI-Agenten mit weitreichenden Rechten bilden eine neue, oft ungemanagte Angriffsfläche innerhalb der Institution.

Coding-Agenten, agentische Workflows und ihre Komponenten (MCP-Server, IDE-Erweiterungen, Skills) erhalten in der Praxis häufig pauschale Rechte, persönliche Benutzerkonten und keinen Audit-Trail. Wenn ein solcher Agent durch Prompt-Injection, kompromittierte Komponenten oder manipulierte Werkzeugdefinitionen übernommen wird, dann handelt der Angreifer mit den Rechten des Agenten innerhalb der legitimen Infrastruktur. Ohne Inventar entsteht zusätzlich Schatten-KI (Shadow AI): Agenten, die außerhalb jeder Sicherheitsbetrachtung betrieben werden. Bestehende Control-Frameworks erfassen diese Angriffsfläche nicht explizit.

3 Anforderungen

Im Folgenden sind die spezifischen Anforderungen des Bausteins MRIS.bd.1 Mythos-resistente Härtung des ISMS aufgeführt. Der ISB ist dafür zuständig, dass alle Anforderungen gemäß dem festgelegten Sicherheitskonzept erfüllt und überprüft werden. Bei strategischen Entscheidungen ist der ISB stets einzubeziehen.

Im IT-Grundschutz-Kompendium sind darüber hinaus weitere Rollen definiert. Sie sollten besetzt werden, insofern dies sinnvoll und angemessen ist.

Zuständigkeiten	**Rollen**
Grundsätzlich zuständig	Informationssicherheitsbeauftragte (ISB)
Weitere Zuständigkeiten	IT-Betrieb, Entwickler

Genau eine Rolle sollte Grundsätzlich zuständig sein. Darüber hinaus kann es noch Weitere Zuständigkeiten geben. Falls eine dieser weiteren Rollen für die Erfüllung einer Anforderung vorrangig zuständig ist, dann wird diese Rolle hinter der Überschrift der Anforderung in eckigen Klammern aufgeführt.

3.1 Basis-Anforderungen

Die folgenden Anforderungen MÜSSEN für diesen Baustein vorrangig erfüllt werden.

MRIS.bd.1.A1 Neubewertung der Wirksamkeitsannahmen bestehender Sicherheitsmaßnahmen (B)

Die Institution MUSS die Wirksamkeits- und Wahrscheinlichkeitsannahmen der implementierten Sicherheitsmaßnahmen gegen die KI-beschleunigte Bedrohungslage neu bewerten. Sicherheitsmaßnahmen, deren Schutzwirkung überwiegend auf Angreiferreibung statt auf einer harten strukturellen Barriere beruht, MÜSSEN dabei identifiziert werden. Die Ergebnisse der Neubewertung MÜSSEN in die Risikoanalyse einfließen. Falls die Schutzwirkung einer Maßnahme als degradiert bewertet wird, dann MUSS über Härtung, Kompensation oder begründete Risikoübernahme entschieden werden. Die Neubewertung MUSS bei wesentlichen Änderungen der Bedrohungslage wiederholt werden.

MRIS.bd.1.A2 Einsatz phishing-resistenter Multi-Faktor-Authentisierung (B)

Alle privilegierten Zugänge MÜSSEN mit phishing-resistenten Verfahren nach FIDO2/WebAuthn (Passkeys oder Hardware-Token) abgesichert werden. Alle extern erreichbaren Dienste MÜSSEN mit phishing-resistenten Verfahren abgesichert werden. Zugänge zu Systemen mit erhöhtem Schutzbedarf MÜSSEN mit phishing-resistenten Verfahren abgesichert werden. SMS-basierte Verfahren und einfache Push-Bestätigungen ohne Number-Matching DÜRFEN für neue Zugänge NICHT eingesetzt werden. Für bestehende Legacy-MFA-Verfahren MUSS ein Ablösungsplan mit Termin dokumentiert werden. Die Registrierungs- und Wiederherstellungsprozesse für Authentisierungsmittel MÜSSEN gegen Social Engineering abgesichert werden.

