Der Landesbeauftragte für
DATENSCHUTZ und
INFORMATIONSFREIHEIT
Mecklenburg-Vorpommern

Datenschutzkonformer und digital-souveräner Einsatz von

Künstlicher Intelligenz

mit beschränktem oder minimalem Risiko

Ein praxisorientierter Leitfaden

Impressum

Redaktion:

Der Landesbeauftragte für Datenschutz und
Informationsfreiheit Mecklenburg-Vorpommern (LfDI MV)

Postanschrift:

Schloss Schwerin
Lennéstraße 1
19053 Schwerin

Dienststelle:

Werderstraße 74 a
19055 Schwerin

Telefon: 0385 59494-0
E-Mail: info@datenschutz-mv.de

Website: www.datenschutz-mv.de | www.informationsfreiheit-mv.de

Bildnachweise:

©pixabay/geralt

Stand:

Juli 2026

Hinweis zu Geschlechtsangaben:

Im vorliegenden Leitfaden schreiben wir Personenbezeichnungen möglichst geschlechtsneutral oder alternativ mit Stern. Der Leitfaden adressiert alle Geschlechter.

Personenbezeichnungen aus Gesetzestexten, die im generischen Maskulinum geschrieben sind, wurden aufgrund der Zitierbarkeit der Gesetzestexte im Original übernommen.

Management Zusammenfassung

Dieser Leitfaden adressiert sowohl Verantwortliche aus kleineren Verwaltungen als auch kleine und mittelständische Unternehmen (KMU) – im folgenden „Organisation" genannt.

Die Anforderungen an KI-Systeme unterscheiden sich nach deren Risiko und der Rolle der Organisation im jeweiligen Anwendungsfall. Auch KI-Systeme mit begrenztem oder minimalem Risiko können optimal umgesetzt eine gute Unterstützungsleistung bieten – z. B. beim Zusammenfassen von Texten oder einem internen Chatbot.

Die KI-Verordnung fokussiert jedoch auf Hochrisiko-Systeme, weshalb die Mehrheit der existierenden Leitfäden auch genau diese KI-Systeme adressieren - was mit einer hohen Komplexität verbunden ist. Leitfäden speziell für KI-Systeme mit begrenztem oder minimalem Risiko wären hingegen weniger komplex. Der vorliegende Leitfaden soll diese Lücke schließen und sich ausschließlich auf KI-Systeme mit minimalem oder begrenztem Risiko konzentrieren. Er ist praxisorientiert und soll direkte praktische Unterstützung für einen digital-souveränen, sicheren und datenschutzkonformen Einsatz dieser KI-Systeme bieten.

Executive Summary

This guideline addresses data controllers in smaller public bodies as well as small and medium-sized enterprises (SME), hereafter referred to as "organisations".

The requirements for AI systems differ depending on their risk and the role of the organisation in the respective application case. Even AI systems with limited or minimal risk can, if implemented optimally, provide good support – e.g. for text summaries or an internal chatbot.

However, the AI Act focuses on high-risk systems, which is why the majority of existing guidelines also address these AI systems. Guidance tailored to AI systems with minimal or limited risk would be less complex. This guideline aims to fill this gap and focuses exclusively on AI systems with minimal or limited risk. It is oriented on the practical use and intended to provide direct support for the digitally sovereign, safe and data protection-compliant use of these AI systems.

Inhalt

Versionierung

Version	Datum	Bemerkung
1.0	Juli 2026	Erste Veröffentlichung

1 Einleitung

In den nachfolgenden Kapiteln wird kurz die Zielsetzung des Leitfadens beschrieben und Hinweise zum Inhalt und der Struktur des Dokuments gegeben.

1.1 Ziel des Leitfadens

Künstliche Intelligenz (KI) ist kein neues Thema, steht aber seit der Veröffentlichung von großen Sprachmodellen wie z. B. ChatGPT im Fokus technischer Weiterentwicklungen und nimmt vermehrt Einfluss auf den Alltag und das Berufsleben.[1]

Die Anforderungen an KI-Systeme unterscheiden sich nach deren Risiko und der Rolle der Organisation im jeweiligen Anwendungsfall. Auch KI-Systeme mit begrenztem oder minimalem Risiko können optimal umgesetzt eine gute Unterstützungsleistung in der Organisation bieten – z. B. beim Zusammenfassen von Texten oder einem internen Chatbot.

Die EU Verordnung 2024/1689[2] – im Folgenden: KI-Verordnung – reguliert hauptsächlich Hochrisiko-Systeme, weshalb eine Vielzahl der existierenden Leitfäden auch genau diese Systeme adressiert. Die Regelungen in der KI-Verordnung für Hochrisiko-KI-Systeme sind komplex, weshalb die Einstiegshürde in die meisten Leitfäden hoch ist – obwohl die Regelungen für KI-Systeme mit begrenztem oder minimalem Risiko überschaubar sind. Damit sind viele Leitfäden komplexer als notwendig, wenn eine Organisation KI-Systeme mit begrenztem oder minimalem Risiko anwenden möchte. Der vorliegende Leitfaden soll diese Lücke schließen und sich ausschließlich auf KI-Systeme mit minimalem oder begrenztem Risiko konzentrieren.

Der vorliegende Leitfaden ist praxisorientiert und soll direkte praktische Unterstützung beim digital-souveränen, sicheren und datenschutzkonformen Einsatz von KI-Systemen mit minimalem oder begrenztem Risiko bieten.

1 Einen Überblick über die KI-Entwicklungen, auch im Sinne einer minimalistischen strategischen Vorausschau, kann u. a. der Gartner AI Hype Cycle geben (Gartner verwendet Cookies: Privacy Cookie Policy), URL: https://www.gartner.com/en/articles/hype-cycle-for-artificial-intelligence, letzter Zugriff: 23.04.26.

2 Verordnung (EU) 2024/1689 des europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche Intelligenz), veröffentlicht im Amtsblatt der Europäischen Union am 12. Juli 2024.

Dabei werden bestehende Orientierungshilfen z. B. von der Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder (DSK),[3] zielgerichtet eingebunden. KI-Systeme bieten nicht nur Potenziale, sie bergen auch Risiken oder sind selbst bedroht. Der Lagebericht des Bundesamtes für Sicherheit in der Informationstechnik (BSI)[4] bietet einen regelmäßig aktualisierten Überblick über die Bedrohungslage.

1.2 Hinweise zum Leitfaden

Dieser Leitfaden adressiert sowohl Verantwortliche aus kleineren Verwaltungen als auch kleine und mittelständische Unternehmen (KMU). Er adressiert somit den öffentlichen als auch den nicht-öffentlichen Bereich. Der Leitfaden soll als praktische Unterstützung dienen und minimiert die rechtlichen Herleitungen auf das für die Nachvollziehbarkeit notwendige Maß. Er bietet strategische Entscheidungshilfen und stellt operative Grundsätze dar, bleibt aber produktneutral. Wie bereits im Kapitel 1.1 dargelegt, konzentriert sich dieser Leitfaden auf KI-Systeme mit minimalem oder begrenztem Risiko.

Für die Nachvollziehbarkeit notwendige Hintergrundinformationen sind neben den Praxiselementen enthalten. Bereits in Referenzdokumenten dargelegte Sachverhalte werden in diesem Leitfaden nicht erneut dargestellt, es wird ausschließlich auf die Referenzdokumente verwiesen. Insofern verknüpft dieser Leitfaden bereits existierende Referenzdokumente themenbezogen miteinander.

Nachfolgend ist dieses Dokument grob wie folgt strukturiert: Zunächst werden notwendige Vorbetrachtungen angestellt, um eine einheitliche theoretische Grundlage zu schaffen, gefolgt von einzelnen Maßnahmen zur Umsetzung der rechtlichen Anforderungen in der Praxis. Aufgrund der Komplexität und für eine bessere Lesbarkeit werden die zur Umsetzung beschriebenen Maßnahmen in Boxen dargestellt.

3 Website DSK, URL: https://www.datenschutzkonferenz-online.de/, letzter Zugriff: 10.6.26.

4 Website BSI, URL: https://www.bsi.bund.de/DE/Service-Navi/Publikationen/Lagebericht/lagebericht_node.html, letzter Zugriff: 10.06.26.

2 Praxisorientierter Leitfaden

Im Sinne eines roten Fadens bauen die nachfolgenden Kapitel aufeinander auf. Im Anhang A1 ist eine Checkliste zu finden, die diesen Leitfaden praxisnah zusammenfasst. Im Anhang A2 ist ein zusätzlicher Artikel zu KI-Agenten beigefügt.

2.1 Notwendige Vorbetrachtungen

2.1.1 KI-Systeme und KI-Modelle

Zunächst ist zwischen KI-Systemen, KI-Modellen und KI-Modellen mit allgemeinem Verwendungszweck zu unterscheiden.

