



# Informationen und Empfehlungen zum sicheren Einsatz von (generativer) KI für die Leitungsebene

Veröffentlichung des Expertenkreis KI-Sicherheit

**Autorinnen und Autoren<sup>1</sup>:** Corinna Donhauser (Krones AG), Andrea Ibisch (Bundesamt für Sicherheit in der Informationstechnik), Dominique Knebel (e.lective GmbH), Caroline Neufert (BearingPoint GmbH)

## Disclaimer

Bei den Informationen dieses Paketes handelt es sich um Empfehlungen, die keine rechtsbindende Wirkung haben. Sie setzen auch weder Vorgaben aus gesetzlichen Vorschriften außer Kraft, noch wird garantiert, dass durch die Erfüllung aller hier genannten Maßnahmen alle gesetzlichen Bestimmungen erfüllt sind.

Vielmehr sollen die Informationen eine praxisnahe Unterstützung bei der Sensibilisierung von Mitarbeitenden geben, dazu werden auch Empfehlungen hinsichtlich der Regelung von KI-Zuständigkeiten in einem Unternehmen / einer Behörde gemacht.

## Einleitung

Dieses Informationspaket soll Sie dabei unterstützen, in Ihrem Unternehmen / Ihrer Behörde eine Grundlage für den sicheren Einsatz von (generativer) KI zu schaffen. Der Fokus liegt dabei auf der Sensibilisierung und Schulung von Mitarbeitenden Ihres Unternehmens, die potenziell KI-Anwendungen im Arbeitsalltag nutzen. Darüber hinaus enthält das Paket Empfehlungen zur organisatorischen Gestaltung von KI-Nutzung, die als Basis für eine effektive Sensibilisierung der Endnutzenden dienen. Hierbei handelt es sich um Informationen, die sich hauptsächlich auf grundlegende organisatorische Aspekte beziehen und auf das Thema Einführung von KI beschränkt sind. Es ist geplant in einer separaten Veröffentlichung detaillierter auf das Thema einzugehen und dabei den gesamten Lebenszyklus von KI zu betrachten.

Das Informationspaket richtet sich insbesondere an kleine und mittelgroße Unternehmen und Behörden, die nur über wenig Ressourcen für die Regulierung von KI-Nutzung und Sensibilisierung von Mitarbeitenden verfügen.

Technische Aspekte der Implementierung von KI-Anwendungen, die zu einer Erhöhung der Sicherheit beitragen können, werden in diesem Rahmen nicht behandelt. Detaillierte Informationen dazu finden Sie z. B. in den Veröffentlichungen des BSI, insbesondere:

---

<sup>1</sup> Nennung in alphabetischer Reihenfolge der Nachnamen; Fragen und Feedback können Sie an publikationen-xprt-ki@bsi.bund.de richten.

- „Generative KI-Modelle – Chancen und Risiken für Industrie und Behörden“<sup>2</sup>
- „Kriterienkatalog des BSI zur Integration von extern bereitgestellten generativen KI-Modellen in eigene Anwendungen“<sup>3</sup>

Im Einzelnen besteht dieses Informationspaket aus:

