

Zur Verantwortung, Relevanz und Zukunftsfähigkeit von Museen und Sammlungen im Zeitalter der Künstlichen Intelligenz

Vorwort

Museen stehen heute an einem neuralgischen Punkt ihrer Entwicklung. Zwischen digitaler Transformation, gesellschaftlichem Wandel und wachsender Erwartung an ihre öffentliche Rolle müssen sie ihre Kernaufgaben neu justieren, ohne ihren Auftrag preiszugeben. Im Zeitalter der Künstlichen Intelligenz (KI) stellt sich diese Herausforderung mit besonderer Dringlichkeit: Was bedeutet es, zu sammeln, zu bewahren, zu erforschen und zu vermitteln, wenn Technologien nicht nur reproduzieren, sondern generieren, interpretieren und kuratieren?

Bereits Walter Benjamin unterstrich in seinem Werk *Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit*, dass technische Reproduzierbarkeit die „Aura“ des Kunstwerks verändert – seine Einzigartigkeit, seine Verortung im Hier und Jetzt. Heute stehen wir vor einer neuen Qualität dieser Entwicklung. KI kann nicht nur reproduzieren, sondern eigenständig Bilder, Texte und Narrative erzeugen. Die Frage nach der Aura verschiebt sich – hin zur Frage nach Authentizität, Autor:innenschaft und Kontext. Museen sind gefordert, nicht nur Bewahrer von Objekten zu sein, sondern auch Vermittler von Bedeutungen in einer Welt, in der Original und Simulation zunehmend schwer unterscheidbar werden.

Dabei darf die technologische Dimension nicht den Blick auf die zentrale Grundlage musealer Arbeit verstellen: die Sammlung. Sammlungen sind keine statischen Archive, sondern dynamische Wissensräume. Ihre Pflege, Erforschung und kritische Erweiterung bilden das Rückgrat jeder Institution. Gerade im KI-Zeitalter gewinnt diese Arbeit an Bedeutung. Denn Algorithmen sind nur so gut wie die Daten, auf denen sie basieren. Wer Sammlungen vernachlässigt, überlässt die kulturelle Deutungshoheit implizit jenen, die Daten aggregieren, ohne sie historisch, ethisch und kuratorisch zu kontextualisieren.

Die im Wiener Manifest des Digitalen Humanismus geforderte menschenzentrierte Gestaltung digitaler Technologien bedeutet für Museen, dass KI nicht als Selbstzweck eingesetzt werden darf, sondern sich an den Werten der Institution messen lassen muss – Transparenz, Inklusion, wissenschaftliche Integrität und gesellschaftliche Verantwortung. Museen können hier als Laboratorien eines digitalen Humanismus wirken, indem sie Technologien kritisch reflektieren und zugleich produktiv einsetzen.

Zugleich bleibt das Museum ein physischer Ort – und gerade darin liegt eine seiner größten Stärken. Zahlreiche Studien belegen die positiven Auswirkungen von Museumsbesuchen auf psychische Gesundheit und Wohlbefinden. In einer zunehmend digitalisierten und beschleunigten Welt bieten Museen Räume der Entschleunigung, der Konzentration und der sinnlichen Erfahrung. Sie ermöglichen Begegnungen – mit Objekten, mit Geschichten und allen voran mit anderen Menschen. Diese Qualität ist nicht durch digitale Anwendungen oder Zugänge ersetzbar, sondern wird durch eben diese noch wichtiger.

Museen sind darüber hinaus Orte der kritischen Auseinandersetzung. Sie können und müssen gesellschaftliche Debatten aufgreifen und auch Stellung beziehen, sei es zum Klimawandel, zu sozialen Ungleichheiten und Ungerechtigkeiten in Gegenwart und Geschichte. Im Kontext von KI

stellt sich die Frage nach Bias, nach Machtstrukturen in Datensätzen oder nach der Rolle von Technologie in demokratischen Gesellschaften. Museen haben die Möglichkeit, diese Themen nicht nur zu illustrieren, sondern zur Diskussion zu stellen – interdisziplinär, evidenzbasiert und öffentlich.

Für Museen und ihre Teams ergibt sich daraus eine strategische und inhaltliche Verantwortung. Strategisch gilt es, Institutionen so aufzustellen, dass sie technologisch anschlussfähig bleiben, ohne ihre Unabhängigkeit zu verlieren. Inhaltlich bedeutet es, Programme zu entwickeln, die sowohl Expertise als auch Haltung zeigen. Museen und ihre Teams benötigen dafür Ressourcen, Weiterbildung und Unterstützung ihrer Stakeholder. Politik und Verwaltung sind gefordert, die langfristige Bedeutung von Museen als öffentliche Infrastruktur anzuerkennen und entsprechend zu fördern.

