Mustervorlagen und Checklisten

zum Buch
„KI-Verordnung und EU AI Act für Unternehmen und Selbständige“

Hartmut Kruger

Arbeitsstand: 18. Juli 2026
Downloadseite: https://www.hartmut-kruger.de/de/ai-act/
Die Vorlagen sind allgemeine Muster und müssen an Unternehmen, Branche, Systeme, Verträge und Datenverarbeitungen angepasst werden.

Inhalt und Verwendung

Dieses Dokument enthält direkt nutzbare Mustertexte, Formulare, Schnellchecks, Entscheidungsleitfäden und Vertragsklauseln. Ergänzend steht eine Excel-Arbeitsmappe für Bestandsverzeichnis, Risikoklassifizierung, Schulung, Freigaben, Vorfälle und den 90-Tage-Plan zur Verfügung. Platzhalter in eckigen Klammern sind vor der betrieblichen Verwendung zu ersetzen.

- AI-Act-Schnellcheck
- Kompakte betriebliche KI-Richtlinie
- Mitarbeiterunterweisung
- Freigabeformular für ein KI-System
- Anbieterfragebogen
- Schulungsnachweis und Kompetenzmatrix
- Kennzeichnungsmuster für KI-Inhalte
- Muster eines jährlichen KI-Checks
- Kompakte Checklisten für den Arbeitsalltag
- Entscheidungsleitfäden für typische Situationen
- Musterklauseln für Dienstleister und Projektverträge

Hinweis: Eine ausgefüllte Vorlage ersetzt keine Prüfung des konkreten Anwendungsfalls. Bei Hochrisiko-KI, biometrischen Anwendungen, Entscheidungen über Personen, besonderen Kategorien personenbezogener Daten oder sicherheitskritischen Funktionen ist zusätzliche fachliche oder rechtliche Unterstützung sinnvoll.

1. AI-Act-Schnellcheck

Der Schnellcheck ermöglicht eine erste Einordnung und ersetzt keine vertiefte Prüfung. Fragen aus roten Bereichen führen zu einem Stopp oder einer gesonderten fachlichen Bewertung.

System

- Erzeugt oder bewertet die Software Texte, Bilder, Empfehlungen, Prognosen oder Entscheidungen?
- Leitet sie aus Eingaben selbständig Ausgaben ab?
- Wurde durch ein Update eine neue KI-Funktion aktiviert?

Rolle

- Nutzen wir nur einen fertigen Dienst?
- Bieten wir die Lösung unter eigener Marke an?
- Haben wir Funktion oder Zweck wesentlich verändert?
- Vertreiben wir ein fremdes System?

Verbote

- Soll die KI Menschen erheblich manipulieren oder Schutzbedürftigkeit ausnutzen?
- Werden Emotionen am Arbeitsplatz aus biometrischen Daten abgeleitet?
- Werden sensible Eigenschaften biometrisch kategorisiert?
- Werden ungezielt Gesichtsbilder gesammelt?

Hochrisiko

- Bewertet oder filtert die KI Bewerber oder Beschäftigte?
- Bildet sie Ranglisten oder Eignungswerte?
- Bewertet sie Kreditwürdigkeit natürlicher Personen?
- Wird sie in kritischer Infrastruktur, Bildung oder Rechtspflege eingesetzt?

Daten

- Enthalten Prompts personenbezogene oder vertrauliche Daten?
- Sind besondere Kategorien, Personalakten oder Gesundheitsdaten betroffen?
- Kann der Zweck mit weniger oder fiktiven Daten erreicht werden?

Anbieter

- Sind Produkt, Tarif, Training, Speicherung, Unterauftragnehmer und Drittlandzugriffe bekannt?
- Besteht ein erforderlicher Auftragsverarbeitungsvertrag?
- Können Daten gelöscht und exportiert werden?

Kontrolle

- Ist eine fachkundige Person benannt?
- Hat sie Ausgangsdaten, Zeit und Ablehnungsrecht?
- Werden Quellen und Berechnungen unabhängig geprüft?

Transparenz

- Interagieren Kunden direkt mit KI?
- Werden realitätsähnliche künstliche Medien veröffentlicht?
- Besteht ein echter redaktioneller Freigabeprozess?

Organisation

- Sind Nutzer geschult?
- Sind Dienstleister einbezogen?
- Bestehen Bestandsverzeichnis, Vorfallweg und Prüftermin?

Gesamtauswertung

Grün: normale Assistenz, freigegebenes Konto, geringe Datenrisiken, wirksame Kontrolle.

Gelb: personenbezogene oder vertrauliche Daten, externe Kommunikation, unklare Anbieterbedingungen oder begrenzte Entscheidungswirkung.

Rot: mögliche verbotene Praxis, Bewertung von Menschen, besondere Daten ohne Prüfung, erhebliche automatisierte Entscheidung oder fehlende Kontrolle.

2. Ausformulierte Muster und Arbeitshilfen

Hinweis zur Verwendung

Die folgenden Muster sind bewusst so formuliert, dass sie von Selbständigen und kleinen Unternehmen ohne umfangreiche Rechtsabteilung als Ausgangspunkt verwendet werden können. Sie müssen an die tatsächlichen Systeme, Daten, Verträge, Beschäftigtenstrukturen und Branchenanforderungen angepasst werden. Ein ausgefülltes Muster ersetzt keine Prüfung des konkreten Anwendungsfalls. Besonders bei Hochrisiko-KI, biometrischen Funktionen, besonderen Kategorien personenbezogener Daten, Berufsgeheimnissen oder erheblichen Entscheidungen sollte fachkundige Unterstützung einbezogen werden.

Die Muster folgen dem Grundsatz der Verhältnismäßigkeit. Ein kleiner Betrieb benötigt klare, gelebte Regeln und nachvollziehbare Entscheidungen, aber keine künstlich aufgeblähte Konzernbürokratie. Die Vorlagen können einzeln übernommen, gekürzt oder mit bestehenden Datenschutz-, IT-Sicherheits- und Qualitätsprozessen verbunden werden.

2.1 Muster einer kompakten betrieblichen KI-Richtlinie

1. Zweck und Grundsatz

Unser Unternehmen setzt künstliche Intelligenz ein, um Beschäftigte bei geeigneten Aufgaben zu unterstützen, Arbeitsabläufe zu verbessern und neue Leistungen zu entwickeln. KI-Systeme werden so verwendet, dass Kunden, Beschäftigte, Geschäftspartner und Unternehmensinformationen geschützt bleiben. Die Verantwortung für betriebliche Entscheidungen, Veröffentlichungen und Arbeitsergebnisse verbleibt bei den zuständigen Menschen. KI-Ausgaben gelten grundsätzlich als Entwürfe, solange sie nicht nach dem vorgesehenen Verfahren geprüft und freigegeben wurden.

2. Geltungsbereich

Diese Richtlinie gilt für Geschäftsführung, Beschäftigte, Praktikanten, freie Mitarbeiter und externe Dienstleister, soweit sie im Auftrag des Unternehmens KI-Systeme verwenden oder KI-generierte Ergebnisse bearbeiten. Erfasst sind eigenständige KI-Dienste, KI-Funktionen in vorhandener Software, Browsererweiterungen, mobile Apps, Programmierschnittstellen, lokale Modelle und durch Dienstleister betriebene Systeme. Die Bezeichnung des Produkts ist unerheblich; maßgeblich ist die tatsächliche Funktion.

3. Zugelassene Systeme und Konten

Für Unternehmens- und Kundendaten dürfen nur die in der jeweils aktuellen Systemfreigabeliste genannten Geschäftskonten, Funktionen und Verwendungszwecke eingesetzt werden. Private oder nicht zentral freigegebene Konten sind für Unternehmensdaten nicht zulässig. Neue Apps, Erweiterungen, Schnittstellen und Agentenfunktionen dürfen erst nach der vorgesehenen Freigabe aktiviert werden. Nutzer prüfen vor jeder Verwendung, ob sie im richtigen Geschäftskonto angemeldet sind.

