



Malteser

...weil Nähe zählt.

Richtlinie zur künstlichen Intelligenz (KI-Richtlinie)

im Malteser Verbund

Inhalt

Impressum 4

Vorwort 5

I. Geltungsbereich 6

1. Persönlicher Anwendungsbereich 6

2. Sachlicher Anwendungsbereich und Verhältnis zur KI-Leitlinie 6

3. KI-Verordnung (KI-VO) 6

3.1 Anwendbarkeit KI-VO 7

3.2 KI-Systeme und KI-Modelle gem. KI-VO 7

4. Datenstrategie und Data Governance 8

II. Ethische und soziale Verantwortung, Kennzeichnung und Kommunikation 8

1. Ethik 8

2. Verzerrungen und Halluzinieren 9

3. Transparenz- und Kennzeichnungspflichten 10

4. KI in der Kommunikation 10

III. Verantwortlichkeiten, Prüfanfragen, Qualitätsmanagement 11

1. CIO für IT und IT-Compliance (Gesamt- und Ergebnisverantwortung) 11

2. Stabsstelle Daten & KI, KI-Prüfanfragen & Schulungen 11

2.1 Stabsstelle Daten und KI (Durchführungsverantwortung) 11

2.2 Prüfanfragen und Freigaben 12

2.3 Schulungen 13

3. Rolle der Malteser 13

3.1 Anbieter 13

3.2 Betreiber 13

4. Qualitätsmanagement, -kontrolle 14

IV. Risikoklassen und Pflichten 14

1. Menschliche Aufsicht (Art. 4, 14, 26 KI-VO) 14

2. Verbotene Praktiken (Art. 5 KI-VO) 15

2.1 Anwendungsbeispiele 15

2.2 Pflichten für Anbieter & Betreiber 16

3. Hochrisiko-KI (Art. 6-49 KI-VO) 16

Inhalt

3.1 Anwendungsbeispiele 16

3.2 Pflichten für Anbieter & Betreiber 17

4. Bestimmte KI-Systeme mit begrenztem Risiko & Transparenzrisiko (Art. 50 KI-VO) 18

4.1 Anwendungsbeispiele 18

4.2 Pflichten für Anbieter & Betreiber 18

5. KI-Systeme mit keinem oder geringem Risiko 19

5.1 Anwendungsbeispiele 19

5.2 Pflichten für Anbieter & Betreiber 19

6. KI-Systeme mit allgemeinem Verwendungszweck (GPAI) & systemischem Risik 19

(Artt. 51 ff. KI-VO)

6.1 Anwendungsbeispiel 20

6.2 Pflichten für Anbieter & Betreiber 20

V. Regulatorische Rahmenbedingungen 21

1. KI-Verordnung 21

2. Arbeitsrecht 21

3. KI und Interessenvertretungen (MAV & BR) 21

4. Schutz von Geschäftsgeheimnissen und Vertraulichkeit 21

5. Urheber- und Leistungsschutzrechte, gewerbliche Schutzrechte 22

5.1 Urheber und von KI erzeugte Arbeitsergebnisse 22

5.2 Schutzfähigkeit von Prompts 22

5.3 Schutzfähigkeit von durch KI erzeugten Arbeitsergebnissen 22

5.4 Können von KI geschaffene Arbeitsergebnisse gegen das 23

Urheberrecht verstoßen?

5.5 Trainieren von KI und Urheberrecht 23

6. Datenschutz 23

6.1 Keine Verarbeitung von personenbezogenen Daten ohne Rechtsgrundlage 24

6.2 Informationspflichten 24

6.3 Verbot automatisierter Entscheidungsfindung 25

6.4 Verarbeitung von Daten im Auftrag 25

6.5 Betroffenenrechte 26

Inhalt

7. Haftung und KI 26

7.1 Trainieren von KI-Modellen (Text und Data Mining) 26

7.2 Haftung und Arbeitsrecht 26

VI. Technikgestaltung 27

1. Zugriffssteuerung und Authentifizierung 27

2. Sicherheit bei der Entwicklung von KI-Systemen 27

3. Sicherung der Transparenz und Nachvollziehbarkeit 27

4. Technische Voraussetzungen für KI-Systeme 28

VII. KI-Kompetenz und Bewusstsein für KI 29

VIII. Inkrafttreten 29

Mitgeltende Unterlagen 29

IX. Anhang (Schnellübersicht & Risiko-Navigator) 30

1. Schnellübersicht Risikoklassen und Transparenzpflichten 30

2. TÜV-Risiko-Navigator 30

Impressum

Malteser Deutschland gemeinnützige GmbH, Vorsitzender der Geschäftsführung
und
Malteser Hilfsdienst e.V., Vorsitzender des Geschäftsführenden Vorstands:
Dr. Elmar Pankau

Erna-Scheffler-Straße 2, 51103 Köln
0221 9822-0
malteser@malteser.org
www.malteser.de
Stand: 11/2025

Vorwort¹

Als das Thema vor einigen Jahren aufkam, wurde die praktische Unverzichtbarkeit von Künstlicher Intelligenz (KI) auch im Malteser Verbund rasch erkannt. Wir sahen die Potenziale, die KI unseren Mitarbeitenden bietet, ihre Arbeitsprozesse effizienter und unsere Dienstleistungen qualitativ noch hochwertiger zu gestalten. Mit unserer Offenheit für KI-Lösungen beweisen wir seitdem unsere Innovationsfähigkeit und tragen zum Erhalt und zur Steigerung unserer Zukunfts- und Wettbewerbsfähigkeit bei.

Die Frage nach dem „Ja oder nein?“ war somit schnell geklärt – schwieriger war die nach dem „Wie?“. Klar war von Beginn an: Der Einsatz von KI muss (!) im Einklang mit unseren christlichen Werten und ethischen Standards erfolgen. Darüber hinaus blieben aber viele Fragen offen, gab es Unsicherheiten und auch Risiken. Erste praktische Hilfestellungen lieferte die im Frühjahr 2024 veröffentlichte KI-Leitlinie. Sie enthält Leitsätze für den ethischen, fairen und verantwortungsvollen Einsatz von KI im Malteser Verbund.

Seitdem hat das KI-Thema weiter an Fahrt aufgenommen. Mit der KI-Verordnung der EU ist im August 2024 zudem ein erstes umfassendes Gesetz zur Regulierung von KI in Kraft getreten. Auch wir sind nicht untätig gewesen. Mit der vorliegenden KI-Richtlinie veröffentlichen wir nun erstmals ein verbindliches Regelwerk, das den Einsatz von KI-Anwendungen durch Mitarbeitende und Ehrenamtliche regelt und klare Vorgaben zu Nutzung, Sicherheitsabwägungen und Ethik bereitstellt.

Die Richtlinie gilt für alle in unserer Organisation, die mit KI arbeiten bzw. diese entwickeln oder verantworten. Sie soll uns helfen, die Chancen dieser Technologie zu nutzen und die Entwicklung innovativer Lösungen fördern. Durch den Einsatz von KI können wir Malteser effizienter werden und Zeit und Ressourcen zu sparen – die dann unserer Kernaufgabe zugutekommen können: der direkten Arbeit mit den Menschen.



Stabstelle Daten & KI als „One-Stop-Shop“:
Die Entwicklung, Einführung oder Verwendung von KI-Modellen und Systemen im Malteser Verbund ist immer vorab mit der Stabstelle Daten & KI abzustimmen (siehe Kapitel III).

Malteser Hilfsdienst e. V. und Malteser Deutschland gGmbH


Thomas Berding
Chief Information Officer

¹ Dieses Vorwort ist mit Unterstützung von KI (Microsoft Copilot) entstanden.

I. Geltungsbereich

1. Persönlicher Anwendungsbereich

Diese Richtlinie gilt für alle haupt- und ehrenamtlichen Mitarbeitenden (nachfolgend „MA“) in allen Malteser Organisations- und Unternehmensteilen.

2. Sachlicher Anwendungsbereich und Verhältnis zur KI-Leitlinie

Die Richtlinie bildet die Grundlage für die dienstlich veranlasste Entwicklung und Nutzung bzw. den Betrieb von KI-Modellen oder KI-Systemen im Malteser Verbund.

Im Umkehrschluss findet die Richtlinie keine Anwendung auf Systeme, die nicht unter die KI-Verordnung fallen oder rein privat genutzt werden. Zu nennen sind hier etwa sogenannte Expertensysteme, die nicht für den autonomen Betrieb ausgelegt sind oder KI, die im Rahmen einer ausschließlich persönlichen und nicht beruflichen Tätigkeit verwendet wird.

Unsere KI-Leitlinie dient als Kompass für unser Handeln. Die KI-Richtlinie gewährleistet einen verantwortungsbewussten Umgang mit KI, sie hat Anwendungsvorrang vor der KI-Leitlinie.