Das Museum der Zukunft wird weder rein analog noch vollständig digital sein. Es wird hybrid, vernetzt und reflexiv agieren. Seine Stärke liegt nicht in der Konkurrenz zu technologischen Plattformen, sondern in seiner spezifischen Kompetenz: der Verbindung von Objekt, Wissen und gesellschaftlicher Relevanz. Gerade im Zeitalter der KI kann das Museum zu einem Ort werden, an dem nicht nur kulturelles Erbe bewahrt, sondern auch Zukunft verhandelt wird. Der Museumsbund Österreich sieht in der Veranstaltung ein zukunftsweisendes Modellformat für grenzüberschreitenden Austausch und bekräftigt hiermit seine Bereitschaft zur ideellen Unterstützung sowie zur fachlichen Einbindung unserer Expertise und unseres Netzwerks.

Empfehlungen für den verantwortungsvollen Einsatz von Künstlicher Intelligenz

Kurzversion

Museen und Sammlungen sind Teil des Bildungsangebots; sie generieren neues Wissen, unterziehen vorhandenes Wissen einer Prüfung und werden der Verantwortung gerecht, gültige Auslegungen von Fragestellungen aus Wissenschaft und Gesellschaft anzubieten, denen die Menschen vertrauen können. KI darf nicht als Ersatz für tatsächliche Wissensarbeit und menschliche Beurteilung angewendet werden, vielmehr ist im Zuge der Implementierung und Anwendung auch ihre Auswirkung kritisch zu hinterfragen.

KI ist ein leistungsfähiges Werkzeug, aber kein Ersatz für menschliches Urteilsvermögen. KI-Anwendungen werden gezielt, verantwortungsvoll und transparent eingesetzt. Die finale Verantwortung für alle KI-gestützten Inhalte und Entscheidungen liegt stets beim Menschen.

Allgemeiner Hinweis

Die nachfolgenden Empfehlungen versteht sich als Grundlage, um für die eigene Institution Richtlinien zu erstellen. Wir empfehlen außerdem eine jährliche Überprüfung der erstellten Richtlinien auf ihre Sinnhaftig- und Gültigkeit.

Diese Empfehlung wurden mit Hilfe der Kolleg:innen aus den österreichischen Landesmuseen erstellt.

Die 7 Grundprinzipien

1. Rechtliche Rahmenbedingungen

- nur genehmigte KI-Systeme einsetzen (interne Tool-Liste beachten)
- Urheber-, Persönlichkeits- und Markenrechte sowie DSGVO einhalten
- KI kann kein Urheber sein – trotzdem können Input-Daten geschützt sein
- Nutzungsbedingungen (AGB) der verwendeten Tools prüfen
- Beschaffung neuer KI-Tools nur nach interner Freigabe

2. Menschliche Kontrolle und Verantwortung

- Der Mensch bleibt immer die wichtigste Instanz („Human in the Loop“)
- KI-Ergebnisse vor Verwendung auf Richtigkeit, Fairness und Kontext prüfen
- keine ungeprüfte Übernahme von KI-Outputs – mindestens eine qualifizierte Person validiert
- KI „halluziniert“: generierte Fakten, Quellen und Daten können falsch sein
- finale Entscheidungen und Verantwortung verbleiben beim Menschen

3. Transparenz und Kennzeichnung

- KI-generierte Inhalte kennzeichnen, wenn KI substantiell zum Ergebnis beigetragen hat
- Kennzeichnung bei Deepfakes, KI-Stimmen, generierten Bildern/Videos verpflichtend (EU AI Act Art. 50)
- Interaktionen mit KI-Chatbots müssen als solche erkennbar sein
- keine Kennzeichnung nötig bei rein assistierender KI (z. B. Rechtschreibkorrektur)

4. Datenschutz, Informationssicherheit und Urheberrecht

- keine personenbezogenen Daten in externe KI-Systeme eingeben ohne Freigabe
- keine Geschäfts- oder Betriebsgeheimnisse, vertraulichen Informationen oder nicht veröffentlichte Forschungsergebnisse einspeisen
- kostenlose KI-Tools können Eingaben für Trainingszwecke speichern – Vorsicht!
- EU-/europäisch gehostete oder interne KI-Lösungen bevorzugen
- bei Eingabe personenbezogener Daten: vorher anonymisieren oder pseudonymisieren

5. Sicherheit und Qualität der Daten

- KI-Ergebnisse mit gesichertem Wissen und/oder mit Fachliteratur abgleichen
- Quellen kritisch hinterfragen – KI kann nicht-existierende Quellen erfinden
- Informationsqualität, Vollständigkeit und Aktualität prüfen
- KI-generierten Code auf Fehler und Sicherheitslücken überprüfen
- Keine Desinformation verbreiten – Plausibilitätskontrolle vor Veröffentlichung