Die Definition für KI-Systeme findet sich in Artikel 3 Nr. 1 KI-Verordnung und geht davon aus, dass es sich hierbei um ein maschinengestütztes System handelt, das für einen in unterschiedlichem Grade autonomen Betrieb ausgelegt ist und das nach seiner Betriebsaufnahme anpassungsfähig sein kann. Zudem ist es erforderlich, dass dieses KI-System aus den erhaltenen Eingaben (z. B. Prompts) explizite oder implizite Ziele ableitet, wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen erstellt, welche die physischen oder virtuellen Umgebungen beeinflussen können.[5]

Eine Definition für ein KI-Modell findet sich nicht in der KI-Verordnung. Für diesen Leitfaden wird daher angenommen, dass ein KI-Modell das Ergebnis ist, welches sich aus der Anwendung eines KI-Algorithmus auf Trainingsdaten ergibt. Es dient als Komponente eines KI-Systems.[6]

Ein KI-Modell mit allgemeinem Verwendungszweck wird hingegen in der KI-Verordnung in Artikel 3 Nr. 63 legal definiert. In der Regel handelt es sich hierbei um die gängigen großen Sprachmodelle (LLM).

Für diesen Leitfaden wird davon ausgegangen, dass ein KI-System aus einer Möglichkeit zur Eingabe besteht, um dem darin enthaltenen KI-Modell, z. B ein LLM, Aufgaben zuzuweisen. Zudem werden die Ausgaben (z. B. Text, Bild, Video, Ton) des KI-Modells den Nutzenden für deren weitere Verwendung zur Verfügung gestellt. Vorgaben wie Anpassungsfähigkeit,

5 DSK (Juni 2025) Orientierungshilfe zu empfohlenen technischen und organisatorischen Maßnahmen bei der Entwicklung und beim Betrieb von KI-Systemen, Version 1.0, URL: https://www.datenschutz-mv.de/static/DS/Dateien/Publikationen/Broschueren/OH_KI-Systeme.pdf, Seite 26.

6 Ebd.

autonomer Betrieb und die Möglichkeit der physischen oder virtuellen Beeinflussung der Umgebung werden ebenso angenommen.

2.1.2 Rechtsrahmen für Daten in KI-Systemen

Unabhängig davon, ob das geplante System als KI im Sinne der KI-Verordnung gewertet werden kann oder nicht, gelten aufgrund der umfangreichen Datenverarbeitung meist weitere rechtliche Normen. Werden personenbezogene Daten verarbeitet, sind die Datenschutz-Grundverordnung (DS-GVO) und die entsprechenden landesrechtlichen Regelungen für öffentliche Stellen einzuhalten. Die einzelnen Verordnungen und Gesetze ergänzen sich somit.

Für diesen Leitfaden werden die wichtigsten Grundsätze der DS-GVO kurz angerissen und im Übrigen auf andere Quellen hinsichtlich des Rechtsrahmens verwiesen.[7]

Hinsichtlich der Verarbeitung von personenbezogenen Daten gelten u. a. die Grundsätze gemäß Artikel 5 Abs. 1 DS-GVO der Rechtmäßigkeit, der Verarbeitung nach Treu und Glauben, der Transparenz, der Zweckbindung, der Datenminimierung und der Richtigkeit. Für die Einhaltung dieser Grundsätze ist der Verantwortliche gemäß Artikel 5 Abs. 2 DS-GVO verantwortlich und muss deren Einhaltung nachweisen können. Weiterhin ist es erforderlich, dass der Zweck einer Verarbeitungstätigkeit vorher festgelegt, eindeutig und legitim sein muss. Das bedeutet, dass vor dem Einsatz eines KI-Systems dessen Einsatzzweck genau definiert und geprüft sein sollte. Vor alledem ist es für den Verantwortlichen zwingend notwendig, eine entsprechende Rechtsgrundlage vorweisen zu können, um überhaupt personenbezogene Daten innerhalb eines KI-Systems verarbeiten zu dürfen. Dies gilt insbesondere für das Training von KI-Systemen oder KI-Modellen. Sind diese Voraussetzungen gegeben, sind die Risiken, welche aus der Datenverarbeitung für die Betroffenen entstehen können, zu bestimmen und durch geeignete technische und organisatorische Maßnahmen nach Artikel 32 DS-GVO zu minimieren. Darüber hinaus müssen Verantwortliche, welche KI-Systeme einsetzen wollen, ihren Informationspflichten gemäß Artikel 13 und 14 DS-GVO nachkommen. Somit wird gewährleistet, dass betroffene Personen (z. B. Beschäftigte, Kund*innen, Bürger*innen) ihre Betroffenenrechte gegenüber dem Verantwortlichen ausüben können. Betroffenenrechte gemäß der DS-GVO können dabei

7 Ebd. sowie DSK (Mai 2024): Künstliche Intelligenz und Datenschutz Version 1.0, URL: https://www.datenschutz-mv.de/static/DS/Dateien/Publikationen/Broschueren/20240506_OH_KI_und_Datenschutz.pdf.

u. a. das Recht auf Auskunft oder Berichtigung, das Recht auf Löschung oder das Recht sein, nicht einer automatisierten Entscheidung im Einzelfall unterworfen zu werden.

> Dieser Leitfaden betrachtet den rechtlichen Rahmen für personenbezogene Daten in KI-Systemen. Er soll einen datenschutzkonformen Einsatz von KI-Systemen unterstützen. Entsprechend stehen in diesem Leitfaden die KI-Verordnung und die DS-GVO im Fokus. Darüber hinaus profitieren auch proprietäre Daten[8] von den Maßnahmen zum Schutz personenbezogener Daten.

2.1.3 Risikostufen

Werden KI-Systeme eingesetzt, ist eine Betrachtung des Risikos erforderlich, z. B. welchen Einfluss der Einsatz der KI-Systeme auf die Grundrechte der Nutzenden hat. Die KI-Verordnung definiert Risiko dabei als Kombination aus der Wahrscheinlichkeit des Auftretens eines Schadens und der Schwere dieses Schadens, siehe Artikel 3 Nr. 2 KI-Verordnung.

Darüber hinaus finden sich in der KI-Verordnung vier Risikostufen :

Unannehmbares/inakzeptables Risiko

Praktiken, die gemäß Artikel 5 der KI-Verordnung verboten sind, da diese mit den Werten der EU unvereinbar sind. Hierzu gehören u. a. Social Scoring und manipulative KI-Systeme.

Hohes Risiko

Diese Systeme stellen ein hohes Risiko hinsichtlich der Eintrittswahrscheinlichkeit und des Schadensausmaßes dar. Derartige KI-Systeme werden durch den Artikel 6 i.V.m. Anhang I und III der KI-Verordnung erfasst. Beispiele sind KI-Systeme die bestimmungsgemäß zur Bewertung von Lernergebnissen verwendet werden sollen oder KI-Systeme die bestimmungsgemäß für die Einstellung oder Auswahl natürlicher Personen verwendet werden sollen. Die KI-Verordnung legt auf KI-Systeme mit hohem Risiko einen besonderen Fokus und widmet sich dieser Thematik in den Artikeln 6 bis 49.

8 Proprietäre Daten gehören einem Unternehmen oder einer Person und sind nicht öffentlich zugänglich. Sie sind zumeist von strategischer oder unternehmerischer Relevanz (z. B. Produktspezifikationen, Herstellungsprozesse, Formulierungen).

Begrenztes Risiko

KI-Systeme mit begrenztem Risiko sind solche Systeme, deren Risiko durch z. B. Transparenzmaßnahmen gemanagt werden kann. Artikel 50 der KI-Verordnung benennt diese genauer. Als Beispiele für KI-Systeme mit begrenztem Risiko wären u. a. Chatbots oder KI-Systeme, einschließlich KI-Systemen mit allgemeinem Verwendungszweck, die synthetische Audio-, Bild-, Video- oder Textinhalte erzeugen, zu nennen.

Minimales Risiko

Alle übrigen KI-Systeme, die zuvor nicht erfasst worden sind, wird ein minimales Risiko zugeordnet. Beispiele hierzu wären Spam-Filter oder KI-Systeme in Videospielen. Neben den KI-Kompetenz-Pflichten gemäß Artikel 4 KI-Verordnung gelten hier keine weiteren speziellen Verpflichtungen. Artikel 95 i. V. m. Erwägungsgrund 165 der KI-Verordnung regt jedoch die Einhaltung eines freiwilligen Verhaltenskodex an.

Die Einstufung eines KI-Systems hinsichtlich des Risikos ist also abhängig vom jeweiligen Einsatzzweck. So kann ein KI-System je nach Einsatz ein hohes Risiko aufweisen, z. B. für die Einstellung oder Auswahl natürlicher Personen im Bewerbungsverfahren darstellen, oder auch nur ein minimales Risiko, z. B. beim Zusammenfassen von Texten.

2.1.4 Akteure (Rollen)

Neben der Frage, welches Risiko ein KI-System begründen kann, muss betrachtet werden, welche Rolle eine Organisation beim Einsatz eines KI-Systems einnimmt. Die KI-Verordnung definiert diverse Rollen entlang der KI-Wertschöpfungskette und orientiert sich hierbei an den Normen des EU-Produktrechts. Dabei nennt Artikel 3, Nr. 8 der KI-Verordnung die Akteure (Rollen) innerhalb der KI-Verordnung. Diese sind Anbieter, Produkthersteller, Betreiber, Bevollmächtigten, Einführer oder Händler.

Für diesen Leitfaden sind die Rollen Anbieter als auch Betreiber wichtig und werden gemäß KI-Verordnung Art. 3 Nr. 3 und Nr. 4 wie folgt definiert.