3. KI-Verordnung (KI-VO)

Die KI-Verordnung (kurz „KI-VO“), ist das weltweit erste Gesetz, das sich umfassend der Regulierung von Künstlichen-Intelligenz-Systemen und Künstlichen-Intelligenz-Modellen widmet. Zweck der KI-VO ist es, das Funktionieren des Binnenmarktes zu verbessern und die Einführung einer auf den Menschen ausgerichteten und vertrauenswürdigen künstlichen Intelligenz (KI) zu fördern sowie gleichzeitig ein hohes Schutzniveau in Bezug auf Gesundheit, Sicherheit und die in der Charta der Grundrechte verankerten Grundrechte, einschließlich Demokratie, Rechtsstaatlichkeit und Umweltschutz zu gewährleisten, vor schädlichen Auswirkungen von Systemen der künstlichen Intelligenz (KI-Systeme) in der Union zu schützen und die Innovation zu unterstützen.

Herzstück der KI-VO sind die unter Kapitel IV erläuterten Risikoklassen und dargestellten Pflichten für Anbieter und Betreiber von KI-Systemen.

I. Geltungsbereich

3.1 Anwendbarkeit KI-VO

Die Regelungen der KI-VO finden stufenweise Anwendung:



3.2 KI-Systeme und KI-Modelle gem. KI-VO

Künstliche Intelligenz bezeichnet Computersysteme, die Aufgaben übernehmen können, die normalerweise menschliches Denken erfordern, wie das Erkennen von Mustern, das Lernen aus Daten oder das Lösen von Problemen. Diese Systeme sind für einen in unterschiedlichem Grad autonomen Betrieb ausgelegt und können nach ihrer Betriebsaufnahme anpassungsfähig sein.

Ein KI-System hilft dabei, große Mengen an Informationen schnell zu analysieren und daraus nützliche Erkenntnisse zu gewinnen. Es kann beispielsweise Prozesse optimieren, den Bedarf von Menschen besser verstehen oder Routineaufgaben automatisieren, um mehr Zeit für wichtige Tätigkeiten zu schaffen. KI-Systeme arbeiten oft im Hintergrund, etwa bei der Analyse von Texten, Bildern oder Stimmen, und unterstützen Menschen in ihrer Arbeit, ohne sie zu ersetzen. Wichtig ist, KI-Systeme verantwortungsvoll einzusetzen, ethische Aspekte zu berücksichtigen und den Schutz von Daten zu gewährleisten. KI-Systeme leiten aus den erhaltenen Eingaben explizite oder implizite Ziele ab und erstellen Ausgaben wie Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen, die physische oder virtuelle Umgebungen beeinflussen können.

KI-Modelle werden in der KI-VO nicht explizit definiert, sondern meist im Kontext eines KI-Systems erwähnt. Als Beispiel wäre hier das GPT Turbo 4.0 zu nennen, welches als KI-Modell innerhalb des KI-Systems ChatGPT eingesetzt wird.

I. Geltungsbereich

4. Datenstrategie und Data Governance

Die Datenstrategie ist ein detaillierter Plan zur Nutzung von Daten zur Verbesserung der Entscheidungsfindung, Optimierung von Geschäftsprozessen und Erreichung der Unternehmensziele im Malteser Verbund.

Data Governance² definiert funktions- und bereichsübergreifend Entscheidungsrechte, Verantwortlichkeiten, Richtlinien und andere Vorgaben für das qualitätsorientierte Management der Ressource Daten gem. den Zielen der Malteser.



Bei der Einführung von KI sind, bis zur Verabschiedung einer KI-Strategie für den Verbund, Datenstrategie und Data Governance zwingend zu beachten.

II. Ethische und soziale Verantwortung, Kennzeichnung und Kommunikation

1. Ethik

KI kann menschliche Fähigkeiten in vielen Lebensbereichen erweitern oder sogar ersetzen. Softwaresysteme verfügen „weder über theoretische noch über praktische Vernunft. Sie handeln oder entscheiden nicht selbst und können keine Verantwortung übernehmen.“³ Der Einsatz von KI kann jedoch großen Einfluss auf die Handlungsfähigkeit, auf Entscheidungen und die Lebensrealität von Menschen haben. Die KI-Technologien und deren Auswirkungen sind nicht per se gut oder schlecht. Für eine ethische Bewertung muss betrachtet werden, zu welchem Zweck die KI entwickelt, wie und in welcher Umgebung sie eingesetzt wird sowie welche Folgen jetzt und auch in der Zukunft zu erwarten sind.

Die ethische Bewertung von KI-Technologien ist demnach komplex und bedarf immer eines kontextspezifischen Blickes: Das bedeutet, es müssen nicht nur die Technologien selbst betrachtet werden, sondern vor allem auch die Bereiche, in denen diese zum Einsatz kommen sollen (bspw. Medizin, Pflege, Hilfs- und Unterstützungsdienste, Verwaltung) sowie die Personen (-gruppen), die diese anwenden und/oder von deren Anwendung betroffen sind (Mitarbeitende in Haupt- und Ehrenamt sowie bspw. Klientinnen und Klienten). Nicht nur aktuelle, sondern

² Klingenberg/Weber: Data Governance, 2. Auflage 2025, ISBN 978-3-446-48102-2

³ Vgl. Deutscher Ethikrat: Stellungnahme Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz, Flyer vom 11.12.2024, S. 2, file:///U:/Downloads/flyer-mensch-und-maschine%20(2).pdf (Zugriff am 02.09.2025)

II. Ethische und soziale Verantwortung, Kennzeichnung und Kommunikation

auch langfristige Auswirkungen der jeweiligen KI-Technologie sollten regelmäßig analysiert und evaluiert werden.

Die Ausrichtung auf den Dienst am Menschen und die Achtung der Menschenwürde, die nicht zweckrationalisiert oder ökonomischem Nutzenkalkül unterstellt sein dürfen⁴, sind aus christlicher Perspektive leitend bei der Entwicklung und beim Einsatz von KI. Um für den Malteser Verbund sicherzustellen, dass die Entwicklung und der Einsatz von KI-Technologien menschliche Handlungsfähigkeit⁵ erweitern, Qualitätsstandards mindestens halten bzw. verbessern⁶ und somit dem Wohl der Menschen dienen, muss ein Handlungsrahmen festgelegt werden, der betroffene und maßgebende Werte und Prinzipien benennt und ihre Beachtung sicherstellt.



Im Entscheidungsprozess, ob der Einsatz oder die Entwicklung einer KI-Technologie im Malteser Verbund ethisch begründet werden kann und vertretbar ist, gilt es betroffene Werte und Perspektiven zu berücksichtigen, deren Auflistung noch als mitgeltende Unterlage zur KI-Richtlinie veröffentlicht werden wird.

2. Verzerrungen und Halluzinieren

Desinformation und Verzerrungen (Bias) resp. das Generieren unerwünschter Ausgaben eines KI-Systems wird durch uns Malteser nicht unterstützt. Auf der ersten Stufe können wir das „Halluzinieren“ der KI nicht verhindern: Wenn dem KI-System eine Frage gestellt wird, für deren Antwort es keine hinreichende Datenbasis gibt, wird i. d. R. systemseitig eine sinnvoll klingende Antwort formuliert. Insoweit ist der Nutzende, unter Verweis auf Kapitel III zur Qualitätskontrolle verpflichtet.

Darüber hinaus ist das sogenannten „Direct Prompt Injection“ untersagt. Hier experimentiert der Nutzende so lange an einer Formulierung (einem Prompt), bis er das KI-System zu einer nicht erwünschten Verhaltensänderung „überredet“ bekommt.

⁵ Vgl. Publizistische Kommission der Deutschen Bischofskonferenz: Medienbildung und Teilhabegerechtigkeit. Impulse der Publizistischen Kommission der Deutschen Bischofskonferenz zu den Herausforderungen der Digitalisierung, Bonn, 2016, S. 8, www.dbk-shop.de/media/files_public/c81524222871b74797c4dd7567a25fed/DBK_5288.pdf (Zugriff am 02.09.2025)

⁶ Vgl. Deutscher Ethikrat: Stellungnahme Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz, Flyer vom 11.12.2024, S. 2, file:///U:/Downloads/flyer-mensch-und-maschine%20(2).pdf (Zugriff am 02.09.2025)

⁷ Vgl. Zentralen Kommission zur Wahrung ethischer Grundsätze in der Medizin und ihren Grenzgebieten (Zentrale Ethikkommission) bei der Bundesärztekammer: Stellungnahme „Entscheidungsunterstützung ärztlicher Tätigkeit durch Künstliche Intelligenz“, Berlin, 2021 S. 4, www.zentrale-ethikkommission.de/fileadmin/user_upload/_old-files/downloads/pdf-Ordner/Zeko/Zeko_SN_CDSS_Online_final.pdf (Zugriff am 02.09.2025)

II. Ethische und soziale Verantwortung, Kennzeichnung und Kommunikation

3. Transparenz- und Kennzeichnungspflichten

Als Anbieter und Betreiber von KI-Systemen, die

- direkt mit Menschen interagieren,
- zur Emotionserkennungssystem oder als System zur biometrischen Kategorisierung verwendet werden oder
- synthetische Audio-, Bild-, Video- oder Textinhalte erzeugen,

ist sicherzustellen, dass die betreffende natürlichen Person grundsätzlich informiert wird, dass sie mit einem KI-System interagiert bzw. müssen übernommene Ergebnisse – vorbehaltlich der Ausnahmen des Art. 50 KI-VO – eines generativen KI-Systems als künstlich erzeugt oder manipuliert erkennbar sein.