6. Ethik, Vielfalt, Inklusion und Nachhaltigkeit

- KI-Inhalte regelmäßig auf Verzerrungen (Bias) und Diskriminierung prüfen
- alle Zielgruppen gleichermaßen berücksichtigen: Herkunft, Geschlecht, Alter, Behinderung
- KI nur einsetzen, wenn ein klarer Mehrwert besteht und Ressourcen geschont werden
- Energieverbrauch und ökologischen Fußabdruck bei KI-Nutzung bedenken
- einfachere Alternativen prüfen, bevor rechenintensive KI-Modelle eingesetzt werden

7. Aus- und Weiterbildung

- vor der KI-Nutzung muss ein KI-Kompetenztraining absolviert werden
- kontinuierliche Weiterbildung ist Pflicht: Technologie entwickelt sich rasch
- Institutionen stellen Schulungsressourcen und E-Learning-Angebote bereit
- Wissen und Erfahrungen institutionsübergreifend teilen

KI-Anwendungsregister / KI-Too-Liste

Ein KI-Anwendungsregister sollte alle eingesetzten, getesteten oder geplanten KI-Anwendungen erfassen. Es dient der Transparenz, Steuerung und Risikobewertung.

Monitoring, Feedback und Weiterentwicklung

Die KI-Richtlinie soll regelmäßig überprüft und weiterentwickelt werden.

Schnellcheck vor jeder KI-Nutzung

Vor dem Einsatz fragen:

- Ist das verwendete KI-Tool genehmigt?
- Enthalten meine Eingaben personenbezogene oder vertrauliche Daten?
- Habe ich das Ergebnis fachlich geprüft und verifiziert?
- Muss ich den KI-Einsatz kennzeichnen?
- Werden Urheberrechte eingehalten?
- Bietet der KI-Einsatz echten Mehrwert?

Verbotene Anwendungen (EU AI Act)

- Social Scoring – KI zur Bewertung von Personen anhand sozialer Daten
- Manipulation und Verhaltenssteuerung ohne Wissen der Betroffenen
- Biometrische Kategorisierung anhand sensibler Merkmale (Religion, Ethnie u.a.)
- Gesichtserkennungs-Datenbanken mit Bildern aus dem Internet
- Einsatz bei kritischen Infrastrukturen ohne aufsichtsrechtliche Freigabe

Verantwortlichkeiten

Wer	Verantwortung
Mitarbeitende	Sorgfältige und regelkonforme Nutzung, Prüfung der Ergebnisse, Meldepflichten einhalten
Leitung / Direktion	Freigabe von KI-Tools, Schulungsangebote, strategische Verankerung der Prinzipien
IT / Datenschutz	Technische Prüfung, Sicherheitsfreigabe, Beratung bei Einzelfällen
alle	keine verbotenen Inhalte eingeben, Kennzeichnungspflichten beachten, bei Unsicherheit nachfragen

Empfehlungen für den verantwortungsvollen Einsatz von Künstlicher Intelligenz

Langversion

Museen und Sammlungen sind Teil des Bildungsangebots; sie generieren neues Wissen, unterziehen vorhandenes Wissen einer Prüfung und werden der Verantwortung gerecht, gültige Auslegungen von Fragestellungen aus Wissenschaft und Gesellschaft anzubieten, denen die Menschen vertrauen können. KI darf nicht als Ersatz für tatsächliche Wissensarbeit und menschliche Beurteilung angewendet werden, vielmehr ist im Zuge der Implementierung und Anwendung auch ihre Auswirkung kritisch zu hinterfragen.

KI ist ein leistungsfähiges Werkzeug, kein Ersatz für menschliches Urteilsvermögen. KI-Anwendungen werden gezielt, verantwortungsvoll und transparent eingesetzt. Die finale Verantwortung für alle KI-gestützten Inhalte und Entscheidungen liegt stets beim Menschen. KI unterstützt – Menschen entscheiden.

Die Empfehlung orientiert sich an nationalen und internationalen Gesetzen und Handlungsanleitungen:

- EU AI Act (Verordnung (EU) 2024/1689)
- Datenschutz-Grundverordnung (DSGVO / Verordnung (EU) 2016/679)
- UNESCO-Empfehlung zur Ethik der Künstlichen Intelligenz
- Nationale Strategien und institutionelle Erfahrungen österreichischer Kultureinrichtungen

Allgemeiner Hinweis

Die nachfolgenden Empfehlungen versteht sich als Grundlage, um für die eigene Institution Richtlinien zu erstellen. Wir empfehlen außerdem eine jährliche Überprüfung der erstellten Richtlinien auf ihre Sinnhaftig- und Gültigkeit.

Diese Empfehlung wurden mit Hilfe der Kolleg:innen aus den österreichischen Landesmuseen erstellt.

