Swico Mustervorlage:

KI-Leitlinie für Mitarbeitende

Einleitung

Die rasante Entwicklung der künstlichen Intelligenz (KI) eröffnet faszinierende neue Möglichkeiten für die Erstellung von Texten, Bildern, Audio- und Videomaterial sowie Programmiercode. KI-Tools können unseren Arbeitsalltag erleichtern und die Effizienz unserer Prozesse steigern. Wir möchten euch daher herzlich einladen, diese innovativen Werkzeuge zu erkunden und dabei unsere Unternehmenswerte und internen Vorgaben (z.B. Markenidentität, Design-Richtlinien) zu beachten. Diese Richtlinie dient als Grundlage für den sicheren und verantwortungsvollen Umgang mit KI.

Es ist wichtig zu verstehen, dass KI-generierte Inhalte auf Wahrscheinlichkeitsrechnungen und Zufallsprozessen basieren, die von den zugrunde liegenden Trainingsdaten beeinflusst werden. Daher sind sowohl die Qualität der Dateneingabe (Prompt) als auch das kritische Hinterfragen der Ergebnisse entscheidend für optimale Resultate.

1. Sorgfaltspflichten

Bei der Nutzung von KI im Arbeitsalltag sind, unabhängig vom konkreten Anwendungsbereich, gewisse Sorgfaltspflichten zu erfüllen, welche nachfolgend dargelegt sind.

1.1 Verantwortlichkeit und Arbeitsweise

- Prompting: Basis für gute Ergebnisse bilden klare Eingabeaufforderungen (Prompt), die strukturierte Arbeitsweise im «Ping-Pong» mit der Maschine und das kritische Hinterfragen der Ergebnisse.
- Verantwortung für die Verwendung des Ergebnisses: Wer ein KI-Tool einsetzt, ist nebst dem Prompt (siehe oben) auch für die Verwendung des Ergebnisses verantwortlich.
- Verantwortung für die Nutzung: Für die Einhaltung gesetzlicher und interner Vorgaben sind die Mitarbeitenden (Anwendungsbereich, Dateneingabe) und das Unternehmen (Gesetz, Verträge, Lizenzen, Nutzungsbedingungen etc.) verantwortlich.
- Verantwortung für die Sicherheit: Der Zugang zum Konto oder zum Chatverlauf muss vor fremdem Zugriff geschützt werden (z.B. sicheres Passwort wählen und im Passwortmanager hinterlegen oder Zwei-Faktor-Authentifizierung aktivieren).

1.2 Ausgewählte Anforderungen mit Blick auf Gesetze

- Geheimhaltungspflichten: Keine Eingaben von internen und sensiblen Informationen, die bspw. das Berufsgeheimnis oder das Geschäftsgeheimnis verletzen – dazu gehören auch vertraglich geschützte Daten von Geschäftspartnern.
  Beispiel: Finanzdaten oder Information zur Strategie.
- Datenschutzgesetz: Keine Eingabe von Daten die vom Datenschutz erfasst sind. Dazu zählen insbesondere auch Personendaten bzw. Daten, die mit einer natürlichen Person in Bezug gebracht werden können.
  Beispiel: Keine Daten der Kontaktperson beim Kunden (Name, Vorname, Adresse, Geburtstag) oder Daten, die den Rückschluss auf eine bestimmte Person

ermöglichen (Eingabe von Angaben, wie Firma, Rolle und Arbeitszeitraum einer Person, die es ermöglichen, diese Person zu bestimmen).

- Urheberrecht: Keine Eingabe von urheberrechtlich geschützten Daten. Einholung des Einverständnisses, wenn Daten oder Inhalte von Dritten genutzt werden.
  Beispiel: Ohne Einwilligung des Rechteinhabers einen journalistischen Text lediglich leicht umformulieren lassen und unter eigenem Namen veröffentlichen.
- Persönlichkeitsrecht: Kein Prompting oder Verwendung von Ergebnissen, die gegen die Rechte von Personen verstossen.
  Beispiel: Analyse und automatische Aussonderung von Bewerbungsdossiers aufgrund von diskriminierenden Bewertungskriterien, wie Ethnie, Namen, Geschlecht, etc.

2. Anforderungen pro Anwendungsbereich

2.1 Anforderungen im Bereich Text

- Der Inhalt wird vor der Publikation auf seinen Wahrheitsgehalt überprüft.
- Der Text enthält keine Vorurteile und Stereotypen.
- Der Text ist sachgerecht, logisch, sprachlich korrekt und gut verständlich.