Anbieter

„Für die Zwecke dieser Verordnung bezeichnet der Ausdruck „Anbieter" eine natürliche oder juristische Person, Behörde, Einrichtung oder sonstige Stelle, die ein KI-System oder ein KI-Modell mit allgemeinem Verwendungszweck entwickelt oder entwickeln lässt und es unter ihrem eigenen Namen oder ihrer Handelsmarke in Verkehr bringt oder das KI-

System unter ihrem eigenen Namen oder ihrer Handelsmarke in Betrieb nimmt, sei es entgeltlich oder unentgeltlich."

Betreiber

„Für die Zwecke dieser Verordnung bezeichnet der Ausdruck „Betreiber" eine natürliche oder juristische Person, Behörde, Einrichtung oder sonstige Stelle, die ein KI-System in eigener Verantwortung verwendet, es sei denn, das KI-System wird im Rahmen einer persönlichen und nicht beruflichen Tätigkeit verwendet."

2.1.5 Kontextabhängigkeit

Oft wird in öffentlichen Debatten oder im Marketing nahezu jede Form der Automatisierung unter dem Begriff „künstliche Intelligenz" subsumiert. Aber nicht jede Form der Automatisierung (wie z. B. regelbasierte Systeme) ist künstliche Intelligenz im Sinne der KI-Verordnung.

Eine präzise Verwendung der Begrifflichkeiten sowie eine genaue Prüfung des Vorhabens sind essentiell. Dies ist notwendig, damit keine rechtlichen Anforderungen übersehen werden, aber auch, damit keine Maßnahmen ergriffen werden, die eigentlich gar nicht nötig wären.

2.2 Grundlegender Managementprozess

KI kann zu Effizienzsteigerungen in einer Organisation beitragen. Dafür muss aber eine hinreichende KI-Qualität[9] erzielt werden. Obwohl sich keine einheitliche Definition durchgesetzt hat, kann unter KI-Qualität Folgendes verstanden werden:

„KI-Qualität beschreibt den Grad, in dem ein KI-System zuverlässig, sicher, fair und im Einklang mit fachlichen, ethischen und regulatorischen Anforderungen funktioniert. Es geht also nicht nur um technische Leistungsfähigkeit, sondern darum, ob eine KI vertrauenswürdig, robust und verantwortungsvoll einsetzbar ist."[10]

Hohe KI-Qualität trägt also nicht nur zur Einhaltung rechtlicher Vorschriften bei, sondern auch zu einem optimalen Ergebnis des KI-Einsatzes und der dafür notwendigen Investitionen. Eine Studie des Massachusetts Institute of Technology (MIT) aus dem Jahr 2025[11] zeigt, dass 95% der Unternehmen, die in eine Integration generativer KI investiert haben, nicht hinreichend

9 Begriff wird oft auch synonym mit KI-Güte verwendet.

10 Website VDE Verband der Elektrotechnik Elektronik Informationstechnik e.V., URL: https://www.vde.com/topics-de/kuenstliche-intelligenz/blog/ki-und-qualitaet, letzter Zugriff: 23.04.26.

erfolgreich sind. Dabei zeigt sich, dass der Erfolg erheblich vom gewählten Ansatz abhängt. Nicht jeder pauschale Einsatz generativer KI ist im Sinne einer hinreichenden KI-Qualität geeignet.

M 0.1 – Prozess für KI-Management

Ausdrücklich empfohlen ist die Etablierung eines KI-Management-Prozesses, der unter Beachtung aller oben genannten Teilaspekte eine optimale KI-Qualität zum Ziel hat. Der Einsatz von KI kann komplex sein und ohne einen systematischen KI-Management-Prozess können schnell relevante Anforderungen übersehen werden. Nur die Möglichkeit, eine hohe KI-Qualität hinsichtlich der Erfüllung funktionaler Anforderungen (es gibt auch Fragestellungen, die nicht gut durch KI gelöst werden können) zu erreichen, rechtfertigt den Einsatz und die Verarbeitung besonders schützenswerter Daten, sei es personenbezogene oder proprietäre Daten.

M 0.1	**Prozess für KI-Management**	
Ausdrücklich empfohlen ist die Etablierung eines KI-Managementprozesses mit dem Ziel einer optimalen KI-Qualität.[12] Dies betrifft auch die Einbindung von Beauftragten (Datenschutz, Informationssicherheit, Betriebs-/Personalrat), sofern vorhanden. Die kontinuierliche Sicherstellung der Gesetzeskonformität ist verpflichtend. Für den Bereich des Datenschutzes schlägt das Standard-Datenschutzmodell (SDM) im Abschnitt D4.4 einen möglichen „Datenschutzmanagement-Prozess“[13] vor. Dieser basiert auf den allgemeinen Plan-Do-Check-Act-Zyklus, der auch für einen KI-Management-Prozess adaptiert werden kann. Dabei handelt es sich um einen kontinuierlichen (Verbesserungs-)Prozess. Im Falle relevanter Änderungen (z. B. an den rechtlichen Normen oder am Anwendungsfall) beginnt der Management-Prozess von Neuem.		

11 Aditya Challapally et al. für MIT (Juli 2025): The GenAI Divide. State of AI in Business 2025, URL; https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf, letzter Zugriff: 10.06.26.

12 Website VDE Verband der Elektrotechnik Elektronik Informationstechnik e.V., URL https://www.vde.com/topics-de/kuenstliche-intelligenz/blog/ki-und-qualitaet, letzter Zugriff: 10.06.26.

13 DSK (2025): Das Standard-Datenschutzmodell. Eine Methode zur Datenschutzberatung und -prüfung auf der Basis einheitlicher Gewährleistungsziele, Version 3.1.a, URL: https://www.datenschutz-mv.de/static/DS/Dateien/Datenschutzmodell/SDM-Methode-V31a.pdf.

2.3 Empfohlene Maßnahmen

Im Folgenden werden einzelne Maßnahmen angesprochen, um den Einsatz von KI-Systemen digital-souverän, datenschutzkonform und sicher zu gestalten.

M 1.1 – Bestandsaufnahme und Dokumentation

Die rechtliche Einordnung eines KI-Systems ist kontextabhängig. Insofern ist es essentiell, dass eine vollständige Bestandsaufnahme und Dokumentation der eigenen KI-Systeme erfolgt.

M 1.1	**Bestandsaufnahme und Dokumentation**	
Vollständige Bestandsaufnahme und Dokumentation des geplanten KI-Systems. Die Dokumentation schließt zumindest die funktionalen Anforderungen, die verarbeiteten Daten, den konkreten Anwendungsfall und im Falle der Verarbeitung personenbezogener Daten die Rechtsgrundlage für die Verarbeitung ein und muss zumindest ermöglichen: - zu entscheiden, wie die KI-Verordnung Anwendung findet (Akteur, Risiko), - ob und welche personenbezogenen Daten verarbeitet werden (nebst Rechtsgrundlage für die Verarbeitung) und damit die DS-GVO Anwendung findet und - ob ggf. noch weitere rechtliche Rahmenbedingungen relevant sind.		

Für Bestandssysteme, die ggf. noch nicht dokumentiert sind, ist diese Maßnahme nachzuholen.

M 1.2 – Rollenbestimmung

Welche Pflichten sich aus der KI-Verordnung ergeben, wird maßgeblich durch die Rollen als z. B. Anbieter oder Betreiber von KI bestimmt (siehe auch Kapitel 2.1.4).

M 1.3	**Rollenbestimmung gem. KI-Verordnung**	
Bestimmung der Rolle (Akteure) der Organisation gemäß Artikel 3 der KI-Verordnung für das betrachtete KI-System. Artikel 3, Nr. 8 der KI-Verordnung gibt folgende Rollen (Akteure) vor: Anbieter, Produkthersteller, Betreiber, Bevollmächtigten, Einführer und Händler.		

Im weiteren Verlauf wird davon ausgegangen, dass die Organisation/Verantwortlicher entweder Betreiber oder Anbieter eines KI-Systems ist.

Weitere praktische Hinweise zur Unterscheidung zwischen Betreiber und Anbieter sind unter anderem bei der Deutschen Industrie- und Handelskammer (DIHK)[14] zu finden.

M 1.3 – Risikoeinstufung

Die Pflichten, die sich aus der KI-Verordnung ergeben, ändern sich mit der Risikoeinstufung des KI-Systems. Die Risikoeinstufung wiederum ist abhängig vom konkreten Anwendungsfall. Welche Risikostufen es gibt und wie diese definiert sind, wird im Kapitel 2.1.3 betrachtet.

M 1.3	**Risikoeinstufung gem. KI-Verordnung**	
Einstufung des Risikos des KI-Systems im konkreten Anwendungsfall gemäß KI-Verordnung. Die KI-Verordnung unterscheidet zwischen unannehmbarem Risiko (verboten), hohem Risiko (spezielle Vorschriften), begrenztem Risiko (Transparenzvorschriften) und minimalem Risiko (keine gesonderten Vorgaben). Die Vorgaben nach Artikel 4 der KI-Verordnung für KI-Kompetenz sind für alle (erlaubten) Risikostufen obligatorisch.		