1. Einführung und Zweck

Die Entwicklung der Künstlichen Intelligenz (KI) eröffnet Organisationen im Kulturbereich neue Chancen. Gleichzeitig bringt der KI-Einsatz erhebliche Anforderungen und Risiken mit sich – insbesondere in den Bereichen Datenschutz, Urheberrecht, wissenschaftliche Qualität, Transparenz, Diskriminierungsfreiheit und Informationssicherheit.

Diese Empfehlung bietet einen Rahmen für den verantwortungsvollen, ethisch fundierten und rechtssicheren Umgang mit KI-Systemen. Sie richtet sich an Mitarbeitende aller Ebenen sowie an die Institutionen selbst.

2. Geltungsbereich

Diese Empfehlung gilt für alle Personen, die im Auftrag oder Namen der Institution handeln, darunter:

- festangestellte Mitarbeitende aller Abteilungen
- freie Dienstnehmer:innen, Praktikant:innen und Werkvertragsnehmer:innen
- externe Kooperationspartner:innen, die KI-basierte Arbeitsbeiträge liefern
- Auftragnehmer:innen, die im Rahmen von Projekten KI nutzen

Sie gilt für alle KI-Systeme, die dienstlich genutzt werden – unabhängig davon, ob die Nutzung intern oder extern erfolgt, kostenpflichtig oder kostenlos ist.

Externe Personen, die der Institution KI-basierte Ergebnisse liefern, müssen ausdrücklich darauf hinweisen, dass und wie KI eingesetzt wurde. Dies gilt auch dann, wenn KI nur unterstützend genutzt wurde.

3. KI-Systeme und ihre Risikostufen (EU AI Act)

Der EU AI Act klassifiziert KI-Systeme nach ihrem Risikopotenzial. Diese Einstufung bestimmt, welche Anforderungen gelten:

Risikostufe	Beschreibung	Beispiele
verboten	inakzeptables Risiko, gesetzlich untersagt	Social Scoring, Manipulation, biometrische Massenüberwachung, Gesichtsdatenbanken aus dem Internet
hoch	erhebliche Risiken, strenge Auflagen	Systeme zur Bewerbungsbeurteilung, Lernbewertung, Sicherheitsanalyse
begrenzt	Interaktion mit Menschen oder Deep-fakes	Chatbots, Emotionserkennung, synthetische Medien
minimal	geringes Risiko, wenig Einschränkungen	Spamfilter, Texterstellung, Übersetzung, Bildgenerierung

4. Die sieben Grundprinzipien

4.1 Rechtliche Rahmenbedingungen

Alle Aktivitäten im Zusammenhang mit KI müssen den gesetzlichen Vorgaben, dem Arbeitsvertrag und den damit verbundenen Geheimhaltungspflichten entsprechen.

Urheberrecht

Das österreichische Urheberrechtsgesetz sieht nur natürliche Personen als Urheber vor – KI kann selbst kein Urheber sein. Die der KI zugeführten und verarbeiteten Daten können jedoch urheberrechtlich geschützt sein. Nicht alle öffentlich verfügbaren Inhalte dürfen auch verwendet werden.

- Eingaben (Texte, Bilder, Datenbanken) auf Urheberrechtsfreiheit prüfen
- erforderliche Berechtigungen und Lizenzvereinbarungen sicherstellen
- AGB der verwendeten KI-Anwendung auf Nutzungsrechte der Outputs prüfen
- Marketing-Materialien und Designs auf bestehende Schutzrechte prüfen

Beschaffung und Freigabe

- nur genehmigte KI-Systeme einsetzen (interne Tool-Liste)
- neue KI-Tools erst nach Freigabe durch Leitung und IT/Datenschutz beschaffen
- IT-Verantwortliche, Datenschutzbeauftragte und Informationssicherheitsbeauftragte frühzeitig einbinden
- rechtliche und sicherheitstechnische Prüfung abwarten, bevor ein System produktiv genutzt wird

Praxistipps

- Faustregel: Wäre es problematisch, wenn diese Information zusammen mit Ihrem Namen öffentlich im Internet erschiene? Dann nicht in ein externes KI-System eingeben.
- Bei Unsicherheit: vor Nutzung Rücksprache mit dem Datenschutzbeauftragten halten

4.2 Menschliche Kontrolle und Verantwortung

KI-Systeme sind leistungsfähig, aber nicht fehlerfrei. Sie basieren auf statistischen Wahrscheinlichkeiten, nicht auf echtem Fachwissen oder Werturteilen. Menschliche Fähigkeiten wie Wahrnehmung, Intuition, Empathie und ethisches Urteil bleiben unverzichtbar.