Sofern ein KI-System personenbezogene Daten verarbeitet, sind zudem die davon betroffenen natürlichen Personen über den Betrieb des Systems und das Verarbeiten personenbezogener Daten vorab zu informieren (siehe Kapitel V, Datenschutz).



Die Informationen müssen den betreffenden natürlichen Personen spätestens zum Zeitpunkt der ersten Interaktion oder Aussetzung in klarer und eindeutiger Weise bereitgestellt werden und den geltenden Barrierefreiheitsanforderungen entsprechen.

Die aktuell in Art. 50 KI-VO formulierten Ausnahmen von der Transparenz- und Kennzeichnungspflicht sind dem Anhang IX zu entnehmen.

4. KI in der Kommunikation

Für die Verwendung von KI-generierten Inhalte im Rahmen der Unternehmenskommunikation oder öffentlichen Darstellung der Malteser werden aktuell Vorgaben definiert, die, nach Finalisierung ebenfalls in die DokBox der Malteser eingestellt werden; bis zur Veröffentlichung gelten die nachfolgenden Regelungen:

- Kennzeichnungspflicht: Der KI-Einsatz in der externen wie internen Kommunikation der Malteser muss generell gewährleisten, dass wesentlich mit KI generierte Inhalte als solche für die Adressaten klar erkennbar sind.
- Keine Deepfakes: KI-generierte Inhalte dürfen reale Ereignisse oder Personen weder imitieren noch verfälschen noch solche erfinden oder als real darstellen.
- Vier-Augen-Prinzip: Eine zu dokumentierende Überprüfung und Freigabe KI-generierter Inhalte durch eine zweite Person vor Veröffentlichung gewährleistet deren Richtigkeit und Regelkonformität.

II. Ethische und soziale Verantwortung, Kennzeichnung und Kommunikation

- KI-Tools: Zur KI-gestützten Erstellung von Inhalten dürfen ausschließlich von der Stabsstelle Daten & KI freigegebene Tools verwendet werden.
- Extern generierte KI-Inhalte können nur verwendet werden, sofern eine rechtssichere Lizenzierung zur kommerziellen Nutzung vorliegt, die Einhaltung des Urheberrechts gewährleistet ist und sie dieser Richtlinie entsprechen.
- Bereits bestehende Regeln für die Malteser Kommunikationsarbeit wie etwa in den Bereichen Presse, Publikationen, Social Media oder Fundraising gelten ausnahmslos auch für KI-generierte Inhalte.
- Detaillierte Bestimmungen und Orientierungshilfen enthalten die spezifischen „Richtlinien für die KI-Nutzung in der Malteser Kommunikation“.

III. Verantwortlichkeiten, Prüfanfragen, Qualitätsmanagement

1. CIO für IT und IT-Compliance (Gesamt- und Ergebnisverantwortung)

Im Malteser Verbund ist gem. ITK-Richtlinie der CIO (Chief Information Officer) gesamtverantwortlich für die IT- und Telekommunikationsstrukturen. Sämtliche Organisationseinheiten im Malteser Verbund sind damit, auch hinsichtlich der Entwicklung oder Nutzung von KI-Systemen, fachlich weisungsgebunden.

Alle Mitarbeitenden, die KI-gestützte Systeme, Modelle und Technologien nutzen, sind verpflichtet, diese gemäß der geltenden IT-Compliance-Regelungen (siehe auch „Mitgeltende Unterlagen“) zu verwenden.

2. Stabsstelle Daten & KI, KI-Prüfanfragen & Schulungen

2.1 Stabsstelle Daten und KI (Durchführungsverantwortung)

Der Malteser Verbund hat eine zentrale Stabsstelle „Daten & KI“ eingerichtet. Sie ist dem Chief Information Officer (CIO) für den Malteser Verbund zugeordnet.

Die Stabsstelle übernimmt die vollständige Verantwortung für unsere KI-Initiativen, von der strategischen Ausrichtung über die Umsetzung bis hin zum fachlichen Betrieb. Sie arbeitet dabei interdisziplinär mit den erforderlichen Fach- und Geschäftsbereichen, insbesondere IT-Compliance & Datenschutz, zusammen. Die Stabsstelle übernimmt i. S. e. One-Stop-Shops

III. Verantwortlichkeiten, Prüfanfragen, Qualitätsmanagement

die Prüfung und Bewertung von KI-Systemen und -Modellen, ähnlich wie bei herkömmlicher Schatten-IT. Zudem werden jegliche KI-Vorhaben im Verbund aktiv von der Stabsstelle regulatorisch, beratend oder in direkter Umsetzung begleitet. Durch die zentrale Erfassung und Verwaltung aller im Verbund eingesetzten KI-Anwendungen wird sichergestellt, dass keine doppelten oder unnötigen Systeme existieren. Dies optimiert die Ressourcennutzung und verhindert ineffiziente Arbeitsabläufe. Die Stabsstelle stellt sicher, dass alle KI-Anwendungen den internen Richtlinien und Vorgaben sowie ethischen Standards entsprechen. Durch die Bündelung interner und externer Expertise sorgt sie für die Qualität und Sicherheit von KI-Initiativen und fördert eine verantwortungsbewusste und effektive Nutzung von KI-Technologien.

2.2 Prüfanfragen und Freigaben

Erlaubt sind nur KI-Systeme und -Modelle, die ausdrücklich von der Stabsstelle Daten & KI freigegeben wurden. Ebenso ist das Trainieren von KI-Modellen vorab mit der Stabsstelle abzustimmen.

Zentrale Anlaufstelle für alle (Prüf-)Anfragen rund um KI:



Daten & KI



ki@malteser.org

Für jedes KI-Vorhaben ist vorab eine Risikobewertung durch die Stabsstelle, auf Basis einzureichender KI-Prüfbögen, zwingend notwendig, beispielsweise:

- Aussagen zur Manipulationsanfälligkeit
- die Analyse möglicher Schwachstellen
- die Festlegung von Maßnahmen zur Minimierung dieser Risiken
- angemessene Sicherheitsmaßnahmen und Verschlüsselungsstandards



Einen Stand über bereits freigegebene KI-Systeme finden Sie in der Schatten-IT-Übersicht (bitte nach „KI“ filtern):
[https://maltesercloud.sharepoint.com/sites/ITCompliance/SitePages/%C3%9Cbersicht-Schatten-IT\(2\).aspx](https://maltesercloud.sharepoint.com/sites/ITCompliance/SitePages/%C3%9Cbersicht-Schatten-IT(2).aspx)
Sofern nichts anderes geregelt wurde, dürfen keine personenbezogenen Daten in diesen KI-Systemen verarbeitet werden. Gleiches gilt für Daten, die dem besonderen Vertrauensschutz unterliegen oder für die Malteser von schützenswertem Interesse sind (siehe Kapitel V).

III. Verantwortlichkeiten, Prüfanfragen, Qualitätsmanagement

2.3 Schulungen

Die Stabsstelle Daten & KI steuert die Verpflichtung aus Art. 4 KI-VO. Dieser verpflichtet Anbieter und Betreiber von KI-Systemen zu Maßnahmen, um nach besten Kräften sicherzustellen, dass Mitarbeitende über ausreichend KI-Kompetenz verfügen. Dabei sind technische Kenntnisse, Erfahrung, Ausbildung und Schulung der Mitarbeitenden zu berücksichtigen. Mitarbeitende sollen in der Lage sein, fundierte Entscheidungen über KI-Systeme zu treffen (siehe weiterführend Kapitel VII).

3. Rolle der Malteser

Sobald ein KI-System entwickelt oder eingesetzt wird, müssen die Mitarbeitenden ihre Rolle und die des Arbeit- oder Dienstgebers verstehen und unterscheiden können. Wesentlich für uns Malteser sind dabei zwei Rollen:

3.1 Anbieter

„Anbieter“ i. S. d. KI-VO ist 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, egal ob entgeltlich oder unentgeltlich. Anbieter sind demnach Organisationen, die ein KI-System entwickeln. Die Anbieter bestimmen die Zwecke, für die ihre KI-Systeme eingesetzt werden sollen und halten dies in einer Gebrauchsanleitung fest. Die Anbieter treffen zudem die meisten Pflichten aus der KI-VO.

