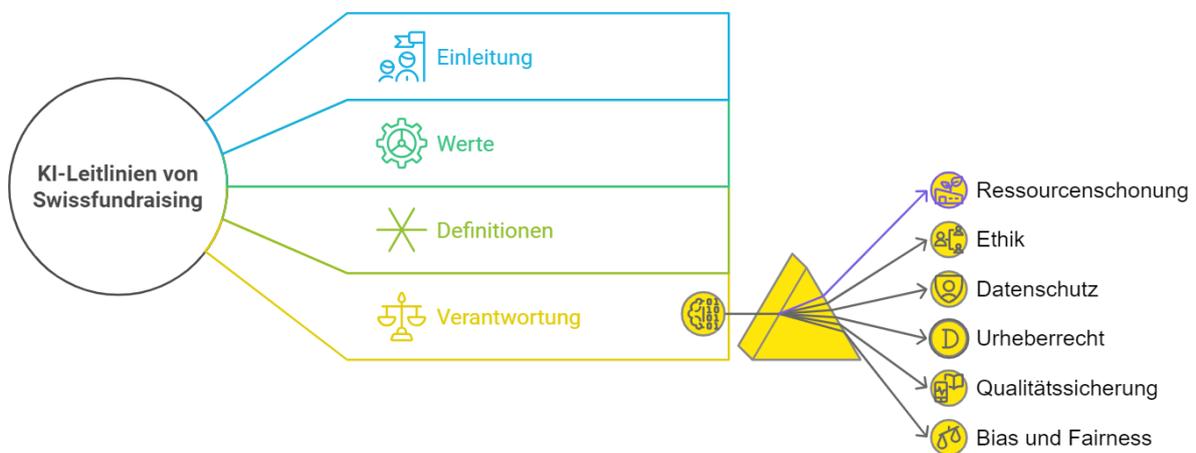


KI-Leitlinien Swissfundraising

Stand: Juni 2025

Mitarbeitende der Arbeitsgruppe KI & Fundraising:

Christine Bill, Janine Braun, Claudia Forcellini, Olivier Graz, Susi Kammergruber, Steff Kerkhof, Arlette Pond, Christian Schipp, Simeon Stiefel, Roger Tinner



1 Einleitung: Ziel und Zweck der Leitlinien	_____	2
2 Definition der Begrifflichkeiten	_____	2
3 Verantwortungsbereiche	_____	3
A) Ethik	_____	3
B) Bias und Fairness in der KI-Nutzung	_____	3
C) Qualitätssicherung und Kompetenzaufbau	_____	4
D) Urheberrecht/Kennzeichnung	_____	4
E) Datenschutz	_____	4
F) Nachhaltiger Einsatz/Ressourcen	_____	5

1 Einleitung: Ziel und Zweck der Leitlinien

Für das Handeln im Fundraising gelten gemäss den ethischen Richtlinien von Swissfundraising fünf übergeordnete Grundprinzipien: Ehrlichkeit, Respekt, Integrität, Professionalität und Transparenz. Diese Prinzipien gelten für Mitglieder des Verbands auch für den Einsatz von Künstlicher Intelligenz (KI) im Fundraising.

Diese Leitlinien bieten eine Orientierung für den verantwortungsvollen Einsatz von KI durch Fundraiser:innen, insbesondere innerhalb einer gemeinnützigen Organisation. Für Mitglieder von Swissfundraising sind sie integraler Bestandteil der Mitgliedschaft. Sie zielen darauf ab, die Potenziale von KI für das Fundraising effektiv zu nutzen, ohne dabei ethische, rechtliche oder ökologische Aspekte zu vernachlässigen. Übergeordnet gelten immer die aktuelle Gesetzgebung und Rechtsprechung, z.B. in Bezug auf Datenschutz oder Urheberrecht.