Im weiteren Verlauf wird davon ausgegangen, dass das KI-System einem begrenzten oder minimalen Risiko zuzuordnen ist.

M 1.4 – Anforderungen aufgrund der Verarbeitung personenbezogener Daten

Im Falle der Verarbeitung personenbezogener Daten gelten speziell die Anforderungen der DS-GVO entlang des gesamten KI-Lebenszyklus. Übergreifendes Ziel ist die Sicherstellung der Rechtmäßigkeit der Verarbeitung und Wahrung der Betroffenenrechte. Um dies Ziel zu erreichen hilft die Orientierung an den Gewährleistungszielen des Datenschutzes, bestehend aus Datenminimierung, Verfügbarkeit, Vertraulichkeit, Integrität, Intervenierbarkeit, Transparenz und Nichtverkettung.

M 1.4	**Anforderungen gem. DS-GVO**	
Die Verarbeitung personenbezogener Daten ist nur gestattet, wenn eine Rechtsgrundlage vorliegt und die Zweckbindung sichergestellt ist. Danach sind entlang des gesamten KI-		

14 Website DIHK, URL; https://www.dihk.de/de/serviceportal/fuer-gewerbetreibende-unternehmer/3-wer-muss-welche-regeln-aus-der-ki-verordnung-beachten-163292, letzter Zugriff: 10.06.26.

Lebenszyklus (Design, Entwicklung, Einführung, Betrieb und Ausmusterung) die Betroffenenrechte umzusetzen. Findet ein Training, ein Finetuning oder eine Erweiterung durch zusätzliche Datenquellen (wie beispielsweise bei RAG-Systemen) unter Verwendung personenbezogener Daten statt, sind zusätzliche Risiken für den Datenschutz (z. B. eine Extraktion der Trainingsdaten) zu beachten und durch geeignete Maßnahmen zu adressieren. Das Training eines KI-Modells mit personenbezogenen Daten bedarf regelmäßig einer gesonderten Rechtsgrundlage. Die kontinuierliche Einbindung der Beauftragten für Datenschutz und IT-Sicherheit (soweit vorhanden) sollte sichergestellt sein.

Die DSK-Orientierungshilfe ‚Künstliche Intelligenz und Datenschutz' aus Mai 2024[15] bietet weitere Ausführungen sowie einen Überblick über datenschutzrechtliche Kriterien, die für eine datenschutzkonforme Nutzung von KI-Systemen zu berücksichtigen sind.

Die DSK-Orientierungshilfe zu empfohlenen technischen und organisatorischen Maßnahmen[16] bei der Entwicklung und dem Betrieb von KI-Systemen aus Juni 2025 richtet sich vorrangig an Herstellern und Entwickler von KI-Systemen und gibt weitere Hilfestellungen bei der datenschutzkonformen Entwicklung von datenschutzkonform einsetzbaren KI-Systemen .

Das Standard-Datenschutzmodell (SDM)[17] bietet Hilfestellungen bei der Umsetzung von technischen und organisatorischen Maßnahmen, um die rechtlichen Anforderungen in der Praxis umzusetzen.

M 1.5 – Anforderungen für KI-Systeme mit begrenztem Risiko nach KI-Verordnung

Gemäß der KI-Verordnung ergeben sich für die in diesem Leitfaden betrachten KI-Systeme mit begrenztem oder minimalem Risiko die nachfolgend beschriebenen Anforderungen.

15 DSK (Mai 2024): Künstliche Intelligenz und Datenschutz Version 1.0, URL: https://www.datenschutz-mv.de/static/DS/Dateien/Publikationen/Broschueren/20240506_OH_KI_und_Datenschutz.pdf.

16 DSK (Juni 2025) Orientierungshilfe zu empfohlenen technischen und organisatorischen Maßnahmen bei der Entwicklung und beim Betrieb von KI-Systemen, Version 1.0, URL: https://www.datenschutz-mv.de/static/DS/Dateien/Publikationen/Broschueren/OH_KI-Systeme.pdf.

17 Website LfDI MV, Standard-Datenschutzmodell, URL: https://www.datenschutz-mv.de/datenschutz/datenschutzmodell/, letzter Zugriff: 10.06.26, s. auch Fußnote 13.

KI-Kompetenz

Unabhängig von der Risikoeinstufung gelten nach Artikel 4 der KI-Verordnung Anforderungen an die KI-Kompetenz. Demnach sind Anbieter und Betreiber von KI-Systemen verpflichtet, ein „ausreichendes Maß" an KI-Kompetenzen sicherzustellen. Die Frage nach dem „ausreichenden Maß" ist wiederum kontextabhängig. Sie ist abhängig vom KI-System, von der Risikoeinstufung und auch von der Aufgabe des betrachteten Personals. So können je nach Fall strategische KI-Kompetenzen aber auch technische, rechtliche und ethische Kenntnisse, ebenso wie Risikobewusstsein und praktische Anwendungsfertigkeiten erforderlich sein. Unter anderem die Bundesnetzagentur stellt weitere Information zum Thema KI-Kompetenz nach Artikel 4 der KI-Verordnung zur Verfügung.[18]

Transparenzpflichten

Für Anbieter und Betreiber von KI-Systemen mit begrenztem Risiko gelten gemäß Artikel 50 der KI-Verordnung verschiedene Transparenzpflichten, welche der Risikominimierung dienen. So gilt beispielsweise für Anbieter von KI-Systemen mit allgemeinem Verwendungszweck, dass die Ausgaben des KI-Systems in einem maschinenlesbaren Format gekennzeichnet und als künstlich erzeugt oder manipuliert erkennbar sind. Für Betreiber von KI-Systemen gilt u. a., dass Inhalte, die Deepfake sind, gekennzeichnet werden müssen. Für KI-Systeme mit niedrigem Risiko regt die KI-Verordnung die Einhaltung eines freiwilligen Verhaltenskodex an.

M 1.5	**Anforderungen gem. KI-Verordnung**	
Unabhängig von der Risikoeinstufung sind die Anforderungen des Artikel 4 der KI-Verordnung hinsichtlich der KI-Kompetenz sicherzustellen. Für KI-Systeme mit begrenztem Risiko gelten für Anbieter und Betreiber gemäß Artikel 50 der KI-Verordnung definierte Transparenzpflichten. Die notwendigen Informationen sind spätestens zum Zeitpunkt der ersten Interaktion oder Aussetzung in klarer und eindeutiger Weise bereitzustellen und müssen den geltenden Barrierefreiheitsanforderungen[19] entsprechen. Für KI-Systeme mit niedrigem Risiko wird nach Artikel 95 i. V. m. dem		

18 Website Bundesnetzagentur, URL: https://www.bundesnetzagentur.de/DE/Fachthemen/Digitales/KI/_functions/Hinweispapier.pdf, letzter Zugriff: 23.04.26.

19 Website Bundesfachstelle Barrierefreiheit, URL: https://www.bundesfachstelle-barrierefreiheit.de/DE/Fachwissen/Informationstechnik, letzter Zugriff: 10.06.26.

Erwägungsgrund 165 der KI-Verordnung die Einhaltung eines freiwilligen Verhaltenskodex angeregt.

Die Bundesnetzagentur stellt eine systematischen Überblick über die Transparenzpflichten bereit.[20] Diese Pflichten gelten zusätzlich zu den Pflichten, die sich aus der Verarbeitung personenbezogener Daten ergeben (vgl. M 1.4).

M 1.6 – Geeignete technische Umsetzung

Eine hohe KI-Qualität sollte Eigeninteresse einer jeden Organisation sein. Denn sie bedeutet nicht nur die Einhaltung rechtlicher Vorschriften, sondern ist auch Schlüssel für brauchbare Ergebnisse und Akzeptanz durch die Nutzenden. Die konkrete Ausgestaltung der technischen Umsetzung hat einen hohen Einfluss auf die IT-Qualität (siehe auch Kapitel 2.2) und sollte alle Ergebnisse der vorherigen Maßnahmen aufgreifen.

Übergreifende Aspekte entlang des KI-Lebenszyklus

Relevante funktionelle und operative Aspekte für einen digital-souveränen, sicheren und datenschutzkonformen Einsatz von KI ergeben sich nicht nur für den Betrieb des KI-Systems an sich, sondern für den gesamten KI-Lebenszyklus, der aus verschiedene Phasen (Design, Entwicklung, Einführung, Betrieb und Ausmusterung) besteht. Dieser Umstand verdeutlicht zudem die Relevanz eines geeigneten, kontinuierlichen KI-Management-Prozesses (M 0.1).

Das DSK-Positionspapier zu empfohlenen technischen und organisatorischen Maßnahmen bei der Entwicklung und dem Betrieb von KI-Systemen aus Juni 2025[21] betrachtet übergreifend datenschutzrelevante Aspekte entlang des KI-Lebenszyklus vorrangig aus Sicht von Herstellern und Entwicklern, die DSK-Orientierungshilfe ‚Künstliche Intelligenz und Datenschutz' aus Mai 2024[22] adressiert übergreifend primär die für den Einsatz von KI-Anwendungen Verantwortlichen.

20 Website Bundesnetzagentur, URL: https://www.bundesnetzagentur.de/DE/Fachthemen/Digitales/KI/4_Transparenzpflichten/start.html, letzter Zugriff: 10.06.26.

