

Projektarbeit
zur Erlangung der Zertifizierung
IHK Cyber Security Advisor

Richtlinie zur Nutzung von KI-Systemen

Abgenommen durch



Zertifiziert durch

IHK Cottbus



Vorgelegt von:	René Kost
Eingereicht am:	20.07.2025
Kurs:	WB CS_09

Teil 1: ISMS – Methodik
Teil 2: Richtlinie zur Nutzung von KI-Systemen

Inhaltsverzeichnis

1. Firmenbeschreibung.....	3
1.1. MDTP AG.....	3
1.2. Bedarf einer Richtlinie zur Nutzung von KI-Systemen.....	3
1.3. IT-Infrastruktur.....	3
2. Schutzbedarfsfeststellung.....	4
3. SOLL-/IST-Abgleich.....	4
4. Risikoanalyse.....	6
5. Fazit.....	7
Tabellenverzeichnis.....	8
Abkürzungsverzeichnis.....	8

1. Firmenbeschreibung

1.1. MDTP AG

Die MDTP AG ist ein kleines Unternehmen mit 51 Mitarbeitenden im Bereich Digitalisierung und Optimierungsberatung von Anlagen und Anlagenteilen im produzierenden Gewerbe. Mithilfe von Prozessdaten wird ein digitaler Zwilling erstellt. Anhand dieses Modells können Problemstellen identifiziert und eine fundierte Beratung zur Optimierung durchgeführt werden.

1.2. Bedarf einer Richtlinie zur Nutzung von KI-Systemen

Der CEO stellte bei der Nutzung eines bekannten KI-Tools fest, dass dabei Kundeninformationen des Unternehmens ausgegeben wurden. Dies war Anlass zur kritischen Reflexion, inwieweit die Informationssicherheit bei der Nutzung von KI-Systemen gewährleistet ist. Daher besteht ein klarer Bedarf an einer Richtlinie zur Nutzung von KI-Systemen, um den sicheren Umgang mit sensiblen Daten sicherzustellen und die Informationssicherheit im Unternehmen zu wahren.

1.3. IT-Infrastruktur

Sensible Bereiche:

- **Physische Server:** Alle Zugriffe auf genehmigte KI-Systeme müssen über die zentralen Server erfolgen. Hier werden sämtliche Verbindungen protokolliert und Nachweisdokumente zur Nutzung geführt.
- **Lokales LLM TwinMe:** Dieses Modell verarbeitet sensible Daten von Interner Geschäfts, Kunden- und Mitarbeiter_innen. Es befindet sich im selben Subnet 172.17.14.0/24 in dem auch sensible Daten von Internen Geschäfts, Kunden- und Mitarbeiter_innen als DBMS vorhanden sind. Diese werden zum Training und Täglichen Arbeit der Geschäftsleitung, Vertrieb, Buchhaltung und HR genutzt.
- **Arbeitsgeräte der IT und Geschäftsführung:** Haben Zugang zum Subnetz 172.17.12.0/28, das ausschließlich für das Genehmigungsverfahren neuer KI-Systeme vorgesehen ist.
- **Zugriffe des Datenschutzbeauftragten:** Auf Blacklists und Whitelists für die Freigabe und zum Blockieren der KI-Systeme.

Nicht-sensible Bereiche:

Alle Arbeitsgeräte außerhalb der Geschäftsführung und IT-Abteilung sind im Mobile Device Management (MDM) registriert. Der Zugriff ist auf KI-Systeme beschränkt, die ausdrücklich in der zentral gepflegten Whitelist freigegeben wurden.

2. Schutzbedarfsfeststellung

Bereich	Daten	Schutzbedarf	Begründung
GP001 Sekretariat	Der Zugriff auf öffentliche Daten ist erlaubt, jedoch darf mit sensiblen Daten nur unter Aufsicht gearbeitet werden.	Normal	Das Risiko durch potenziell über ein KI-System geleakte Daten wird als gering eingeschätzt, wodurch auch das Reputationsrisiko und die Wahrscheinlichkeit rechtlicher Schadensersatzforderungen als niedrig bewertet werden.
GP002 Entwicklung	Arbeiten mit Standard-CAD-Daten, eigenen Konstruktionen und Simulationen sowie die Bearbeitung von Teilstücken durch KI-Agenten.	Hoch	Entwicklungsdaten, die möglicherweise von KI-Tools gespeichert wurden, könnten im Falle von Industriespionage zu einem Wettbewerbsnachteil führen.
GP003 Geschäftsführung	Haben Zugang zu allen Daten	Sehr hoch	Es besteht ein Risiko für die Informationssicherheit, da sensible Daten versehentlich über Kopier- und Einfügevorgänge (Copy & Paste) durch Mitarbeitende in KI-Systeme übertragen werden könnten.
GP004 Buchhaltung	Umgang mit sensiblen Kundendaten und Zugriff auf das DBMS mit Kundendaten	Sehr hoch	
GP005 Vertrieb		Sehr hoch	
GP006 IT-	Zugang zu allen Daten, Konfigurationen	Sehr hoch	