MRIS.bd.1.A3 Vorhalten unveränderlicher Datensicherungen mit regelmäßigen Wiederherstellungstests (B) [IT-Betrieb]

Für alle geschäftskritischen Datenbestände MUSS mindestens eine unveränderliche oder vom Netz getrennte Sicherungskopie vorgehalten werden. Die Datensicherung SOLLTE nach dem 3-2-1-1-0-Prinzip aufgebaut werden. Die Backup-Infrastruktur MUSS über eigene, von der Produktionsumgebung getrennte Zugriffswege und Berechtigungen verwaltet werden. Wiederherstellungstests MÜSSEN mindestens vierteljährlich durchgeführt werden. Die Wiederherstellungstests MÜSSEN dokumentiert werden. Falls ein Wiederherstellungstest fehlschlägt, dann MUSS die Ursache unverzüglich behoben werden.

MRIS.bd.1.A4 Aufbau einer verhaltensbasierten Angriffserkennung mit Ereigniskorrelation (B)

Die Institution MUSS über Verfahren zur Erkennung von Angriffen anhand von Verhaltensauffälligkeiten und der Korrelation zusammenhängender Ereignisse verfügen. Die Erkennungsregeln MÜSSEN sich an einem anerkannten Katalog von Angriffstechniken (MITRE ATT&CK) orientieren. Die Abdeckung der wichtigsten Angriffstechniken MUSS gemessen werden. Die gemessene Abdeckung MUSS regelmäßig überprüft werden. Threat-Hunting SOLLTE als feste Funktion mit dokumentierter Methodik etabliert werden. Falls kein eigenes Security Operations Center betrieben wird, dann SOLLTE ein Managed-Detection-and-Response-Dienst mit vergleichbarem Leistungsumfang beauftragt werden.

MRIS.bd.1.A5 Einführung von Workload-Identitäten und identitätsbasierter Zugriffskontrolle (B) [IT-Betrieb]

Privilegierte Service-Konten und Workloads MÜSSEN mit kryptografisch verifizierbaren, kurzlebigen Identitäten ausgestattet werden. Die Kommunikation zwischen kritischen Diensten MUSS anhand dieser Workload-Identität autorisiert werden. Die Netzwerklage allein DARF NICHT als Autorisierungsgrundlage dienen. Langlebige, im Klartext gespeicherte gemeinsame Geheimnisse DÜRFEN auf kritischen Kommunikationspfaden NICHT eingesetzt werden. Die Kommunikation zwischen produktiven Diensten SOLLTE durchgängig mit mTLS abgesichert werden. Die Umstellung SOLLTE stufenweise erfolgen, beginnend mit den privilegierten Workloads.

MRIS.bd.1.A6 Erstellung eines kryptografischen Inventars und einer Post-Quantum-Migrationsstrategie (B)

Die Institution MUSS ein Inventar aller eingesetzten kryptografischen Verfahren, Schlüssellängen und Implementierungen führen. Falls Daten mit langer Vertraulichkeitsdauer verarbeitet werden, dann MUSS eine dokumentierte Migrationsstrategie zu quantensicheren Verfahren vorliegen. Falls Daten mit langer Vertraulichkeitsdauer übertragen werden, dann SOLLTEN im Übergang hybride Verfahren eingesetzt werden. Inventar und Strategie MÜSSEN mindestens jährlich oder anlassbezogen überprüft werden. Bei Neubeschaffungen SOLLTE die Unterstützung quantensicherer oder austauschbarer Verfahren als Anforderung aufgenommen werden.

MRIS.bd.1.A7 Geregelter Einsatz von KI-Agenten (B)

Die Institution MUSS ein Inventar aller produktiv eingesetzten KI-Agenten und agentischen Komponenten führen. Falls KI-Agenten produktiv eingesetzt werden, dann MÜSSEN diese wie privilegierte Systeme behandelt werden. Jeder produktive KI-Agent MUSS eine technische, auf die erforderlichen Rechte begrenzte Identität erhalten. Aktionen von KI-Agenten mit Schreib- oder Netzwerkrechten MÜSSEN lückenlos und der auslösenden menschlichen Identität zuordenbar protokolliert werden. Für jeden produktiven KI-Agenten MUSS eine benannte verantwortliche Person mit Abschaltmöglichkeit festgelegt werden. Irreversible Aktionen eines KI-Agenten MÜSSEN eine menschliche Freigabe erfordern. Agentische Komponenten wie MCP-Server, IDE-Erweiterungen und Skills DÜRFEN NUR über eine gepflegte Freigabeliste eingesetzt werden.