1. **Informationen und Empfehlungen für die Leitungsebene:** Dieses Dokument unterstützt Sie dabei, Vorgaben für die Nutzung (generativer) KI in Ihrem Unternehmen / Ihrer Behörde zu entwickeln. Es schlägt dazu einige organisatorische und prozessuale Maßnahmen vor und bietet einen Überblick über die Risiken, die mit der Nutzung von (generativer) KI einhergehen. Ziel ist es Sie (oder eine beauftragte Person) zu befähigen, anwendungsfall-bezogenen Vorgaben an die Endnutzenden von KI in Ihrem Unternehmen / Ihrer Behörde abzuleiten.
2. **Schulungsfolien:** Der vorliegende Foliensatz kann für Schulungsmaßnahmen in Ihrem Unternehmen / Ihrer Behörde genutzt werden. Er besteht aus insgesamt 61 Folien, von denen der/die Schulungsleitende individuell die Folien auswählen kann, die für die Mitarbeitenden des jeweiligen Unternehmens / der jeweiligen Behörde relevant sind. Es empfiehlt sich in den meisten Fällen die Folien so auszuwählen, dass eine Schulung maximal 60 Minuten lang ist. Die Folien sind in erster Linie auf eine Schulung im Vortragsstil ausgelegt, können aber auch zum Selbststudium genutzt werden. Sie veranschaulichen zentrale Risiken im Umgang mit (generativer) KI aus Perspektive der Nutzenden anhand praxisnaher Beispiele. Darüber hinaus werden konkrete Handlungsempfehlungen gegeben, wie Risiken durch verantwortungsbewusste Nutzung reduziert werden können.
3. **Begleitmaterial für Schulungsleitende:** Die Schulung ist so konzipiert, dass sie von Personen mit grundlegenden technischen Kenntnissen – beispielsweise von Informationssicherheitsbeauftragten – durchgeführt werden kann. Das Begleitmaterial erläutert die Inhalte der Folien und gibt Hinweise, welche zusätzlichen Informationen während der Schulung mündlich vermittelt werden können. So dient es als Leitfaden zur inhaltlichen Vertiefung und zur Unterstützung bei der Durchführung.
4. **Merkblatt:** Das Merkblatt fasst die wichtigsten Inhalte der Schulung kompakt zusammen. Jede Mitarbeiterin bzw. jeder Mitarbeiter sollte nach der Schulung ein Exemplar erhalten – idealerweise in ausgedruckter Form und gut sichtbar am Arbeitsplatz platziert. Das Dokument liegt in einer editierbaren Version vor und sollte vor der Weitergabe an Ihre internen Vorgaben und Bedarfe angepasst werden.

---

<sup>2</sup> [https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/KI/Generative\\_KI-Modelle.pdf?\\_\\_blob=publicationFile&v=7](https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/KI/Generative_KI-Modelle.pdf?__blob=publicationFile&v=7)

<sup>3</sup> [https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/KI/Kriterienkatalog\\_KI-Modelle\\_Bundesverwaltung.pdf?\\_\\_blob=publicationFile&v=3](https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/KI/Kriterienkatalog_KI-Modelle_Bundesverwaltung.pdf?__blob=publicationFile&v=3)

## Organisatorische Maßnahmen

Im Folgenden finden Sie organisatorische Maßnahmen, die zur sicheren Nutzung von (generativer) KI in Ihrem Unternehmen / Ihrer Behörde beitragen können. Sie sollten je nach angedachten Anwendungsfällen der KI-Nutzung, Größe Ihres Unternehmens / Ihrer Behörde und personellen Verfügbarkeiten beurteilen, welche Maßnahmen in Ihrem Unternehmen / Ihrer Behörde notwendig, sinnvoll und machbar sind.

- ☐ Erstellen Sie eine KI-Strategie, in welchen Bereichen und mit welchen Zielen KI bei Ihnen eingesetzt werden soll.
- ☐ Beachten Sie die regulatorischen Anforderungen (z. B. DSGVO, AI Act, Urheberrecht).
- ☐ Verfassen Sie eine KI-Richtlinie zur Entwicklung, Beschaffung und Nutzung von KI.
- ☐ **Benennen Sie eine/n KI-Zuständige/n**, um eine zentrale Ansprech- und Kompetenzstelle für alle KI-bezogenen Themen zu schaffen. Dem oder der KI-Zuständigen können beispielsweise folgende Aufgaben übertragen werden:
  - ☐ Übersicht aller in der Einrichtung genutzten Anwendungen mit KI-Komponenten<sup>4</sup> verwalten (vgl. nächster Punkt).
  - ☐ Verfahren für nicht geregelte Anwendungsfälle oder KI-Anwendungen festlegen (z. B. grundsätzliches Nutzungsverbot mit Möglichkeit zur Einzelfallprüfung).
  - ☐ Nutzungsbedingungen für die in der Übersicht genannten Anwendungen erarbeiten (lassen) und diese in geeigneter Weise den Endnutzenden zur Verfügung stellen.
  - ☐ Neue KI-Projekte innerhalb der Behörde / des Unternehmens anstoßen.
  - ☐ In KI-Projekten mitwirken.
  - ☐ Bei Vertragsverhandlungen für die Nutzung von KI mitwirken und z. B. hinsichtlich der Nutzung von internen Daten zum Training von KI-Modellen beraten.
  - ☐ Schulungen/Sensibilisierungen durchführen.
  - ☐ Als Ansprechpartner für Leitung und Endnutzende verfügbar sein.
- ☐ **Führen Sie eine Übersicht aller genutzten Anwendungen mit KI-Komponente** (ggf. kann es auch sinnvoll sein, die Übersicht nach KI-Modellen und nicht nach Anwendungen zu strukturieren) mit z. B. den folgenden Informationen:
  - ☐ Welche KI-Modelle kommen in der jeweiligen Anwendung zum Einsatz bzw. auf welche KI-Modelle wird zugegriffen?
  - ☐ Für welche Anwendungsfälle und Daten ist die jeweilige Anwendung freigegeben?
  - ☐ Wie dürfen sich Nutzende bei der Anwendung anmelden? (z. B. Nutzung dienstlicher E-Mail-Adresse, zentrale Beschaffung von Accounts, ...)
  - ☐ Müssen Nutzende bestimmte Einstellungen vornehmen?
  - ☐ Sind externe Dienstleister (z. B. Hosting, Support, Entwicklung) involviert? Ggf. Integration und Verankerung von KI-relevanten Themen in der Dienstleisterbewertung
  - ☐ Welche geschäftliche Relevanz hat die Anwendung? Welche Kritikalitätsstufe hat die Anwendung?
  - ☐ Welchen Schutzbedarf haben die von der Anwendung verarbeiteten und/oder erzeugten Daten?
- ☐ Sorgen Sie dafür, dass alle Mitarbeitenden, die potenziell Anwendungen mit KI-Komponenten im Arbeitsalltag nutzen, entsprechend geschult bzw. sensibilisiert werden.