Table 1: Schutzbedarfsfeststellung

3. SOLL-/IST-Abgleich

Pflicht	S	I	Status	Verantwortliche Person	Zeitraum
SI01 Mitarbeiterschulungen	4	4	Umgesetzt: Mitarbeitende müssen vor der Nutzung eines KI-Systems eine Schulung zum sicheren und verantwortungsvollen Umgang absolvieren. Für die fortlaufende Nutzung sind regelmäßige Auffrischungsschulungen erforderlich. Abweichungen: Keine.	IT-Abteilung: zuständig für Unterlagen Personalabteilung: zuständig für die regelmäßige Durchführung der Schulungen	Beim Onboarding und dann jedes halbe Jahr ein mal
SI02 Datenkategorisierung und Implementierung eines Data Loss Prevention (DLP) Systems	4	3	Teilweise erfüllt: Daten werden gemäß ihrer Kategorien den jeweils zugelassenen KI-Systemen zugeordnet. Sensible Daten aus der Buchhaltung, Geschäftsführung und dem Personalwesen dürfen ausschließlich mit dem unternehmenseigenen lokalen KI-Agenten TwinMe verarbeitet werden. Abweichungen: Die bestehenden Trainingsdaten des Systems müssen	IT-Datenschutzbeauftragte Person	Die Trainingsdaten werden monatlich aktualisiert

			erweitert und teilweise korrigiert werden, um eine präzisere Verarbeitung und Kategorisierung sicherzustellen.		
SI03 Kontrolle und Reglementierung von KI-Diensten	4	3	<p>Teilweise erfüllt: Die Zugriffsberechtigungen zu KI-Systemen werden durch Black- und Whitelists geregelt.</p> <p>Abweichungen: Permanente Neuerungen müssen kontinuierlich in die Listen eingepflegt und aktualisiert werden.</p>	IT-Datenschutzbeauftragte Person	Monatliches Update der Listen
SI04 Prozesshaftigkeit	4	4	<p>Umgesetzt: In jedem Prozessschritt ist eine verantwortliche Person integriert, die sowohl einen Plausibilitätscheck vornimmt als auch die rechtliche Verantwortlichkeit trägt.</p>	<p>Mitarbeiter der Fachabteilung: für die Richtigkeit des Ergebnisses</p> <p>IT-Datenschutzbeauftragte Person: für die Log Dateien</p>	Bei jeder Verwendung eines KI-Ergebnisses
SI05 Audit	4	4	<p>Umgesetzt: Es werden regelmäßige Audits durchgeführt, dokumentiert und kontinuierlich nach weiteren Lösungen und Verbesserungen gesucht.</p> <p>Abweichungen: Keine.</p>	IT-Administration	Jährlich

Table 2: SOLL-/IST-Abgleich

Legende:

- 4 = Alle Maßnahmen in diesem Bereich wurden umgesetzt. Keine weiteren Schritte nötig
- 3 = Es wurde nicht alle Maßnahmen umgesetzt, der Großteil ist jedoch bereits vorhanden.
- 2 = Es sind mehrere Mängel und nicht umgesetzte Mängel vorhanden
- 1 = Es wurden keine Maßnahmen getroffen

4. Risikoanalyse

Stufe	Bezeichnung	Auswirkung (prozentual)	Finanzieller Schaden (EUR)
1	Sehr niedrig	0 % – 1,5 %	bis 15.000 €
2	Niedrig	1,6 % – 5 %	bis 50.000 €
3	Akzeptabel	5,1 % – 9 %	bis 90.000 €
4	Mittel	9,1 % – 20 %	bis 200.000 €
5	Hoch	20,1 % – 60 %	bis 600.000 €
6	Sehr hoch	60,1 % – 100 %	über 600.000 €

Table 3: Bewertung nach wirtschaftlichem Schweregrad (SLE)

Die freie Verfügbarkeit von KI-Tools verleitet Mitarbeitende dazu, diese in ihrer täglichen Arbeit auszuprobieren und gezielt einzusetzen. Dies kann zu einem unachtsamen Umgang mit Kundendaten führen und damit folgende Schwachstellen verursachen.

