



## Leitlinie

Prager Straße 287  
A-1210 Wien  
Tel.: +43-1-319 60 43 410  
Fax: +43-1-319 60 42 499  
Email: [office@adra.at](mailto:office@adra.at)

# KI-Richtlinie für den ethisch-christlichen Einsatz Künstlicher Intelligenz

ADRA Österreich

	Datum:	Unterschrift Vorsitzender
Beschluss Vorstand		R. Schwab
Revision 1		
Revision 2		
Revision 3		
Revision 4		

### Inkrafttreten: 01. August 2024

Die vorliegende KI-Richtlinie ist eine verbindliche Entscheidung des ADRA Österreich Vorstandes und gilt für alle Mitarbeiterinnen und Mitarbeiter sowie nationale und internationale Partner. Jeder beteiligte Mitarbeiter ist für deren Einhaltung verantwortlich.

## INHALTSVERZEICHNIS

1	EINLEITUNG .....	3
1.1	Präambel.....	3
1.2	Geltungsbereich.....	3
1.3	Definitionen .....	3
2	Respekt und Schutz der Menschenwürde.....	3
3	Transparenz und Rechenschaftspflicht .....	4
4	Datenschutz und Sicherheit .....	4
5	Verantwortung und ethisch-christliches Verhalten .....	4
6	Gemeinwohl und soziale Gerechtigkeit .....	5
7	Nachhaltigkeit und Umweltbewusstsein .....	5
8	Glaubwürdigkeit gegenüber Spendern.....	5
9	Dialog und Lernkultur .....	6
10	SCHLUSBESTIMMUNG .....	6
11	Besondere Hinweise für die Verwendung von ChatGPT .....	6
11.1	Datenschutz und Inhaltsschutz .....	6
11.2	Umgang mit Ergebnissen.....	6
11.3	Transparenz und Urheberrecht.....	7
11.4	Nutzung durch Mitarbeitende .....	7
11.5	Beispiele für zulässige Verwendung bei ADRA .....	7
12	ANHANG: Verzeichnis KI-Tools bei ADRA Österreich.....	8

# 1 EINLEITUNG

---

## 1.1 Präambel

Diese KI-Richtlinie dient als ethisch-christlicher Orientierungsrahmen für den verantwortungsvollen Einsatz von Künstlicher Intelligenz (KI) bei ADRA Österreich. Als Teil einer internationalen christlichen Hilfsorganisation verpflichten wir uns, technologische Innovationen wie KI stets im Dienst des Menschen, der Gemeinschaft und der Schöpfung zu nutzen.

Ziel ist es, KI so einzusetzen, dass sie den Schwächsten dient, Hilfe effizienter macht und dabei unsere Werte – Menschenwürde, Transparenz, Integrität, Solidarität und Gerechtigkeit – wahrt.

## 1.2 Geltungsbereich

Diese Richtlinie gilt für:

- Alle Mitarbeitenden von ADRA Österreich (Vollzeit, Teilzeit, Ehrenamtliche)
- Den Vorstand von ADRA Österreich
- Vertraglich gebundene Dienstleister im Auftrag von ADRA Österreich
- Internationale ADRA-Partner (englische Version in Erstellung)

## 1.3 Definitionen

**Künstliche Intelligenz (KI):** Ein Teilgebiet der Informatik, das Maschinen in die Lage versetzt, Aufgaben zu übernehmen, die menschliche Intelligenz erfordern – z. B. Spracherkennung, Bilderkennung, Entscheidungsfindung, Textgenerierung.

**KI-Tools:** Softwarelösungen, Plattformen oder Systeme, die auf KI basieren (z. B. ChatGPT, Canva AI, Adobe Firefly). Eine aktuelle Liste befindet sich im Anhang dieser Richtlinie.

**KI-generierte Inhalte:** Texte, Bilder, Grafiken, Videos oder Audiodateien, die ganz oder teilweise durch KI erstellt wurden.

# 2 Respekt und Schutz der Menschenwürde

---

**Menschenzentrierung:** AI-Systeme müssen so gestaltet und eingesetzt werden, dass sie die Menschenwürde und die Rechte aller betroffenen Personen respektieren und schützen. Dies beinhaltet die Wahrung der Autonomie und Würde aller Individuen.

