Trust, Risk and Security Management (TRiSM), Reifegrad & offene Standards: Einordnung und Umsetzungsleitfaden für Kliniken

03.11.2025, Olaf Dunkel, htpp://www.olafdunkel.de

Executive Summary

Viele Krankenhäuser pilotieren derzeit generative Künstliche Intelligenz (GenKI). Ohne tragfähige Governance bleiben Nutzen, Skalierbarkeit und Prüfpfade jedoch fragil. Dieser Bericht ordnet das von Gartner geprägte Konzept Trust, Risk and Security Management (TRiSM, Vertrauens-, Risiko- und Sicherheitsmanagement) und das Gartner-AI-Reifegradmodell im Verhältnis zu offenen Referenzwerken ein: dem Artificial Intelligence Risk Management Framework (AI RMF, Risikomanagementrahmen für Künstliche Intelligenz) des National Institute of Standards and Technology (NIST, US-Bundesinstitut für Standards und Technologie) sowie der ISO/IEC 42001 als Artificial Intelligence Management System (AIMS, Managementsystem-Norm für KI). Ergänzend werden die KI-Prinzipien der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) berücksichtigt. Kernaussage: Kliniken profitieren am stärksten, wenn sie offene Prozessstandards (NIST AI RMF, ISO/IEC 42001) mit TRiSM-Kontrollen als Code in Continuous Integration/Continuous Delivery (CI/CD, kontinuierliche Integration und Auslieferung) und Model Operations (ModelOps, operative Bereitstellung und Überwachung von Modellen) kombinieren.

1. Zielsetzung und Kontext

Ziel ist eine quellenkritische, praxisnahe Einordnung zentraler Governance-Rahmenwerke für Künstliche Intelligenz (KI) im Krankenhauskontext. Der Bericht adressiert Versorgungs- und Administrationsprozesse, in denen große Sprachmodelle (LLMs, Large Language Models) und klassische ML-Modelle eingesetzt werden. Er baut auf öffentlich zugänglichen Primärquellen zu NIST AI RMF, ISO/IEC 42001 und OECD auf und referenziert veröffentlichte Beschreibungen zu TRiSM und dem Gartner-Reifegradmodell.

2. Frameworks im Überblick

2.1 TRiSM – Trust, Risk and Security Management (Gartner)

TRiSM bezeichnet einen von Gartner geprägten Ansatz, der Vertrauen, Risiko und Sicherheit für KI-Systeme integriert steuert. Zentrale Elemente sind Richtlinien, Kontrollen, Monitoring und Nachweise über den gesamten Modelllebenszyklus – von der Datenbewirtschaftung über

Training und Evaluierung bis zum Betrieb. In der Praxis wirksam wird TRiSM, wenn Kontrollen in CI/CD-Pipelines als Code automatisiert und in ModelOps dauerhaft überwacht werden. Nachteilig sind proprietäre Veröffentlichungsformen und wechselnde Darstellungslogiken (Säulen/Schichten), die Vergleichsstudien erschweren.

2.2 AI-Reifegradmodell (Gartner)

Das Gartner-Reifegradmodell ordnet Organisationen entlang von Dimensionen wie Strategie, Daten, Technologie und Fähigkeiten ein. Es dient als Standortbestimmung und zur Identifikation nächster Entwicklungsschritte. Für Kliniken ist es hilfreich, um Investitionen zu priorisieren, ersetzt aber keine auditierbaren Nachweise zur Wirksamkeit von Kontrollen.

2.3 NIST AI RMF – Artificial Intelligence Risk Management Framework

Das NIST AI RMF strukturiert das Management von KI-Risiken in die Funktionen Govern, Map, Measure und Manage. Es ist quelloffen, anpassbar und bietet eine gemeinsame Sprache für Rollen im Haus und bei Lieferanten. Stärken sind die Transparenz und die Anschlussfähigkeit an andere NIST-Publikationen; Limitationen ergeben sich aus der Notwendigkeit, sektorspezifische Interpretationen (z. B. Klinik) eigenständig zu konkretisieren.