Die Leitlinien geben einen klaren Rahmen für Datenschutz, Fairness, Qualitätssicherung und Nachhaltigkeit vor. So fördern sie Transparenz und Vertrauen bei den Mitarbeitenden der Organisation sowie bei externen Anspruchsgruppen. Die Leitlinien unterstützen Fundraiser:innen und Organisationen dabei, die Vorteile von KI mit den eigenen Werten und Grundsätzen in Einklang zu bringen und ihre ethische Verantwortung wahrzunehmen. Die Leitlinien schaffen also eine Grundlage, die sowohl Innovation als auch Verantwortung ermöglicht.

Dieses Dokument unterscheidet zwischen Vorgaben, die unbedingt zu erfüllen sind («muss»), solchen, deren Umsetzung aus Sicht von Swissfundraising wünschenswert ist («soll»), und solchen, die als Empfehlung («kann») zu betrachten sind. Es ist Organisationen freigestellt, weitergehende Regelungen für den Einsatz von KI zu beschliessen.

2 Definition der Begrifflichkeiten

Künstliche Intelligenz (KI) ist eine Technologie, die Computern beibringt, aus Daten zu lernen und Aufgaben zu erledigen, die normalerweise menschliches Denken erfordern, wie zum Beispiel Sprache zu verstehen oder Muster zu erkennen.

Generative KI erstellt aus Trainingsdaten neue Inhalte wie Texte, Bilder oder Videos. Bekannte Tools dafür sind etwa ChatGPT, Midjourney oder Runway ML.

Machine Learning oder maschinelles Lernen (ML) ist ein Teilbereich der KI, bei dem Algorithmen verwendet werden, um innerhalb eines Datensatzes Muster zu erkennen und Vorhersagen zu treffen – ohne dafür explizit programmiert zu werden. Daten können aus Zahlen, Text oder Fotos bestehen.

LLM (Large Language Model) ist ein KI-Modell, das mit grossen Mengen an Daten (Text, Bilder, Audio, Video, etc.) trainiert wird, um Outputs zu generieren, die auch von einem Menschen stammen könnten, oder Spracheingaben zu verstehen und darauf zu reagieren.

Bias sind Verzerrungen oder Vorurteile in Daten oder KI-Modellen. Sie können zu ungerechten oder diskriminierenden Ergebnissen führen, wenn sie nicht beachtet werden.

Unter **Halluzinationen** werden überzeugend formulierte Resultate einer KI verstanden, die objektiv falsch oder nicht durch Trainingsdaten gerechtfertigt sind.

3 Verantwortungsbereiche

A) Ethik

Der Einsatz von KI im Fundraising erfordert eine ethische Herangehensweise, die sicherstellt, dass die Technologie im Einklang mit den Werten und Grundsätzen der Organisation genutzt wird. **KI sollte als den Menschen unterstützendes Werkzeug dienen, während wichtige Entscheidungen stets in menschlicher Verantwortung bleiben.** Der Schutz der Menschenwürde, der Respekt vor individuellen Rechten sowie die Vermeidung von Diskriminierung müssen dabei gewährleistet sein.

Es ist entscheidend, dass KI-Systeme keine Vorurteile oder ungerechtfertigten Annahmen verstärken. **Algorithmen können nur so neutral sein wie die Daten, mit denen sie trainiert wurden.** Deshalb müssen sowohl die eingesetzten Modelle als auch ihre Ergebnisse regelmässig überprüft werden, um unethische Verzerrungen zu vermeiden. Es ist essenziell, dass Mitarbeitende der Organisationen für ethische Fragestellungen sensibilisiert und dazu ermutigt werden, potenzielle Risiken oder problematische Anwendungen zu hinterfragen und Ergebnisse einer KI zu verifizieren.