KI kann derzeit nicht oder nur sehr eingeschränkt:

- empathisch sein oder Emotionen wirklich nachempfinden
- Sarkasmus, Ironie und Humor verlässlich verstehen
- moralische und ethische Urteile treffen
- Fakten zuverlässig von Lügen unterscheiden
- Halluzinationen vermeiden – also das Erfinden plausibel klingender, aber falscher Informationen

Pflicht zur Prüfung

KI-basierte Inhalte dürfen erst dann übernommen werden, wenn sie von mindestens einer qualifizierten Person auf Richtigkeit, Faktentreue, Fairness und Kontext überprüft wurden. Die finale Verantwortung liegt immer bei den Mitarbeitenden.

Praktische Leitfragen bei der Verwendung von KI-Ergebnissen:

- Ist das Ergebnis für den Verwendungszweck relevant und trifft es das Ziel?
- Ist es konsistent mit dem aktuellen Stand des Fachwissens?
- Gibt es nachprüfbare Quellen für die generierten Informationen?
- Enthält das Ergebnis Verzerrungen oder diskriminierende Elemente?
- Wurde der Inhalt einer Plausibilitätskontrolle unterzogen?

4.3 Transparenz und Kennzeichnung

KI-basierte Technologien sind so weit fortgeschritten, dass es schwierig werden kann zu unterscheiden, ob ein Inhalt von einem Menschen oder einer KI erstellt wurde. Transparenz schützt das Vertrauen der Öffentlichkeit, Besucher:innen und Mitarbeitenden.

Eine Kennzeichnung gemäß Kennzeichnungspflicht (EU AI Act, Art. 50) ist verpflichtend, wenn:

- KI-generierte Inhalte können eine täuschende Wirkung haben, insbesondere wenn sie als authentische Werke oder Aussagen einer Person erscheinen
- synthetische Medien erstellt wurden (Deepfakes, KI-generierte Stimmen, digitale Kunstwerke), die nicht offensichtlich als künstlich erkennbar sind
- Texte generiert wurden, die die Öffentlichkeit über Angelegenheiten von öffentlichem Interesse informieren sollen
- ein KI-Chatbot direkt mit Personen interagiert (muss vor Interaktionsbeginn kommuniziert werden)

Keine Kennzeichnungspflicht

Wenn KI lediglich assistierend eingesetzt wurde, z. B. für Rechtschreibkorrekturen, rein interne Recherchen oder Zusammenfassungen, die noch erheblich überarbeitet wurden.

Mögliche Kennzeichnungsformen

- Im Text oder Bild: „Dieser Inhalt wurde mit KI-Unterstützung erstellt und durch [Name] geprüft.“
- Metadaten in digitalen Dateien ergänzen
- Wasserzeichen oder visuelle Overlays bei Bild-/Videoinhalten
- Hinweis in Fußzeile, Bildunterschrift oder E-Mail-Signatur

4.4 Datenschutz, Informationssicherheit und Urheberrecht

Der Einsatz von KI-Tools, die auf externen Datenpools basieren, birgt das Risiko, dass eingegebene Daten von Dritten eingesehen oder für das Training des Modells verwendet werden.

Verbotene Eingaben (ohne explizite Freigabe):

- personenbezogene Daten: Name, Adresse, E-Mail, Geburtsdatum, Sozialversicherungsnummer, IP-Adressen, biometrische Daten
- vertrauliche Unternehmensinformationen: Finanzplanung, strategische Pläne, unveröffentlichte Ausstellungskonzepte
- Geschäfts- und Betriebsgeheimnisse
- nicht veröffentlichte Forschungsergebnisse oder wissenschaftliche Daten
- Passwörter, Zugangsdaten, interne Systemdaten
- urheberrechtlich geschützte Drittinhalte ohne Berechtigung

Datenklassifizierung als Orientierung

Kategorie	Beispiele	KI-Eingabe erlaubt?
öffentliche Daten	bereits veröffentlichte Webinhalte, Pressemitteilungen	ja
interne Daten	interne Dokumentationen, Veranstaltungsplanungen	eingeschränkt
vertrauliche Daten	Finanzzahlen, strategische Planungen, Personalangelegenheiten, unveröffentlichte Forschung	nur nach Freigabe
streng vertraulich	personenbezogene Mitarbeitendendaten, Rechtsdokumente	nein

Technische Schutzmaßnahmen

- Datenanonymisierung oder Pseudonymisierung vor KI-Verarbeitung
- interne oder europäisch gehostete KI-Systeme bevorzugen
- bei Datenverarbeitung außerhalb des EWR: DSGVO Art. 44-50 beachten
- Opt-out bei kostenfreien Tools aktivieren (Training der Modelle mit Eingaben)

4.5 Sicherheit und Qualität der Daten

KI-Systeme basieren auf Wahrscheinlichkeitsmodellen, nicht auf echtem Fachwissen. Sie haben keinen Zugriff auf Echtzeitinformationen und sind nur so zuverlässig wie die Daten, mit denen sie trainiert wurden. Das Phänomen der Halluzination – das Erfinden plausibel klingender, aber falscher Informationen – ist ein systemimmanentes Risiko.