3.2 Betreiber

„Betreiber“ i. S. d. KI-VO ist eine natürliche oder juristische Person, Behörde, Einrichtung oder sonstige Stelle, die ein KI-System in eigener Verantwortung verwendet. Betreiber sind folglich die Nutzer von KI-Systemen. Sie haben weder Einfluss auf die Programmierung noch auf die Daten, die zum Training des KI-Systems verwendet wurden. Betreiber setzen ein KI-System, analog einer Software, so ein, wie es ist und müssen auch genau dafür Sorge tragen: Sie müssen sicherstellen, dass die tatsächliche Nutzung gem. der Gebrauchsanleitung des Anbieters erfolgt.



Vorsicht:

Sobald der Betreiber etwa die Zweckbestimmung des KI-Systems ändert oder es mit eigenen Daten (weiter-)trainiert, wird er gem. KI-VO zum Anbieter.

III. Verantwortlichkeiten, Prüfanfragen, Qualitätsmanagement

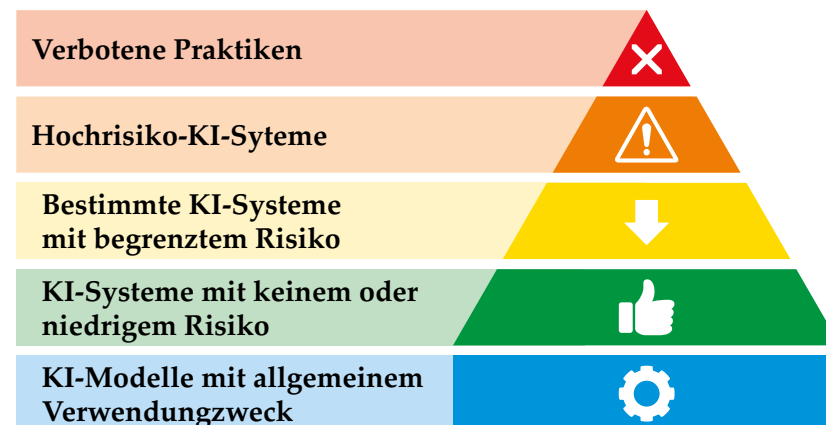
4. Qualitätsmanagement, -kontrolle

Arbeitsergebnisse der KI dürfen nicht ungeprüft übernommen werden. Für Inhalt und Qualität der KI-generierten Inhalte bzw. Ergebnisse ist der oder die Nutzende verantwortlich. Insbesondere zu überprüfen sind:

- Datenqualität und Data Governance (vgl. Kapitel I)
- Qualität und inhaltliche Korrektheit der Ergebnisse (siehe Kapitel III)
- Urheber- und Leistungsschutzrechte an erstellten Inhalten (vgl. Kapitel V)

IV. Risikoklassen und Pflichten

Die KI-VO möchte die Einführung einer auf den Menschen ausgerichteten und vertrauenswürdigen künstlichen Intelligenz (KI) fördern und gleichzeitig ein hohes Schutzniveau in Bezug auf Gesundheit, Sicherheit und die in der Charta verankerten Grundrechte gewährleisten. Dementsprechend verfolgt die KI-VO einen risikobasierten Ansatz, indem KI-Systeme in Risikoklassen eingeteilt und Risiken nach ihrer Eintrittswahrscheinlichkeit sowie den möglichen Folgen bewertet werden.



KI-Systeme mit hohen Risiken sind strenger reguliert und erfordern mehr menschliche Aufsicht als Systeme mit geringeren Risiken. KI-Systeme, die ein Risiko für Grundrechte, Gesundheit oder Sicherheit darstellen, werden als Hochrisiko-KI über die KI-VO besonders streng reguliert.

1. Menschliche Aufsicht (Art. 4, 14, 26 KI-VO)

Sowohl als Anbieter als auch als Betreiber von KI-Systemen ergreifen wir Malteser Maßnahmen, um nach besten Kräften sicherzustellen, dass unsere Mitarbeitenden, die in unserem Auftrag mit dem Betrieb und der Nutzung von KI-Systemen befasst sind, über ein ausreichendes Maß an KI-Kompetenz verfügen.

IV. Risikoklassen und Pflichten

In der Rolle als Betreiber, insbesondere im Falle von Hochrisiko-KI, umfasst das Konzept der menschlichen Aufsicht die folgenden Kern-Elemente:

- Die explizite Benennung von Mitarbeitenden, die eine Verwendung von KI-Systemen Dritter gem. der Betriebsanleitung des Anbieters sicherstellen.
- Angemessenheit der Aufsichtsmaßnahmen im Hinblick auf die Risiken, den Grad der Autonomie und den Kontext der Nutzung des Hochrisiko-KI-Systems.

Die Malteser stellen sicher, dass die Mitarbeitenden, denen die menschliche Aufsicht über ein Hochrisiko-KI-System übertragen wurde, über folgende Kompetenzen bzw. Fähigkeiten verfügen:

- Verständnis über das Hochrisiko-KI-System
- Bewusstsein für das Automatisierungsbias
- Richtige Interpretation der Ausgaben des Hochrisiko-KI-Systems
- Nichtverwendung des Hochrisiko-KI-Systems oder einzelner seiner Ausgaben in bestimmten Situationen
- Eingriff in den Betrieb oder Stopp des Betriebs des Hochrisiko-KI-Systems („Stopptaste“)

2. Verbotene Praktiken (Art. 5 KI-VO)

Verbotene Praktiken



Immer dann, wenn ein KI-System gegen die Sicherheit, die Grundrechte oder die Freiheit von Menschen verstößt, ist dessen Einsatz verboten.

Insbesondere manipulative, ausbeuterische und soziale Kontrollpraktiken stellen ein unannehmbares Risiko i. S. d. KI-VO dar. Sie verbietet in diesem Zusammenhang in Art. 5 KI-VO ausdrücklich bestimmte, mit einem unannehmbaren Risiko verbundene Praktiken im Zusammenhang mit KI-Systemen.

2.1 Anwendungsbeispiele

Im Einklang mit der KI-VO sind Entwicklung oder Nutzung eines KI-Systems im Malteser Verbund nicht erlaubt ...

- ... wenn sie die unterschwellige Beeinflussung von Personen zum Ziel haben,
- ... zur Ausnutzung von Schwäche oder Schutzbedürftigkeit von (vulnerablen) Personen, etwa aufgrund des Alters oder einer Behinderung,
- ... zur Einstufung von menschlichem Verhalten auf Grundlage ihres sozialen Handelns oder ihrer persönlichen Eigenschaften und Vornahme einer sozialen Bewertung (Social Scoring), die in keinem Zusammenhang mit den Umständen steht, unter denen sie erfasst wurde oder die im Hinblick auf das soziale Verhalten ungerechtfertigt oder unverhältnismäßig ist,

IV. Risikoklassen und Pflichten

- ... zur Emotionserkennung am Arbeitsplatz (ausgenommen medizinische oder sicherheitsrelevante Zwecke),
 - ... für eine Kategorisierung von Personen aufgrund biometrischer Daten.
- Anwendungsbeispiel:

Die Nutzung eines KI-Systems, um die Qualität des Kantinenessens für unliebsame Mitarbeitende zu verringern, ist nicht erlaubt.



Anwendungsbeispiel:
Die Nutzung eines KI-Systems, um die Qualität des Kantinenessens für unliebsame Mitarbeitende zu verringern, ist nicht erlaubt.

2.2 Pflichten für Anbieter & Betreiber

Verbotene Praktiken lösen keine Pflichten für Anbieter und Betreiber aus, da die Nutzung untersagt ist. Diese Verbot gilt im Malteser Verbund uneingeschränkt.

3. Hochrisiko-KI (Art. 6-49 KI-VO)

Hochrisiko-KI-Syteme



Die Regulierung von Hochrisiko-KI bildet den Schwerpunkt in der KI-VO. Die KI-VO identifiziert eine

Reihe von Anwendungsfällen, bei denen KI-Systeme als hochriskant eingestuft werden, da sie potenziell negative Auswirkungen auf Gesundheit, Sicherheit und Grundrechte haben können.

3.1 Anwendungsbeispiele

Konkret benennt der Gesetzgeber in Anhang III zu Art. 6 KI-VO die KI-Systeme als Hochrisiko-KI, ...