2.4 ISO/IEC 42001 – Artificial Intelligence Management System (AIMS)

ISO/IEC 42001 definiert Anforderungen an ein Managementsystem für KI (AIMS). Analog zu ISO 9001/27001 kombiniert die Norm Politiken, Rollen, Ziele, Prozesse und Nachweise. Für Kliniken bietet sie einen auditierbaren Rahmen, um Governance und Compliance – etwa gegenüber der Datenschutz-Grundverordnung (DSGVO) – methodisch zu verankern.

2.5 OECD-KI-Prinzipien

Die OECD-Prinzipien betonen Werte wie Inklusivität, Sicherheit, Transparenz und Verantwortlichkeit. Sie sind nicht auditierbar, bieten aber normative Leitplanken und können als Prüf-Checkliste in Projekten genutzt werden.

3. Vergleich der Rahmenwerke (kompakt)

Kriterium	**TRiSM (Gartner)**	**Gartner Reifegrad**	**NIST AI RMF**	**ISO/IEC 42001 (AIMS)**	**OECD-Prinzipien**
Zielbild	Operative Kontrollen & Monitoring über den Modelllebenszyklus	Standortbestimmung & Entwicklungsfahrplan	Risikomanagementrahmen (Govern-Map-Measure-Manage)	Auditierbares Managementsystem für KI	Normative Leitplanken (Werte, Rechte)
Art der Quelle	Proprietär (Paywall, Berichte)	Proprietär (Paywall, Berichte)	Offen (Leitfaden, Profile)	Norm (zertifizierbar)	Empfehlungen (soft law)

Umsetzungsnähe	Hoch bei ‚Kontrollen als Code' in CI/CD & ModelOps	Mittel – Orientierung, keine Nachweise	Mittel – erfordert Klinik-Spezifizierung	Hoch – prozess- und auditfokussiert	Niedrig – Leitplanken
Nachweisführung	Betriebslogs, Tests, Metriken	Reifegradberichte	Risikoregister, Maßnahmenpläne	Policies, Rollen, KPIs, Audits	Selbstverpflichtungen
Typische Nutzung im KH	Betriebliche Absicherung von GenKI-Workflows	Planung & Priorisierung	Gemeinsame Sprache für Lieferanten & Fachbereiche	Zertifizierungsfähige Governance-Struktur	Wertebasierte Plausibilisierung

4. Anwendung im Krankenhaus (DE)

Für den Krankenhausbetrieb empfiehlt sich eine kombinierte Vorgehensweise: Offene Prozessstandards (NIST AI RMF, ISO/IEC 42001) definieren Rollen, Nachweise und Auditpfade; TRiSM ergänzt operative Kontrollen für GenKI- und ML-Workflows, insbesondere in CI/CD-Pipelines und ModelOps. Rechtliche Flanken sind u. a. die EU-Verordnung über Künstliche Intelligenz (EU AI Act) und die Richtlinie über Maßnahmen für ein hohes gemeinsames Cybersicherheitsniveau (NIS2).

4.1 Governance-Setup

- RACI-Matrix (Responsible–Accountable–Consulted–Informed) für KI-Lebenszyklus etablieren.
- Zentrales Inventar aller KI-Systeme (Zweck, Daten, Modelle, Lieferanten).
- Policy-Set: Datenqualität, Bias-Prüfung, Sicherheit, Zulassung, Betrieb, Abschaltung (Sunset).
- Lieferanten-Due-Diligence inkl. Nachweisen zu Datenschutz (DSGVO) und Sicherheit.

4.2 Kontrollen als Code & Betrieb

- CI/CD-Gates: automatisierte Tests (z. B. Robustheit, Prompt-Leckage, PII-Filter – personenbezogene Identifikatoren).
- ModelOps-Monitoring: Drift, Datenqualität, Performance, Kosten; Alarmierung & Rollback.
- Dokumentation: Model Cards, Datenblätter, Audit-Trails; reproduzierbare Freigaben.