ID	Schwachstelle	Auswirkung	Maßnahme	SLE	ARO	ALE
R001	Verletzung der Vertraulichkeit durch Datenabfluss, bei dem unbeabsichtigt sensible Daten in KI-Systemen verwendet werden die nicht reglementiert sind.	Verstöße gegen grundlegende Datenschutzverordnungen im privaten und juristischen Bereich können zu Schadenszahlungen und Reputationsverlust führen.	SI02 SI03	3	24	72
R002	Die Herkunft von Informationen oder Daten ist nicht eindeutig nachvollziehbar oder überprüfbar	Fehlerhafte Konstruktionen und falsche juristische Texte wirken sich negativ auf die Produktentwicklung und Verträge aus. Schlechte Vertragsbedingungen sowie Reputationsschäden sind die Folgen	SI04	4	12	48
R003	Fehlende Schulung der Mitarbeiter führt dazu, dass diese nicht sensibel genug mit Unternehmensdaten umgehen und diese unachtsam an KI's weitergeben. Dadurch wird das Risiko R001 verstärkt.	Mitarbeiter werden unbeabsichtigt zu Whistleblowern des Unternehmens.	SI01	1	600	600

Table 4: Analyse der Risiken

5. Fazit

Bei der Analyse der MDTP AG wird deutlich, dass im Unternehmen ein solides Gesamtkonzept vorhanden ist. Dennoch zeigt sich, dass der Umgang mit KI-Systemen ein entscheidender Faktor für die zukünftige Wettbewerbsfähigkeit ist. Der sinnvolle und kontrollierte Einsatz von KI's mit einer angestrebten Fehlertoleranz von unter 20% - ist wesentlich, um Effizienzvorteile nachhaltig zu nutzen.

Gleichzeitig besteht bei einer unachtsamen oder nicht reglementierten Nutzung von KI, insbesondere durch fehlende Richtlinien und ungeschultes Personal, ein erhebliches Risiko eines Datenabflusses. Ein solcher Vorfall kann für das Unternehmen existenzbedrohend sein.

Daher ist es von zentraler Bedeutung, klar definierte und verbindliche Nutzungsregeln für den Einsatz von KI-Systemen zu erarbeiten und konsequent umzusetzen.

Tabellenverzeichnis

Table 1: Schutzbedarfsfeststellung.....	5
Table 2: SOLL-/IST-Abgleich.....	6
Table 3: Bewertung nach wirtschaftlichem Schweregrad (SLE).....	7
Table 4: Analyse der Risiken.....	7

Abkürzungsverzeichnis

Abkürzung	Bedeutung
ALE	Annualized Loss Expectancy (Jährlich erwarteter Verlust)
ARO	Annualized Rate of Occurrence (Jährliche Eintrittswahrscheinlichkeit)
DBMS	Datenbankmanagementsystem
HR	Human Resources (Personalabteilung)
KI	Künstliche Intelligenz
MDMP	Mobile Device Management
MDTP	Machine Digital Twin Production
SLE	Service Level Expectation (Erwartete Leistungsebene)

Richtline zu Nutzung von KI-Systemen

Dokumentenklassifizierung

Allgemeine	
Titel	Richtline zur Nutzung von KI-Systemen
Dokumentebene	Grundlagen
Kategorie	IT-Sicherheit
Sicherheitsklassifizierung	SK-1 (Intern), UY-2(ext. Mitarbeiter)
Referenz Dokumente	ISO 27001, ISO 23894, BSI-Grundschutz
Verantwortlichkeiten	
Autor	René Kost
Hauptverantwortlicher	Klaus Müller / IT-Leiter
Ansprechpartner	René Kost / ISB
E-Mail, Telefon	Rene.kost@MDTP.de 0176411965165
Gültigkeiten	
In Kraft seit	20.07.2025
In Kraft gesetzt durch	Max Müller / CEO
Überarbeitungsintervall	6 Monate
Nächste Überarbeitung	20.01.2026
Erstellt am	19.07.2025

Dokumentenhistorie

Version	Änderung	Datum	Autor	Prüfer
0.0	Entwurf	02.06.25	René Kost	Max Müller
1.0	Initiale Fassung	15.06.25	René Kost	Max Müller
1.1	Aufnahme KI-Agent PTC Creo+	02.07.25	René Kost	Max Müller
1.2	Aufnahme KI-Agent Fusion 360	19.07.25	René Kost	Max Müller

V1.2 / 19.07.2025	Richtlinie zur Nutzung von KI-Systemen	Klassifizierung: Grundlagen
-------------------	--	--------------------------------

Inhaltsverzeichnis

1. Geltungsbereich.....	4
2. Zweck.....	4
3. Allgemeine Verantwortlichkeiten.....	4
4. Regelung.....	4
4.1. Genehmigte KI-Systeme.....	4
4.2 Nutzungsregeln für KI-Systeme.....	4
4.2.1 Nutzungsregeln für <i>Aleph Alpha</i>	5
4.2.2 Nutzungsregeln für CAD-KI-Agenten.....	5
4.2.3 Nutzungsregeln des lokalen LLM-Agenten <i>TwinMe</i>	5
4.3 Zugriffsrechte.....	5
4.4 Prompting.....	6
4.5 Prozesssicherheit und Human-in-the-Loop.....	6
4.5.1. Fehleranalyse.....	7
4.5.2. Qualitätssicherung & Bug-Bounty-Programm.....	7
5. Pflichten des Systembetreibers.....	7
5.1 IT-Administration.....	7
5.2 Datenschutzbeauftragter.....	8
6. Schulung der Mitarbeitenden.....	8
7. Genehmigung von Ausnahmen.....	8
8. Berichterstattung und Eskalation.....	9
9. Gültigkeit, Verwaltung und Überprüfung.....	9
10. Folgen bei Verstößen.....	9
Literaturverzeichnis.....	11