**Nicht-Diskriminierung:** Der Einsatz von AI darf nicht zu Diskriminierung oder Vorurteilen aufgrund von Geschlecht, ethnischer Herkunft, Alter, Religion, Behinderung oder anderen persönlichen Merkmalen führen. Algorithmen und Modelle müssen regelmäßig überprüft und angepasst werden, um Vorurteile zu minimieren und Fairness zu gewährleisten.

### 3 **Transparenz und Rechenschaftspflicht**

---

**Offenlegung:** Wir verpflichten uns, offen über den Einsatz von AI-Systemen zu informieren. Dies beinhaltet eine klare und verständliche Erklärung darüber, wie und warum AI-Systeme genutzt werden, welche Daten verwendet werden und welche potenziellen Auswirkungen sie haben können. Werden AI generierte Bilder oder Grafiken verwendet, so ist das anzugeben bzw. als Bildnachweis z.B. Canva oder andere Tools, welche das Bild erzeugt haben.

**Rechenschaft:** Alle AI-Systeme müssen nachvollziehbar und auditierbar sein. Entscheidungen, die durch AI getroffen werden, sollen überprüfbar und erklärbar sein, um sicherzustellen, dass sie im Einklang mit unseren ethischen Prinzipien stehen.

### 4 **Datenschutz und Sicherheit**

---

**Datenschutz:** Der Schutz der Privatsphäre und personenbezogener Daten hat höchste Priorität. Wir sammeln und verarbeiten nur die Daten, die unbedingt notwendig sind, und sorgen für deren sichere Speicherung und Verarbeitung in Übereinstimmung mit den geltenden Datenschutzgesetzen und -richtlinien. Es dürfen keine Personenbezogene Daten ((Name, Fotos, Fallgeschichten) in öffentlich zugängliche AI-Tools wie z.B. ChatGPT eingegeben werden.

**Sicherheit:** AI-Systeme müssen sicher und robust gegen Missbrauch, Manipulation und Angriffe sein. Wir verpflichten uns, kontinuierlich Maßnahmen zur Sicherstellung der Sicherheit unserer AI-Systeme zu ergreifen, einschließlich regelmäßiger Sicherheitsüberprüfungen und -updates. Eingesetzte KI-Systeme müssen vor externem Missbrauch und interner Manipulation geschützt sein. Es finden regelmäßige Sicherheits- und Ethik-Audits statt.

Ergänzend zu dieser KI-Richtlinie sind die bestehenden Datenschutzrichtlinien zu beachten, insbesondere die Richtlinie Datenschutz, Prozess Datenschutz und Verfahrensverzeichnis.

### 5 **Verantwortung und ethisch-christliches Verhalten**

---

**Verantwortung:** Jede Person innerhalb der Organisation, die an der Entwicklung, Implementierung oder Nutzung von AI beteiligt ist, trägt die Verantwortung für den ethisch-christlichen Einsatz dieser Technologie. Dies umfasst die Pflicht, mögliche ethische Herausforderungen zu erkennen und zu adressieren. Es ist ein:e KI-Verantwortliche:r bei ADRA Österreich zu benennen, der oder die Monitoring, Schulungen und Weiterentwicklung der Richtlinie koordiniert.

**Ethik-Training:** Alle Mitarbeiterinnen und Mitarbeiter von ADRA sollen zukünftig regelmäßig in ethischen Fragen und den Auswirkungen von AI geschult. Diese Schulungen sollen das Bewusstsein für die Verantwortung und die ethischen Herausforderungen im Zusammenhang mit AI stärken und sicherstellen, dass alle Beteiligten die Prinzipien dieses Kodex verstehen und einhalten. Die aktuell bevorzugte Ausbildungsplattform bei ADRA ist Udemy. Alle Mitarbeiterinnen und Mitarbeiter sind dazu aufgerufen zumindest alle zwei Jahre hierzu einen geeigneten Trainingskurs zu absolvieren. Neue Mitarbeitende erhalten eine Einweisung in diese Richtlinie.

## 6 Gemeinwohl und soziale Gerechtigkeit

---

**Gemeinwohl:** Der Einsatz von AI soll dem Wohle der Gemeinschaft dienen und positive soziale Auswirkungen haben. AI-Projekte sollen so konzipiert und durchgeführt werden, dass sie zur Lösung gesellschaftlicher Probleme beitragen und das allgemeine Wohl fördern. Barrierefreie Kommunikation ist hierzu sicherzustellen (siehe ADRA Leitfaden zur Barrierefreiheit).