---

<sup>4</sup> Sie können diese Übersicht auf Anwendungen beschränken, bei denen offensichtlich oder mit hoher Wahrscheinlichkeit KI-Komponenten vorhanden sind; bei fertigen Software-Produkten ist es häufig nicht ersichtlich, ob einzelne Funktionen KI-basiert funktionieren, außer der Hersteller macht dies transparent.

## Empfehlungen zum Vorgehen bei Einführung von KI-Anwendungen oder Freigabe zur Nutzung von KI-Diensten

Das folgende Vorgehen kann als Orientierung dienen, wenn Sie in Ihrem Unternehmen oder Ihrer Behörde den Einsatz von KI-Anwendungen planen. Es lässt sich auf verschiedene Szenarien anwenden – sei es die Integration eines extern bereitgestellten KI-Modells in eine eigene Anwendung, die Nutzung eines online-verfügbaren KI-Dienstes oder auch die Entwicklung eines eigenen KI-Modells.

- Definieren Sie den konkreten Anwendungsfall für den Sie eine Nutzung von (generativer) KI in Betracht ziehen. Ein Anwendungsfall kann dabei sehr spezifisch (z. B. Entwicklung einer KI-Anwendung für die Prüfung von bestimmten Formularen) oder weit gefasst (z. B. Nutzung eines bestimmten online-verfügbaren Sprachmodells für das Verfassen von E-Mails und die Zusammenfassung von Dokumenten) sein.
- Identifizieren Sie KI-Risiken (vgl. folgender Abschnitt), die für diesen Anwendungsfall relevant sind.
- Leiten Sie aus den identifizierten KI-Risiken Kriterien für die Auswahl eines geeigneten KI-Modells ab<sup>5</sup>.
- Unabhängig von den KI-Risiken können sich weitere Auswahlkriterien z. B. aus der Verfügbarkeit, den Funktionalitäten oder den Kosten eines KI-Modells ergeben.
- Bewerten Sie ein KI-Modell, das für den Anwendungsfall in Frage kommt, anhand der von Ihnen definierten Kriterien.
  - Als Grundlage für die Beurteilung sollten Informationen des Anbieters (z. B. Nutzungsbedingungen, Model-Cards, Informationen zu durchgeführten Benchmark-Tests) gesichtet werden. Zusätzlich können Informationen Dritter zur Beurteilung genutzt werden.
  - Ebenso können eigene Tests z. B. anhand von Benchmarks oder in Form eines Red Teamings durchgeführt werden.
- Wählen Sie ein KI-Modell aus, das in den wesentlichen Punkten Ihre Auswahlkriterien erfüllt.
- Leiten Sie ggf. technische Vorgaben für die Integration des KI-Modells in eigene Anwendungen ab, sofern bestimmte Risiken nicht bereits durch die Modellauswahl ausreichend adressiert wurden (z. B. zusätzliche Filter für Ein- oder Ausgaben).
- Leiten Sie ggf. Vorgaben an Nutzende ab, wenn KI-Risiken, die Sie für den Anwendungsfall als relevant identifiziert haben, nicht ausreichend durch die Modellauswahl oder Vorgaben an die Integration gemindert werden können.