Qualitätssicherung in der Praxis

- alle KI-Ergebnisse vor Verwendung inhaltlich prüfen und verifizieren
- Quellen, die KI nennt, eigenständig auf Existenz und Korrektheit überprüfen
- mit Fachliteratur, Expert:innenwissen und vertrauenswürdigen Quellen abgleichen
- Desinformation aktiv vermeiden: Plausibilitätskontrolle vor jeder Veröffentlichung
- KI-generierten Code auf Fehler, Sicherheitslücken und Effizienz prüfen
- besondere Sorgfalt bei wissenschaftlichen Inhalten, historischen Fakten und Ausstellungstexten

Tool-Auswahl

Generalistische Modelle (z.B. ChatGPT, Gemini) sind nicht für wissenschaftliche Präzision oder historische Recherche optimiert. Für spezialisierte Aufgaben eignen sich dedizierte Fachanwendungen besser. Die Wahl des richtigen Tools ist Teil der Qualitätssicherung.

4.6 Ethik, Vielfalt, Inklusion und Nachhaltigkeit

Ethische Grundsätze

KI-Systeme können nicht selbst entscheiden, was ethisch richtig oder falsch ist. Verzerrungen entstehen durch fehlerhafte oder unausgewogene Trainingsdaten und können zu Diskriminierung führen.

- KI-Inhalte regelmäßig auf Verzerrungen (Bias) und Diskriminierung überprüfen
- Stereotype und problematische Darstellungen erkennen und korrigieren
- Prompts bewusst steuern, um inklusive und faire Ergebnisse zu fördern
- Fachexpertise zur Prüfung diskriminierungssensibler Inhalte einbeziehen

Inklusion

Alle Zielgruppen sollen gleichermaßen von KI-gestützten Angeboten profitieren können, unabhängig von Herkunft, Geschlecht, Alter, Behinderung oder Sprachkenntnissen.

- KI zur Verbesserung der Barrierefreiheit nutzen (z. B. automatische Untertitel, Leichte Sprache)
- Mehrdeutigkeiten unterschiedlicher Kulturen und Gesellschaftsgruppen beachten
- Inklusion als aktives Ziel in KI-Projekten verankern

Nachhaltigkeit und Ressourceneffizienz

- KI nur dann einsetzen, wenn ein klarer Mehrwert besteht
- einfachere, energieeffizientere Methoden zuerst prüfen
- passendes KI-System für die Aufgabe wählen – komplexe Modelle nur wenn nötig
- Energieverbrauch und ökologischen Fußabdruck bei der Tool-Auswahl bedenken
- KI-Einsatz im Sinne der UN Sustainable Development Goals gestalten

4.7 Aus- und Weiterbildung

KI-Anwendungen sind auf das Wissen und die Expertise der Anwender:innen angewiesen. Ähnlich wie bei einem neuen Werkzeug muss der kompetente Umgang mit KI erst erlernt werden.

Verpflichtungen der Mitarbeitenden

- vor der erstmaligen KI-Nutzung ein KI-Kompetenztraining absolvieren
- sich über aktuelle Entwicklungen und institutionsinterne Richtlinien auf dem Laufenden halten
- gelerntes verantwortungsbewusst im Arbeitsalltag anwenden
- Erfahrungen und Erkenntnisse im Team und überinstitutionell teilen

Verantwortung der Institution

Gemäß EU AI Act (Art. 4) sind Institutionen verpflichtet, ausreichende KI-Kompetenz ihrer Mitarbeitenden sicherzustellen.

Dazu gehören:

- regelmäßige Schulungsangebote und E-Learning-Module
- interne Anlaufstellen für KI-Fragen (z. B. Datenschutz-Support, IT, KI-Arbeitsgruppen)
- Förderung des institutionsübergreifenden Austauschs
- Schulungsinhalte: KI-Grundlagen, Datenschutz, Urheberrecht, Ethik, Bias, Informationssicherheit

5. Freigabeprozess für KI-Tools

Bevor ein neues KI-System beschafft oder erstmals dienstlich genutzt wird, ist folgender Prozess einzuhalten:

Schritt 1 – Bedarf identifizieren: Welches Problem soll die KI lösen? Gibt es einfachere Alternativen?