4. Zulässige Standardanwendungen

Ohne zusätzliche Einzelfreigabe können die ausdrücklich freigegebenen Systeme insbesondere für allgemeine Textentwürfe, sprachliche Überarbeitung eigener Inhalte, Gliederungen, Ideensammlungen, fiktive Beispiele, Zusammenfassungen freigegebener Unterlagen und einfache Tabellenhilfen verwendet werden. Voraussetzung ist, dass keine roten Daten verarbeitet, keine Personen bewertet und keine Inhalte ungeprüft extern veröffentlicht werden. Die jeweils freigegebene Funktion und der dokumentierte Zweck sind einzuhalten.

5. Nicht zulässige oder besonders freigabepflichtige Anwendungen

Ohne dokumentierte Sonderprüfung sind insbesondere unzulässig: die Eingabe von Passwörtern, API-Schlüsseln oder anderen Zugangsdaten; die Verarbeitung vollständiger Personalakten, Gesundheitsdaten oder zentraler Geschäftsgeheimnisse in allgemeinen KI-Diensten; die Bewertung, Rangordnung oder automatische Entscheidung über Bewerber, Beschäftigte, Kunden oder andere natürliche Personen; biometrische oder emotionale Auswertungen; die automatische Veröffentlichung externer Inhalte; sowie Agentenaktionen mit Zahlungen, Vertragsänderungen, Datenlöschung oder unkontrolliertem Versand. Möglicherweise verbotene Praktiken sind unverzüglich zu stoppen und an die verantwortliche Stelle zu melden.

6. Datenampel und Datenminimierung

Vor jeder Eingabe ist zu prüfen, welche Informationen für die Aufgabe tatsächlich benötigt werden. Grüne Daten können in freigegebenen Systemen grundsätzlich verwendet werden. Gelbe Daten benötigen eine Prüfung von Zweck, Konto, Rechtsgrundlage, Geheimhaltung und erforderlicher Datenmenge. Rote Daten dürfen nur nach ausdrücklicher Freigabe in einer dafür geeigneten technischen Umgebung verarbeitet werden. Wenn ein Ausschnitt, eine anonymisierte Arbeitskopie oder fiktive Daten ausreichen, darf nicht die vollständige Originaldatei verwendet werden. Kommentare, Metadaten, versteckte Tabellenblätter und Zugangsdaten sind vor einem Upload zu entfernen.

7. Kontrolle und Freigabe von Ergebnissen

KI-Ergebnisse werden auf Richtigkeit, Vollständigkeit, Aktualität, Quellen, Zahlen, Rechte Dritter und Eignung für den konkreten Zweck geprüft. Wichtige Quellen sind im Original zu öffnen; Berechnungen sind nachvollziehbar zu kontrollieren. Für rechtlich, steuerlich, technisch, finanziell oder personell bedeutsame Inhalte ist eine fachkundige Person erforderlich. Die Freigabestufe richtet sich nach der möglichen Auswirkung eines Fehlers. Eine Person darf eine Ausgabe nur freigeben, wenn sie Zugang zu den Ausgangsdaten besitzt, ausreichend Zeit hat und den Vorschlag tatsächlich ändern oder ablehnen kann.

8. Externe Inhalte und Kennzeichnung

Texte, Bilder, Audio- und Videoinhalte dürfen erst nach der vorgesehenen Tatsachen-, Rechte- und Kennzeichnungsprüfung veröffentlicht werden. Künstliche Produktmerkmale, erfundene Referenzen, Testimonials oder Zertifizierungen sind unzulässig. Bei Kundenchatbots muss die KI-Interaktion erkennbar sein. Realitätsnahe künstliche oder wesentlich manipulierte Inhalte werden entsprechend den gesetzlichen und betrieblichen Regeln offengelegt. Technische Herkunftsmarkierungen dürfen nicht ohne sachlichen Grund entfernt werden.

9. Personalentscheidungen

KI darf im Personalbereich ohne Sonderfreigabe nur für eindeutig administrative oder sprachliche Assistenz verwendet werden. Bewertung, Filterung, Rangordnung, Profiling, Leistungsüberwachung, Aufgabenverteilung und andere entscheidungsrelevante Funktionen benötigen eine dokumentierte Prüfung nach AI Act, Datenschutz-, Arbeits- und Gleichbehandlungsrecht. Emotionserkennung anhand biometrischer Daten ist am Arbeitsplatz grundsätzlich ausgeschlossen, soweit keine eng geprüfte gesetzliche Ausnahme vorliegt.

10. Vorfälle und Meldung

Fehlübermittlungen, kompromittierte Konten, falsche wesentliche Ausgaben, unerwartete Datenzugriffe, mögliche Diskriminierung, fehlende Kennzeichnung und nicht freigegebene Anwendungen sind unverzüglich an [Ansprechpartner/Funktion] zu melden. Die sofortige Schadensbegrenzung hat Vorrang. Zugangsdaten werden gesperrt, Inhalte soweit möglich gelöscht, betroffene Veröffentlichungen korrigiert und Datenschutz- oder Informationspflichten geprüft. Eine unverzügliche und vollständige Meldung wird bei der Bewertung des Vorfalls berücksichtigt.

11. Schulung und Aktualisierung

Nutzer nehmen an den für ihre Rolle vorgesehenen KI-Kompetenzmaßnahmen teil. Neue Beschäftigte erhalten eine Grundinformation; bei Rollenwechsel, neuen Systemen oder wesentlichen Funktionen erfolgt eine ergänzende Einweisung. Die Richtlinie, Systemfreigabeliste und Datenampel werden mindestens jährlich sowie nach wesentlichen Änderungen oder Vorfällen geprüft. Die jeweils aktuelle Fassung ist unter [Ablageort] verfügbar.

2.2 Muster einer Mitarbeiterunterweisung

Ziel der Unterweisung

Nach der Unterweisung sollen die Teilnehmer die freigegebenen Systeme erkennen, Daten vor der Eingabe einordnen, typische Fehler von KI-Ausgaben verstehen und wissen, wann eine Freigabe oder Meldung erforderlich ist. Die Schulung soll keine Angst vor KI erzeugen. Sie soll einen sicheren und produktiven Einsatz ermöglichen.

Einführungstext für die Schulung

Künstliche Intelligenz kann Texte, Bilder, Zusammenfassungen und Vorschläge schnell erzeugen. Sie versteht unsere Arbeit jedoch nicht wie ein verantwortlicher menschlicher Sachbearbeiter. Eine Antwort kann überzeugend klingen und trotzdem falsch, unvollständig oder veraltet sein. Aus diesem Grund behandeln wir KI als Assistenzwerkzeug. Wir nutzen ihre Geschwindigkeit, behalten aber Daten, Prüfung und Entscheidung unter menschlicher Kontrolle.

Die fünf verbindlichen Grundregeln

- Nur freigegebene Geschäftskonten und Funktionen verwenden. Private Konten, unbekannte Apps und selbst installierte Erweiterungen dürfen keine Unternehmens- oder Kundendaten verarbeiten.
- Vor jedem Prompt die Datenampel anwenden. Zugangsdaten und rote Informationen niemals in einen allgemeinen KI-Dienst eingeben.
- Nur die für die Aufgabe notwendige Information bereitstellen. Wenn ein Ausschnitt, eine anonymisierte Kopie oder ein fiktives Beispiel genügt, darf nicht die vollständige Datei hochgeladen werden.
- Jede KI-Ausgabe zunächst als Entwurf behandeln. Tatsachen, Quellen, Zahlen und wichtige Formulierungen vor Verwendung prüfen.
- Fehler, falsche Uploads und unerwartete Zugriffe sofort melden. Ein früh gemeldeter Fehler lässt sich häufig begrenzen; ein verschwiegener Fehler kann sich vergrößern.

Typische Halluzinationen

KI-Systeme können Quellen, Urteile, Zahlen, Produktfunktionen oder Zitate erfinden. Sie können verschiedene Länder, Zeitstände oder Dokumentversionen vermischen. Eine professionelle Sprache ist kein Beweis für Richtigkeit. Bei wichtigen Informationen ist die Originalquelle zu öffnen. Bei Zahlen sind Ausgangsdaten und Rechenweg zu kontrollieren. Wenn das System eine Quelle nicht zuverlässig belegen kann, darf die Aussage nicht übernommen werden.