4.3 Kennzahlen (KPIs) & Evidenz

Empfohlene Key Performance Indicators (KPIs, Leistungskennzahlen) für Steuerung und Nachweise:

KPI	**Definition**	**Ziel/Orientierung**
Time-to-Approval	Durchlaufzeit von Änderung bis Freigabe	−30 % ggü. Basis
Drift-Alerts pro Monat	Anzahl/Schweregradsumme	Abnehmend, < definiertem

	erkannter Modell-Drifts	Schwellwert
Incident-Rate KI	KI-bezogene Betriebs-/Sicherheitsvorfälle je Monat	Abnehmend, 0 kritische A-Vorfälle
Audit-Findings geschlossen	Anteil geschlossener Feststellungen im Quartal	≥ 90 %
ROI-Surrogat	Produktivitäts-/Qualitätsmetriken je Use-Case	Positiver Trend

5. 90-Tage-Umsetzungsfahrplan

Zeitraum	**Maßnahmen**
Tage 1–30	- RACI, Inventar, Policy-Minimalkorpus - Auswahl Pilot-Use-Case (klinisch oder administrativ) - CI/CD-Grundlagen & Logging etablieren
Tage 31–60	- TRiSM-Kontrollen als Code implementieren (Bias-/Sicherheits-Tests) - ModelOps-Monitoring (Drift, Kosten, Performance) - Lieferanten-Nachweise einsammeln (DSGVO, Sicherheit)
Tage 61–90	- Interne Abnahme + Red-Team-Übungen - KPI-Baseline & Reporting aufsetzen - Auditplan (ISO/IEC 42001-kompatibel) erstellen

6. Hauptrisiken & Gegenmaßnahmen

- Komplexität/Überfrachtung: Umfang modularisieren; zuerst kritische Kontrollen.
- Lieferanten-Lock-in: Offene Schnittstellen/Kontrakte fordern; Exit-Strategie vertraglich sichern.
- Datenqualität: Data-Stewardship, Validierungen, Schulungen.
- Akzeptanz: Frühe Einbindung von Klinik, IT-Sicherheit, Datenschutz, Betriebsrat; transparente KPIs.

7. Fazit

Governance ist Beschleuniger – nicht Bremse –, wenn sie automatisiert, messbar und auditierbar umgesetzt wird. Die Kombination aus NIST AI RMF und ISO/IEC 42001 als offenem Prozess- und Managementsystem mit den operativen TRiSM-Kontrollen in CI/CD und ModelOps

liefert Kliniken eine robuste, skalierbare Grundlage für den sicheren Einsatz von GenKI und klassischen ML-Modellen.

8. Quellen (kurz)

- Gartner (2025): AI TRiSM – Konzepte und Kontrollen.
- NIST (2023): AI Risk Management Framework 1.0.
- ISO/IEC (2023): 42001 – Artificial Intelligence Management System (AIMS).
- OECD (2024): Aktualisierte KI-Prinzipien.
- OCEG (2024): Erklärbarkeit und Governance bei GenKI.

Anhang A – Glossar der Abkürzungen

AI RMF – Artificial Intelligence Risk Management Framework (NIST-Rahmenwerk für KI-Risikomanagement).

AIMS – Artificial Intelligence Management System (Managementsystem für KI gemäß ISO/IEC 42001).

CI/CD – Continuous Integration/Continuous Delivery (kontinuierliche Integration und Auslieferung von Software).

DSGVO – Datenschutz-Grundverordnung (EU-Verordnung zum Schutz personenbezogener Daten).

EU AI Act – EU-Verordnung über Künstliche Intelligenz (Regelwerk für Entwicklung und Einsatz von KI in der EU).

GenKI – Generative Künstliche Intelligenz (KI-Systeme, die Inhalte erzeugen, z. B. Text, Bild, Code).

KPIs – Key Performance Indicators (Leistungskennzahlen zur Steuerung und Messung).

LLM – Large Language Model (großes Sprachmodell).

ModelOps – Model Operations (operative Bereitstellung, Überwachung und Steuerung von Modellen).

NIS2 – Richtlinie über Maßnahmen für ein hohes gemeinsames Cybersicherheitsniveau in der EU.

NIST – National Institute of Standards and Technology (US-Bundesinstitut für Standards und Technologie).

OECD – Organisation für wirtschaftliche Zusammenarbeit und Entwicklung.

PII – Personally Identifiable Information (personenbezogene Identifikatoren).

RACI – Responsible–Accountable–Consulted–Informed (Verantwortlichkeitsmatrix).

ROI – Return on Investment (Wirtschaftlichkeitsmaß).

TRiSM – Trust, Risk and Security Management (Vertrauens-, Risiko- und Sicherheitsmanagement; Gartner).