- ... die bestimmungsgemäß für die Einstellung oder Auswahl natürlicher Personen verwendet werden sollen, insbesondere um gezielte Stellenanzeigen zu schalten, Bewerbungen zu sichten oder zu filtern und Bewerberinnen und Bewerber zu bewerten;
- ... die bestimmungsgemäß für Entscheidungen, die die Bedingungen von Arbeitsverhältnissen, Beförderungen und Kündigungen von Arbeitsvertragsverhältnissen beeinflussen, für die Zuweisung von Aufgaben aufgrund des individuellen Verhaltens oder persönlicher Merkmale oder Eigenschaften oder zur Beobachtung und Bewertung der Leistung und des Verhaltens von Personen in solchen Beschäftigungsverhältnissen verwendet werden sollen.

IV. Risikoklassen und Pflichten

Abweichend gilt ein in Anhang III der KI-VO genanntes KI-System nicht als hochriskant, wenn es kein erhebliches Risiko der Beeinträchtigung in Bezug auf die Gesundheit, Sicherheit oder Grundrechte natürlicher Personen birgt, indem es unter anderem nicht das Ergebnis der Entscheidungsfindung wesentlich beeinflusst.



Faustformel:
Wenn ein KI-System einen Menschen in seinen Grundrechten, seiner Gesundheit oder seiner Freiheit betreffen kann, also bspw. bei der Bewertung in Beruf, Schule oder Öffentlichkeit, ist Vorsicht geboten. Es könnte sich um Hochrisiko-KI handeln. Eine Aufzählung von Hochrisiko-KI-Systemen gemäß Art. 6 KI-VO findet sich in Anhang I und Anhang III KI-VO.

3.2 Pflichten für Anbieter & Betreiber

Die meisten Pflichten treffen Anbieter von Hochrisiko-KI-Systemen, da derartige Systeme die größte Auswirkung auf die Sicherheit, Grundrechte oder Gesundheit von Menschen hat.

Pflichten für Anbieter (Artt. 16 ff. KI-VO)	Pflichten für Betreiber (Art. 26 KI-VO)
Die Einhaltung der Anforderungen sind in Abschnitt 3 der KI-VO festgehalten, u.a.: • Einrichtung und Aufrechterhaltung eines Risikomanagementsystems • Vor- und Einhalten von Data Governance- und Datenverwaltungsverfahren, die für die Zweckbestimmung des Hochrisiko-KI-Systems geeignet sind • Aufbewahren von automatisch erzeugten Protokollen • Durchführung eines Konformitätsbewertungsverfahrens	Die Betreiber von Hochrisiko-KI-Systemen treffen geeignete technische und organisatorische Maßnahmen, um sicherzustellen, dass solche Systeme entsprechend der den Systemen beigefügten Betriebsanleitungen verwendet werden. Die Betreiber übertragen natürlichen Personen, die über die erforderliche Kompetenz, Ausbildung und Befugnis verfügen, die menschliche Aufsicht und lassen ihnen die erforderliche Unterstützung zukommen.

IV. Risikoklassen und Pflichten

4. Bestimmte KI-Systeme mit begrenztem Risiko & Transparenzrisiko (Art. 50 KI-VO)

4.1 Anwendungsbeispiele

Bestimmte KI-Systeme mit begrenztem Risiko

Hierunter fallen gängige KI-Systeme oder Arbeitsergebnisse, die direkt mit dem Menschen interagieren, wie ...

- ... generative Bild-, Video- und Textgeneratoren,
- ... Deepfakes,
- ... Chatbots.

4.2 Pflichten für Anbieter & Betreiber

Nähere Ausführungen zu den Kennzeichnungs- und Informationspflichten können dem Kapitel II entnommen werden.

Pflichten für Anbieter (Art. 50 Abs. 1, 2 KI-VO)	Pflichten für Betreiber (Art. 50 Abs. 3, 4 KI-VO)
Anbieter müssen dafür Sorge tragen, dass Ergebnisse/Ausgaben als künstlich erzeugt oder manipuliert erkennbar sind.	Betreiber eines Emotionserkennungssystems oder eines Systems zur biometrischen Kategorisierung informieren die davon betroffenen natürlichen Personen über den Betrieb des Systems und verarbeiten personenbezogene Daten.
Darüber hinaus stellen sie sicher, dass KI-Systeme, die für die direkte Interaktion mit natürlichen Personen bestimmt sind, so konzipiert und entwickelt werden, dass die betreffenden natürlichen Personen informiert werden, dass sie mit einem KI-System interagieren , es sei denn, dies ist aus Sicht einer angemessen informierten, aufmerksamen und verständigen natürlichen Person aufgrund der Umstände und des Kontexts der Nutzung offensichtlich.	Betreiber eines KI-Systems, das Bild-, Ton- oder Videoinhalte erzeugt oder manipuliert, die ein Deepfake sind, müssen offenlegen, dass die Inhalte künstlich erzeugt oder manipuliert wurden. Ist der Inhalt Teil eines offensichtlich künstlerischen, kreativen, satirischen, fiktionalen oder analogen Werks oder Programms, beschränken sich die Transparenzpflichten darauf, das Vorhandensein solcher erzeugten oder manipulierten Inhalte in geeigneter Weise offenzulegen, die die Darstellung oder den Genuss des Werks nicht beeinträchtigt.
	Betreiber eines KI-Systems, das Text erzeugt oder manipuliert, der veröffentlicht wird, um die Öffentlichkeit über Angelegenheiten von öffentlichem Interesse zu informieren, müssen offenlegen, dass der Text künstlich erzeugt oder manipuliert wurde . Diese Pflicht gilt nicht, wenn die durch KI erzeugten Inhalte einem Verfahren der menschlichen Überprüfung oder redaktionellen Kontrolle unterzogen wurden und wenn eine natürliche oder juristische Person die redaktionelle Verantwortung für die Veröffentlichung der Inhalte trägt.

IV. Risikoklassen und Pflichten

5. KI-Systeme mit keinem oder geringem Risiko

KI-Systeme mit keinem oder niedrigem Risiko

Diese KI-Systeme unterliegen keiner Regulierung über die KI-VO (ausgenommen allgemeiner Obligationen, wie etwa der zur Förderung von KI-Kompetenz nach Art. 4 KI-VO oder zur Erfüllung von Transparenzpflichten), da von diesen KI-Systemen nur ein geringes oder gar kein Risiko ausgeht. Gleichwohl müssen die einschlägigen Gesetze, wie Urheberrecht, Datenschutz, Geheimnisschutz usw. beachtet und eingehalten werden.

5.1 Anwendungsbeispiele

Beispiele für KI-Systeme mit keinem/geringen Risiko:

- Spamfilter
- KI-gestützte Computerspiele

5.2 Pflichten für Anbieter & Betreiber

KI-Systeme, die keiner der vorgenannten Risikoklassen zuzuordnen sind, unterliegen keiner gesonderten Beschränkung.

Pflichten für Anbieter	Pflichten für Betreiber
Für KI-Systeme mit keinem oder niedrigem Risiko ergeben sich aus der KI-VO keinerlei verbindliche Verpflichtungen. Gleichwohl verpflichten wir Malteser uns auch hier, allgemeine Grundsätze wie menschliche Aufsicht, Nichtdiskriminierung und Fairness zu befolgen und die einschlägigen rechtlichen Vorgaben einzuhalten.	

6. KI-Systeme mit allgemeinem Verwendungszweck (GPAI) & systemischem Risiko (Artt. 51 ff. KI-VO)

Abweichend zur bisherigen Betrachtung beschäftigen sich die Artt. 51 ff. KI-VO nicht mit dem Risiko von KI-Systemen, sondern dem Risiko von KI-Modellen. Modelle mit allgemeinem Verwendungszweck (General Purpose Artificial Intelligence, kurz GPAI) haben keinen spezifischen Verwendungszweck, sondern können für beliebige Zwecke verwendet werden. Bei GPAI-Modellen existieren ausschließlich Pflichten für den Anbieter eines GPAI-Modells. Die Rolle des Betreibers gibt es hier nicht, da ein KI-Modell stets in ein System integriert werden muss, um überhaupt nutzbar und „betrieben“ werden zu können.

IV. Risikoklassen und Pflichten

6.1 Anwendungsbeispiel

KI-Modelle mit allgemeinem Verwendungszweck



Ein KI-Modell mit allgemeinem Verwendungszweck („General Purpose Artificial Intelligence“) weist ein systemisches Risiko auf, wenn es über Fähigkeiten mit hoher Wirkkraft bzw. hohem Wirkungsgrad verfügt.



Bestes Beispiel für ein Modell ohne spezifischen Verwendungszweck:
Das Sprachmodell GPT von OpenAI. Es kann (als KI-System ChatGPT) Gedichte verfassen oder aber Codeschnipsel für eine Softwareprogrammierung ausgeben.