3.2 Standard-Anforderungen

Gemeinsam mit den Basis-Anforderungen entsprechen die folgenden Anforderungen dem Stand der Technik für diesen Baustein. Sie SOLLTEN grundsätzlich erfüllt werden.

MRIS.bd.1.A8 Erzeugung von Software-Stücklisten und Build-Provenance-Nachweisen (S) [Entwickler]

Für jede selbst erstellte und weitergegebene Software SOLLTE automatisch in der Build-Pipeline eine Software-Stückliste (SBOM) in einem standardisierten Format erzeugt werden. Die Stückliste SOLLTE auch die transitiven Abhängigkeiten erfassen. Die erfassten Komponenten SOLLTEN automatisiert gegen Schwachstellendatenbanken abgeglichen werden. Falls eine neue Schwachstelle bekannt wird, dann MUSS die eigene Betroffenheit anhand der Stücklisten innerhalb von 24 Stunden ermittelbar sein. Für kritische Artefakte SOLLTE eine manipulationssichere Build-Provenance vorgehalten werden. Von kritischen Softwarelieferanten SOLLTE die Bereitstellung einer Stückliste vertraglich eingefordert werden.

MRIS.bd.1.A9 Integration automatisierter Sicherheitstests in die Entwicklungs-Pipeline (S) [Entwickler]

Sicherheitsprüfungen des Quellcodes, der Abhängigkeiten und der laufenden Anwendung SOLLTEN automatisiert in der Entwicklungs-Pipeline bei jeder Änderung ausgeführt werden. Schwerwiegende Funde SOLLTEN die Übernahme in den Hauptzweig blockieren. KI-generierter Code SOLLTE als eigene Risikokategorie zusätzlich geprüft werden. Falls eine aktiv ausgenutzte Schwachstelle in einer eingesetzten Komponente bekannt wird, dann SOLLTE der Patch für internet-exponierte Systeme innerhalb von 24 Stunden ausgerollt werden.

MRIS.bd.1.A10 Automatisierung der Vorfallsreaktion mit abgesicherten Playbooks (S)

Für eindeutige Vorfallstypen SOLLTEN vordefinierte, automatisierte Reaktions-Playbooks vorliegen. Die Eindämmung eindeutiger Vorfälle SOLLTE automatisiert innerhalb von Minuten eingeleitet werden. Mehrdeutige oder schwerwiegende Fälle SOLLTEN gezielt an dafür zuständige Personen eskaliert werden. Falls eine automatische Aktion eine große Schadwirkung entfalten kann, dann MUSS vor der Ausführung eine menschliche Freigabe erfolgen. Die Automatisierungsschicht selbst SOLLTE gehärtet werden, insbesondere durch authentisierte Auslöser, vollständigen Audit-Trail und Rate-Limits. Übungen für mehrere parallele Vorfälle SOLLTEN mindestens halbjährlich durchgeführt werden. Falls kein eigenes Security-Team vorhanden ist, dann KANN ein Managed-Detection-and-Response-Dienst mit vertraglich vereinbarter Eindämmungszeit genutzt werden.

MRIS.bd.1.A11 Einrichtung einer kontinuierlichen automatisierten Wirksamkeitsprüfung (S)

Die Einhaltung der im Statement of Applicability geführten Sicherheitsmaßnahmen SOLLTE laufend und automatisiert gegen festgelegte Soll-Zustände geprüft werden. Sicherheitsvorgaben SOLLTEN, wo möglich, als ausführbare Regeln (Policy-as-Code) formuliert werden. Diese Regeln SOLLTEN in die Bereitstellungs- und Betriebsprozesse eingebunden werden. Abweichungen SOLLTEN automatisch eine definierte Reaktion auslösen. Der Anteil der automatisiert geprüften Sicherheitsmaßnahmen SOLLTE gemessen werden. Dieser Anteil SOLLTE schrittweise erhöht werden.