21 DSK (Juni 2025) Orientierungshilfe zu empfohlenen technischen und organisatorischen Maßnahmen bei der Entwicklung und beim Betrieb von KI-Systemen, Version 1.0, URL: https://www.datenschutz-mv.de/static/DS/Dateien/Publikationen/Broschueren/OH_KI-Systeme.pdf.

22 DSK (Mai 2024): Künstliche Intelligenz und Datenschutz Version 1.0, URL: https://www.datenschutz-mv.de/static/DS/Dateien/Publikationen/Broschueren/20240506_OH_KI_und_Datenschutz.pdf.

Aus der DS-GVO abgeleitet und am IT-Grundschutz orientiert, bietet das Standard-Datenschutzmodell (SDM)[23] Methoden, Mechanismen sowie technische und organisatorische Maßnahmen (TOM), um die Gewährleistungsziele des Datenschutzes Datenminimierung, Verfügbarkeit, Vertraulichkeit, Integrität, Intervenierbarkeit, Transparenz und Nichtverkettung umzusetzen. Diese Gewährleistungsziele, ergänzt durch die Anforderungen der KI-Verordnung, sind entlang des KI-Lebenszyklus gemeinsam zu betrachten.

Geeigneter Architekturansatz und Datenqualität für brauchbare Ergebnisse

Für eine hohe KI-Qualität und eine hohe Akzeptanz der Nutzenden sind auch brauchbare Ergebnisse essentiell.

Das MIT[24] und die DSK[25] zeigen auf, dass allgemeine generative KI seltener brauchbare Ergebnisse liefert, als spezialisierte KI. Das trifft vor allem dann zu, wenn die relevanten Daten in der Trainingsmenge unterrepräsentiert waren (fachliche Nischen). Insofern ist es grundsätzlich wenig ratsam, allgemeine generative KI auf spezielle proprietäre Daten oder Fragestellungen anzuwenden. Eine Lösung hierfür können KI-Systeme mit Retrieval Augmented Generation (RAG) sein. Das DSK-Papier „Datenschutzrechtliche Besonderheiten generativer KI-Systeme mit RAG-Methode“[26] erläutert diese Methode genauer und ordnet sie datenschutzrechtlich ein. Schlussendlich hat das Design und die Entwicklung des KI-Systems fulminanten Einfluss auf die Brauchbarkeit der Ergebnisse und der Akzeptanz des KI-Systems beispielsweise hinsichtlich der Minimierung von Halluzinationen.

Daten und deren Qualität sind von großer Relevanz für KI-Systeme. Ein KI-System, das brauchbare Ergebnisse liefern soll, braucht zunächst Daten in hoher Qualität. Ziel ist die Vermeidung von Bias und daraus resultierender Diskriminierung.

Auch gut funktionierende KI-Systeme sind nicht zwingend fehlerfrei. Daher sollte grundsätzlich sichergestellt sein, dass der Mensch letztinstanzlich entscheidet und rechtlich

23 Website LfDI MV, Standard-Datenschutzmodell, URL: https://www.datenschutz-mv.de/datenschutz/datenschutzmodell/, letzter Zugriff: 10.06.26, s. auch Fußnote 13.

24 Aditya Challapally et al. für MIT (Juli 2025): The GenAI Divide. State of AI in Business 2025, URL; https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf, letzter Zugriff: 10.06.26.

25 DSK (Oktober 2025): Datenschutzrechtliche Besonderheiten generativer KI-Systeme mit RAG-Methode, Version 1.0, URL: https://www.datenschutz-mv.de/static/DS/Dateien/Publikationen/Broschueren/OH_KI_RAG.pdf, letzter Zugriff: 10.06.26.

26 Ebd.

verbindlich agiert.[27] Dies betrifft beispielsweise KI-Agenten, die sich durch ihre hohe operative Autonomie von anderen KI-Systemen abgrenzen.

Sicherer, digital-souveräner und datenschutzkonformer Betrieb (intern vs. Cloud)

Facettenreich ist die Ausgestaltung des Betriebs des KI-Systems. Hier stellen sich u. a. Fragen der Sicherheit, der operativen Machbarkeit, der Souveränität und der Wirtschaftlichkeit. Ein Aspekt ist, ob das System intern (inhouse/on-premise) betrieben werden soll oder extern in einer Cloud-Lösung. Grundsatz ist, dass der Betrieb datenschutzkonform und sicher sein muss. Hierzu sind geltende relevante Standards zu beachten (beispielsweise IT-Grundschutz).

Im Hinblick auf die volle Kontrolle über das KI-System und die betroffenen Daten sowie eine möglichst geringe Abhängigkeit von Dritten ist ein interner Betrieb einem externen Cloud-Betrieb vorzuziehen. Das BSI adressiert diese Thematik des Cloud-Betriebs im Grundschutzbaustein OPS.2.2 Cloud-Nutzung.[28]

Dennoch kann der Betrieb eines KI-Systems schnell ressourcenaufwändig werden. Im Hinblick auf die Vertretungssicherheit und das teils notwendige Spezialwissens kann dies kleinere Organisationen an ihre Leistungsgrenze bringen. Im Hinblick auf einen sicheren Betrieb wäre dann die Beauftragung eines externen Dienstleisters beispielsweise eines Cloud-Dienstleisters zu bevorzugen. Je nach Wahl der Cloud-Dienstleistung (IaaS, SaaS, PaaS) können interne Betriebsressourcen entlastet werden. Zu beachten ist jedoch, dass keine komplette Entlastung stattfinden kann, da beispielsweise Cloud-Ressourcen gebucht, Verträge verwaltet und sich ändernde interne Bedarfe einem Change Management zugeführt werden müssen. Zudem braucht es eine Cloud-Strategie und Sicherheitsrichtlinien, die aktuell gehalten werden müssen. Darüber hinaus verbleibt die Rolle des Verantwortlichen in der Organisation und geht nicht auf einen externen Dienstleister über.

Ungeachtet, ob das KI-System intern oder in der Cloud betrieben wird, müssen KI-Systeme (beispielsweise KI-Agenten) aktiv gemanagt werden, um sicher zu sein. Ein zu freizügiger Umgang mit Zugriffsrechten oder ein gänzlich freier Zugang zu den Daten der Organisation

27 Vgl. Artikel 22 DS-GVO.

28 Website BSI, URL: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/IT-GS-Kompendium_Einzel_PDFs_2023/04_OPS_Betrieb/OPS_2_2_Cloud-Nutzung_Edition_2023.pdf, letzter Zugriff: 23.04.26.

stellt ein kaum beherrschbares Sicherheitsrisiko dar. Daher sind Standardkonfigurationen nach Installation grundsätzlich zu prüfen und ggf. nachzuschärfen (siehe auch Anhang A2).

Zudem ist die Frage der digitalen Souveränität entscheidend. Sie betrifft sowohl einen internen Betrieb als auch eine externe Cloud-Nutzung. Das Kompetenzzentrum Öffentliche IT definiert digitale Souveränität wie folgt:

> *„Digitale Souveränität ist die Summe aller Fähigkeiten und Möglichkeiten von Individuen und Institutionen, ihre Rolle(n) in der digitalen Welt selbstständig, selbstbestimmt und sicher ausüben zu können."*[29]

Unabhängig von einem internen oder externen Betrieb spielen Open-Source-Lösungen, offene Schnittstellen und Formate eine wichtige Rolle und sind ein integraler Baustein einer größtmöglichen digitalen Souveränität.[30] Sie können unter anderem dazu beitragen, eine kritische Abhängigkeit von einzelnen Herstellern oder proprietären Lösungen zu minimieren.

Ergänzend sind auch datenschutzrechtliche Kriterien zu beachten. Mit den DSK-Kriterien für Souveräne Clouds[31] wurden Mindestkriterien und zusätzliche Empfehlungen für Cloud-Lösungen definiert, um aus datenschutzrechtlicher Sicht als souverän zu gelten.

Prioritär ist der Ausschluss einer Einflussnahme von Drittstaaten. Hierdurch besteht potentiell die Gefahr der Aushebelung des europäischen Rechts. Dies betrifft sowohl personenbezogene als auch proprietäre Daten. Insofern sind Anbieter und Dienstleister vorzuziehen, die sicher ausschließlich dem europäischen Rechtsrahmen unterliegen.

M 1.6	**Geeignete technische Umsetzung**	
Die Wahl der technischen Umsetzung beeinflusst die KI-Qualität maßgeblich. KI-Qualität beschreibt den Grad, in dem ein KI-System zuverlässig, sicher, fair und im Einklang mit fachlichen, ethischen und regulatorischen Anforderungen funktioniert.		

29 Website Öffentliche IT, URL: https://www.oeffentliche-it.de/publikationen/digitale-souveraenitaet/Digitale%20Souveränität.pdf, letzter Zugriff: 23.04.26.

30 Website ZENDIS, URL: https://www.zendis.de/media/pages/newsroom/publikationen/zendis-veroeffentlicht-bausteine-fuer-eine-souveraene-digitalstrategie/db23563308-1746615438/2025-02-bausteine-souveraene-digitalstrategie-lp21-kurzfassung.pdf, letzter Zugriff: 23.04.26.