Ein weiterer wichtiger Aspekt ist die Transparenz: **Adressat:innen sollten zweckmässig darüber informiert werden, wie und warum KI-Systeme Entscheidungen getroffen oder Inhalte generiert haben.** Schliesslich sollte der Einsatz von KI nicht nur rechtliche Vorgaben einhalten, sondern auch den höchsten Standards von Respekt, Integrität, Professionalität und Transparenz entsprechen. Das gewährleistet nicht nur die Qualität der Ergebnisse, sondern stärkt auch das Vertrauen in die Organisation und deren Umgang mit KI.

B) Bias und Fairness in der KI-Nutzung

Ein verantwortungsvoller KI-Einsatz erfordert die systematische Auseinandersetzung mit Bias und Fairness, da Verzerrungen in Daten oder Algorithmen zu diskriminierenden Ergebnissen führen können. Um diese Risiken zu minimieren, ist eine kritische Analyse der Trainingsdaten und der zugrunde liegenden Modelle notwendig. Einseitige oder unvollständige Datensätze können die Ergebnisse verfälschen und müssen daher frühzeitig identifiziert und korrigiert werden.

Die Organisation sollte klare Richtlinien festlegen, um sicherzustellen, dass KI-Modelle so neutral und inklusiv wie möglich arbeiten. Das kann durch Tests mit diversifizierten Datensätzen, die regelmässige Überprüfung der Ergebnisse auf Stereotypen und die kontinuierliche Evaluation der Algorithmen erreicht werden. Mitarbeitende sollten befähigt werden, Halluzinationen und potenzielle Verzerrungen zu erkennen und Strategien zu deren Reduzierung zu entwickeln (siehe dazu auch Kapitel C).

Transparenz und Verantwortung in den Entwicklungs- und Einsatzprozessen von KI sind entscheidend für faire und diskriminierungsfreie Ergebnisse. Die Etablierung eines Feedback-Mechanismus sowie eine fortlaufende Sensibilisierung für Bias tragen dazu bei, Schwachstellen in den Modellen zu identifizieren und kontinuierlich Verbesserungen vorzunehmen. So wird gewährleistet, dass KI nicht nur effizient, sondern auch ethisch und sozial verträglich eingesetzt wird.

C) Qualitätssicherung und Kompetenzaufbau

Um die Verbreitung von Fehlinformationen oder Verzerrungen zu vermeiden, ist eine systematische Qualitätsprüfung von KI-generierten Inhalten essenziell. **Vor der internen oder externen Nutzung sollten alle erstellten Inhalte auf inhaltliche Richtigkeit sowie die Einhaltung der Grundsätze Fairness und Diversität geprüft werden.** Besonders bei sensiblen oder öffentlich relevanten Inhalten wird das Vier-Augen-Prinzip empfohlen. Ein klar definierter Prüfprozess sollte sicherstellen, dass die Ergebnisse konsistent und verlässlich sind.

Parallel dazu ist der Kompetenzaufbau der Mitarbeitenden ein zentraler Faktor für den verantwortungsvollen Einsatz von KI. Schulungen sollten neben technischen Grundlagen auch Themen wie ethische Risiken, Bias-Vermeidung (siehe Kapitel B) und effektive Methoden wie Prompt Engineering umfassen. **Ziel ist es, dass Mitarbeitende KI-generierte Inhalte kritisch bewerten, ihre Stärken und Schwächen erkennen und sie gezielt für ihre Fundraisingtätigkeit einsetzen können.** Durch kontinuierliche Weiterbildung und die Integration von Qualitätssicherungsprozessen wird nicht nur die Arbeitsqualität und Fundraisingeffizienz erhöht, sondern auch das Vertrauen in den KI-Einsatz innerhalb der Organisation gestärkt.

D) Urheberrecht/Kennzeichnung

Beim Einsatz von KI ist die Beachtung des Urheberrechts von zentraler Bedeutung, da viele KI-Systeme mit Daten trainiert wurden und werden, die urheberrechtlich geschützt sein können. **Mitarbeitende sollten keine geschützten Inhalte in KI-Tools hochladen, es sei denn, es liegt eine ausdrückliche Genehmigung (Lizenz) vor.** Ebenso ist eine sorgfältige Prüfung der generierten Inhalte unerlässlich, um sicherzustellen, dass keine Urheberrechte verletzt werden oder geschützte Werke unrechtmässig reproduziert werden.