Schritt 2 – Abstimmung mit der Abteilungsleitung und Information an IT / Datenschutz

Schritt 3 – Vorprüfung durch IT und Datenschutzbeauftragten
(technisch, rechtlich, sicherheitstechnisch)

Schritt 4 – Entscheidung durch die Direktion

Schritt 5 – Aufnahme in die offizielle Tool-Liste; Schulung der Nutzer:innen vor der Erstnutzung

Schritt 6 – Regelmäßige Evaluierung und Aktualisierung der Freigabe

Bei der Bewertung eines neuen Tools sind folgende Kriterien besonders relevant:

- physischer Serverstandort und rechtlicher Firmensitz (EU-Lösungen bevorzugen)
- Datenschutzkonformität und Opt-out-Möglichkeiten für Training
- Möglichkeit zum Selbst-Hosting oder internen Betrieb
- Risikoeinstufung gemäß EU AI Act

6. Spezifische Anwendungsfälle und Hinweise

6.1 Texterstellung und Übersetzung

- generierte Texte vor Veröffentlichung fachlich prüfen und überarbeiten
- keine internen, vertraulichen Texte (Verträge, Protokolle, Personalunterlagen) vollständig über externe KI-Systeme übersetzen
- urheberrechtlich geschützte Texte nicht ohne Genehmigung als Eingabe verwenden
- Übersetzungen stichprobenartig durch menschliche Lektor:innen prüfen

6.2 Bild-, Audio- und Videoinhalte

- KI-generierte Bilder mit erkennbaren Personen erfordern vorherige Einwilligung
- keine Täuschung des Publikums über den Einsatz von KI bei Kunstproduktionen
- Deepfakes und KI-veränderte Bewegtbilder sind klar zu kennzeichnen
- bei historischen Dokumenten: Original-Aussage nicht durch KI verändern (nur Retusche, Farbkorrektur u. a. erlaubt)
- KI-generierte Stimmen für Audioinhalte: sorgfältige Auswahl, keine Kinderstimmen bei sensiblen Themen

6.3 Sammlungsmanagement und Forschung

- KI-Ergebnisse sind als Arbeitsgrundlagen zu verstehen und müssen fachlich geprüft werden
- unveröffentlichte Forschungsergebnisse und nicht freigegebene Sammlungsdaten nicht in externe KI-Systeme eingeben

- KI-gestützte Metadatengenerierung und Objektbeschreibungen auf wissenschaftliche Richtigkeit prüfen
- Quellenkritik besonders bei historischen Rekonstruktionen anwenden

6.4 Marketing und Kommunikation

- KI-generierte Marketinginhalte können Markenrechte verletzen: prüfen!
- zwischen aufmerksamkeitsorientierten Marketinginhalten und wissenschaftlich gesicherten Inhalten unterscheiden
- personalisierte Werbung: Datenschutzrechtliche Anforderungen beachten

6.5 Interne Prozesse und Verwaltung

- KI-gestützte Besprechungsprotokolle: bei vertraulichen oder personenbezogenen Gesprächen besondere Sorgfalt und vorige Zustimmung der Anwesenden
- KI zur Optimierung von Arbeitsabläufen nutzen, aber keine Entscheidungen über Personen delegieren
- KI-generierte Auswertungen und Kennzahlen prüfen, bevor sie Grundlage für Entscheidungen werden

7. Verbotene Anwendungen

Absolutes Verbot

Folgende Anwendungen sind gemäß EU AI Act generell verboten und dürfen unter keinen Umständen eingesetzt werden:

- Social Scoring-Systeme: KI zur Bewertung oder Einschätzung von Personen anhand ihres Sozialverhaltens
- Manipulation: KI-Systeme, die darauf ausgelegt sind, Personen durch unterschwellige Techniken zu manipulieren
- Verhaltensauswertung: Systeme zur Analyse des Verhaltens natürlicher Personen ohne deren Wissen und Einwilligung
- Biometrische Kategorisierung: Systeme zur Kategorisierung von Personen anhand sensibler Merkmale (Religion, politische Ansichten, Gesundheit, Sexualität, Ethnie)
- Gesichtserkennungs-Datenbanken: Systeme, die Datenbanken zur Gesichtserkennung mit Bildern aus dem Internet oder Überwachungsmaterial aufbauen
- Echtzeit-Fernüberwachung im öffentlichen Raum (nur mit strengen Ausnahmeregelungen und behördlicher Genehmigung)

8. Checkliste für die tägliche KI-Nutzung

Vor jeder KI-Nutzung im dienstlichen Kontext:

- Genehmigung: Ist das verwendete KI-Tool auf der internen genehmigten Liste?
- Datenschutz: Enthält meine Eingabe personenbezogene oder vertrauliche Daten? Wenn ja – anonymisieren oder Tool nicht verwenden.

- Urheberrecht: Sind meine Eingaben urheberrechtsfrei oder habe ich die notwendigen Berechtigungen?
- Qualität: Habe ich das Ergebnis fachlich geprüft und mit verlässlichen Quellen verglichen?
- Kennzeichnung: Muss ich den KI-Einsatz kennzeichnen? (substanzieller Beitrag der KI = Kennzeichnung erforderlich)
- Ethik: Enthält das Ergebnis Verzerrungen, Stereotypen oder diskriminierende Elemente?
- Mehrwert: Bietet der KI-Einsatz einen klaren Mehrwert? Gibt es einfachere Alternativen?
- Nachhaltigkeit: Ist der Energieeinsatz für diesen Zweck verhältnismäßig?