V1.2 / 19.07.2025	Richtlinie zur Nutzung von KI-Systemen	Klassifizierung: Grundlagen
-------------------	--	--------------------------------

1. Geltungsbereich

Diese Richtlinie betrifft alle internen und externen Mitarbeitenden. Auch beauftragte Dienstleister sind daran gebunden, sofern dies abgestimmt wurde.

2. Zweck

Ziel dieser Richtlinie ist es, für den verantwortungsvollen Umgang mit KI-Tools zu sensibilisieren und so den Sicherheitsstandard im Umgang mit diesen Anwendungen zu erhöhen.

3. Allgemeine Verantwortlichkeiten

HR-Manager: Verantwortlich für die Sicherstellung der Einhaltung sowie der Dokumentationspflicht von Schulungen im Umgang mit KI-Systemen für alle Beschäftigten.

Mitarbeitende: Verantwortlich für die Einhaltung der Richtlinie und einen sensiblen sowie verantwortungsvollen Umgang mit KI-Tools.

IT-Personal: Siehe Kapitel 5.

4. Regelung

4.1. Genehmigte KI-Systeme

Für die interne Nutzung kann entsprechend der Zugriffsrechte (siehe Abschnitt 4.3) der lokale LLM-Agent *TwinMe* verwendet werden. Als externes Tool steht derzeit laut Whitelist *Aleph Alpha* zur Verfügung. Im Bereich der Entwicklung sind die KI-Agenten der CAD-Programme auf den SaaS-Plattformen *Fusion 360* und *PTC Creo+* genehmigt. Die Nutzung anderer KI-Systeme muss vorab durch die IT genehmigt werden [1].

4.2 Nutzungsregeln für KI-Systeme

Die Nutzung ist nur auf von der IT autorisierten MDM-Geräten und ausschließlich während der Arbeitszeit von *8:00 bis 20:00 Uhr* gestattet. Die Mitarbeitenden müssen einen nachweisbaren und aktuellen Status der KI-Grundschulung [2] vorweisen können. Die Verbindung zu genehmigten externen KI-Systemen darf ausschließlich über den firmeneigenen Server erfolgen. Außerhalb des Büros ist eine abhörsichere und verschlüsselte *VPN-Verbindung* zwingend erforderlich. Ohne VPN oder ohne ein autorisiertes Endgerät ist die Nutzung von KI-Systemen mit Unternehmensdaten untersagt. Backups von Chatverläufen, die Unternehmensdaten enthalten, sind nicht erlaubt. *Daten*, welche *juristische und private Personen* betreffen, dürfen nur von den Abteilungen Buchhaltung, HR, Vertrieb und Geschäftsführung unter Einhaltung der Richtlinien [3][4][5] verarbeitet werden. Für *interne Prozess- und Geschäftsdaten* ist ausschließlich das lokale LLM *TwinMe* zugelassen. *Entwicklungsdaten* mit dem *Label Standard* dürfen unter den Regeln aus Abschnitt 4.2.3. genutzt werden. Eine Validierung gemäß Abschnitt 4.5.1. ist für die Verwendung zwingend erforderlich.

V1.2 / 19.07.2025	Richtlinie zur Nutzung von KI-Systemen	Klassifizierung: Grundlagen
-------------------	--	--------------------------------

4.2.1 Nutzungsregeln für *Aleph Alpha*

Vor jeder Nutzung ist der Kontext auf sensible oder personenbezogene Daten mittels des installierten Filtersystems *PromptSecurity* zu prüfen. Das Ergebnis darf erst nach einer Plausibilitäts- und Richtigkeitsprüfung übernommen werden. Die Nutzung ist fortlaufend im Logfile zu dokumentieren unter: */BUSSNISS/AI_RULE/USE/<Name_Mitarbeiter>.md*.

Das Logfile muss folgende Informationen enthalten:

Bearbeitetes Dokument: <Name>
Datum: <Datum>
Verwendetes KI-System: <KI-System>

Werden Ergebnisse der KI-Bildsynthese verwendet, ist dies zusätzlich am Anfang der Bildunterschrift mit dem Vermerk „KI-generiert“ zu kennzeichnen.

4.2.2 Nutzungsregeln für CAD-KI-Agenten

Die *KI-Agenten* der SaaS-Plattformen *Fusion 360* und *PTC Creo+* dürfen von der *Entwicklungsabteilung* genutzt werden, sofern sie auf selektiver Ebene der bestehenden Konstruktionszeichnung eingesetzt werden. Dabei ist die Nutzung ausschließlich bis Layer 2 erlaubt. Zusätzlich müssen sämtliche Größenverhältnisse vor der Verarbeitung durch die KI um mindestens den *Faktor 11 verschleiert* werden.