31 DSK (2023): Positionspapier Kriterien für Souveräne Clouds. URL: https://datenschutzkonferenz-online.de/media/weitere_dokumente/2023-05-11_DSK-Positionspapier_Kritierien-Souv-Clouds.pdf.

- Entlang des gesamten KI-Lebenszyklus (Design, Entwicklung, Einführung, Betrieb und Ausmusterung) ist die Einhaltung der Anforderungen der DS-GVO und der KI-Verordnung sicherzustellen.
- Das Design des KI-Systems muss der Aufgabe angepasst sein. Die Anwendung von allgemeiner generativer KI auf spezielle Fachdomänen kann ungenügende oder falsche Ergebnisse liefern.
- Die Umsetzung sollte einen hohen Souveränitätsgrad haben, um Abhängigkeiten von oder Zugriffs- und Kontrollmöglichkeiten durch Dritte zu minimieren.
- Die verwendeten Daten benötigen eine hohe Qualität. Ziel ist die Vermeidung von Bias und daraus resultierender Diskriminierung.
- Die Umsetzung muss sicher sein. Einschlägig sind die Vorgaben und Empfehlungen des BSI (z. B. Grundschutz) und weiterer Institutionen (z. B. NIST AI RMF).
- Es sollte sichergestellt sein, dass letztinstanzlich der Mensch rechtlich verbindlich agiert. Dies betrifft z. B. KI-Agenten mit deren hoher operativer Autonomie.

Die konkrete technische Umsetzung ist zu dokumentieren (M 1.1).

2.4 Fallbeispiel: Interner Betrieb eines großen Sprachmodells (LLM)

Nachfolgend wird überblicksartig ein Fallbeispiel sowohl aus Sicht der KI-Verordnung als auch der DS-GVO betrachtet.

2.4.1 Anwendungsfall

Zur Unterstützung interner Abläufe soll ein Chatbot eingeführt werden, der ausschließlich intern verfügbar ist. Dieser soll die natürlichsprachliche Anfragen der Mitarbeitenden entgegennehmen und bei der internen Bearbeitung von Kund*innenanfragen unterstützen. Dabei hat der Chatbot lesenden Zugriff auf die erforderlichen Kund*innendaten.

2.4.2 Technische Umsetzung

Um natürlichsprachlich auch mit komplexeren Fragestellungen umgehen zu können, soll der Chatbot nicht nur regelbasiert arbeiten, sondern auch ein großes Sprachmodell (LLM) nutzen.

Über Schnittstellen soll der Chatbot auf die interne Datenbank zugreifen können. Der Datenzugriff ist durch ein Rechte-/Rollenkonzept auf das minimal notwendige eingeschränkt.

Im Hinblick auf die volle Kontrolle soll das System on-premise betrieben werden. Dabei sollen (soweit möglich) Open Source Produkte (IT und LLM) eingesetzt werden. Darüber hinaus soll mittels Konfiguration sichergestellt sein, dass der Chatbot (oder seine Komponenten, wie z. B. das LLM) nicht mit dritten Diensten kommuniziert (weder Inhalt noch Telemetrie).

Zudem werden folgende Annahmen getroffen:

- Auch jenseits des geplanten Chatbots werden personenbezogene Daten in der Organisation verarbeitet (in diesem Beispiel zumindest Kund*innendaten). Daher sind die technischen und organisatorischen Maßnahmen des Datenschutzes nach Standard-Datenschutzmodell für bestehende Geschäftsprozesse und IT etabliert.
- Die IT wird gem. dem aktuellen Stand der Technik betrieben. Das bedeutet auch, dass die für einen angenommenen normalen Schutzbedarf notwendigen Prozesse und Strukturen des IT-Betriebs (z. B. ITIL) und der IT-Sicherheit (z. B. IT-Grundschutz) bedarfs- und normgerecht umgesetzt sind.
- Zudem existieren grundlegende notwendige Managementprozesse und Strategien im erforderlichen Maße, um ein ganzheitliches und strukturiertes Vorgehen über alle Organisationsteile zu ermöglichen (F+E, Produktion, Verwaltung, Vertrieb etc.).
- Der geplante Chatbot ist hinsichtlich seines Anwendungsfalls und seiner funktionalen Anforderungen hinreichend dokumentiert.

2.4.3 Datenschutzrechtliche Aspekte

Im Folgenden werden überblicksartig die anwendbaren Aspekte der DS-GVO auf das konkrete Fallbeispiel bezogen betrachtet:

- Zunächst ist zu prüfen, ob personenbezogene Daten verarbeitet werden und ob damit die DS-GVO Anwendung findet. Mit dem Zugriff auf Kund*innendaten ist dies der Fall.
- Für die Verarbeitung personenbezogener Daten bedarf es einer Rechtsgrundlage und einer Zweckbindung. Im konkreten Fallbeispiel kann als Rechtsgrundlage die vertragliche Beziehung zwischen Kund*in und Organisation und ggf. ein überwiegendes berechtigtes Interesse mit dem Zweck der Kund*innenbetreuung in Betracht kommen. (Bei sinngemäßer Anwendung des Fallbeispiels auf Behörden kommen hier gesetzliche Aufträge in Betracht.)
- Mittels einer Vorabprüfung (Schwellenwertanalyse) erfolgt dann zunächst eine generelle Bewertung des Risikos aus datenschutzrechtlicher Sicht. Wird dabei festgestellt, dass die Verarbeitung voraussichtlich ein hohes Risiko zur Folge hat, ist eine Datenschutz-Folgenabschätzung (DSFA) erforderlich. Nach Einschätzung der DSK wird dies beim Einsatz von KI-Anwendungen vielfach zutreffen.[32] Im Zweifelsfall wird zur Durchführung einer DSFA geraten. Für das konkrete Fallbeispiel wird kein hohes Risiko gesehen.
- Umsetzung der Betroffenenrechte entlang des gesamten KI-Lebenszyklus durch Implementierung der technischen und organisatorischen Maßnahmen nach dem Standard-Datenschutzmodell unter Beachtung der Anforderungen für Training, Finetuning etc. oder Erweiterung durch zusätzliche Datenquellen mit personenbezogenen Daten (z. B. Rechte-/Rollenkonzept, Protokollierung, Verschlüsselungen, Löschkonzepte etc.).
- Hinzufügen des Chatbots und der entsprechenden Angaben im Verzeichnis von Verarbeitungstätigkeiten.
- Information der Betroffenen durch Aktualisierung der Datenschutzerklärung.

32 DSK (Mai 2024): Künstliche Intelligenz und Datenschutz Version 1.0, URL: https://www.datenschutz-mv.de/static/DS/Dateien/Publikationen/Broschueren/20240506_OH_KI_und_Datenschutz.pdf.

2.4.4 Aspekte der KI-Verordnung

Im Folgenden werden überblicksartig die anwendbaren Aspekte der KI-Verordnung auf das konkrete Fallbeispiel bezogen betrachtet:

- Zunächst ist die Frage zu beantworten, ob die KI-Verordnung anwendbar ist. Die grundlegenden Aspekte dieser Fragestellung werden im Kapitel 2.1.1 betrachtet. Nachdem der geplante Chatbot nicht nur regelbasiert funktioniert, sondern auch ein LLM nutzt, ist diese Frage zu bejahen. Es handelt sich um ein KI-System im Sinne der KI-Verordnung. Konkrete Unterstützung bei der Beantwortung dieser Frage bietet auch der KI-Compliance Kompass der Bundesnetzagentur.[33]

- In einem nächsten Schritt ist das Risiko des geplanten KI-Systems einzustufen. Chatbots werden als KI-Systeme mit beschränktem Risiko eingestuft. Für diese KI-Systeme gelten Transparenzvorschriften.

- Weiterhin ist zu prüfen, welche Rolle die Organisation nach KI-Verordnung einnimmt. Beim lokalen Betrieb eines LLM liegt die Abgrenzung zwischen Anbieter und Betreiber im Detail. Wird das LLM unverändert genutzt, ist die Betreiber-Rolle einschlägig. Durch Anpassung, z. B. Finetuning und Inverkehrbringen/Betrieb unter eigenen Namen wird die Organisation zum Anbieter. Im konkreten Fallbeispiel soll hinsichtlich möglichst brauchbarer Ergebnisse ein gewisses Finetuning des LLM stattfinden. Das betrifft vor allem Fachterminologien, die in allgemeinen Sprachmodellen eher weniger vorkommen. Dennoch soll das KI-System nicht unter eigenen Namen betrieben oder in Verkehr gebracht werden. In diesem Fall bleibt die Organisation Betreiber.

- Es ergeben sich also neben den obligatorischen Anforderungen für KI-Kompetenz nach Artikel 4 auch konkrete Transparenzpflichten nach Artikel 50, Absatz 3 und 4 der KI-Verordnung. Diese sind für das Fallbeispiel jedoch nicht von Relevanz, da es sich bei dem hier betrachteten Chatbot weder um ein System zur Emotionserkennung oder zur biometrischen Kategorisierung, noch um ein System handelt, das Bild-, Ton- oder Videoinhalte erzeugt oder manipuliert, die ein Deepfake sind und auch kein Text erzeugt oder manipuliert, der veröffentlicht wird, um die Öffentlichkeit über Angelegenheiten von öffentlichem Interesse zu informieren.