Anwendung erlaubt (Beispiele)	**Anwendung nicht erlaubt (Beispiele)**
Effizienz: Zusammenfassen oder analysieren einer Bedienungsanleitung	**Datenschutz:** Zusammenfassen vertraulicher Inhalte oder von Geschäftsgeheimnissen
Inklusion: Erstellen, optimieren oder übersetzen von Texten, z.B. Rechtschreibung, einfache Sprache	**Vorurteile und Stereotypen / Wahrheitsgehalt:** Übernahme von Texten ohne Qualitätskontrolle
Kreativität: Vorschläge für Titel von Kampagnen oder Planungs- und Strukturierungsaufgaben	**Wahrheitsgehalt / Qualität:** Nutzen als einzige oder als wissenschaftliche Quelle
Effektivität: Überführen von Informationen in Tabellen oder Grafiken	**Qualität:** Erstellen umfangreicher Aufgaben ohne Überprüfung auf Vollständigkeit, Richtigkeit, Logik
Effektivität: Inspirationsquelle für erste Ideen, zum Beispiel für eine öffentliche Präsentation	**Datenschutz:** Übersetzen von Arbeitszeugnissen
Datenschutz: Erstellen eines Arbeitszeugnisses mit anonymisierten Personendaten	
Urheberrecht: Datenanalyse in umfangreichen, urheberrechtlich geschützten Unterlagen, für die das Unternehmen die Einwilligung des Urhebers hat	

2.2 Anforderungen im Bereich Bild

- Das Ergebnis passt zum Inhalt oder Verwendungszweck (Sujetwahl, Stil).
- Das Ergebnis enthält keine Vorurteile und Stereotypen (Diskriminierung).
- Für Bilder, die exklusiv verwendet werden sollen, wird auf KI verzichtet, weil kein urheberrechtlicher Schutz möglich ist.
- Es werden keine Bilder mit der Absicht erstellt und verwendet, um das Publikum / die Öffentlichkeit zu täuschen.
- Das Bild wird gut sichtbar gekennzeichnet nach dem Muster «KI-generiertes Bild mit Midjourney».

Anwendung erlaubt (Beispiele)	Anwendung nicht erlaubt (Beispiele)
Effektivität: Erstellen von Bildern, Design-Entwürfen oder Produktbildern	**Vorurteile und Stereotypen:** Verwenden eines Bildes, das Vorurteile enthält
Keine Täuschung: Verwenden eines fotorealistischen Bildes mit Kennzeichnung	**Täuschungsgefahr:** Verwenden eines fotorealistischen Bildes ohne Kennzeichnung
Effektivität: Bearbeiten eines Bildes, z.B. indem ein nicht relevantes Element entfernt wird	**Manipulation:** Verändern der Aussagekraft als Ergebnis der Bildbearbeitung
Persönlichkeitsrechte: Bearbeiten von Bildern von realen Personen, deren Einverständnis dafür vorliegt	

2.3 Anforderungen im Bereich Audio und Video

- Der Inhalt wird vor der Publikation auf Wahrheitsgehalt und Aussagekraft überprüft.
- Das Ergebnis enthält keine Vorurteile und Stereotypen (Diskriminierung).
- Es wird kein Audio- und Videomaterial (inklusive Deepfakes) erstellt oder verwendet, um das Publikum / die Öffentlichkeit zu täuschen.
- Beim Audio- und Videomaterial wird gut sichtbar gekennzeichnet, dass es mit KI erzeugt wurde. Muster für die Kennzeichnung: «KI-generiertes Video mit Synthesia». Ausnahme von der Kennzeichnungspflicht: Musik und Geräusche werden nicht gekennzeichnet.

Anwendung erlaubt (Beispiele)	Anwendung nicht erlaubt (Beispiele)
Inklusion: Vertonung von Text, z.B. auf der Website	**Täuschungsgefahr:** Einsetzen einer künstlich erzeugten Stimme ohne Deklaration
Inklusion: Transkription von gesprochener Sprache in Text, z.B. für Untertitel	**Täuschungsgefahr / Persönlichkeitsrecht:** Die Stimme oder Person wird einer real existierenden Person nachempfunden

2.4 Anforderungen im Bereich Programmiercode

- Die internen Sicherheits- und Governance-Richtlinien für Codes bzw. Coding werden eingehalten.
- Der Code wird vor der Publikation auf Fehler und Sicherheitslücken überprüft.
- Einzelne Parameter oder die Entscheidungslogik führen nicht zu einer Diskriminierung der Nutzer oder Kunden.

Anwendung erlaubt (Beispiele)	Anwendung nicht erlaubt (Beispiele)
Effizienz: Optimieren und Vervollständigen von Code in Open Source Projekten	**Datenschutz:** Verwenden von Code, der Personen- oder Unternehmensdaten enthält
Effektivität: Debugging der eigenen Codebasis	**Urheberrecht:** Verwenden von Code, der via Lizenz geschützt oder der geheim ist
Effektivität: Inspirationsquelle für spezifische Fragen	

3. Ausblick

Aufgrund der schnell voranschreitenden technologischen Entwicklung verändern sich die Möglichkeiten und Vertragsbedingungen der einzelnen KI-Tools zurzeit stark. Deshalb wird dieses Dokument regelmässig auf seine Aktualität und Nützlichkeit überprüft. Es ist wichtig, dass immer die aktuelle Version dieser Richtlinie verwendet wird.

Wer ein Risiko vermutet, einen Fehler gemacht hat oder etwas Wichtiges melden möchte, kann sich an die interne Kontaktperson wenden.

Interne Kontaktperson:	((Vorname und Nachname))
Stand der KI-Leitlinie:	((Datum / Version der Richtlinie))

Diese Mustervorlage für eine KI-Leitlinie unterliegt der Creative Commons Lizenz. Dies bedeutet, dass die Inhalte übernommen und auf die spezifischen Bedürfnisse angepasst werden können unter der Bedingung der Namensnennung.
Lizenz: Namensnennung 4.0 International (CC BY 4.0)