**Zugangsgerechtigkeit:** Wir setzen uns dafür ein, dass die Vorteile von AI allen zugänglich gemacht werden. Niemand darf aufgrund von sozioökonomischen, geografischen oder anderen Barrieren vom Zugang zu den positiven Auswirkungen von AI ausgeschlossen werden. Dies beinhaltet auch die Bemühungen, die digitale Kluft zu verringern und allen Gemeinschaftsmitgliedern die notwendigen Fähigkeiten und Ressourcen zur Nutzung von AI zur Verfügung zu stellen.

## 7 Nachhaltigkeit und Umweltbewusstsein

---

**Nachhaltiger Einsatz:** AI-Systeme sollen so gestaltet und eingesetzt werden, dass sie nachhaltig sind und die Umwelt so wenig wie möglich belasten. Wir verpflichten uns, die ökologischen Auswirkungen unserer AI-Aktivitäten zu bewerten und zu minimieren.

**Umweltbewusstsein:** Wir verpflichten uns, die ökologischen Auswirkungen unserer AI-Systeme zu minimieren und umweltfreundliche Praktiken zu fördern. Dies umfasst die Reduktion des Energieverbrauchs und die Förderung von Recycling und Wiederverwendung in unseren AI-Projekten.

Folgende Parameter sind zu berücksichtigen:

- KI-Systeme sollen mit möglichst geringem Energieaufwand betrieben werden.
- Bei der Tool-Auswahl wird auf energieeffiziente Anbieter geachtet.
- Cloud-Lösungen sollen Server mit nachweislicher Nachhaltigkeit bevorzugen.

## 8 Glaubwürdigkeit gegenüber Spendern

---

**Ehrliche Kommunikation:** Wir verpflichten uns, ehrlich und transparent über den Einsatz von AI und die damit erzielten Ergebnisse zu kommunizieren. KI-generierte Inhalte dürfen niemals eine falsche Realität erzeugen. Insbesondere bei der Verwendung von AI-generierten Bildern müssen wir klarstellen, ob diese Bilder authentisch oder künstlich erstellt wurden. ADRA verwendet nur in begründeten Ausnahmefällen Stockfotos oder KI generierte Fotos. Die Authentizität hat höchste Priorität. Wir verwenden lieber ein Originalfoto in weniger guter Qualität als ein realitätsfernen Stockfoto.

**ADRA verwendet keine erfundenen Erfolgsgeschichten oder Notlagen.**

**Integrität:** Der Einsatz von AI-generierten Bildern darf nicht dazu verwendet werden, die Realität zu verzerren oder die Spender in die Irre zu führen. Alle visuellen Darstellungen müssen wahrheitsgetreu sein und die tatsächlichen Bedingungen und Aktivitäten von ADRA und ADRA's Partnern widerspiegeln.

Es dürfen keine von KI erzeugten bzw. erfundenen Need- oder Successstory verwendet werden. Noch dürfen uns übermittelte Bilder oder Geschichten durch KI so verändert werden, dass sie nicht mehr der

Realität entsprechen. Es dürfen rote Augen, störende Haarsträhnen oder Pickel entfernt werden, aber keine Zähne gebleicht oder Emotionen, insbesondere durch den Gesichtsausdruck verändert werden, z.B. darf aus einem traurigen Kind, kein freudestrahlendes

## 9 Dialog und Lernkultur

---

**Offener Dialog:** Wir fördern den offenen Dialog mit der Öffentlichkeit, Stakeholdern und Fachleuten, um gemeinsam ethische Standards für AI zu entwickeln und kontinuierlich zu verbessern. Dies umfasst regelmäßige Konsultationen und Feedback-Schleifen, um sicherzustellen, dass unsere KI-Praktiken den Erwartungen und Bedürfnissen der Gemeinschaft entsprechen.