KI-generierte Komponenten dürfen nur bis Layer 2 verwendet werden. Sie sind im Layout eindeutig zu *kennzeichnen und im Entwicklungslog* des Projektmanagementsystems unter Angabe der verantwortlichen Mitarbeitenden in Verbindung mit dem Hinweis „KI-Nutzung“ zu dokumentieren. Generierte sicherheitskritische Komponenten müssen gemäß dem Vier-Augen-Prinzip validiert werden. Die Nutzung von KI auf *Layer 3 ist untersagt*.

4.2.3 Nutzungsregeln des lokalen LLM-Agenten *TwinMe*

TwinMe ist ein internes Produkt der IT-Abteilung mit Zugriff auf sensible Unternehmensdaten. Der Agent wird ausschließlich innerhalb eines separaten Subnetzes (172.17.14.0/24) betrieben, das speziell für Unternehmensdaten vorgesehen ist. Die Nutzung des Systems wird automatisch mitgeloggt. Bei Verstößen gegen die Prompt-Regeln (siehe Abschnitt 4.4) oder beim Versuch, unautorisierte Daten abzufragen, greift Abschnitt 10 der Richtlinie.

4.3 Zugriffsrechte

Entsprechend den Nutzungsregeln gemäß *Abschnitt 4.2* dürfen die nachfolgend genannten Abteilungen die jeweils zugewiesenen KI-Systeme unter den genannten Bedingungen nutzen.

IT-Abteilung hat vollständigen Zugriff auf alle genehmigten KI-Systeme. Nicht genehmigte Systeme dürfen ausschließlich testweise und nur innerhalb eines abgesicherten Subnetzes (172.17.12.0/26) eingesetzt werden.

Geschäftsleitung hat Zugriff auf sämtliche genehmigte KI-Systeme. Die Nutzung nicht genehmigter Systeme ist nur unter Aufsicht des IT- und Datenschutzbeauftragten erlaubt.

Teamleiter der Entwicklungsabteilung besitzen die *gleichen Rechte wie Mitarbeitende der Entwicklung*. Zusätzlich dürfen sie den lokalen LLM TwinMe nutzen und sind berechtigt, entsprechende Freigaben zur Nutzung dieses Systems für ihre Teammitglieder zu erteilen.

V1.2 / 19.07.2025	Richtlinie zur Nutzung von KI-Systemen	Klassifizierung: Grundlagen
-------------------	--	--------------------------------

Mitarbeitende der Entwicklungsabteilung dürfen die CAD-KI-Agenten der SaaS-Plattformen *Fusion 360* und *PTC Creo+* sowie das KI-System *Aleph Alpha* im vollen Umfang verwenden. Der lokale LLM TwinMe darf eigenständig genutzt werden, sofern eine sechsmonatige Freigabe über das Formular MDTP-F-001 durch den zuständigen Teamleiter vorliegt.

Vertrieb / Buchhaltung / HR ist die Nutzung von KI-Agenten innerhalb der eingesetzten Softwareprodukte untersagt; diese Funktionen müssen deaktiviert sein. Die Nutzung des lokalen LLM TwinMe ist erlaubt. Auf Antrag eines Mitarbeitenden kann durch die Geschäftsleitung eine zeitlich begrenzte (sechsmonatige) Nutzung von *Aleph Alpha* genehmigt werden.

Sekretariat / Facility Management ist der Einsatz von KI-Agenten in Softwareprodukten nicht gestattet. Die Nutzung des lokalen LLM TwinMe sowie des KI-Systems *Aleph Alpha* ist nur unter Aufsicht einer dafür qualifizierten Person erlaubt.

4.4 Prompting

Nicht erlaubt ist die Eingabe folgender Inhalte in KI-Systeme: personenbezogene Daten von Mitarbeitenden oder Kund*innen, Geodaten, Inhalte aus gespeicherten oder verwendeten Dokumenten, Planungsdaten, Zugangsdaten sowie Konfigurationsdaten der eingesetzten Systeme. Ebenfalls untersagt ist die Weitergabe vertraulicher oder geschäftskritischer Informationen. Darüber hinaus ist der gültige Verhaltenskodex [6] zwingend zu beachten.

Zulässig sind allgemeine, anonymisierte oder abstrahierte Inhalte, beispielsweise zur Textstrukturierung, Sprachprüfung oder Ideengenerierung. Ebenso erlaubt ist die Nutzung öffentlich zugänglicher Daten wie Gesetzestexte, Normen oder veröffentlichte wissenschaftliche Ergebnisse.