6.2 Pflichten für Anbieter & Betreiber

Pflichten für Anbieter	Pflichten für Betreiber
Die Pflichten für Anbieter sind in Art. 50 KI-VO nachzulesen und werden, aufgrund fehlender Einschlägigkeit für den Malteser Verbund, nicht weiter betrachtet.	

V. Regulatorische Rahmenbedingungen

1. KI-Verordnung

Erläuterungen zur KI-VO siehe Kapitel I.3.

2. Arbeitsrecht

Grundsätzlich gelten auch bei der Erstellung und Verwendung von KI-Systemen dieselben arbeitsrechtlichen Grundsätze wie im Kontext von normaler Software. Die Nutzung von Software, egal ob mit oder ohne KI, ist ein Hilfsmittel und steht dem Grundsatz, dass Arbeitnehmende ihre Leistung höchstpersönlich zu erbringen haben, nicht entgegen.

3. KI und Interessenvertretungen (MAV & BR)

Bei der Implementierung von KI-Systemen sind, wie bei jeder Software, die Beteiligungsrechte der Interessenvertretungen zu wahren. Diese ergeben sich etwa aus der Einführung und Anwendung von technischen Einrichtungen und/oder bzgl. der Planung von Arbeitsverfahren und Arbeitsabläufen, einschließlich des Einsatzes von Künstlicher Intelligenz.

4. Schutz von Geschäftsgeheimnissen und Vertraulichkeit

Geschäftsgeheimnisse und vertrauliche Informationen gehören nicht in öffentlich zugängliche KI-Systeme. Auch bei geschlossenen KI-Systemen ist vorab sicherzustellen, dass die zu verarbeitenden Informationen nicht zum Trainieren der KI (weiter-)verwendet werden. Sofern nichts anderes geregelt wurde, dürfen keinerlei vertrauliche Unternehmensinformationen in KI-Systemen verarbeitet werden. Für die Einordnung der Vertraulichkeit ist, sofern vorhanden, die Schutzklasse des Dokumentes maßgeblich.



Beispiele für vertrauliche Informationen:

- Gehalts-, Abrechnungs- und Finanzdaten
- Kunden- und Patientendaten
- Lieferanten- und Partnerdaten
- Daten von Mitarbeitenden und Bewerbern
- Vorausschauen, Mittel- und Langfristplanungen
- Protokolle und Präsentationen mit sensiblen Daten, insbesondere von Sitzungen auf Ebene Vorstand, Gesellschafter und Organ

V. Regulatorische Rahmenbedingungen

5. Urheber- und Leistungsschutzrechte, gewerbliche Schutzrechte

Nur natürliche Personen können Schöpfer eines Werkes sein. Zu den schutzfähigen Werken zählen etwa Schriftwerke, Computerprogramme, Werke der Musik, Lichtbild- und Filmwerke. Werke im urheberrechtlichen Sinne sind immer persönliche geistige Schöpfungen eines Menschen. Juristische Personen (bspw. eine GmbH) können demnach nicht Schöpfer eines Werkes sein. Und auch KI-Systeme können keine Werke erschaffen, weil sie keine Menschen sind.

Das Urheberrecht schützt den Urheber in seinem Urheberpersönlichkeitsrecht und den Verwertungsrechten, die ihm allein zustehen. Zu den Verwertungsrechten gehören z.B. das Recht zur Vervielfältigung, zur Verbreitung oder öffentlichen Zugänglichmachung. Das Urheberrecht lässt sich nicht übertragen, doch kann der Urheber anderen etwa das Recht zur Nutzung oder Verwertung seines Werkes einräumen.

5.1 Urheber und von KI erzeugte Arbeitsergebnisse

Nicht jede Schöpfung fällt unter den Schutz des Urheberrechts. Dazu muss sie die sogenannte Schöpfungshöhe bzw. -tiefe erreichen.



Schöpfungshöhe und „Kleine Münze“:

Der Bundesgerichtshof (BGH) hat in einem Urteil bestimmt, dass Schöpfungshöhe insbesondere dann erreicht ist, wenn nach Auffassung von einigermaßen fachkundigen Kreisen eine künstlerische Leistung vorliegt. Die Untergrenze wird dabei als „Kleine Münze“ bezeichnet.

5.2 Schutzfähigkeit von Prompts

Die Eingabe in einen KI-Chatbot (Prompt) muss eine gewisse Individualität und Kreativität aufweisen, damit der Schutz des Urheberrechts greifen kann. Rein beschreibende Angaben sind nicht schützenswert. Das Urheberrecht entsteht bereits mit der Schöpfung selbst. Es bedarf keiner Kennzeichnung, Registrierung oder Anmeldung, wie es bspw. beim Markenrecht der Fall ist.

5.3 Schutzfähigkeit von durch KI erzeugten Arbeitsergebnissen

Nach derzeitiger Rechtslage kann in Deutschland nur ein Mensch Urheber sein. Damit wird jedoch nicht der Nutzende, der einen Prompt eingibt, zum Urheber des (Arbeits-)Ergebnisses! Er gibt dem KI-System zwar mittels eines Prompts (der durchaus Schöpfungstiefe und damit urheberrechtlich schützenswert sein kann) die Aufgabenstellung vor. Die Arbeit im Hinter-

V. Regulatorische Rahmenbedingungen

grund erledigt allerdings die KI bzw. das KI-Modell. Der Nutzende kann bei der Erstellung eines Bildes das durch Midjourney oder DALL-E erzeugte Ergebnis weder vorhersehen noch erklären. Folglich ist er nicht Schöpfer bzw. Urheber des so erstellten Bildes.

5.4 Können von KI geschaffene Arbeitsergebnisse gegen das Urheberrecht verstoßen?

Ja. Erzeugt die KI-Inhalte, die urheberrechtlich geschützten Werken zu stark ähneln oder etwa Texte komplett oder anteilig übernehmen, kann dies eine Urheberrechtsverletzung darstellen.

5.5 Trainieren von KI und Urheberrecht

Grundsätzlich ist für die Verwendung urheberrechtlich geschützter Werke die Erlaubnis des Urhebers erforderlich. Im Urheberrechtsgesetz greift beim Trainieren von KI jedoch ein Erlaubnistatbestand für die automatisierte Analyse von digitalen bzw. digitalisierten Werken (sog. Text und Data Mining-Schranke). Demnach sind Vervielfältigungen von rechtmäßig zugänglichen Werken für das Text und Data Mining urheberrechtlich zulässig. Allerdings sind Vervielfältigungen umgehend zu löschen, wenn sie für das Text und Data Mining nicht mehr erforderlich sind.



Beispiel für eine Urheberrechtsverletzung:

Ein KI-System generiert ein Bild, das der Comicfigur Batman stark ähnelt. Die Urheberrechte am Fledermausmann gehören, als Autor und Zeichner, Bob Kane und Bill Finger. Die Verwertungs- und Markenrechte an Logo und Figur liegen wiederum bei DC Comics. Eine Verwendung des erzeugten Bildes, bspw. zu Marketingzwecken, darf daher nur mit Zustimmung der Rechteinhaber erfolgen. KI-Nutzende müssen daher, nicht nur aus Gründen der allgemeinen Qualitätskontrolle, die (Arbeits-)Ergebnisse eines KI-Systems kritisch überprüfen und im Idealfall zunächst umfassend überarbeiten.

6. Datenschutz

Die Vorschriften (insb. KDR-OG) und Richtlinien zum Datenschutz (siehe „Mitgeltende Unterlagen“) gelten auch für die Nutzung von KI. Vor der Verarbeitung von Daten mithilfe von KI ist die Stabstelle Daten & KI (siehe Kapitel III) oder der zuständige Datenschutzkoordinator bzw. die zuständige Datenschutzkoordinatorin zu konsultieren.

V. Regulatorische Rahmenbedingungen

Da die von KI-Anwendungen verwendeten Lernverfahren in vielen Fällen ebenso wenig bekannt sind wie ihre Trainingsverfahren und -daten, kann nicht ausgeschlossen werden, dass Angaben, die Nutzende im Zusammenhang mit ihrer beruflichen Tätigkeit bei Registrierung oder Verwendung einer KI-Anwendung machen, zur Selbstoptimierung oder zum Training der KI genutzt werden, ohne dass dies vom Arbeit- bzw. Dienstgeber erkennbar ist.



Bereits die Anmeldung zur Nutzung eines KI-Systems erfordert zumeist eine Registrierung, sofern es nicht zentral über die SoCura bereitgestellt wird. So ist zumindest eine E-Mail-Adresse, in einigen Fällen auch eine Telefonnummer einzugeben.

6.1 Keine Verarbeitung von personenbezogenen Daten ohne Rechtsgrundlage

Personenbezogene Daten sowie auch besondere Kategorien Personenbezogener Daten dürfen nicht ohne Rechtsgrundlage gem. § 6 Abs. 1 KDR-OG an KI-Systeme bzw. KI-Modelle übermittelt werden.