**Kontinuierliche Verbesserung:** Wir überprüfen regelmäßig unsere AI-Richtlinien und -Praktiken, um sicherzustellen, dass sie den neuesten ethischen und technologischen Standards entsprechen. Dies beinhaltet auch die Bereitschaft, aus Fehlern zu lernen und unsere Ansätze kontinuierlich zu verbessern. Uns ist bewusst, dass KI-Systeme noch am Anfang ihrer Entwicklung stehen und sich fortlaufend wandeln und sich so Risikoeinschätzungen immer wieder verändern werden. Erfahrungen und Learnings aus Projekten mit KI-Einsatz werden dokumentiert.

## 10 SCHLUSBESTIMMUNG

---

Dieser Richtlinie ist verbindlich für alle Mitglieder von ADRA Österreich und den internationalen ADRA Partnerbüros. Verstöße gegen diese Richtlinie werden ernst genommen und entsprechend behandelt. Unser Ziel ist es, durch den verantwortungsvollen Einsatz von AI das Vertrauen der Gemeinschaft und unserer Spender zu gewinnen und zu erhalten. Wir glauben, dass ein ethisch-christlicher Ansatz beim Einsatz von AI entscheidend ist, um positive soziale Veränderungen zu bewirken und das Wohl aller Beteiligten zu fördern.

Diese Richtlinie wurde teilweise unter Einsatz von ChatGPT erstellt. Es ist eine Prüfung durch das Research Institute in Wien geplant.

## 11 Besondere Hinweise für die Verwendung von ChatGPT

---

ChatGPT ist eines der meistgenutzten KI-Tools bei ADRA Österreich. Um die sichere, transparente und ethisch verantwortungsvolle Nutzung zu gewährleisten, gelten folgende besondere Richtlinien:

### 11.1 Datenschutz und Inhaltsschutz

- **Keine personenbezogenen Daten eingeben:** Namen, Adressen, E-Mail-Kontakte, Fallgeschichten mit identifizierbaren Personen oder sensible Informationen (z. B. zu Gesundheit, Religion, Finanzen) dürfen nicht in ChatGPT eingegeben werden.
- **Keine Originaldokumente hochladen,** die persönliche oder sensible Informationen enthalten.

### 11.2 Umgang mit Ergebnissen

- **Inhalte immer prüfen:** Die von ChatGPT generierten Texte sind Vorschläge und können sachlich oder rechtlich falsch sein. Alle Inhalte sind kritisch zu prüfen, insbesondere bei rechtlichen oder

projektbezogenen Themen. Idealerweise im Prompt eingeben, das ChatGPT die Quellen hinzufügen soll.

- **Kein Ersatz für fachliche Expertise:** ChatGPT darf keine fachliche Beratung (z. B. juristische Einschätzungen, Projektentscheidungen) ersetzen, sondern lediglich als Unterstützung genutzt werden.
- **Keine automatisierte Entscheidungsfindung:** ChatGPT soll nicht verwendet werden, um Entscheidungen über Menschen oder Projekte zu treffen. Es dient der Textunterstützung, nicht der Projektsteuerung.

### 11.3 Transparenz und Urheberrecht

- **Kennzeichnung von KI-Inhalten:** Bei externen Publikationen die überwiegend mit ChatGPT erstellt wurden, sollte vermerkt werden: „Erstellt mit Unterstützung durch KI (ChatGPT)“.
- **Urheberrecht beachten:** ChatGPT kann Ideen liefern, aber keine Garantie für urheberrechtsfreie Inhalte geben. Alle generierten Texte, Slogans, Gedichte etc. müssen vor Veröffentlichung eigenständig geprüft und ggf. redigiert werden.

### 11.4 Nutzung durch Mitarbeitende

- **Verantwortungsbewusster Einsatz:** Die Nutzung von ChatGPT soll stets gemäß dieser Richtlinie und ethischen Gesichtspunkten erfolgen.
- **Schulung und Aufklärung:** Jede:r Nutzer:in soll wissen, wie das Tool funktioniert, welche Daten verarbeitet werden und wie man es ethisch korrekt einsetzt. Hierzu werden regelmäßige Kurzschulungen angeboten.
- **Keine Delegation sensibler Kommunikation:** Öffentlichkeitswirksame oder sensible Texte (z. B. Presseinformationen, Briefe an Behörden, Texte für Begünstigte) dürfen nicht ungeprüft über ChatGPT erstellt oder versendet werden.