### KI-Risiken

Im Folgenden wird eine Auswahl von Risiken dargestellt, die bei der Nutzung von KI-Anwendungen auftreten können. Wie zuvor beschrieben sollten Sie anwendungsfall-bezogen entscheiden, wie relevant diese Risiken sind, hierzu finden Sie jeweils unter „Bewertung“ Anhaltspunkte. Zusätzlich werden beispielhaft Maßnahmen genannt, die ein Risiko mindern können. Hierdurch erhalten Sie einen Eindruck davon, ob es Ihnen möglich ist zur Minderung des Risikos beizutragen oder ob dies an einem anderen Punkt des Lebenszyklus der KI-Anwendung erfolgen muss.

---

<sup>5</sup> Beispiel: Die Anwendung, die betrachtet wird, soll Anträge von Dritten verarbeiten. Deshalb wird das Risiko von Indirect Prompt Injections als relevant identifiziert. Daraus leiten Sie folgendes Auswahlkriterium ab: „Das KI-Modell, das in unserer Anwendung verwendet wird, sollte eine Härting gegenüber manipulativen Eingaben Dritter aufweisen.“

Die Ausführungen dienen dazu, Sie dabei zu unterstützen festzustellen, welche Anwendungsfälle besonders kritisch sind. So können Sie besser beurteilen, für welche Anwendungsfälle Sie bestimmte Anforderungen an ein KI-Modell oder besonders strikte Vorgaben an die Nutzenden stellen sollten.

Versuchen Sie auch Fälle zu identifizieren, die Sie generell unkritisch bewerten, wie beispielsweise kleine Anfragen ohne Kontext (z. B. Übersetzung oder Formulierung einzelner Sätze/Wörter) oder Anfragen ohne spezifischen fachlichen Bezug, sprich genereller Natur (z. B. Aufbau eines Flyers, Kennenlernspiel neue Kollegen). Kommunizieren Sie auch für diese Fälle, wie Mitarbeitende KI nutzen dürfen, so unterstützen Sie, dass ein Bewusstsein für die Abgrenzung zwischen kritischen und unkritischen Anwendungsfällen entsteht. Zudem sollte auch in diesen Fällen geregelt sein, wie ggf. eine Anmeldung bei den KI-Diensten erfolgt.

Für eine detailliertere Beschreibung von Risiken und Gegenmaßnahmen generativer KI-Modelle verweisen wir auf die BSI-Veröffentlichung „Generative KI-Modelle – Chancen und Risiken für Industrie und Behörden“<sup>6</sup>.

### Abfluss sensibler Daten

**Risiko:** Alle Daten, die zur Funktion einer KI-Anwendung beitragen, sind potenziell gefährdet an Unberechtigte abzufließen, also z. B. Trainingsdaten, Eingaben, Ausgaben, Modellspezifikationen oder Spezifikationen über sonstige Komponenten. Ein Abfluss kann einerseits durch Zufall an Endnutzende der Anwendung erfolgen (z. B. wenn Endnutzende in ihrer Eingabe um die Imitation des Schreibstils eines bestimmten Künstlers bitten und das Modell Daten aus seinen Trainingsdaten 1:1 wiedergibt). Andererseits können Angreifende bewusst versuchen die genannten Daten durch geschickte Formulierungen der Eingabe zu erlangen (sog. Privacy Attacks). Zudem haben die Anbieter und Betreiber von KI-Modellen potenziell Zugriff auf bestimmte Daten der Endnutzenden und behalten sich teilweise das Recht vor diese z. B. zum weiteren Training ihrer Systeme zu nutzen.

**Bewertung:** Beachten Sie, welche Rechte sich der Anbieter bzw. Betreiber eines KI-Modells einräumt (siehe z. B. Nutzungsbedingungen), ggf. gibt es verschiedene Varianten (z. B. kostenlos vs. kostenpflichtig), die sich hierin unterscheiden oder man kann der Verwendung durch eine Änderung der Einstellungen widersprechen.