KI-generierte Inhalte, wie Texte, Bilder oder Videos, sind klar und eindeutig zu kennzeichnen, um Transparenz zu gewährleisten. Eine mögliche Formulierung lautet: «Dieser Inhalt wurde mit Unterstützung von KI erstellt.» Bei visuellen oder audiovisuellen Medien ist eine genaue Angabe der verwendeten Tools oder Plattformen erforderlich, z.B. «KI-Bild: DALL-E/Name der Organisation». Texte, die lediglich bearbeitet, aber nicht substantiell verändert wurden, benötigen in der Regel keine Kennzeichnung.¹

Durch klare Vorgaben zur Kennzeichnung und die Einhaltung des Urheberrechts wird das Vertrauen in die Organisation gestärkt und rechtliche Risiken werden minimiert. Es ist daher entscheidend, dass Mitarbeitende über diese Regeln informiert und bei der sicheren und rechtlich konformen Nutzung von KI unterstützt werden. Regelmässige Überprüfungen und Schulungen zu diesen Themen helfen, eine einheitliche und verantwortungsvolle Praxis sicherzustellen.

E) Datenschutz

Der Schutz personenbezogener Daten ist ein zentraler Aspekt beim Einsatz von KI und muss höchste Priorität haben. **Um den Schutz der Privatsphäre zu gewährleisten, dürfen in**

¹ Eine Kennzeichnungspflicht gilt für alle Inhalte, die gemäss Art. 50 EU AI Act davon betroffen sind: <https://artificialintelligenceact.eu/de/article/50/>

offenen Systemen ausschliesslich anonymisierte Daten verarbeitet werden, und der Betrieb eines geschlossenen Systems muss in der Datenschutzerklärung entsprechend transparent gemacht werden. Es ist unerlässlich, dass Datenschutzgesetze und organisationsinterne Richtlinien konsequent eingehalten werden.

Bei der Nutzung von KI-Tools ist darauf zu achten, dass Eingaben nicht für das Training der Modelle gespeichert oder weiterverwendet werden. Viele Plattformen bieten entsprechende Einstellungen an, die sicherstellen sollen, dass Daten nicht in die Datenbanken der Anbieter gelangen – auch wenn das hier, wie überall sonst auch, keinen 100-prozentigen Schutz vor Datenmissbrauch durch die Anbieter darstellt. Die Nutzung von freien oder ungesicherten Versionen von KI-Tools ist zu vermeiden, da diese oft nicht den notwendigen Schutz bieten.

Zusätzlich sollten Mitarbeitende sensibilisiert und geschult werden, um Risiken bei der Nutzung von KI zu erkennen und zu minimieren. Die Einhaltung des Prinzips der Datensparsamkeit – nur notwendige Daten zu nutzen – ist dabei essenziell. **Durch klare Vorgaben und die kontinuierliche Überprüfung der eingesetzten Tools und Prozesse kann die Organisation sicherstellen, dass personenbezogene Daten umfassend geschützt werden und vertrauensvoll mit KI umgegangen wird.**

F) Nachhaltiger Einsatz/Ressourcen

Der Einsatz von KI sollte stets unter Berücksichtigung ökologischer Aspekte erfolgen, da Systeme wie LLMs (Large Language Models) einen hohen Energieverbrauch (vorwiegend im Training) aufweisen. Es empfiehlt sich, wenn immer möglich energieeffiziente Tools oder solche zu bevorzugen, die mit erneuerbaren Energien betrieben werden, wie zum Beispiel herkömmliche Suchmaschinen. Dabei sollte der Nutzen des KI-Einsatzes stets gegen den Ressourcenverbrauch abgewogen werden.