### 11.5 Beispiele für zulässige Verwendung bei ADRA

- ✓ Ideen für Social Media Posts oder Überschriften
- ✓ Vorschläge für Formulierungen in Projektberichten
- ✓ Unterstützung beim Schreiben von Infotexten
- ✓ Verbesserung des Schreibstils oder Umformulierung in leicht verständliche Sprache und bei Übersetzung von fremdsprachlichen Texten
- ✓ Brainstorming für neue Kampagnenideen oder Projekttitle
- ✗ Texte über reale Personen erstellen oder interpretieren
- ✗ Erfolgs- oder Needgeschichten erfinden
- ✗ Entscheidungen über die Auswahl von Begünstigten durch ChatGPT vorstrukturieren
- ✗ Informationen über ADRA-Partner oder sensible Projektdaten eingeben
- ✗ Verwendung als alleinige Quelle bei rechtliche Einschätzungen

## 12 ANHANG: Verzeichnis KI-Tools bei ADRA Österreich

Version vom: 05.06.2025

Name der Anwendung	Ort der Anwendung	Beschreibung	Verwendete Technologien	Anwendungsbereich	Datenkategorien	Gefahren	Verantwortlicher
Deepl - gratis Version	<a href="http://www.deepl.com">www.deepl.com</a>	Onlinedienst für maschinelle Übersetzung, der von dem Unternehmen DeepL mit Sitz in Köln betrieben wird	Natural Language Processing (NLP), Machine Learning (ML), Deep Learning (DL)	Übersetzung unterschiedlicher Texte im Bereich Programm-Management, Marketing und Fundraising	unterschiedlich	Eingabe personenbezogener Daten, Copyrightverletzungen	Programm-Mgt/FR
Duden – Bezahl- version	<a href="http://www.duden.de">www.duden.de</a>	Online Rechtschreibe Assistent des Dudenverlags dem führenden deutschsprachigen Rechtschreibwörterbuch	Natural Language Processing (NLP), Machine Learning (ML),	Rechtschreibe- und Grammatikprüfung diverser deutschsprachiger Texte	unterschiedlich	Eingabe personenbezogener Daten, Copyrightverletzungen	Programm-Mgt/FR
CANVA - NGO- Version	<a href="http://www.canva.com">www.canva.com</a>	Canva wurde 2013 gestartet und ist eine Plattform für Online-Design und visuelle Kommunikation	Natural Language Processing (NLP), Machine Learning (ML), Deep Learning (DL), Computer Vision, recommendation Engines, generative Modelle	Erstellung von Designs vor allem für den online Bereich, inkl. Grafik und Bildbearbeitung	unterschiedliche Text- und Bilddaten	Eingabe personenbezogener Daten, Copyrightverletzungen, Verletzung Rechte am eigenen Bild	FR
Chat-GPT - gratis Version	<a href="http://www.chatgpt.com">www.chatgpt.com</a>	ist ein im November 2022 vorgestellter Chatbot des US-amerikanischen Unternehmens OpenAI, der in der Lage ist, mit Nutzern über textbasierte Nachrichten und Bilder zu kommunizieren. Die Grundlage von ChatGPT ist ein Large Language Model. Dabei handelt es sich um ein sehr leistungsfähiges Sprachmodell, das mit einer Vielzahl von Textdokumenten trainiert wurde.	Transformer, Deep Learning, Natural Language Processing (NLP), Reinforcement Learning from Human Feedback (RLHF), Transfer Learning, Contextual Embeddings	Erstellung von Postings, Berichten, Konzepten und unterschiedlichen Dokumenten, Textvorschläge, Zusammenfassungen von Texten, Übersetzungen von Texten, Korrektur von Texten	unterschiedlich	Eingabe personenbezogener Daten, Copyrightverletzungen, Fehlerhafte Antworten von Chatgpt	FR
Adobe Premiere Pro - Bezahl- version	Lokal bei ADRA installiert	Adobe Premiere Pro ist eine leistungsstarke Videobearbeitungssoftware, die verschiedene Technologien und Funktionen nutzt, um eine umfassende und effiziente Bearbeitungserfahrung zu bieten	Non-linear Editing (NLE), Real-Time Rendering, GPU Acceleration, Machine Learning, Advanced Color Grading Tools, Audio Editing Tools, Multi-Camera Editing	Übersetzung von Untertiteln von einer Sprache zu einer anderen und Erstellung von Untertiteln	unterschiedliche Videodateien	Copyrightverletzungen, Verletzung Rechte am eigenen Bild	FR