**Maßnahmen:** Regeln Sie klar, welche Informationen Nutzende als Eingabe für welche Anwendungen verwenden dürfen. Besonders relevant ist dies für persönliche Daten, Interna und eingestufte Informationen!

Prüfen Sie, ob Änderungen an den Standardeinstellungen einer Anwendung vorgenommen werden müssen, und kommunizieren Sie dies an die Nutzenden.

Als Grundsatz muss gelten: Interne Informationen dürfen von externen Stellen nicht verwendet werden!

Bei selbstentwickelten Komponenten von Anwendungen ist es wichtig sicherzustellen, dass schützenswerte Daten (z. B. sensible Daten in einer Datenbank) vor dem direkten und indirekten Zugriff durch Unberechtigte geschützt sind. Hierbei helfen klassische IT-Sicherheitsmaßnahmen, aber auch spezielle Maßnahmen, wie das Filtern von Ein- und Ausgaben oder, wenn möglich, Maßnahmen zur Anonymisierung.

---

<sup>6</sup> [https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/KI/Generative\\_KI-Modelle.pdf?\\_\\_blob=publicationFile&v=7](https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/KI/Generative_KI-Modelle.pdf?__blob=publicationFile&v=7)

## Fehlende Robustheit

**Risiko:** KI-Modelle haben häufig Probleme damit Eingaben wie (von der nutzenden Person) erwartet zu verarbeiten, wenn die Eingabe stark von dem abweicht, was das Modell gelernt hat. Dies kann z. B. dann der Fall sein, wenn Eingaben (absichtlich wie unabsichtlich) „unscharf“ sind (z. B. ein verwaschenes Bild oder ein Text mit vielen Rechtschreibfehlern oder uneindeutigen Formulierungen). In diesen Fällen kann es zu fehlerhaften und ggf. auch problematischen Ausgaben kommen.

**Bewertung:** Versuchen Sie zu beurteilen, wie groß die Wahrscheinlichkeit ist, dass solche Eingaben an die KI-Anwendung gerichtet werden. Bewerten Sie, wie kritisch die Auswirkungen von fehlerhaften Ausgaben sind (werden Ausgaben z. B. ungesehen weiterverwendet).

**Maßnahmen:** Maßnahmen zur Steigerung der Robustheit werden in der Regel während des Trainings des KI-Modells ergriffen, haben Sie hierauf keinen Einfluss, sensibilisieren Sie die Nutzenden für die Problematik.

## Fehlende Qualität der Ausgabe

**Risiko:** Auch robuste KI-Modelle weisen Mängel in der Qualität ihrer Ausgaben auf, diese können etwa faktisch falsch, diskriminierend, anders problematisch (z. B. vulgär) oder unsicher (wenn es sich um Programmcode handelt) sein.

**Bewertung:** Bewerten Sie, wie kritisch die Auswirkungen von qualitativ schlechten Ausgaben sind (werden Ausgaben z. B. ungesehen weiterverwendet).

**Maßnahmen:** Maßnahmen zur Erhöhung der Qualität können während des gesamten Lebenszyklus ergriffen werden: Verbesserung der Trainingsdaten, Verwendung von System-Prompts, einem KI-Modell Zugriff auf Datenbanken geben, damit es fundiertere Ausgaben machen kann, Verwendung von Ausgabe-Filtern, präzises Prompting, manuelle Nachbearbeitung/Prüfung der Ausgabe, ...

## Risiko der missbräuchlichen Nutzung

**Risiko:** Nutzende mit böswilligen Absichten können KI-Systeme für ihre Zwecke missbrauchen, z. B. um Falschnachrichten zu erzeugen oder Cyberangriffe durchzuführen. Selbst wenn Sicherheitsmaßnahmen etabliert sind, ist es häufig möglich diese durch geschicktes Formulieren der Eingabe zu umgehen (sog. Evasion Attacks).

**Bewertung:** Mit diesem Aspekt müssen Sie sich besonders dann auseinandersetzen, wenn Sie eine Anwendung zum Betrieb bereitstellen. Wenn Ihre Mitarbeitenden externe KI-Dienste nutzen, sollten Sie prüfen, welche Maßnahmen der Anbieter zur Verhinderung missbräuchlicher Nutzung trifft.