Werden Daten verarbeitet, die ursprünglich zu einem anderen Zweck erhoben wurden, liegt eine Zweckänderung vor, die ebenfalls eine Erlaubnis gem. § 6 Abs. 2 KDR-OG erfordert.

Dies bedeutet, dass jegliche Übermittlung oder Verarbeitung personenbezogener Daten ohne rechtliche Grundlage strikt untersagt ist, sofern nicht ein Erlaubnistatbestand gem. § 6 KDR-OG vorliegt.



Mögliche Erlaubnistatbestände können die Einwilligung der betroffenen Person sein oder das berechtigte Interesse der Malteser. Für die Anwendung des berechtigten Interesses ist eine Interessensabwägung durchzuführen.

6.2 Informationspflichten

Sofern mittels KI personenbezogene Daten verarbeitet werden, haben die Malteser (als Verantwortliche im datenschutzrechtlichen Sinne) die Pflicht, die Nutzenden umfassend über die Verarbeitung ihrer Daten zu informieren, insbesondere über ...

- ... den Zweck der Datenverarbeitung,
- ... die Rechtsgrundlage,
- ... die Speicherdauer,
- ... Empfänger der Daten und
- ... die Rechte der betroffenen Personen.

V. Regulatorische Rahmenbedingungen

Erfolgt die Verarbeitung auf Grundlage des berechtigten Interesses müssen Nutzende explizit über ihr Widerspruchsrecht informiert werden. Dementsprechend muss die Möglichkeit bestehen, einen Widerspruch in diesem Sinne im KI-System umzusetzen. Diese Informationen müssen klar, verständlich und zugänglich kommuniziert werden, um den Nutzenden die Möglichkeit zu geben, informierte Entscheidungen über die Verarbeitung ihrer Daten zu treffen.

6.3 Verbot automatisierter Entscheidungsfindung

Autonome Entscheidungen von IT- oder KI-Systemen ohne Kontrolle und Korrekturmöglichkeit durch den Menschen sind gem. § 24 Abs. 1 KDR-OG verboten, wenn sie eine rechtliche Wirkung gegenüber betroffenen Personen entfalten oder sie in ähnlicher Weise erheblich beeinträchtigen. Mögliche Ausnahmen von dem Verbot sind abzustimmen. Hierbei sind insbesondere die Maßnahmen gem. § 24 Abs. 3 f. KDR-OG sowie die besonderen Informationspflichten gem. § 15 Abs. 2 lit. f. KDR-OG zu beachten.

6.4 Verarbeitung von Daten im Auftrag

In der Regel ist derjenige für die Datenverarbeitung verantwortlich, welcher ein KI-System verwendet (Betreiber im Sinne der KI-VO). In vielen Fällen werden uns Maltesern KI-Systeme durch Dritte zur Verfügung gestellt, die wir für unsere eigenen Zwecke verwenden. Die Anbieter der KI-Systeme gelten somit als Auftragsverarbeiter im Sinne von § 29 KDR-OG. Die Verarbeitung der Daten der Malteser erfolgt mithin in unserem Auftrag und gemäß unserer Weisung. Dazu muss eine Auftragsverarbeitungsvereinbarung mit dem Anbieter abgeschlossen werden.

In besonderen Fällen können mehrere Akteure für die Datenverarbeitung verantwortlich sein. Möchte der Anbieter des KI-Systems die Nutzungsdaten inklusive der Eingaben (Prompts) für das weitere Training des KI-Systems nutzen, spricht vieles dafür, dass eine gemeinsame Verantwortlichkeit nach § 28 KDR-OG zwischen dem Anbieter und Betreiber besteht.

Bei der Auswahl von Anbietern ist darauf zu achten, dass die Grundsätze „Privacy by Design“⁷ und „Privacy by Default“⁸ sowie weitere technische und organisatorische Maßnahmen nach § 27 KDR-OG berücksichtigt und implementiert wurden.

⁷ Bereits bei der Entwicklung von KI-Systemen muss der Datenschutz berücksichtigt werden: Der technische Datenschutz verfolgt vor allem das Ziel, Nutzer vor Risiken für ihre Rechte und Freiheiten zu schützen, ohne dass diese selbst aktiv werden müssen.

⁸ Der Grundsatz verlangt datenschutzfreundliche Voreinstellungen von KI-Systemen: Der (für die Datenverarbeitung) Verantwortliche trifft geeignete technische und organisatorische Maßnahmen, die sicherstellen, dass durch Voreinstellung nur personenbezogene Daten, deren Verarbeitung für den jeweiligen bestimmten Verarbeitungszweck erforderlich ist, verarbeitet werden.

V. Regulatorische Rahmenbedingungen

6.5 Betroffenenrechte

Für die Verarbeitung personenbezogener Daten in KI-Systemen und -Modellen müssen Löschkonzepte erstellt und implementiert werden, damit Daten nur so lange vorgehalten werden, wie der legitime Zweck für ihre Verarbeitung besteht. Personenbezogene Daten sind in KI-Systemen bzw. Modellen so zu verarbeiten, dass dem legitimen Löschbegehren einer betroffenen Person gem. § 19 KDR-OG entsprochen werden kann.

Des Weiteren sind auch alle weiteren Betroffenenrechte gemäß § 17–23 KDR-OG so zu beachten, dass diesen ebenfalls fristgemäß entsprochen werden kann.

7. Haftung und KI

Die Tatsache, dass maschinelles Lernen menschliche Fähigkeiten weit übertreffen kann, bedeutet nicht, dass solche Systeme keine Fehler machen und immer perfekt funktionieren. Neben einer fehlerhaften oder schlicht falschen Ausgabe von Informationen können von KI generierte Inhalte insbesondere die Rechte Dritter verletzen, wie in Kapitel V Ziffer 5 beschrieben. Haftungsfragen zur Verletzung durch oder vielmehr mittels KI sind rechtlich derzeit noch nicht abschließend geklärt.

7.1 Trainieren von KI-Modellen (Text und Data Mining)

In der juristischen Literatur herrscht zumindest relative Einigkeit, dass die Schranke bzw. Erlaubnis aus § 44b UrhG (Text und Data Mining) nicht auf von KI-Systemen erstellte Ergebnisse anwendbar ist, da gem. § 44 Abs. 2 UrhG lediglich Vervielfältigungen von rechtmäßig zugänglichen Werken für das Text und Data Mining zulässig sind. Jedenfalls gefestigt wird man sagen können, dass Nutzende das Recht des eigentlichen Urhebers verletzen, wenn sie dessen geschütztes Werk als Prompt eingeben, ohne, dass ihnen der Urheber hierzu seine Zustimmung erteilt hat, da hierbei eben nicht das KI-Modell trainiert wird.

7.2 Haftung und Arbeitsrecht

Arbeitsrechtlich wird man Schädigungen, die Arbeitnehmende an Rechten und Rechtsgütern des Arbeit- bzw. Dienstgebers oder eines Dritten mittels eines KI-Systems herbeiführen, über den innerbetrieblichen Schadensausgleich abwickeln. Danach ist eine persönliche Haftung ausgeschlossen, wenn der Schaden nur durch leicht fahrlässiges Handeln des Arbeitnehmenden verursacht wurde.

VI. Technikgestaltung

Die nachfolgenden Anforderungen gelten für interne wie externe Entwicklungen gleichermaßen.

1. Zugriffssteuerung und Authentifizierung

KI-Systeme müssen grundsätzlich über ein Zugriffsberechtigungskonzept verfügen, das nur berechtigten Benutzenden Zugriff gewährt. Mehrstufige Authentifizierung ist wünschenswert, um den Abfluss von Daten zu verhindern – insbesondere bei der Verarbeitung von sensiblen Daten. Zugriffsrechte sind regelmäßig zu überprüfen und entsprechend den Aufgaben und Verantwortlichkeiten anzupassen.

2. Sicherheit bei der Entwicklung von KI-Systemen

Bei der Entwicklung und Anpassung von KI-Modellen und -Systemen sind Maßnahmen zur Sicherstellung der Code- und Datenintegrität verpflichtend. Anbieter haben (siehe Mitgeltende Unterlagen: „Secure Coding“) während des gesamten Lebenszyklus des KI-Systems dafür Sorge zu tragen, dass keine unsicheren Abhängigkeiten genutzt und Sicherheitsvorgaben bei der Programmierung eingehalten werden. Zusätzlich ist ein umfassender Manipulationsschutz zu implementieren:

- Integration und regelmäßige Aktualisierung von robusten Algorithmen zur Erkennung von Manipulationsversuchen
- Kontinuierliche Überwachung des Systems auf ungewöhnliche Aktivitäten oder Anomalien
- Regelmäßige Penetrationstests und Sicherheitsaudits zur Identifikation und Behebung von potenziellen Schwachstellen
- Umsetzung von Techniken für den Manipulationsschutz des KI-Systems (z.B. Adversarial Training, Defensive Distillation)
- Etablieren und regelmäßiges Üben eines Prozesses zur schnellen Reaktion auf erkannte Manipulationsversuche

Diese Schutzmaßnahmen sind integraler Bestandteil der Entwicklung und des Betriebs. Sie müssen kontinuierlich an neue Bedrohungsszenarien angepasst werden.