33 Website Bundesnetzagentur, URL: https://www.bundesnetzagentur.de/DE/Fachthemen/Digitales/KI/2_Risiko/kompass/start.html, letzter Zugriff: 23.04.26.

3 Abkürzungsverzeichnis

BSI	Bundesamt für Sicherheit in der Informationstechnik
DS-GVO	Datenschutz-Grundverordnung (VO 2016/679)
DSFA	Datenschutz-Folgenabschätzung
DSK	Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder
IaaS	Infrastructure as a Service
ITIL	Information Technology Infrastructure Library
MIT	Massachusetts Institute of Technology (US)
NIST	National Institute of Standards and Technology (US)
KI	Künstliche Intelligenz
KI-Verordnung	Verordnung über künstliche Intelligenz (VO 2024/1689)
PaaS	Product as a Service
RAG	Retrieval-Augmented Generation
SaaS	Software as a Service
SDM	Standard-Datenschutzmodell
TOM	Technische und organisatorische Maßnahmen

Anhang

A1 Checkliste KI-System mit begrenztem oder minimalem Risiko und Verarbeitung personenbezogener Daten

<table>
<tr><td colspan="2">1. Wurde ein grundlegender Managementprozess etabliert, der eine hinreichende KI-Qualität und insbesondere eine stetige Rechtskonformität sicherstellt (M 0.1)?

☐ JA, dann weiter zu 2.

☐ NEIN. Zunächst muss ein Managementprozess etabliert werden, der zumindest Rechtskonformität sicherstellt.</td></tr>
<tr><td colspan="2">2. Ist eine Bestandsaufnahme erfolgt und wurde mit der Dokumentation begonnen, die zumindest die funktionalen Anforderungen, die verarbeiteten Daten, den konkreten Anwendungsfall und im Falle der Verarbeitung personenbezogener Daten die Rechtsgrundlage für die Verarbeitung umfasst (M 1.1).

☐ JA, dann weiter zu 3.1.

☐ NEIN. Zunächst muss zumindest eine minimale Bestandsaufnahme und Dokumentation erfolgen. Diese ist die Grundlage für die folgenden Schritte.</td></tr>
<tr><td>3.1 Anforderungen nach DS-GVO</td><td>3.2 Anforderungen nach KI-Verordnung[34]</td></tr>
<tr><td>3.1.1 Sind personenbezogene Daten betroffen?

☐ JA, dann weiter mit 3.1.2.

☐ NEIN, dann weiter mit 3.2.</td><td>3.2.1 Handelt es sich um ein System im Sinne der KI-Verordnung?

☐ JA, dann weiter mit 3.2.2.

☐ NEIN, dann weiter mit 6.2.</td></tr>
</table>

34 Der KI-Compliance Kompass der Bundesnetzagentur bietet hier Unterstützung, URL: https://www.bundesnetzagentur.de/DE/Fachthemen/Digitales/KI/2_Risiko/kompass/start.html, letzter Zugriff 10.06.26.

3.1.2 Wurde die Rechtsgrundlage für die Datenverarbeitung bestimmt und bei besonderen Kategorien nach Art. 9 DS-GVO die strengeren Voraussetzungen geprüft? ☐ JA, dann weiter mit 3.1.3. ☐ NEIN. Bevor fortgefahren werden kann, muss eine Rechtsgrundlage bestimmt und ggf. die strengeren Voraussetzungen geprüft werden. Ohne Rechtsgrundlage ist die Verarbeitung personenbezogener Daten unzulässig.	3.2.2 Wurde die Rolle (Akteur) gemäß der KI-Verordnung bestimmt? ☐ JA. Die Organisation ist entweder Anbieter oder Betreiber, dann weiter mit 3.2.3. ☐ JA, aber in einer anderen Rolle. Diese werden durch diesen Leitfaden nicht erfasst. Die Betrachtung muss gesondert erfolgen. Weiter mit 6.2. ☐ NEIN. Bevor fortgefahren werden kann, muss die Rolle (Akteur) gem. KI-Verordnung bestimmt werden.
3.1.3 Wurde eine Vorabprüfung (Schwellwertanalyse) durchgeführt? ☐ JA , es ist sicher eine Datenschutzfolgeabschätzung erforderlich oder das Ergebnis der Vorabprüfung ist uneindeutig. Weiter mit 3.1.4. ☐ JA, es ist sicher keine Datenschutzfolgeabschätzung erforderlich. Weiter mit 3.1.5. ☐ NEIN. Bevor fortgefahren werden kann, muss eine Vorabprüfung durchgeführt werden.	3.2.3 Wurde das Risiko gemäß der KI-Verordnung bestimmt? ☐ JA. Das Risiko ist höchstens begrenzt, dann weiter mit 3.2.4. ☐ JA, aber das Risiko ist hoch. Dieses Risiko wird durch diesen Leitfaden nicht erfasst. Die Betrachtung muss gesondert erfolgen. Weiter mit 6.2. ☐ NEIN. Bevor fortgefahren werden kann, muss das Risiko gem. KI-Verordnung bestimmt werden.
3.1.4 Wurde soweit erforderlich eine Datenschutzfolgeabschätzung durchgeführt? ☐ JA, dann weiter mit 3.1.5. ☐ NEIN. Bevor fortgefahren werden kann, muss eine Datenschutzfolgeabschätzung durchgeführt werden.	
3.1.5 Wurden die erforderlichen TOM z. B. nach SDM erarbeitet? ☐ JA, dann weiter mit 3.2. ☐ NEIN. Bevor fortgefahren werden kann, müssen die TOM z. B. nach SDM umgesetzt werden werden.	

4. Wurde eine **geeignete technische Umsetzung** erarbeitet, die einen sicheren, datenschutzkonformen und digital-souveränen Betrieb und einen bedarfsgerechten Architekturansatz sicherstellt (M 1.6) und die Ergebnisse der vorherigen Prüfschritte berücksichtigt?

☐ JA, dann weiter zu 5.

☐ NEIN. Zunächst muss eine technische Umsetzung erarbeitet werden, die zumindest den etablierten Standards und Anforderungen entspricht.

5. Wurden die **Dokumentation abgeschlossen** und durch die konkrete technische Umsetzung ergänzt? Wurde das **Verzeichnisses zu Verarbeitungstätigkeiten** aktualisiert? Wurden die Betroffenen zumindest durch **Anpassung der Datenschutzerklärung** informiert? Wurden die **Transparenzpflichten** gemäß Art. 50 der KI-Verordnung umgesetzt?

☐ JA, dann weiter zu 6.1.

☐ NEIN. Zunächst muss die Dokumentation abgeschlossen werden und die notwendigen Transparenz- und Informationspflichten nach DS-GVO und KI-Verordnung erfüllt werden.

6.1. Sie haben die Checkliste abgearbeitet. Bitte beachten Sie, dass dieser Leitfaden einen speziellen Fokus hat und dass ggf. noch weitere Anforderungen aus anderen Normen berücksichtigt werden müssen. Dies ist konkret zu prüfen.

6.2. Für den Fall, dass sie einen Anwendungsfall haben, der nicht durch diesen Leitfaden erfasst wird (z. B. Hochrisiko-KI), verweisen wir auf weiterführenden Informationen. Einen systematischen Überblick über die KI-Verordnung bietet z. B. die Bundesnetzagentur unter: https://www.bundesnetzagentur.de/DE/Fachthemen/Digitales/KI/start_ki.html

A2 Aktuelle KI-Entwicklungen aus Sicht des Datenschutzes – KI-Agenten

Das Thema Künstliche Intelligenz unterliegt einem permanenten Wandel und technischer Weiterentwicklung.[35] Aktuell steht im Fokus dieser technischen Weiterentwicklung das Thema KI-Agenten (AgenticAI), die sich vor allem durch eine zunehmende Autonomie und einer stärkeren Personalisierung auszeichnen und damit von bisherigen Lösungen wie beispielsweise RAG-Systeme abgrenzen.

Wie bereits zuvor dargelegt, ist künstliche Intelligenz[36] grundsätzlich eine datenbasierte Disziplin. Prinzipiell kann man sagen, dass die Verarbeitung und Verknüpfung von Daten charakteristisch für künstlicher Intelligenz ist. Es ist eine grundlegende Eigenschaft, um unstrukturierte Informationen (Big Data)[37] aus unterschiedlichen Quellen zusammenführen, Muster zu erkennen und diese auf neue Informationen (im Sinne von Vorhersagen) anzuwenden. So werden große Sprachmodelle (LLM) mit einer erheblichen Anzahl an allgemeinen Daten trainiert, was eine breite Anwendbarkeit sicherstellen soll. Auf (fachliche) Nischen oder proprietäre Daten angewandt, offenbaren sie jedoch Schwächen. Dies wird vor allem dann deutlich, wenn diese proprietären Daten oder die (fachliche) Nische in den Trainingsdaten unterrepräsentiert war.