Praxisübung 1: Datenampel

Ordnen Sie folgende Beispiele ein und begründen Sie Ihre Entscheidung: eine veröffentlichte Produktbeschreibung; eine Kunden-E-Mail mit Name und Reklamation; eine vollständige Bewerbung;

ein API-Schlüssel; eine fiktive Umsatzliste; eine interne Preisuntergrenze. Besprechen Sie anschließend, wie gelbe Informationen durch Kürzung, Pseudonymisierung oder Abstraktion sicherer verwendet werden können.

Praxisübung 2: Quellenprüfung

Die Teilnehmer erhalten einen KI-Text mit einer richtigen amtlichen Quelle, einer veralteten Frist und einer erfundenen Fundstelle. Die Aufgabe besteht nicht darin, anhand des Sprachstils zu raten. Jede tragende Aussage wird in der Originalquelle überprüft. Anschließend wird dokumentiert, welche Fehler gefunden und wie sie korrigiert wurden.

Praxisübung 3: Kritischer Anwendungsfall

Ein Mitarbeiter möchte 40 Bewerbungen hochladen und eine Rangliste erstellen lassen. Die Teilnehmer sollen erkennen, dass nicht nur Datenschutz betroffen ist. Die Funktion kann Personen wesentlich beeinflussen und eine Hochrisiko-Prüfung erfordern. Die richtige Reaktion ist nicht, den Prompt geschickter zu formulieren, sondern die Anwendung zu stoppen und die zuständige Stelle einzubeziehen.

Abschluss der Unterweisung

Jeder Teilnehmer nennt zum Abschluss ein erlaubtes Beispiel aus seinem Arbeitsbereich, ein Beispiel für gelbe Daten und einen Fall, der vor Nutzung freigegeben werden muss. Offene Fragen werden gesammelt und in die nächste Aktualisierung der Richtlinie oder Schulung übernommen. Die Teilnahme wird mit Datum, Inhalt und Zielgruppe dokumentiert.

2.3 Muster eines Freigabeformulars für ein KI-System

Das Formular kann als Word-Dokument, Ticket oder digitales Formular umgesetzt werden. Wichtig ist, dass der Antragsteller die konkrete Funktion und nicht nur den Produktnamen beschreibt.

A. Antrag

- Beantragendes Team / Ansprechpartner: _____
- System, Anbieter und Tarif: _____
- Konkrete KI-Funktion: _____
- Geplanter Verwendungszweck: _____
- Vorgesehene Nutzer und Anzahl: _____
- Geplanter Start und Testdauer: _____

B. Daten und Systeme

- Welche Eingaben werden verarbeitet? Text, Datei, Bild, Audio, Datenbank oder verbundene Anwendung?
- Sind personenbezogene Daten betroffen? Wenn ja: welche Personen und Datenkategorien?
- Sind besondere Kategorien, Bewerbungen, Personalakten oder Geschäftsgeheimnisse betroffen?
- Welche Systeme werden verbunden? Bestehen Lese-, Schreib-, Versand- oder Löschrechte?
- Kann der Zweck mit weniger, anonymisierten oder fiktiven Daten erreicht werden?

C. Wirkung und Entscheidung

- Welche Ausgabe entsteht: Entwurf, Zusammenfassung, Bewertung, Rangfolge, Prognose, Entscheidung oder Aktion?
- Kann die Ausgabe natürliche Personen wesentlich beeinflussen?
- Findet Profiling, biometrische Verarbeitung oder Emotionserkennung statt?
- Interagieren Kunden oder Beschäftigte unmittelbar mit dem System?
- Wer prüft die Ausgabe und kann sie ablehnen oder das System stoppen?

D. Anbieter und Vertrag

- Ist ein Geschäftskonto vorhanden?
- Werden Eingaben oder Ausgaben zum Training verwendet?
- Welche Speicherfristen, Unterauftragnehmer und Drittlandzugriffe bestehen?
- Ist ein Auftragsverarbeitungsvertrag erforderlich und vorhanden?
- Wie werden Updates, Vorfälle, Export und Löschung geregelt?

E. Vorläufige Bewertung

- Rolle nach AI Act: Betreiber / Anbieter / Importeur / Händler / unklar
- Risikoeinstufung: normale Assistenz / Transparenz / Hochrisiko-Prüfung / möglicherweise verboten
- Datenschutz: keine personenbezogenen Daten / Prüfung abgeschlossen / Datenschutz-Folgenabschätzung prüfen
- Schutzstufe der Daten: öffentlich / intern / vertraulich / streng vertraulich

F. Entscheidung

- □ freigegeben
- □ eingeschränkt freigegeben mit folgenden Auflagen: _____
- □ nur Pilot mit fiktiven oder bereinigten Daten
- □ weitere Anbieter-, Datenschutz- oder Rechtsprüfung erforderlich
- □ abgelehnt / vorläufig gestoppt
- Verantwortliche Person, Datum und nächster Prüftermin: _____

2.4 Ausführlicher Fragenkatalog für Anbieter

Der Fragenkatalog dient nicht dazu, jeden kleinen Textassistenten mit einem umfangreichen Vergabeverfahren zu belasten. Er wird nach Risiko gekürzt oder erweitert. Für eine Personal-, Finanz-, Sicherheits- oder biometrische Anwendung sollten die Antworten schriftlich und prüfbar vorliegen.

System, Rolle und Verwendungszweck

- Welche juristische Person ist Anbieter und Vertragspartner?
- Welches Modell und welche Systemkomponenten werden verwendet?
- Wie lautet der dokumentierte vorgesehene Verwendungszweck?
- Wie stuft der Anbieter das System nach dem AI Act ein und warum?
- Welche Funktionen sind getrennt zu betrachten und welche sind standardmäßig aktiviert?
- Welche Änderungen des Kunden können eine neue Anbieterrolle oder Hochrisikoeinstufung auslösen?

Daten und Datenschutz

- Welche Eingabe-, Ausgabe-, Konto-, Nutzungs- und Protokolldaten werden verarbeitet?
- Werden Daten für Training, Verbesserung, Sicherheitsprüfung oder Support verwendet?
- Welche Aufbewahrungs- und Löschfristen gelten für aktive Daten, Backups und Protokolle?
- Welche Unterauftragnehmer werden eingesetzt und wie werden Änderungen angekündigt?
- In welchen Staaten werden Daten gespeichert oder zugänglich gemacht?
- Welche Verfahren unterstützen Auskunft, Berichtigung, Löschung und Datenexport?

Funktion, Qualität und Grenzen

- Welche bekannten Fehlertypen und Leistungsgrenzen bestehen?
- Wie wurden Genauigkeit, Robustheit und mögliche Verzerrungen getestet?
- Für welche Zielgruppen, Sprachen und Situationen ist das System nicht geeignet?
- Wie werden Unsicherheit und fehlende Informationen angezeigt?
- Kann der Kunde die verwendeten Quellen, Eingaben und wesentlichen Verarbeitungsschritte nachvollziehen?
- Welche Auswirkungen haben Modell- und Funktionsupdates auf bisherige Tests?

Menschliche Kontrolle und Protokolle

- Welche menschliche Aufsicht sieht der Anbieter vor?
- Kann eine Empfehlung geändert, abgelehnt und das System gestoppt werden?
- Welche Protokolle stehen dem Betreiber zur Verfügung und wie lange?
- Lassen sich automatische Aktionen begrenzen, bestätigen oder rückgängig machen?
- Welche Schnittstellen und Berechtigungen sind erforderlich?

Sicherheit, Vorfälle und Betrieb

- Welche Verschlüsselung, Identitäts- und Berechtigungsmechanismen werden eingesetzt?
- Welche unabhängigen Sicherheits- oder Compliance-Nachweise liegen vor?
- Wie und innerhalb welcher Frist werden Sicherheits- und KI-Vorfälle gemeldet?
- Welche Service-Level, Wiederherstellungszeiten und Supportkanäle gelten?
- Kann der Kunde neue Funktionen zentral deaktivieren und vorab testen?