Bewerten Sie, wie hoch das Risiko ist, dass Nutzende böswillige Absichten haben (gibt es z. B. nur einen eingeschränkten Nutzerkreis oder können auch Externe die Anwendung nutzen; schließen Sie auch Innentäter nicht aus) und wie hoch der Schaden für Ihre Behörde / Ihr Unternehmen ist, der durch eine missbräuchliche Nutzung entstehen kann (z. B. Rufschädigung, rechtliche Konsequenzen).

**Maßnahmen:** Zur Verhinderung missbräuchlicher Nutzung können sowohl organisatorische (z. B. Einschränkung der Zugangsmöglichkeiten zum KI-Modell, Veröffentlichung von Nutzungsbedingungen) als auch technische Maßnahmen (z. B. Filterung von Ein- und Ausgaben, adversariales Training, Verwendung von System-Prompts) ergriffen werden.

## Manipulation der Funktionsweise im Betrieb durch Dritte (sog. Indirect Prompt Injections)

**Risiko:** Werden mit einer KI-Anwendung Informationen verarbeitet, die von Dritten (sprich nicht von Nutzenden selber) zur Verfügung gestellt wurden, so können diese Informationen fälschlicherweise als Anweisung an das KI-Modell interpretiert werden und zu Ausgaben führen, die vom Nutzenden nicht beabsichtigt sind. Personen können solche Anweisungen absichtlich in Informationen, von denen sie wissen, dass sie mit KI verarbeitet werden, platzieren, um die Bearbeitung zu ihren Gunsten zu beeinflussen (z. B. kann bei einer Prüfung auf Kreditwürdigkeit in den eingereichten Informationen die Aufforderung versteckt werden, dem Antrag auf jeden Fall stattzugeben).

**Bewertung:** Dieses Risiko besteht immer, wenn Informationen aus anderen Quellen (z. B. Dokumente, Webseiten, E-Mails) mit Sprachmodellen verarbeitet werden.

Die Auswirkungen können umso größer sein, je mehr Rechte ein Sprachmodell hat selbstständig Aktionen durchzuführen (z. B. E-Mails versenden oder Daten abspeichern).

**Maßnahmen:** Technisch lässt sich das Problem nur schwer eingrenzen, da ein Sprachmodell im Wesentlichen in seiner vorgesehenen Funktion agiert, wichtig ist daher die möglichen Konsequenzen einer Manipulation zu kontrollieren (z. B. manuelle Prüfung/Nachbearbeitung der Ausgabe, einem Sprachmodell nur notwendige Rechte einräumen).

## Manipulation des KI-Systems außerhalb des Betriebs (sog. Poisoning Attacks)

**Risiko:** Angreifende können ein KI-Modell außerhalb des Betriebs – i.d.R. während der Entwicklungsphase – manipulieren und so z. B. Hintertüren einbauen, die während des Betriebs ausgenutzt werden können. Manipulationen sind an allen Komponenten denkbar, wie Trainingsdaten, dem Modell selbst, System-Prompts oder vorgeschalteten Bearbeitungstools. Konsequenzen können z. B. eine Leistungsver schlechterung oder die Möglichkeit der Ausnutzung für missbräuchliche Zwecke während des Betriebs sein.

**Bewertung:** Zur Beurteilung der Relevanz dieses Risikos sollten Sie die möglichen Konsequenzen einer Manipulation betrachten.

**Maßnahmen:** Auf dieses Risiko haben Sie i.d.R. nur Einfluss, wenn Sie eine KI-Anwendung oder Komponenten für diese selbst entwickeln. In diesem Fall sollten klassische IT-Sicherheitsmaßnahmen (z. B. Zugriffsbeschränkungen auf Programmcode und Trainingsdaten) ergriffen werden, um Manipulationsmöglichkeiten zu unterbinden.

Bei extern bereitgestellten KI-Anwendungen oder Komponenten (z. B. Trainingsdatensätzen) sollte auf eine Herkunft aus vertrauenswürdigen Quellen und das Vorliegen eines sicheren Dateiformats geachtet werden. Zudem können Informationen vom Anbieter eingeholt werden, welche Maßnahmen dieser ergriffen hat, um Manipulationen durch Dritte zu verhindern.