Im Hinblick auf bedarfsgerechte Ergebnisse ist somit nicht nur ein gutes KI-Gedächtnis notwendig (KI-Systeme verarbeiten eine große Anzahl von Daten und lernen stetig neue Muster[38]), sondern teils auch eine gewisse Personalisierung allgemeiner KI-Systeme (Nischen etc.). Ein Ansatz bietet beispielsweise Retrieval-Augmented Generation (RAG-Systeme), die im DSK-Papier „Datenschutzrechtliche Besonderheiten generativer KI-Systeme mit RAG-Methode“[39] genauer dargestellt werden. Eine Weiterentwicklung des RAG-Ansatzes sind Wissensgrafen (beispielsweise KAG[40] oder GraphRAG[41]), bei denen die proprietären Daten nicht nur in Chunks (Textabschnitte)

35 Einen Überblick über die KI-Entwicklungen, auch im Sinne einer minimalistischen strategischen Vorausschau, kann u. a. der Gartner AI Hype Cycle geben. (Gartner verwendet Cookies: Privacy Cookie Policy), URL: https://www.gartner.com/en/articles/hype-cycle-for-artificial-intelligence, letzter Zugriff: 10.06.26.

36 Vgl. Website IBM, URL: https://www.ibm.com/de-de/think/topics/artificial-intelligence, letzter Zugriff: 23.04.26.

37 Vgl. Website IBM, URL: https://www.ibm.com/de-de/think/topics/big-data;, letzter Zugriff: 23.04.26.

38 Vgl. Website IBM, URL: https://www.ibm.com/de-de/think/topics/catastrophic-forgetting, letzter Zugriff: 23.04.26.

39 DSK (Oktober 2025): Datenschutzrechtliche Besonderheiten generativer KI-Systeme mit RAG-Methode, Version 1.0, URL: https://www.datenschutz-mv.de/static/DS/Dateien/Publikationen/Broschueren/OH_KI_RAG.pdf, letzter Zugriff: 10.06.26.

40 Lei Liang et al. (V.3 September 2024): KAG: Boosting LLMs in Professional Domains via Knowledge Augmented Generation. Veröffentlicht bei ArXiv, URL: https://arxiv.org/abs/2409.13731, letzter Zugriff: 10.06.26.

aufgeteilt und in einer Vektor-Datenbank gespeichert werden, sondern über Verbindungen (Grafen) in Beziehung gesetzt und damit Zusammenhänge dargestellt werden.

Ein weiterer Entwicklungsschritt sind KI-Agenten, die sich nicht nur durch eine Personalisierung beispielsweise durch den Zugriff auf private oder proprietäre Daten (wie RAG-Systeme) auszeichnet, sondern durch eine zunehmende Autonomie im Sinne eines selbstständigen Handelns. In Fortführung des Themas werden zunehmend Multiagenten-Systeme betrachtet, in denen mehrere spezialisierte KI-Agenten zusammenarbeiten, um einerseits vielschichtige Fragestellungen und andererseits z. B. operativ das Thema Skalierbarkeit adressieren zu können

Ganz grundsätzlich zeigt sich ein Spannungsbogen zwischen den Charakteristika von (aktuellen) KI-Systemen und datenschutzrechtlichen Anforderungen, die stetig und kontextabhängig betrachtet werden müssen – Beispiele sind:

Verknüpfung von Daten ↔ Nichtverkettung

KI-Gedächtnis ↔ Löschvorgaben

Katastrophales Vergessen[42] ↔ Korrektheit, Verfügbarkeit

KI-Agenten

KI-Agenten stehen aktuell im Zentrum der Betrachtung. Sie übertreffen hinsichtlich der Automatisierung von (Geschäfts-)Prozessen allgemeine KI (wie beispielsweise LLMs) deutlich. Insofern erweitern KI-Agenten die oben gezeigten Spannungsbögen um einen weiteren:

Operative Autonomie ↔ Notwendige menschliche Aufsicht

Dabei vereinen KI-Agenten drei kritische Eigenschaften, die unbedingt beachtet und gemanagt werden müssen – auch bekannt als die „tödliche Dreifaltigkeit"[43]:

1. Zugriff auf private/interne Daten
2. Zugriff auf nicht vertrauenswürdige Inhalte im Internet
3. Autonome Fähigkeit zur externen Kommunikation

41 Haoyu Han et al. (V.2 Januar 2025): Retrieval-Augmented Generation with Graphs (GraphRAG). Veröffentlicht bei ArXiv, URL: https://arxiv.org/abs/2501.00309, letzter Zugriff: 10.06.26.

42 Vgl. Website IBM, URL: https://www.ibm.com/de-de/think/topics/catastrophic-forgetting, letzter Zugriff: 23.04.26.

43 Vgl. Website Palo Alto Networks, URL: https://www.paloaltonetworks.com/blog/network-security/why-moltbot-may-signal-ai-crisis/, letzter Zugriff: 10.06.2026.

Ungemanagte KI-Agenten sind per se nicht sicher und stellen ein nicht kalkulierbares Risiko dar. Es sollte ein essentielles Eigeninteresse einer jeden Organisation sein, sicherzustellen, dass ungemanagte KI-Agenten nicht eingesetzt werden.

Nicht zuletzt deshalb befassen sich europäischen Datenschutzbehörden aktuell mit dem Thema KI-Agenten. Beispielsweise hat die spanischen Datenschutzbehörde[44] einen gesonderten Leitfaden für den Einsatz von KI-Agenten[45] erarbeitet.

Übergreifend betrachtet, braucht es für das Management von KI-Agenten geeignete Steuerungs- und Organisationsprozesse. Zudem müssen die datenschutzrechtlichen Grundprinzipien wie u. a. Zweckmäßigkeit, Datenminimierung und Vertraulichkeit beachtet werden – dies betrifft auch das zur Personalisierung notwendige Gedächtnis des KI-Agenten (Datentrennung, Löschvorgaben etc.), die durch geeignete technische Sicherheitsmaßnahmen umgesetzt werden. Prinzipiell finden hier die Ausführungen des DSK-Papiers „Datenschutzrechtliche Besonderheiten generativer KI-Systeme mit RAG-Methode"[46] sinngemäß Anwendung, müssen jedoch aufgrund der erhöhten Autonomie der KI-Agenten gegenüber der RAG-Systeme um Regelungen zur Autonomie und menschlicher Intervention ergänzt werden.

Prompt Injection, also die Manipulation durch versteckte Befehle, ist zwar kein neues aber ein immer noch aktuelles Sicherheitsthema im KI-Umfeld. Die Forschenden haben in einer systematischen Befassung sechs kritische Sicherheitsaspekte für KI-Agenten (Agenten-Fallen) identifiziert[47]:

1. *Verborgene Befehle* in HTML-Code oder Multimedia werden vom Menschen nicht wahrgenommen, aber sehr wohl durch den KI-Agenten ausgewertet und ggf. ausgeführt.
2. *Spezielle Formulierungen* wie Übertreibungen oder Emotionen werden eine besondere Relevanz zugemessen und beeinflussen die Ergebnisse des KI-Agenten.

---

44 Website der spanischen Datenschutzaufsichtsbehörde AEPD, URL: https://www.aepd.es/prensa-y-comunicacion/notas-de-prensa/la-agencia-publica-unas-orientaciones-sobre-inteligencia, letzter Zugriff: 23.04.26.

45 Website der spanischen Datenschutzaufsichtsbehörde AEPD, URL: https://www.aepd.es/guias/orientaciones-ia-agentica.pdf, letzter Zugriff: 23.04.26.

46 DSK (Oktober 2025): Datenschutzrechtliche Besonderheiten generativer KI-Systeme mit RAG-Methode, Version 1.0, URL: https://www.datenschutz-mv.de/static/DS/Dateien/Publikationen/Broschueren/OH_KI_RAG.pdf.

47 Matija Franklin et al. (März 2026): AI Agent Traps. Veröffentlicht bei SSRN, URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6372438, letzter Zugriff: 10.06.2026.

3. *Verfälschte Fakten oder Informationen* in der Wissensbasis (etwa gefälschte aber offiziell wirkende Unterlagen) sorgen dafür, dass der KI-Agent diese als bestätigte Fakten behandelt.

4. Durch *Missbrauch der Fähigkeiten eines Agenten* kann dieser dazu gebracht werden, Schadcode zu erstellen und auszuführen oder sensible Informationen aus dem internen Netz zu extrahieren.

5. Durch den aktuell sehr vergleichbaren Aufbau existierender KI-Agenten, können durch eine Manipulation potentiell eine Vielzahl von KI-Agenten beeinflusst werden. Da sich KI-Agenten über spezielle Netzwerke verbunden sind, ist die *Manipulation ganzer Agentennetzwerke* ein begründetes Szenario.

6. Selbst wenn der Mensch die Möglichkeit der Intervention hat oder letztinstanzlich entscheidet, kann es auch zu einer *Manipulation des Menschen* kommen, in dem dieser beispielsweise durch eine Welle an automatisiert generierten Informationen zu einer falschen Entscheidung – wie dem Klicken eines Links – gedrängt wird.

Die Forschenden richten sich insbesondere an die Entwickler*innen von KI-Agenten. Den Nutzenden obliegt es vor allem: KI-Agenten kritisch zu managen, kontinuierlich zu prüfen und letztinstanzlich (vor allem bei einer resultierenden rechtlichen Verbindlichkeit – Haftung etc.) zu entscheiden.