QuestionPro - gratis Version	<a href="http://www.questionpro.com">www.questionpro.com</a>	QuestionPro ist eine umfassende Umfrage- und Marktforschungsplattform, die es Unternehmen, Organisationen und Forschern ermöglicht, Online-Umfragen zu erstellen, zu verteilen und zu analysieren.	Natural Language Processing (NLP), Machine Learning (ML), Sentiment Analysis, Chatbots und Conversational AI, Predictive Analytics, Data Visualization, Recommendation Systems	Erstellung von Spenderumfragen	Es werden personenbezogene Daten wie Namen, E-Mail-Adressen, Adressen, Telefonnummern erfasst	Mißbrauch erfasster personenbezogener Daten im Falle eines Datenlecks	FR
Mailchimp - Bezahlversion	<a href="http://www.mailchimp.com">www.mailchimp.com</a>	Mailchimp ist eine Plattform für E-Mail-Marketing und Marketing-Automatisierung	Predictive Analytics, Segmentierung und Zielgruppenanalyse, Automatisierte Kampagnen, Sprach- und Textanalyse, Automatisierte Reporting und Analyse	Erstellung von E-Mail-Newslettern an Spender	Es werden personenbezogene Daten wie Namen und E-Mail-Adressen erfasst	Mißbrauch erfasster personenbezogener Daten im Falle eines Datenlecks	FR
Stripe	<a href="http://www.stripe.com">www.stripe.com</a>	Stripe, ein führender Zahlungsdienstleister, verwendet verschiedene KI- und Machine-Learning-Tools, um seine Dienstleistungen zu verbessern und effizienter zu gestalten.	NLP (Natural Language Processing), Betrugsprävention, Risikomanagement, Datenanalyse, Vorhersage und Prognosen,	Zahlungsverkehrsabwicklung für online Spenden an ADRA	Sehr kritische personenbezogene Daten und Zahlungsverkehrsdaten	Bei Datenleck, hoher Imageverlust, da kritische Zahlungsverkehrsdaten missbraucht werden können	Finanzen
PayPal	<a href="http://www.paypal.com">www.paypal.com</a>	Paypal, ein führender Zahlungsdienstleister, verwendet verschiedene KI- und Machine-Learning-Tools, um seine Dienstleistungen zu verbessern und effizienter zu gestalten.	Automatisierte Entscheidungsfindung, NLP (Natural Language Processing), Betrugsprävention,	Zahlungsverkehrsabwicklung für online Spenden an ADRA über PayPal	Sehr kritische personenbezogene Daten und Zahlungsverkehrsdaten	Bei Datenleck, hoher Imageverlust, da kritische Zahlungsverkehrsdaten missbraucht werden können	Finanzen
Excire	<a href="http://www.excire.com">www.excire.com</a>	KI basierte SW zum finden, organisieren und auswählen von Fotos.	CNNs: Convolutional Neural Networks	Marketing & Programme, finden von Projektfotos	Unterschiedliche Bild- und Videodateien	Verletzung Rechte am eigenen Bild. Sicherheit da lokal gespeichert.	Programm-Mgt/FR
BMD	<a href="http://www.bmd.com">www.bmd.com</a>	Automatisierte Belegverarbeitung, Belegtrennung und Erstellung von Buchungstexten	Machine Learning (ML) und andere Technologien	Buchhaltung	Kritische Finanzdaten	Mißbrauch von Finanzdaten bei Datenleck. BMD ist ISO 27001 zertifiziert.	Finanzen
MS-Teams		Intelligente Zusammenfassungen von Meetings, Live-Übersetzungen, KI Algorithmen zu Verbesserung der Sprachqualität	NLP (Natural Language Processing), Speech AI, LLMs basierend auf OpenAI	Videokonferenzsystem	Abhängig vom geteilten Inhalt	Ungewollte Offenlegung personenbezogener und vertraulicher Daten. Unbefugter Datenzugriff durch Fehlkonfigurationen (öffentliche Teams & Gastzugänge)	tbd