Vertragsende und Haftung

- In welchem Format können Daten, Protokolle und Konfigurationen exportiert werden?
- Wie wird die vollständige Löschung nach Vertragsende nachgewiesen?
- Welche Nutzungsrechte gelten für Ein- und Ausgaben?
- Welche Haftungsgrenzen und Unterstützungen bei Ansprüchen Dritter bestehen?
- Welche Fristen und Rechte gelten bei wesentlichen Vertragsänderungen?

2.5 Muster eines KI-Vorfallplans

Zweck

Der Vorfallplan sorgt dafür, dass Fehler schnell begrenzt und nicht aus Unsicherheit verschwiegen werden. Er gilt für Datenschutz-, Sicherheits-, Qualitäts-, Fairness-, Transparenz- und Aktionsvorfälle. Die Meldung ist auch dann erforderlich, wenn noch unklar ist, ob tatsächlich ein Rechtsverstoß vorliegt.

Sofortmaßnahmen des Nutzers

1. Weitere Eingaben, Veröffentlichungen oder automatische Aktionen stoppen.
2. Betroffenes System, Konto, Zeitpunkt, Prompt, Datei und Ausgabe sichern oder nachvollziehbar dokumentieren. Keine Beweise vorschnell löschen.
3. Bei Zugangsdaten oder Schlüsseln: sofort sperren, ersetzen und IT informieren.
4. Chat, Datei oder Freigabelink soweit möglich deaktivieren oder löschen, ohne die notwendige Beweissicherung zu verhindern.
5. Den Vorfall unverzüglich an [Kontakt] melden und keine eigenständigen Zusagen an Kunden oder Betroffene machen, sofern dies nicht zur unmittelbaren Gefahrenabwehr notwendig ist.

Erstbewertung durch die verantwortliche Stelle

- Welche Daten, Personen, Kunden, Systeme und Veröffentlichungen sind betroffen?
- Konnte ein Dritter auf Inhalte zugreifen oder wurden Daten in ein nicht freigegebenes System übertragen?
- Ist eine falsche oder diskriminierende Entscheidung entstanden?
- Wurden Passwörter, Geschäftsgeheimnisse, besondere personenbezogene Daten oder sicherheitsrelevante Informationen betroffen?
- Müssen Datenschutzbeauftragter, IT-Sicherheit, Geschäftsführung, Kunde, Versicherer, Behörde oder betroffene Personen einbezogen werden?

Begrenzung und Kommunikation

Zugriffe werden entzogen, Funktionen deaktiviert und fehlerhafte Veröffentlichungen korrigiert. Die Kommunikation beschreibt gesicherte Tatsachen und vermeidet Spekulation. Bei personenbezogenen Daten werden Meldungen nach Artikel 33 und 34 DSGVO geprüft. Bei möglichen AI-Act-Verstößen

werden zuständige Fachstellen und Anbieter einbezogen. Vertrags- und Berufsgeheimnisse können zusätzliche Informationspflichten auslösen.

Ursachenanalyse und Abschluss

Nach der akuten Phase wird untersucht, ob eine fehlende Alternative, unklare Richtlinie, falsche Berechtigung, unzureichende Schulung, Anbieteränderung oder vorsätzliche Umgehung Ursache war. Maßnahmen können technische Sperren, neue Freigaben, Nachschulung, Vertragsänderung oder Beendigung des Systems sein. Der Vorfall wird erst abgeschlossen, wenn Verantwortlicher, Maßnahmen, offene Risiken und Prüftermin dokumentiert sind.

Mindestfelder des Vorfallprotokolls

- Vorfallnummer, Datum, meldende Person und System
- Beschreibung, betroffene Daten und Personen
- Sofortmaßnahmen und Zeitpunkt
- Risikobewertung und erforderliche Meldungen
- Anbieter- und Behördenkontakte
- Ursache, Korrekturmaßnahme und Abschlussdatum
- Überprüfung der Wirksamkeit

2.6 Muster für Schulungsnachweis und Kompetenzmatrix

Schulungsnachweis

- Unternehmen / Organisation: _____
- Bezeichnung der Maßnahme: _____
- Datum, Dauer und Schulungsform: _____
- Zielgruppe und teilnehmende Rollen: _____
- Behandelte Systeme und Funktionen: _____
- Inhalte: KI-Grundlagen / Datenampel / Halluzinationen / Quellenprüfung / Datenschutz / Personal / Medien / Sicherheit / Vorfälle
- Praktische Übungen und festgestellte Fragen: _____
- Verantwortliche oder schulende Person: _____
- Nächste Aktualisierung oder anlassbezogener Trigger: _____

Beispiel einer Kompetenzmatrix

Geschäftsführung: Kennt Rollen, wesentliche Risiken, Freigaben, Ressourcen und Eskalation. Sie muss keine tägliche Promptpraxis besitzen, aber Anbieter- und Einsatzentscheidungen sachgerecht treffen können.

Normale Nutzer: Können freigegebene Konten erkennen, die Datenampel anwenden, Ausgaben als Entwurf behandeln und Fehler melden.

Power-User: Können neue Anwendungsfälle strukturiert testen, Daten minimieren, Vorlagen entwickeln und Grenzen dokumentieren. Sie besitzen nicht automatisch rechtliche Freigabebefugnis.

Fachliche Freigeber: Können Ausgangsdaten und Quellen prüfen, eine Ausgabe ablehnen und Verantwortung für die finale Verwendung übernehmen.

Personal: Erkennt Hochrisiko- und Verbotsfälle, Profiling, Diskriminierungsrisiken und Grenzen biometrischer Funktionen.

Marketing: Prüft Produktwahrheit, Urheber-, Marken- und Persönlichkeitsrechte sowie Kennzeichnung künstlicher Medien.

IT und Administration: Versteht Konten, Berechtigungen, Verbindungen, Protokolle, Prompt Injection, Agentenrechte, Updates und Vorfallreaktion.

Externe Dienstleister: Kennen die für den Auftrag freigegebenen Systeme, Daten, Prüfpflichten, Unterauftragnehmerregeln und Meldewege.

Bewertung der Maßnahme

Die Schulung gilt nicht allein aufgrund der Anwesenheit als abgeschlossen. Bei normalen Nutzern genügt eine kurze Verständniskontrolle. Bei kritischen Aufgaben sollte eine praktische Übung zeigen, dass die Person Datenrisiken, Systemgrenzen und Ablehnungsmöglichkeiten tatsächlich anwenden kann. Festgestellte Unsicherheiten werden als Schulungsbedarf, nicht als persönliches Versagen dokumentiert.

2.7 Musterformulierungen für Transparenz und Kennzeichnung

Kundenchatbot – Kurzfassung

Sie kommunizieren mit einem KI-gestützten Assistenten.

Kundenchatbot – erweiterte Fassung

Sie kommunizieren mit einem KI-gestützten Assistenten. Die Antworten können unvollständig oder fehlerhaft sein. Bei komplexen, persönlichen oder verbindlichen Fragen können Sie die Übergabe an einen Mitarbeiter verlangen.

KI-generierte Illustration

Diese Abbildung wurde mit generativer KI erstellt und dient der Veranschaulichung. Sie zeigt kein tatsächlich ausgeführtes Projekt und keine verbindliche Produkteigenschaft.

KI-verändertes reales Bild

Das Ausgangsbild wurde mit KI wesentlich verändert. Die gekennzeichneten Elemente entsprechen nicht vollständig der realen Aufnahme.

Virtuelle Einrichtung oder Produktumgebung

Virtuelle KI-Visualisierung. Einrichtung und Umgebung sind beispielhaft und nicht Bestandteil des Angebots.

Künstliche Stimme

Die Sprecherstimme wurde künstlich erzeugt. Die dargestellte Person hat diesen Text nicht selbst eingesprochen.

Künstlich erzeugtes Video einer realen Person

Bild und Stimme dieser Darstellung wurden mit KI erzeugt. Es handelt sich nicht um eine authentische Aufnahme des gezeigten Ereignisses.