3.3 Anforderungen bei erhöhtem Schutzbedarf

Im Folgenden sind für diesen Baustein exemplarische Vorschläge für Anforderungen aufgeführt, die über dasjenige Schutzniveau hinausgehen, das dem Stand der Technik entspricht. Die Vorschläge SOLLTEN bei erhöhtem Schutzbedarf in Betracht gezogen werden. Die konkrete Festlegung erfolgt im Rahmen einer individuellen Risikoanalyse.

MRIS.bd.1.A12 Einsatz signierter Container-Images und geschützter Ausführungsumgebungen (H) [IT-Betrieb]

Produktive Container SOLLTEN ausschließlich aus signierten, vor dem Start geprüften Images aus kontrollierten Registries gestartet werden. Die Prüfung SOLLTE technisch über die Plattform durchgesetzt werden, etwa mit Admission-Controller-Policies. Falls Workloads einen besonders hohen Schutzbedarf haben, dann SOLLTEN geschützte Ausführungsumgebungen (Confidential Computing) mit Remote-Attestation eingesetzt werden.

MRIS.bd.1.A13 Umsetzung einer nachweisbaren Mandantentrennung (H) [IT-Betrieb]

Falls mandantenfähige Dienste betrieben werden, dann MÜSSEN Daten, Netzwerke und Rechenressourcen je Mandant getrennt werden. Die Datentrennung SOLLTE mit mandantenspezifischen Schlüsseln umgesetzt werden. Die Wirksamkeit der Trennung SOLLTE durch regelmäßige, möglichst automatisierte Tests nachgewiesen werden. Falls ein mandantenübergreifender Zugriff festgestellt wird, dann MUSS dieser unverzüglich unterbunden werden.

MRIS.bd.1.A14 Durchführung bedrohungsgeleiteter Penetrationstests mit KI-Szenarien (H)

Die Wirksamkeit der Schutzmaßnahmen SOLLTE regelmäßig durch bedrohungsgeleitete Penetrationstests (Threat-Led Penetration Testing) geprüft werden. Die Testszenarien SOLLTEN KI-typische Vorgehensweisen einschließen, insbesondere fragmentierte Angriffsketten, parallele Multi-Vector-Angriffe und KI-gestütztes Spear-Phishing. Zwischen den Testzyklen SOLLTEN gemeinsame Übungen von Angriffs- und Verteidigungsseite (Purple Team) stattfinden. Festgestellte kritische Funde SOLLTEN innerhalb eines vereinbarten Zeitfensters behoben werden.

4 Weiterführende Informationen

Das Framework „Mythos-resistente Informationssicherheit (MRIS)" (Version 1.9) enthält die vollständige Wirksamkeitsbewertung aller 93 Controls der ISO/IEC 27002:2022 gegen die KI-beschleunigte Bedrohungslage, die Lückenanalyse gegenüber BSI C5:2026, NIST, DORA, CRA und NIS2 sowie den Katalog der dreizehn Mythos-Härtungs-Controls mit Reifegrad-Stufen (CC BY-SA 4.0).

Der „MRIS Implementation Guide" (Version 1.5) beschreibt zu jedem Mythos-Härtungs-Control die organisatorische und technische Umsetzung, Umsetzungsbeispiele, Reifegrad-Pfade, Kennzahlen und typische Fehler. Er ist die inhaltliche Grundlage der Umsetzungshinweise zu diesem Baustein.

Das BSI stellt mit dem Kriterienkatalog C5:2026 Anforderungen an Cloud-Dienste bereit, die mehrere Anforderungen dieses Bausteins framework-seitig stützen, insbesondere zu Kryptografie, Entwicklung und Mandantentrennung. Die NIST-Standards FIPS 203, FIPS 204 und FIPS 205 spezifizieren die finalisierten quantensicheren Verfahren. Die NIS2-Durchführungsverordnung (EU) 2024/2690, der Cyber Resilience Act und DORA enthalten verbindliche Anforderungen, die sich mit einzelnen Anforderungen dieses Bausteins überschneiden.