4.5 Prozesssicherheit und Human-in-the-Loop

Alle Ergebnisse, die durch ein KI-System erzeugt werden, müssen durch *fachliche Expertise validiert* werden. Als fachlich qualifiziert gilt, wer mindestens ein Jahr in dem jeweiligen Anwendungsbereich tätig ist und innerhalb der letzten sechs Monate eine Schulung gemäß [7] erfolgreich abgeschlossen hat. *KI-generierte Ergebnisse dürfen nicht ohne menschliche Prüfung übernommen werden.* Autonomes Handeln der KI ist verboten. Bei der Nutzung von KI-Ergebnissen in *Geschäfts- oder Vertragstexten ist eine rechtliche Prüfung* durch unsere juristische Vertretung, die Kanzlei *Schüssel & Schlüssel AG*, verpflichtend. Bei *Unsicherheiten* hinsichtlich der Qualität oder Korrektheit eines KI-Ergebnisses ist dieses zu *verwerfen*.

Fehlerhafte Ausgaben von KI-Systemen müssen dokumentiert werden. Die Dokumentation erfolgt unter folgendem

Pfad: */BUSINESS/AI_RULE/ERROR/<DATUM_ABTEILUNG_FEHLERNAME>.md*

Zu dokumentieren sind:

Prompt: <Prompt>
KI-Antwort: <Ki-Ergebnis>
Fehlerbeschreibung: <Fehlerbeschreibung>

Die Fehlerdokumentation noch am selben Tag vorzunehmen. Falls das lokale LLM *TwinMe* fehlerhafte oder halluzinierte Ausgaben liefert, ist der Vorfall innerhalb von drei Tagen

V1.2 / 19.07.2025	Richtlinie zur Nutzung von KI-Systemen	Klassifizierung: Grundlagen
-------------------	--	--------------------------------

zusätzlich dem Datenschutzbeauftragten zu melden - inklusive Verweis (Link) auf das erstellte Dokument.

4.5.1. Fehleranalyse

Bei fehlerhaften Ergebnissen des lokalen LLM *TwinMe* sind der Datenschutzbeauftragte und ein IT-Administrator verpflichtet, innerhalb von drei Tagen die dem Prompt zugrunde liegenden Trainingsdaten zu analysieren. Sollten die Trainingsdaten korrekt sein, aber bei mindestens zehn Wiederholungen desselben Prompts weiterhin fehlerhafte Ausgaben auftreten, ist innerhalb von weiteren drei Tagen der Prompt selbst zu analysieren. Dabei muss ermittelt werden, welche Keywords zur falschen Ausgabe geführt haben. Die identifizierten Keywords sind im LLM temporär mit dem Label no output zu versehen. Dadurch wird sichergestellt, dass Nutzer*innen bei Verwendung dieser Keywords keine fehlerhaften Ergebnisse mehr erhalten. Stattdessen erscheint die standardisierte Rückmeldung: *KeywordError - No Output*.

Die Regelung bleibt so lange aktiv, bis der zugrunde liegende Fehler vollständig behoben wurde. Ein Fehler gilt als behoben, wenn bei mindestens zehn Wiederholungen des fehlerverursachenden Prompts keine fehlerhaften Ausgaben mehr auftreten. Die Prüfung ist durch den Datenschutzbeauftragten durchzuführen. Erst nach erfolgreicher Validierung kann das no output - Label wieder aufgehoben werden, sodass Prompts mit den betreffenden Keywords wieder regulär verarbeitet werden. Spezifizierter Maßnahmen zum Umgang und Debugging von Falschen Ergebnissen bei *TwinMe* sind unter [8] zu finden.

Für das KI-System *Aleph Alpha* sowie die CAD-KI-Agenten *Fusion 360* und *PTC Creo+* ist der Datenschutzbeauftragte verpflichtet, vierteljährlich die Nutzungsprotokolle mit den dokumentierten Fehlerfällen abzugleichen. Wird dabei ein *Fehlertrend von 20%* oder mehr im Verhältnis zur Gesamtnutzung festgestellt, *kann der Datenschutzbeauftragte die Nutzung des betroffenen KI-Systems untersagen* und es auf die interne Blacklist setzen. Eine erneute Freigabe ist nur nach formeller Genehmigung gemäß Richtlinie [1] zulässig.

Der Datenschutzbeauftragte ist zudem für die Archivierung aller manuellen und automatisierten Nutzungs- sowie Fehlerprotokolle über einen Zeitraum von zehn Jahren verantwortlich.

4.5.2. Qualitätssicherung & Bug-Bounty-Programm

Zur weiteren Qualitätssicherung und Prävention von Datenlecks wird durch den IT-Administrator jährlich am 01.12. ein Bug-Bounty-Programm auf der Plattform Intigriti ausgeschrieben. Dieses läuft für drei Monate als Pay-per-Bug-Modell mit einem Preisgeld von 1.000€ pro bestätigtem Fund sensibler Daten, die in Verbindung mit MDTP gebracht werden können. Genauere Beschreibungen des Programmablauf sind unter [9] enthalten.