Redaktionell geprüfter Fachtext

KI-Werkzeuge wurden bei Strukturierung und sprachlicher Bearbeitung unterstützend eingesetzt. Inhalt, Quellen und rechtliche Aussagen wurden redaktionell geprüft und von der genannten verantwortlichen Person freigegeben.

Fiktives Beispiel

Fiktives KI-Beispiel. Personen, Unternehmen und Ereignisse sind erfunden und dienen ausschließlich der Erläuterung.

Hinweis zur Auswahl

Die Formulierung muss den tatsächlichen Arbeitsprozess beschreiben. Ein Hinweis darf weder menschliche Prüfung behaupten, die nicht stattgefunden hat, noch einen künstlichen Inhalt als bloße technische Optimierung verharmlosen. Bei Deepfakes oder anderen gesetzlichen

Offenlegungspflichten muss der Hinweis klar, rechtzeitig und wahrnehmbar sein. Ein bloßer Eintrag im Impressum oder versteckte Metadaten genügen für eine sichtbare Offenlegung regelmäßig nicht.

2.8 Muster eines jährlichen KI-Checks

1. Organisation

- Sind Verantwortlichkeiten, Ansprechpartner und Vertretungen aktuell?
- Ist die Richtlinie bekannt und leicht auffindbar?
- Funktioniert der Freigabe- und Vorfallweg in der Praxis?

2. Bestand und Rollen

- Sind alle Systeme, Apps, Erweiterungen und Dienstleister erfasst?
- Wurden nicht mehr verwendete Konten und Schnittstellen entfernt?
- Sind Anbieter-, Betreiber-, Händler- oder Importeurrollen noch zutreffend?

3. Zwecke und Risiken

- Haben sich Verwendungszweck, betroffene Personen oder Entscheidungswirkung geändert?
- Sind neue Hochrisiko- oder Verbotsfragen entstanden?
- Wird eine ursprünglich administrative Funktion inzwischen wertend verwendet?

4. Daten und Datenschutz

- Sind Datenampel, Rechtsgrundlagen, Speicherfristen und Betroffeneninformationen aktuell?
- Wurden neue Datenquellen, besondere Kategorien oder umfangreiches Profiling eingeführt?
- Sind notwendige Datenschutz-Folgenabschätzungen und Auftragsverarbeitungsverträge vorhanden?

5. Anbieter und Technik

- Haben sich Tarif, Modell, Trainingsbedingungen, Unterauftragnehmer oder Datenstandorte geändert?
- Sind Mehrfaktor-Authentifizierung, Rollen, Protokolle und Löschung weiterhin wirksam?
- Wurden neue Agentenrechte, Apps oder automatische Funktionen freigeschaltet?

6. Qualität und menschliche Kontrolle

- Welche Fehler, Halluzinationen, Fehlalarme oder Beschwerden traten auf?
- Haben Freigeber ausreichend Zeit, Fachkenntnis und Ablehnungsbefugnis?
- Werden Quellen, Zahlen und Originalunterlagen tatsächlich geprüft?

7. Transparenz und Veröffentlichungen

- Sind Chatbots und direkte KI-Interaktionen korrekt erkennbar?
- Werden realitätsnahe künstliche Inhalte und öffentliche Informationstexte angemessen behandelt?
- Bleiben technische Herkunftsinformationen erhalten?

8. Kompetenz und Dienstleister

- Sind neue Nutzer, Rollenwechsel und externe Auftragnehmer geschult?
- Welche Fragen oder wiederkehrenden Fehler müssen in die nächste Schulung aufgenommen werden?
- Sind vertragliche KI-Regeln bei Dienstleistern noch aktuell?

9. Wirtschaftlichkeit und Exit

- Erreicht das System den erwarteten Nutzen unter Einbeziehung des Prüfaufwands?
- Bestehen unnötige Lizenzen oder Anbieterabhängigkeiten?
- Sind Datenexport, Löschung und manueller Ersatzprozess getestet?

10. Managemententscheidung

Für jedes wichtige System wird entschieden: unverändert fortführen, erweitern, einschränken, Anbieter wechseln, vertieft prüfen oder beenden. Offene Maßnahmen erhalten Verantwortlichen, Fälligkeit und Nachweis. Der jährliche Check endet nicht mit einer allgemeinen Aussage „alles in Ordnung“, sondern mit konkreten Entscheidungen für den nächsten Prüfzeitraum.

3. Kompakte Checklisten für den Arbeitsalltag

Die Checklisten sind als letzte Kontrolle vor einer Handlung gedacht. Sie ersetzen nicht die ausführliche Prüfung, helfen aber dabei, typische Fehler im entscheidenden Moment zu vermeiden. Unternehmen können sie in das Intranet übernehmen, als einseitige Merkblätter ausgeben oder in digitale Freigabeprozesse integrieren.

3.1 Vor jedem Prompt

- ☐ Bin ich im freigegebenen Geschäftskonto angemeldet?
- ☐ Ist die Aufgabe für dieses System und diesen Zweck freigegeben?
- ☐ Welche kleinste Informationsmenge wird benötigt?
- ☐ Enthält die Eingabe personenbezogene, vertrauliche oder rote Daten?
- ☐ Kann ich Namen, Werte oder Projektdetails entfernen?
- ☐ Ist klar, wie das Ergebnis geprüft und verwendet wird?
- ☐ Muss eine fachkundige Person oder Freigabestelle beteiligt werden?

Erst wenn Zweck, Konto, Daten und Kontrolle zusammenpassen, sollte die Eingabe abgesendet werden. Bei Unsicherheit wird nicht improvisiert, sondern die zuständige Stelle gefragt.

3.2 Vor dem Upload einer Datei

- ☐ Benötige ich wirklich die vollständige Datei oder genügt ein Ausschnitt?
- ☐ Sind Kommentare, Änderungsverfolgung oder frühere Versionen enthalten?
- ☐ Gibt es ausgeblendete Spalten, Tabellenblätter oder Folien?
- ☐ Enthält die Datei Passwörter, API-Schlüssel, interne Pfade oder Sicherheitsinformationen?
- ☐ Sind Namen, Kontaktdaten und andere Identifikatoren erforderlich?
- ☐ Darf die Datei vertraglich und urheberrechtlich an den Anbieter übertragen werden?
- ☐ Wurde eine bereinigte Arbeitskopie erstellt und eindeutig gekennzeichnet?

Das Original bleibt im vorgesehenen Unternehmenssystem. Die bereinigte Kopie wird nach dem festgelegten Zweck und der vorgesehenen Frist gelöscht.

3.3 Vor der Freigabe eines neuen KI-Systems

- ☐ Sind Produkt, Tarif, Anbieter und konkrete Funktion eindeutig benannt?
- ☐ Welche Rolle besitzt das Unternehmen nach AI Act und DSGVO?
- ☐ Welche Daten, Personen und Entscheidungen sind betroffen?
- ☐ Liegt eine verbotene Praxis oder ein Anhang-III-Bereich nahe?
- ☐ Sind Trainingsnutzung, Speicherung, Unterauftragnehmer und Drittlandzugriffe geklärt?
- ☐ Gibt es geeignete Sicherheits-, Protokoll- und Löschfunktionen?
- ☐ Ist menschliche Kontrolle praktisch möglich?
- ☐ Sind Nutzer, Schulung, Auflagen und nächster Prüftermin festgelegt?

Eine Freigabe ist auf Funktion, Zweck, Nutzer und Daten zu begrenzen. Der Produktname allein ist keine ausreichende Freigabebeschreibung.

3.4 Vor einer externen Veröffentlichung

- ☐ Sind Namen, Daten, Preise, Fristen, Produktmerkmale und Zitate richtig?
- ☐ Wurden tragende Quellen im Original geprüft?
- ☐ Sind Ausgangsmaterial und Ergebnis urheber-, marken- und persönlichkeitsrechtlich verwendbar?
- ☐ Wirkt ein künstliches Bild, Audio oder Video wie eine echte Aufnahme?
- ☐ Ist eine gesetzliche oder freiwillige Kennzeichnung erforderlich?
- ☐ Sind technische Herkunftsinformationen erhalten?