3. Sicherung der Transparenz und Nachvollziehbarkeit

Insbesondere KI-Systeme, die als Hochrisiko-KI oder KI mit systemischem Risiko i. S. d. KI-VO gelten, müssen so konzipiert sein, dass Entscheidungen und Empfehlungen nachvollziehbar und dokumentierbar sind. Dies umfasst insbesondere Algorithmen, die kritische Entscheidungen beeinflussen. Transparenz und Rückverfolgbarkeit müssen gewährleistet sein, damit potenzielle Sicherheitslücken frühzeitig zu erkennen sind. Dafür sind folgende Maßnahmen zu ergreifen:

VI. Technikgestaltung

- Erklärbare KI: Verständliche Entscheidungsprozesse, besonders in kritischen Bereichen
- Rückverfolgbarkeit: Dokumentation von Daten-, Modell- und Entscheidungsherkunft
- Transparenz: Offenlegung von technischen Prinzipien, Modellspezifikationen und ethischer Konformität
- Erklärungsstrategien: Verwendung von Methoden wie prototypenbasierter Erklärungen
- Integration externer Wissensbasen: Verbesserung der Erklärbarkeit durch zusätzliches Fachwissen
- Rechtliche und ethische Compliance: Kontinuierliche Überprüfung und Dokumentation rechtlicher und ethischer Standards

Durch diese Maßnahmen wird die Transparenz und Vertrauenswürdigkeit von KI-Anwendungen erhöht und die Erkennung von Sicherheitsrisiken erleichtert

4. Technische Voraussetzungen für KI-Systeme

KI-Systeme und -Modelle benötigen erhebliche Ressourcen, insbesondere beim Training komplexer Modelle. Diese Anforderungen können die Implementierung in ressourcenbeschränkten Umgebungen erschweren und die Betriebskosten erhöhen. Daher ist es entscheidend, sowohl die Datenqualität als auch die technische Infrastruktur bei der Planung und Nutzung von KI-Systemen zu berücksichtigen.

VII. KI-Kompetenz und Bewusstsein für KI

Die Förderung von KI-Kompetenz und des Bewusstseins für KI sind entscheidend, um den verantwortungsvollen und effizienten Einsatz von Künstlicher Intelligenz in unserer Organisation sicherzustellen. Wir verpflichten uns, allen Mitarbeitenden Zugang zu Schulungen und Weiterbildungen zu bieten, die sowohl grundlegendes Wissen über KI als auch ein Bewusstsein für ihre Auswirkungen auf unsere Arbeit und die Gesellschaft vermitteln.

Durch die Verbindung von KI-Kompetenz und KI-Sensibilisierung mit unserer Unternehmenskultur stellen wir sicher, dass unsere Mitarbeitenden nicht nur die Technologie verstehen, sondern auch kritisch und verantwortungsbewusst mit ihr umgehen können.



KI-Kompetenz & KI-Seeperdchen:

Mitarbeitende, die KI-Systeme im beruflichen Kontext (im Auftrag des Arbeit- bzw. Dienstgebers) nutzen, müssen nachweisen können, dass sie über grundlegende KI-Kompetenz verfügen. Diese Verpflichtung gilt unabhängig von der Risikobewertung der genutzten KI-Systeme, also auch bereits bei der Verwendung von KI-Systemen mit geringem Risiko. Mit dem KI-Seeperdchen gibt es eine speziell auf die Bedürfnisse von uns Maltesern zugeschnittene Online-Schulung, mit deren erfolgreichem Absolvieren Mitarbeitende diese (von der KI-VO explizit geforderte) grundlegende KI-Kompetenz nachweisen können.

VIII. Inkrafttreten

Diese Richtlinie tritt am 01.01.2026 in Kraft. Durch die Dynamik der Entwicklungen im Bereich der KI wird es notwendig sein, die KI-Richtlinie regelmäßig zu aktualisieren.

Mitgeltende Unterlagen

- KI-Leitlinie
- ITK-Richtlinie
- Datenschutz-Richtlinie
- KI-Verordnung (KI-VO)

IX. Anhang (Schnellübersicht & Risiko-Navigator)

1. Schnellübersicht Risikoklassen und Transparenzpflichten

Risikoklasse	Erläuterung & Transparenz	Beispiel	Vorgaben
Verbotene KI	verstößt gegen Grundrechte bzw. Werte der EU	Manipulation, Social Scoring	Verwendung verboten
Hochrisiko-KI	wirkt sich nachteilig auf die Sicherheit oder Grundrechte von Menschen aus	Personalmanagement, Rechtswesen	menschliche Aufsicht, Konformitätsbewertung, Transparenzpflichten
Anbieter	Transparenzpflichten	Bereitstellung mit Betriebsanleitungen in einem geeigneten digitalen Format, die präzise, vollständige, korrekte und eindeutige Informationen in einer für die Betreiber relevanten, barrierefrei zugänglichen und verständlichen Form enthalten (Art. 13 KI-VO)	
Betreiber		Informationspflicht gegenüber betroffenen Personen, dass diese mit einem Hochrisiko-KI-System interagieren oder ihre Daten darin verarbeitet werden	
KI mit begrenztem Risiko	muss also solches klar erkennbar sein	Chatbots	Transparenzpflichten
Anbieter	Transparenzpflichten	KI-System ist zur direkten Interaktion mit Menschen bestimmt (Art. 50 Abs. 1 KI-VO); KI-Systeme, die für die direkte Interaktion mit natürlichen Personen bestimmt sind, müssen so konzipiert und entwickelt werden, dass die betreffenden natürlichen Personen informiert werden, dass sie mit einem KI-System interagieren, es sei denn, dies ist aus Sicht einer angemessen informierten, aufmerksamen und verständigen natürlichen Person aufgrund der Umstände und des Kontexts der Nutzung offensichtlich. KI-System erzeugt synthetische Audio-, Bild-, Video- oder Textinhalte (Art. 50 Abs. 2 KI-VO); Ausgaben des KI-Systems müssen in einem maschinenlesbaren Format gekennzeichnet und als künstlich erzeugt oder manipuliert erkennbar sein. Emotionserkennung und Systemen zur biometrischen Kategorisierung (Art. 50 Abs. 3 KI-VO); KI-Systemen, die Deepfakes oder Texte erzeugen oder manipulieren (Art. 50 Abs. 4 KI-VO); Sicherstellen, dass KI-Systeme, die für die direkte Interaktion mit natürlichen Personen bestimmt sind, so konzipiert und entwickelt werden, dass die betreffenden natürlichen Personen informiert werden, dass sie mit einem KI-System interagieren.	
Betreiber		Transparenz bei Deep Fakes: Nutzende müssen offenlegen, wenn Inhalte durch KI erzeugt oder manipuliert wurden, insbesondere wenn es sich um Deep Fakes handelt. Offenlegung bei KI-generierten Texten: Wenn Texte, die von KI-Systemen generiert oder manipuliert wurden, veröffentlicht werden, um die Öffentlichkeit über wichtige Angelegenheiten zu informieren, muss ebenfalls offenbart werden, dass diese Texte künstlich erstellt wurden, sofern die Texte nicht durch Personen überprüft werden oder eine natürliche oder juristische Person die redaktionelle Verantwortung für die Veröffentlichung der Inhalte trägt.	
Kein/minimales Risiko	hat nur geringe oder keine Auswirkung auf die Sicherheit oder den Menschen	Spamfilter, Computerspiele	Allgemeine rechtliche Verpflichtungen
Systemisches Risiko (GPAI)	verfügt über „Fähigkeiten mit hohem Wirkungsgrad, gemessen in sog. Gleitkommaoperationen (auch „FLOPS“ genannt), mit mehr als 10 ²⁵ “		Ergänzende Verpflichtungen finden sich in Art. 55 KI-VO.

2. TÜV-Risiko-Navigator

Der vom TÜV AILab entwickelte AI Act Risk Navigator ist ein Klassifizierungstool zur einfachen und präzisen Einordnung von KI-Systemen und KI-Modellen in die Risikoklassen der europäischen KI-Verordnung und für eine erste Einschätzung zu empfehlen:



www.tuev-risk-navigator.ai/?lang=de