Das Programm gilt ausschließlich für externe LLMs. Im Falle eines nachgewiesenen Datenlecks muss das betroffene KI-System unverzüglich gesperrt und auf die Blacklist gesetzt werden. Eine Freischaltung ist nur nach erneuter Genehmigung zulässig [1].

V1.2 / 19.07.2025	Richtlinie zur Nutzung von KI-Systemen	Klassifizierung: Grundlagen
-------------------	--	--------------------------------

5. Pflichten des Systembetreibers

5.1 IT-Administration

Die IT-Administration ist für die kontinuierliche Überwachung und das Monitoring aller KI-Zugriffe und -Aktivitäten verantwortlich. Zusätzlich obliegt ihr die Konfiguration und Installation aktueller Updates von *PromptSecurity* auf den Endgeräten sowie der Betrieb, die Bereitstellung und Einbindung neuer Trainingsdaten in das lokale LLM *TwinMe*.

5.2 Datenschutzbeauftragter

Der Datenschutzbeauftragte führt die Beurteilung und Freigabe der KI-Systeme durch und ist für die Sperrung von KI-Systemen bei bekannt gewordenen Sicherheitslücken zuständig. Eine tägliche Überprüfung auf Security Breaches (einfache Online Recherche nach Meldungen) der genehmigten KI-Systeme ist durchzuführen. Bei Entdeckung einer Sicherheitslücke muss das betroffene System nach [1] gesperrt werden, wobei eine Freigabe erst nach erneuter Genehmigung erfolgt [1].

Die Verantwortung für die Führung der Nutzungslogs für KI-Aktivitäten sowie die Genehmigung der Trainingsdaten für das lokale LLM TwinMe liegt ebenfalls beim Datenschutzbeauftragten. Monatlich ist ein Abgleich der automatisierten Logs nicht-interner aber genehmigter KI-Systeme den manuellen Nutzungslogs der Mitarbeitenden zur Überprüfung. Somit soll eine Regeleinhaltung über die Dokumentationspflicht gewährleistet werden. Bei Verstoß gegen die Dokumentationspflicht muss der betroffene Mitarbeitende erneut an der KI-Nutzungsschulung teilnehmen, bevor eine Wiederzulassung durch den Datenschutzbeauftragten erfolgt.

6. Schulung der Mitarbeitenden

Die *allgemeine Nutzung* erforderlichen Schulungen von KI-Systemen ist in *Abschnitt 4.2 [2]* sowie zur bereichsspezifischen *Verwendung* in *Abschnitt 4.3 [7]* geregelt.

IT-Administratoren und Datenschutzbeauftragte tragen besondere Verantwortung, um dem stetigen Wandel im Bereich Künstliche Intelligenz gerecht zu werden. Für diesen Personenkreis wird daher ein *Weiterbildungsbudget in Höhe von 2.500€* pro Person und Jahr bei unserem offiziellen Bildungspartner *KI & Tech Learning* bereitgestellt. Für entsprechende Weiterbildungsmaßnahmen kann eine Freistellung von der regulären Tätigkeit für bis zu fünf Arbeitstage beantragt werden. Die Teilnahme an einer Weiterbildung muss im Vorfeld beim CEO beantragt und genehmigt werden (siehe [10]).

7. Genehmigung von Ausnahmen

Die Nutzung nicht genehmigter KI-Systeme ist grundsätzlich untersagt und nur in einem speziell gesicherten Subnetz (172.17.12.0/28) zulässig. IT-Mitarbeitenden ist die Verwendung nicht genehmigter KI-Systeme zu Test- und Validierungszwecken ausschließlich über folgende IP-Bereiche erlaubt:

Administratoren: 172.17.12.1 - 172.17.12.10

Datenschutzbeauftragte: 172.17.12.11 - 172.17.12.20

V1.2 / 19.07.2025	Richtlinie zur Nutzung von KI-Systemen	Klassifizierung: Grundlagen
-------------------	--	--------------------------------

Die *Geschäftsführung* darf nicht genehmigte KI-Systeme ebenfalls verwenden, jedoch nur unter Aufsicht eines Administrators oder Datenschutzbeauftragten und ausschließlich über die IP-Adressen 172.17.12.21 - 172.17.12.30.

Externen Dienstleistern kann die Nutzung nicht genehmigter KI-Systeme im Ausnahmefall genehmigt werden. Voraussetzung ist ein *nachvollziehbar begründetes Interesse*, das gegenüber dem *Datenschutzbeauftragten* *schriftlich darzulegen* ist. Der Datenschutzbeauftragte kann daraufhin die Nutzung des entsprechenden KI-Systems für eine festgelegte Dauer genehmigen.