- ☐ Hat eine benannte Person die finale Fassung und nicht nur einen Entwurf freigegeben?
- ☐ Ist dokumentiert, welche Version veröffentlicht wurde?

Ein KI-Hinweis heilt keine falsche Tatsachenbehauptung. Produktwahrheit und Rechteprüfung bleiben immer erforderlich.

3.5 Vor dem Einsatz im Personalbereich

- ☐ Handelt es sich nur um Text- oder Terminassistenz oder werden Personen bewertet?
- ☐ Werden Bewerbungen analysiert, gefiltert, gerankt oder mit Punktwerten versehen?
- ☐ Findet Profiling oder eine Ableitung persönlicher Eigenschaften statt?
- ☐ Werden Gesicht, Stimme, Bewegung oder andere biometrische Daten ausgewertet?
- ☐ Kann die Funktion unter Anhang III oder ein Verbot fallen?
- ☐ Sind Datenschutz, AGG, Artikel 22 DSGVO und Beteiligungsrechte geprüft?
- ☐ Kann die menschliche Entscheidung unabhängig von der KI getroffen werden?
- ☐ Sind Anbieterunterlagen, Protokolle und Löschfristen vorhanden?

Bei Bewertung, Rangordnung, biometrischer Analyse oder unklarer Wirkung wird die Funktion nicht ohne vertiefte Prüfung verwendet.

3.6 Vor der Aktivierung eines Kundenchatbots

- ☐ Erkennen Kunden zu Beginn, dass sie mit KI kommunizieren?
- ☐ Greift der Bot nur auf aktuelle und freigegebene Informationen zu?
- ☐ Sind Garantie, Preis, Vertrag, Zahlung und Beschwerden als sensible Themen definiert?
- ☐ Gibt es eine erreichbare menschliche Übergabe?
- ☐ Werden nur notwendige Kundendaten abgefragt und angemessen authentifiziert?
- ☐ Welche Aktionen darf der Bot selbst ausführen?
- ☐ Sind Protokollierung, Speicherfrist und Qualitätsstichproben festgelegt?
- ☐ Wie wird ein fehlerhafter oder veralteter Inhalt schnell korrigiert?

Ein Chatbot sollte nicht mehr versprechen oder entscheiden dürfen, als das Unternehmen zuverlässig kontrollieren kann.

3.7 Vor der Aktivierung eines KI-Agenten

- ☐ Welche Daten darf der Agent lesen?
- ☐ Welche Systeme darf er verändern?
- ☐ Darf er E-Mails versenden, Dateien löschen, Zahlungen oder Verträge vorbereiten?
- ☐ Welche Aktionen benötigen eine menschliche Bestätigung?
- ☐ Sind Betrags-, Mengen-, Empfänger- und Zeitgrenzen eingerichtet?
- ☐ Werden alle Aktionen nachvollziehbar protokolliert?
- ☐ Gibt es Rückgängig-Funktion, Testumgebung und Notabschaltung?
- ☐ Wurde Prompt Injection mit manipulierten Dokumenten getestet?

Aktionsrechte werden nach dem Prinzip der geringsten Berechtigung vergeben. Ein Agent, der einen Entwurf erstellen kann, benötigt nicht automatisch Versand- oder Schreibrechte.

3.8 Nach einem Modell- oder Anbieterupdate

- ☐ Hat sich Modell, Verwendungszweck, Datenzugriff oder Trainingsbedingung geändert?
- ☐ Sind neue Apps, Agenten, Transkriptionen oder Bewertungsfunktionen aktiviert worden?
- ☐ Haben sich Unterauftragnehmer, Datenstandorte oder Speicherfristen geändert?
- ☐ Sind frühere Qualitäts- und Risikotests weiterhin aussagekräftig?
- ☐ Funktionieren technische Kennzeichnungen und Protokolle weiterhin?
- ☐ Müssen Richtlinie, Freigabeliste oder Schulung aktualisiert werden?
- ☐ Kann die neue Funktion deaktiviert oder auf die frühere Version zurückgestellt werden?

Wesentliche Änderungen werden wie eine neue oder veränderte Anwendung behandelt und nicht automatisch durch die ursprüngliche Produktfreigabe gedeckt.

3.9 Bei einer versehentlichen Datenübermittlung

- ☐ Weitere Verarbeitung stoppen und den Vorfall sofort melden.

- □ System, Konto, Datei, Zeitpunkt und betroffene Daten dokumentieren.
- □ Chat, Datei oder Freigabelink soweit möglich deaktivieren oder löschen.
- □ Passwörter, Tokens oder Schlüssel unverzüglich sperren und ersetzen.
- □ Anbieter, mögliche Empfänger und Unterauftragnehmer feststellen.
- □ Datenschutz-, Vertrags-, Geheimhaltungs- und Informationspflichten prüfen.
- □ Fehlerhafte externe Inhalte korrigieren oder zurückrufen.
- □ Ursache und Verbesserungsmaßnahme dokumentieren.

Die schnelle und vollständige Meldung ist wichtiger als der Versuch, den Fehler zu verbergen. Sie ermöglicht Löschung, Sperrung und rechtzeitige Information.

3.10 Jährliche Managemententscheidung

- □ Welche KI-Systeme sind geschäftskritisch geworden?
- □ Wo entsteht nach Abzug des Prüfaufwands ein messbarer Nutzen?
- □ Welche Fehler, Beschwerden und Vorfälle treten wiederholt auf?
- □ Welche Systeme beeinflussen Menschen oder verarbeiten sensible Daten?
- □ Sind Rollen, Verträge, Berechtigungen und Schulungen aktuell?
- □ Welche Anbieterabhängigkeiten und Exit-Risiken bestehen?
- □ Welche Systeme sollen erweitert, eingeschränkt, ersetzt oder beendet werden?
- □ Welche Rechts- und Behördenänderungen müssen umgesetzt werden?

Die jährliche Bewertung endet mit konkreten Entscheidungen, Verantwortlichen und Fristen. Ein System wird nicht allein deshalb fortgeführt, weil es bereits eingeführt wurde.

4. Entscheidungsleitfäden für typische Situationen

Die folgenden Leitfäden sind für Situationen gedacht, in denen eine schnelle betriebliche Entscheidung erforderlich ist. Sie verbinden die ausführlichen Kapitel mit einem klaren Ablauf und können als Grundlage für interne Arbeitsanweisungen verwendet werden.

4.1 Ein Mitarbeiter möchte ein neues KI-Werkzeug testen

Zuerst wird geklärt, ob bereits ein freigegebenes System dieselbe Aufgabe erfüllt. Ist dies der Fall, wird grundsätzlich die vorhandene Lösung verwendet. Fehlt die Funktion, beschreibt der Mitarbeiter den konkreten Zweck, die Nutzer, die benötigten Daten und die erwartete Ausgabe. Ein kleiner Test kann schnell zugelassen werden, wenn ausschließlich fiktive oder öffentliche Daten verwendet werden, keine externe Veröffentlichung erfolgt und das System keine Aktionen in anderen Anwendungen ausführen darf.

Sobald personenbezogene oder vertrauliche Daten, regelmäßige Nutzung, mehrere Nutzer oder externe Kommunikation vorgesehen sind, folgt die normale Anbieter- und Datenschutzprüfung. Bewertet das Werkzeug Menschen, verarbeitet es biometrische Daten, verbindet es umfangreiche Datenquellen oder führt es irreversible Aktionen aus, ist eine vertiefte Prüfung erforderlich. Das Ergebnis wird als freigegeben, eingeschränkt, offen oder abgelehnt dokumentiert. Ein einfaches und schnelles Verfahren verhindert, dass sinnvolle Ideen in private Schatten-KI ausweichen.