9. Verantwortlichkeiten und Rollen

Rolle	Verantwortlichkeiten
Mitarbeitende	Regelkonforme, sorgfältige Nutzung; Ergebnisse prüfen; Meldepflichten einhalten; Schulungen absolvieren; bei Unsicherheit fragen
direkte:r Vorgesetzte:r	Sicherstellen, dass Mitarbeitende geschult sind; KI-Bedarfe koordinieren; Nutzung beobachten und steuern
IT / Datenschutz	Technische Prüfung neuer Tools; Sicherheitsfreigaben erteilen; Datenschutzkonformität sicherstellen; Anlaufstelle für Fragen
Direktion	Freigabe von KI-Tools; strategische Verankerung der Richtlinie; Schulungsressourcen bereitstellen; Richtlinie regelmäßig aktualisieren
KI-Arbeitsgruppe (wo vorhanden)	Bestandsaufnahme der KI-Nutzung; Diskussion von Chancen und Risiken; Vorbereitung von Entscheidungsgrundlagen; Pilotprojekte begleiten

10. KI-Anwendungsregister / KI-Tool-Liste

Ein KI-Anwendungsregister sollte alle eingesetzten, getesteten oder geplanten KI-Anwendungen erfassen. Es dient der Transparenz, Steuerung und Risikobewertung.

Folgende Daten sollten in dem Register erfasst und jährlich geprüft werden:

Name des Tools oder der Anwendung	
Zweck der Nutzung	
zuständige Abteilung	
verantwortliche Person	
interne oder öffentliche Nutzung	
verarbeitete Datenarten	
sensible Daten	ja/nein
personenbezogene Daten	ja/nein
externer Anbieter	ja/nein
Datenschutzprüfung erforderlich	ja/nein
fachliche Prüfung erforderlich	ja/nein

Risikoeinschätzung	
Freigabestatus	
Datum der letzten Prüfung	
nächste Evaluierung	

11. Monitoring, Feedback und Weiterentwicklung

Die KI-Richtlinie soll regelmäßig überprüft und weiterentwickelt werden.

Eine Überprüfung sollte mindestens einmal jährlich erfolgen sowie zusätzlich bei:

- neuen rechtlichen Anforderungen
- Einführung neuer KI-Systeme
- sicherheitsrelevanten Vorfällen
- Beschwerden oder Fehlfunktionen
- wesentlichen Änderungen in Arbeitsprozessen
- neuen Erkenntnissen zu Risiken oder Bias
- Änderungen bei Anbieterbedingungen

Für produktiv genutzte KI-Systeme empfiehlt sich ein einfaches Monitoring:

- Werden Ergebnisse fachlich geprüft?
- Gibt es wiederkehrende Fehler?
- Treten diskriminierende oder verzerrende Antworten auf?
- Werden Kennzeichnungspflichten eingehalten?
- Gibt es Datenschutz- oder Sicherheitsvorfälle?
- Sind Nutzer:innen ausreichend informiert?
- Sind Schulungen aktuell?
- Ist das Tool weiterhin geeignet?

12. Glossar

Begriff	Definition
Bias (Verzerrung)	Systematische Fehler in KI-Ergebnissen durch einseitige Trainingsdaten, die zu ungerechten oder diskriminierenden Aussagen führen können.
Deepfake	Realistisch wirkende, KI-generierte oder -manipulierte Medieninhalte, die Personen oder Ereignisse verfälschen
DSGVO	Datenschutz-Grundverordnung (EU) 2016/679 – regelt den Schutz personenbezogener Daten in der EU
EU AI Act	EU-Verordnung 2024/1689 zur Regulierung von KI-Systemen mit Risiko-Klassifizierung, Transparenzpflichten und Verboten
Generative KI	KI-Systeme, die neue Inhalte (Text, Bild, Audio, Video) auf Basis von Trainingsdaten erzeugen

Halluzination	KI-generierte Falschaussagen oder erfundene Informationen, die plausibel klingen, aber nicht der Realität entsprechen
Human in the Loop	Prinzip, bei dem Menschen im KI-Prozess beteiligt bleiben und die finale Entscheidung treffen
KI-Kompetenz	Fähigkeit, KI-Systeme verantwortungsvoll zu verstehen, zu nutzen und zu bewerten personenbezogene Daten Alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen
Pseudonymisierung	Verarbeitung personenbezogener Daten so, dass Identitäten nur mit separaten Zusatzinformationen wiederhergestellt werden können
Trainingsdaten	Daten, mit denen ein KI-Modell trainiert wird. Qualität und Ausgewogenheit sind entscheidend für faire Ergebnisse.
Transparenz	Offenlegung, wann und wie KI eingesetzt wird, einschließlich Zweck, Funktionsweise und Grenzen