Die Nutzung durch externe Dienstleister ist ausschließlich unter Aufsicht des Datenschutzbeauftragten, der die Genehmigung erteilt hat, und nur innerhalb des IP-Bereichs 172.17.12.31 - 172.17.12.60 zulässig. Jede Nutzung externer KI-Systeme ist verpflichtend mitlaufend zu protokollieren. Hierfür ist das folgende Skript zu verwenden:

/BUSINESS/AI_RULE/SCRIPTS/extern_log.sh

Die dabei erzeugten Protokolle werden im Verzeichnis

/BUSINESS/AI_RULE/LOGS_EXTERN

gespeichert und sind für einen Zeitraum von zehn Jahren aufzubewahren.

Die Verantwortung für die sichere und vollständige Archivierung dieser Protokolle liegt beim Datenschutzbeauftragten.

8. Berichterstattung und Eskalation

Eskalations- und Qualitätssicherungsmaßnahmen sind in Abschnitt 4.5 beschrieben.

Sollte der begründete Verdacht bestehen, dass eine Mitarbeitender Unternehmensdaten außerhalb der hier definierten Regeln, insbesondere auf privaten Endgeräten mit KI-Systemen, verwendet, ist dies unverzüglich dem Datenschutzbeauftragten zu melden - per E-Mail oder telefonisch.

V1.2 / 19.07.2025	Richtlinie zur Nutzung von KI-Systemen	Klassifizierung: Grundlagen
-------------------	--	--------------------------------

9. Gültigkeit, Verwaltung und Überprüfung

Diese Richtlinie tritt am 24.07.2025 in Kraft. Die Verantwortung für Pflege, regelmäßige Aktualisierung und Überwachung dieses Dokuments obliegt dem internen Informationssicherheitsbeauftragten (ISB) als Dokumenteneigentümer.

Jährlich ist ein Audit durchzuführen, bei dem stichprobenartig mindestens zehn Mitarbeiterinnen und Mitarbeiter auf ihr grundlegendes Verständnis im Umgang mit Künstlicher Intelligenz entsprechend dieser Richtlinie sowie Dokument [2] geprüft werden. Die Überprüfung dient auch der Sicherstellung der in Abschnitt 4.5 festgelegten Qualitätsstandards. Die Audit-Ergebnisse sind vollständig zu dokumentieren und auszuwerten, um wiederkehrende Muster und Entwicklungen zu identifizieren. Basierend auf dieser Analyse erstellt der ISB einen Verbesserungsplan mit konkreten Maßnahmen zur Stärkung der IT-Sicherheit und zur Prävention künftiger Risiken.

Sämtliche daraus resultierenden Maßnahmen sind verbindlich in die Richtlinie zu integrieren und in der überarbeiteten Fassung zu veröffentlichen. Der ISB trägt die Verantwortung für Planung, Durchführung und Überwachung der Audits sowie für die Umsetzung der daraus abgeleiteten Sicherheitsmaßnahmen. Zur umfassenden Information aller Mitarbeitenden wird die finale Version dieser Richtlinie prioritär per E-Mail an alle Beschäftigten versendet.

10. Folgen bei Verstößen

Verstöße gegen diese Nutzungsrichtlinie gelten als Verletzung der Sicherheitsbestimmungen der MDTP AG und können arbeitsrechtliche Maßnahmen (durch den Datenschutzbeauftragten in Absprache mit dem CEO) zur Folge haben. Je nach Schwere des Verstoßes sind diese bis hin zur Kündigung des Arbeitsverhältnisses möglich. Zusätzlich können bei Regelverstößen die Zugriffsrechte auf KI-Systeme zeitweise oder permanent entzogen werden. Sämtliche Verstöße werden schriftlich erfasst und dem Informationssicherheitsbeauftragten (ISB) zur weiterführenden Bearbeitung übermittelt.



Berlin, den 19.07.2025

Ort, Datum

René Kost

CEO

V1.2 / 19.07.2025	Richtlinie zur Nutzung von KI-Systemen	Klassifizierung: Grundlagen
-------------------	--	--------------------------------

Literaturverzeichnis

- [1] - MDTP-RL-007-033 [KI-System-Genehmigungsrichtlinie]
- [2] - MDTP-SW-007-021 [KI-Grundlagen]
- [3] -MDTP-RL-007-013 [Anwendung_DSGVO]
- [4] - MDTP-RL-007-012 [Anwendung_TTDSG]
- [5] - MDTP-RL-007-011 [Anwendung_GeschGehG]
- [6] - MDTP-BU-003-001[Verhaltens- und Gleichberechtigungskodex]
- [7] - MDTP-SW-011-[KI-Bias-Erkennung und ethischen Aspekten]
- [8] - MDTP-RL-007-040[KI-Prüfungsroutine und Fehleranalyse – TwinMe]
- [9] - MDTP-AW-003-013[Qualitätssicherung externer KI mit Bug Bountry]
- [10] – MDTP-SW-020-[KI-Weiterbildung von Mitarbeitenden der KI-Abteilung]