4.2 Eine interne Datei soll durch KI ausgewertet werden

Die verantwortliche Person bestimmt zunächst die konkrete Fragestellung. Anschließend wird geprüft, welche Informationen dafür tatsächlich erforderlich sind. Eine vollständige Datei ist nur selten notwendig. Namen, Kontaktdaten, Preise, Projektnamen, Kommentare, versteckte Spalten, frühere Versionen und Metadaten werden entfernt, soweit sie für den Zweck nicht benötigt werden. Für die Analyse wird eine eindeutig bezeichnete Arbeitskopie erstellt; das Original bleibt unverändert im vorgesehenen Ablagesystem.

Danach wird die Datei nach der Datenampel bewertet. Grüne Daten können im freigegebenen Dienst genutzt werden. Gelbe Daten benötigen Rechtsgrundlage, Geschäftskonto, Datenminimierung und gegebenenfalls Kunden- oder Vertragsprüfung. Rote Daten werden nur in einer speziell freigegebenen Umgebung verarbeitet. Nach der Analyse wird das Ergebnis mit der Ausgangsdatei verglichen, die

Arbeitskopie nach der vorgesehenen Frist gelöscht und die finale fachlich geprüfte Auswertung im normalen Dokumentensystem gespeichert.

4.3 Ein KI-Inhalt soll veröffentlicht werden

Vor der Veröffentlichung wird zwischen Text, Illustration, bearbeitetem Foto, künstlicher Stimme und realitätsnahem Video unterschieden. Danach werden Tatsachen, Quellen, Produktmerkmale, Preise, Rechte an Vorlagen, reale Personen, Marken und mögliche Täuschungswirkung geprüft. Ein künstliches Bild, das wie ein tatsächliches Referenzprojekt aussieht, benötigt eine deutlichere Einordnung als eine abstrakte Illustration. Ein erfundenes Testimonial darf nicht als echte Kundenstimme erscheinen.

Im nächsten Schritt wird die Transparenzfrage beantwortet. Bei Deepfakes, direkter KI-Interaktion und bestimmten öffentlichen Informationstexten können gesetzliche Pflichten greifen. Außerhalb dieser Fälle kann eine freiwillige Kennzeichnung sinnvoll sein. Die Formulierung muss den tatsächlichen Prozess beschreiben. Abschließend gibt eine benannte Person die finale Datei frei. Spätere inhaltliche Änderungen, der Austausch einer Stimme oder das Entfernen einer Kennzeichnung lösen eine neue Prüfung aus.

4.4 Eine KI soll eine menschliche Entscheidung unterstützen

Zuerst wird beschrieben, welche Entscheidung betroffen ist und welche Folgen sie für die Person haben kann. Eine sprachliche Vorbereitung ist anders zu behandeln als ein Punktwert, eine Rangliste oder eine Prognose. Bei Beschäftigung, Kreditwürdigkeit, Versicherung, Bildung oder anderen sensiblen Bereichen ist die Hochrisikoeinstufung zu prüfen. Zusätzlich gelten Datenschutz, Gleichbehandlung und mögliche Grenzen ausschließlich automatisierter Entscheidungen.

Die menschliche Kontrolle wird anschließend praktisch gestaltet. Die verantwortliche Person benötigt Ausgangsdaten, Fachkenntnis, Zeit und das Recht, den Vorschlag ohne Nachteil abzulehnen. Sie darf nicht nur die oberen Plätze einer Rangliste sehen oder hunderte Vorschläge innerhalb weniger Sekunden bestätigen müssen. Kriterien, Abweichungen und finale Entscheidung werden angemessen dokumentiert. Ist eine unabhängige menschliche Entscheidung nicht realistisch, darf das System nicht als bloße Assistenz bezeichnet werden.

4.5 Ein Anbieter kündigt ein großes Update an

Das Unternehmen vergleicht bisherige und neue Funktion. Zu prüfen sind Modell, Zweck, Datenzugriff, Speicherbedingungen, Unterauftragnehmer, automatische Aktionen, Protokolle und technische Kennzeichnungen. Besonders wichtig sind Funktionen, die standardmäßig aktiviert werden. Eine neue Transkription, Personalbewertung oder Agentenaktion wird nicht allein aufgrund der bisherigen Produktfreigabe verwendet.

Wesentliche Änderungen werden zunächst in einer Testumgebung mit geeigneten Daten geprüft. Der Anbieter muss erläutern, welche früheren Qualitäts- und Sicherheitsnachweise weiterhin gelten. Nutzer erhalten bei Bedarf eine Kurzunterweisung. Ist ein Rückwechsel nicht möglich oder ändern sich Daten- und Vertragsbedingungen wesentlich, werden Exit- und Anbieterwechseloptionen bewertet. Die neue Version erhält erst nach dokumentierter Entscheidung den produktiven Freigabestatus.

4.6 Ein KI-System soll beendet oder ersetzt werden

Die Beendigung beginnt mit einer Bestandsaufnahme: Benutzerkonten, Schnittstellen, gespeicherte Chats, Dateien, Protokolle, Vorlagen, Unterauftragnehmer und abhängige Arbeitsprozesse. Wichtige Geschäftsdokumente werden in geeigneten Formaten exportiert. Proprietäre Systemanweisungen oder Vorlagen werden soweit vertraglich erlaubt gesichert. Anschließend werden automatische Aktionen gestoppt und ein manueller oder alternativer Prozess aktiviert.

Konten und Zugriffe werden entzogen, Apps und Tokens deaktiviert und Löschung beim Anbieter veranlasst. Datenschutz- und Aufbewahrungspflichten sind dabei zu beachten: Nicht jede Information darf sofort gelöscht werden, aber Daten dürfen auch nicht ohne Zweck beim Anbieter verbleiben.

Richtlinie, Bestandsverzeichnis, Verfahrensdokumentation und Schulungsunterlagen werden aktualisiert. Nach Abschluss wird geprüft, ob Schnittstellen tatsächlich geschlossen und vereinbarte Löschungen bestätigt wurden.

4.7 Ein schwerwiegender Fehler wird erst nach externer Verwendung entdeckt

Das Unternehmen stoppt weitere Verwendung und stellt fest, welche Fassung an welche Empfänger gelangt ist. Falsche Rechts-, Preis-, Sicherheits- oder Produktangaben werden nicht stillschweigend in der internen Datei korrigiert; die externe Wirkung muss aktiv begrenzt werden. Je nach Fall werden Webseiten angepasst, Nachrichten zurückgerufen, Kunden informiert oder ein menschlicher Ansprechpartner bereitgestellt. Bei Personenbezug, Geheimnissen oder Sicherheitsinformationen werden zusätzlich Datenschutz- und Vorfallpflichten geprüft.

Die Untersuchung trennt unmittelbare Ursache und organisatorische Ursache. Der unmittelbare Fehler kann eine Halluzination sein; organisatorisch können fehlende Quellenprüfung, automatische Veröffentlichung oder unklare Verantwortung dahinterstehen. Korrekturmaßnahmen beziehen deshalb Prozess, Berechtigung und Schulung ein. Der Vorfall wird erst geschlossen, wenn externe Berichtigung, interne Ursache und Wirksamkeitskontrolle dokumentiert sind.

5. Musterklauseln für Dienstleister und Projektverträge

Die folgenden Klauseln sind Formulierungsbeispiele und keine allgemeingültigen Vertragsbedingungen. Sie müssen an Leistung, Rolle, Daten, Branche und bestehende Verträge angepasst werden. Bei sensiblen oder wirtschaftlich bedeutenden Aufträgen sollte eine rechtliche Prüfung erfolgen. Die Klauseln sind so aufgebaut, dass sie auch einzeln in Leistungsbeschreibungen oder Anlagen übernommen werden können.

5.1 Offenlegung des KI-Einsatzes

Der Auftragnehmer informiert den Auftraggeber vor Beginn der Leistung darüber, ob und für welche Arbeitsschritte KI-Systeme eingesetzt werden. Die Information bezeichnet mindestens die konkrete Funktion, den Anbieter oder die Art des Systems, die verarbeiteten Datenkategorien und den vorgesehenen Umfang menschlicher Prüfung. Änderungen des eingesetzten Systems oder eine Erweiterung des Verwendungszwecks sind vor ihrer produktiven Nutzung mitzuteilen, wenn sie Datenzugriff, Qualität, Rechte oder Risikoeinstufung wesentlich beeinflussen können.

5.2 Zulässige Systeme und Konten

Unternehmens-, Kunden- und Projektdaten des Auftraggebers dürfen nur in den ausdrücklich vereinbarten Systemen und geschäftlichen Konten verarbeitet werden. Private Benutzerkonten, nicht freigegebene Browsererweiterungen und kostenlose Dienste ohne dokumentierte Unternehmensfreigabe sind nicht zulässig. Der Auftragnehmer stellt sicher, dass eingesetzte Personen individuelle Konten, angemessene Zugriffssicherung und die für ihre Aufgabe erforderlichen Berechtigungen besitzen.

5.3 Datenminimierung und ausgeschlossene Informationen

Der Auftragnehmer verarbeitet nur die Informationen, die für den vereinbarten Arbeitsschritt tatsächlich erforderlich sind. Vollständige Originaldateien sind zu vermeiden, wenn ein bereinigter Ausschnitt, pseudonymisierte Angaben oder fiktive Testdaten ausreichen. Passwörter, Zugangsschlüssel, besondere Kategorien personenbezogener Daten, vollständige Personalakten und als streng vertraulich gekennzeichnete Geschäftsgeheimnisse dürfen nur nach vorheriger ausdrücklicher Freigabe und in einer dafür geeigneten technischen Umgebung verarbeitet werden.

5.4 Eigene Trainings- und Verbesserungszwecke

Der Auftragnehmer darf Eingaben, Dateien, Ausgaben und Projektdaten nicht für eigene Modelltrainings-, Produktverbesserungs-, Demonstrations- oder Referenzzwecke verwenden, sofern der Auftraggeber dem nicht zuvor ausdrücklich und dokumentiert zugestimmt hat. Eine technische

Speicherung zur vereinbarten Leistungserbringung, Sicherheit oder Fehleranalyse bleibt nur im vertraglich festgelegten Umfang zulässig. Der Auftragnehmer informiert über Aufbewahrungsfristen und Löschmöglichkeiten.

5.5 Menschliche Prüfung und fachliche Verantwortung

KI-generierte oder KI-bearbeitete Ergebnisse gelten zunächst als Entwürfe. Der Auftragnehmer prüft sie durch eine fachlich geeignete Person auf Richtigkeit, Vollständigkeit, Quellen, Zahlen, Rechte Dritter und Eignung für den vereinbarten Zweck. Eine rein formale Bestätigung oder automatische Weiterleitung genügt nicht. Bei rechtlich, steuerlich, technisch, personell oder sicherheitsbezogen bedeutsamen Inhalten wird die verantwortliche Fachperson benannt. Der Einsatz von KI vermindert die vereinbarte fachliche Sorgfalt nicht.

5.6 Entscheidungen über natürliche Personen

KI-Systeme dürfen ohne gesonderte schriftliche Freigabe nicht zur Bewertung, Filterung, Rangordnung, Profilbildung oder automatisierten Entscheidung über Bewerber, Beschäftigte, Kunden oder andere natürliche Personen eingesetzt werden. Biometrische oder emotionale Auswertungen sind ausgeschlossen, soweit keine ausdrücklich geprüfte und dokumentierte gesetzliche Grundlage vorliegt. Der Auftragnehmer informiert unverzüglich, wenn eine zunächst administrative Funktion tatsächlich eine erhebliche Entscheidungswirkung entfaltet.

5.7 Rechte an Vorlagen und Ergebnissen

Der Auftragnehmer verwendet nur Ausgangsmaterialien, für die die erforderlichen Nutzungs-, Bearbeitungs- und Übermittlungsrechte bestehen. Er prüft KI-Ergebnisse im angemessenen Umfang auf erkennbare Übernahmen fremder Werke, Marken, Designs, Personen oder lizenzpflichtiger Programmbestandteile. Die vertraglich eingeräumten Nutzungsrechte an Ergebnissen gelten nur im Rahmen der tatsächlich verfügbaren Rechte. Bekannte Unsicherheiten und verwendete fremde Lizenzbestandteile sind dem Auftraggeber vor Übergabe mitzuteilen.

5.8 Transparenz und Kennzeichnung

Soweit gesetzliche, vertragliche oder redaktionell festgelegte Transparenzpflichten bestehen, nimmt der Auftragnehmer die erforderliche Kennzeichnung vor oder stellt dem Auftraggeber die dafür notwendigen Informationen und technischen Herkunftsdaten bereit. Kennzeichnungen und maschinenlesbare Herkunftsinformationen dürfen nicht ohne Zustimmung entfernt werden. Realitätsnahe künstliche Darstellungen, künstliche Stimmen, Deepfakes und automatisiert erzeugte öffentliche Informationstexte sind besonders zu prüfen.

5.9 Unterauftragnehmer und Datenstandorte

Der Einsatz weiterer KI-, Hosting-, Transkriptions- oder Analysedienste als Unterauftragnehmer bedarf der vereinbarten Information oder Genehmigung. Der Auftragnehmer führt eine aktuelle Liste der beteiligten Unternehmen, Aufgaben und Verarbeitungsorte. Änderungen werden so früh mitgeteilt, dass der Auftraggeber Datenschutz-, Geheimhaltungs- und Sicherheitsfolgen prüfen und die vorgesehenen Rechte ausüben kann.

5.10 Sicherheits- und KI-Vorfälle

Der Auftragnehmer meldet dem Auftraggeber unverzüglich unzulässige Datenübermittlungen, kompromittierte Konten, wesentliche falsche oder diskriminierende Ausgaben, unerwartete Datenzugriffe, fehlerhafte automatische Aktionen, fehlende Kennzeichnungen und den Verdacht einer verbotenen oder nicht freigegebenen Anwendung. Die Meldung enthält Zeitpunkt, System, betroffene Daten, mögliche Empfänger, Sofortmaßnahmen und Ansprechpartner. Der Auftragnehmer unterstützt bei Untersuchung, Löschung, Berichtigung und erforderlichen Meldungen.

5.11 Änderungen, Updates und Prüfungsrechte

Wesentliche Modell-, Funktions-, Vertrags- oder Datenflussänderungen werden vor produktiver Nutzung mitgeteilt. Der Auftraggeber kann eine erneute Prüfung, einen begrenzten Test, zusätzliche Schutzmaßnahmen oder die Deaktivierung der betroffenen Funktion verlangen. Als Nachweise können technische Dokumentation, Sicherheitsberichte, Protokolle, Unterauftragnehmerlisten und angemessene unabhängige Prüfberichte vereinbart werden. Ein Vor-Ort-Audit ist nur erforderlich, wenn Risiko und Vertragsbedeutung dies rechtfertigen.

5.12 Vertragsende und Löschung

Nach Ende des Auftrags beendet der Auftragnehmer Zugriffe und Schnittstellen, gibt vereinbarte Daten und Ergebnisse in einem nutzbaren Format heraus und löscht verbleibende Projektinformationen nach den festgelegten Fristen, soweit keine gesetzlichen Aufbewahrungspflichten entgegenstehen. Die Löschung umfasst aktive Systeme, verbundene Apps und im angemessenen Umfang Sicherungskopien. Auf Wunsch wird die Durchführung bestätigt. Eigenständig gespeicherte Prompt- oder Vorlagenbestände des Auftraggebers dürfen nicht ohne Vereinbarung weiterverwendet werden.

Hinweis zur Vertragsgestaltung

Die Klauseln sollten mit Leistungsbeschreibung, Datenschutzvereinbarung, Vertraulichkeit, Haftung und Vergütung abgestimmt werden. Besonders wichtig ist eine realistische Risikoverteilung. Ein Dienstleister kann nicht garantieren, dass ein allgemeines KI-System niemals fehlerhaft arbeitet; er kann jedoch geeignete Systeme auswählen, Daten begrenzen, Ergebnisse fachlich prüfen, Änderungen mitteilen und Vorfälle zuverlässig unterstützen. Genau diese überprüfbaren Pflichten sollten im Vertrag im Vordergrund stehen.