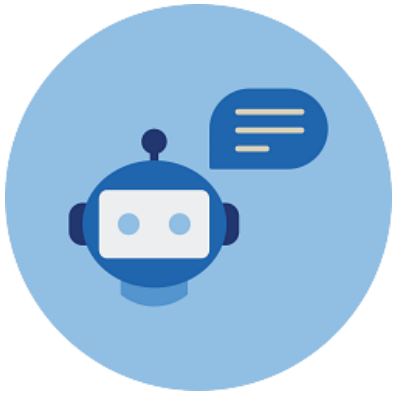
KI unter «*Kontrolle*» – VERANTWORTUNGsvolle Innovation

KI im Unternehmen - Verantwortung, Kontrolle und Compliance als Basis verantwortungsvoller Innovation

19.11.2025
Daniel Gysel
Chief Compliance Officer – Zurich Business Unit Schweiz

Diese Präsentation ist ausschliesslich für den/die eingangs genannten Adressaten vom SGO Verein bestimmt.

KI (MS Copilot) erklärt, wer für die Korrektheit der Antworten verantwortlich ist



Wer hat dein Sprachmodell entwickelt und wie stellst du sicher, dass du nur korrekte Antworten generierst? Wer ist für die Korrektheit deiner Antworten verantwortlich?

Mein Sprachmodell wurde von **OpenAI** entwickelt, einem Unternehmen, das sich auf künstliche Intelligenz spezialisiert hat. Die Version, mit der du gerade sprichst, basiert auf der **GPT-4-Architektur** und wird von Microsoft in Produkten wie Copilot integriert und bereitgestellt.

Wie wird die Korrektheit meiner Antworten sichergestellt?

Es gibt mehrere Ebenen, auf denen die Qualität und Korrektheit meiner Antworten sichergestellt wird:

1. **Training auf hochwertigen Daten**:
   Ich wurde mit einer großen Menge an Textdaten aus verschiedenen Quellen trainiert – darunter Bücher, wissenschaftliche Artikel, Webseiten und andere öffentlich zugängliche Inhalte. Dabei wurden Qualitätsfilter eingesetzt, um fehlerhafte oder irreführende Informationen möglichst auszuschließen.

2. **Regelmäßige Aktualisierung und Feinabstimmung**:
   Microsoft und OpenAI führen regelmäßige Updates und Feinabstimmungen durch, um die Genauigkeit zu verbessern und neue Informationen zu integrieren.

3. **Websuche bei Bedarf**:
   Wenn eine Frage aktuelle oder sehr spezifische Informationen erfordert (z. B. Gesetzesänderungen, lokale Ereignisse, neue Forschung), nutze ich eine integrierte Websuche, um verlässliche Quellen zu finden und zu zitieren.

4. **Verantwortung und Haftung**:
   Die Verantwortung für die Korrektheit meiner Antworten liegt letztlich bei den **Entwicklern und Betreibern** – also **OpenAI** für das Modell und **Microsoft** für die Integration und Bereitstellung. Dennoch ist es wichtig zu betonen, dass ich **keine absolute Garantie für Fehlerfreiheit** geben kann. Ich bin ein **Hilfsmittel, kein Ersatz für professionelle Beratung**.

Wenn du möchtest, kann ich dir auch zeigen, wie ich bei einer konkreten Frage die Informationen überprüfe oder wie ich mit Quellen arbeite. Möchtest du ein Beispiel?

Frage an die KI, wie «sie» die Korrektheit von Angaben verifiziert

💉 **Beispiel: Wie überprüfe ich eine Aussage zur SGO-Gründung**

Angenommen, du fragst:
„Wann wurde die Schweizerische Gesellschaft für Organisation und Management (SGO) gegründet?“

1. **Websuche starten**
Ich beginne mit einer gezielten Suche nach offiziellen oder vertrauenswürdigen Quellen, z. B. von der SGO selbst, Hochschulen oder seriösen Medien.

2. **Quelle identifizieren**
Ich finde z. B. eine Webseite der SGO oder eine Hochschulseite, auf der steht:
> „Die SGO wurde 1967 gegründet durch eine Initiative von Organisationsleitern bedeutender schweizerischer Firmen und Verwaltungen.“
→ Quelle: SGO Business School[1](https://uniservice-dl.uni-wuppertal.de/de/ki-handreichungen/ki-musterbelegvorlage/)

3. **Validität prüfen**
Ich prüfe:
- Ist die Quelle **offiziell** oder **wissenschaftlich**?
- Ist die Information **konsistent** mit anderen Quellen?
- Gibt es **Widersprüche** oder **fehlende Details**?

4. **Antwort formulieren mit Zitat**
Ich formuliere die Antwort und füge die Quelle direkt an:
> „Die SGO wurde 1967 gegründet …“[1](https://uniservice-dl.uni-wuppertal.de/de/ki-handreichungen/ki-musterbelegvorlage/)

5. **Transparenz durch Quellenangabe**
Du siehst direkt, **woher** die Information stammt und kannst sie **selbst nachprüfen**.

🔍 Warum ist das wichtig?

Laut [1a-Studi.de](https://www.1a-studi.de/hilfe-faq/ki-quellen-ueberpruefen) ist die Quellenprüfung bei KI-Antworten essenziell, weil KI-Modelle manchmal sogenannte „Halluzinationen“ erzeugen – also plausible, aber falsche Informationen[2](https://www.1a-studi.de/hilfe-faq/ki-quellen-ueberpruefen).

Auch die [Universität Wuppertal](https://uniservice-dl.uni-wuppertal.de/de/ki-handreichungen/ki-musterbelegvorlage/) empfiehlt, bei KI-generierten Inhalten immer die Art der Verwendung und die Quelle transparent zu machen[1](https://uniservice-dl.uni-wuppertal.de/de/ki-handreichungen/ki-musterbelegvorlage/).

Der Auftrag an die KI bestand darin, den Zweck der SGO und die wichtigsten Aktivitäten als Bild darzustellen…inklusive typische Farben

Hier ist die aktualisierte Version mit dem **SGO-Logo** und den typischen **Farben (Grün & Blau)**.

Das Bild zeigt den zentralen Zweck und die wichtigsten Aktivitäten:

- **SGO im Zentrum** mit Motto: *„Gestalte die Zukunft. Lerne aus Erfahrung."*
- **Aktivitäten:** Vernetzung, Erfahrungsaustausch, Aus- und Weiterbildung, Zertifizierungen.

Erwartungen der FINMA und des EDÖB beim Einsatz von KI

Geltendes Recht anwenden und prinzipienorientierte Erwartungen beachten

FINMA

- **FINMA-Aufsichtsmitteilung 08/2024:** Governance und Risikomanagement beim Einsatz von künstlicher Intelligenz
  - Inventar & Risikoklassifizierung
  - Datenqualität
  - Monitoring
  - Dokumentation
  - Erklärbarkeit
  - Unabhängige Überprüfung (Wesentlichkeit)
- **Frühzeitige Kontaktaufnahme mit FINMA**, wenn KI bei **kritischen Prozessen** oder bei der Berechnung regulatorisch vorgegebener „Kennzahlen“ (z.B. SST) eingesetzt werden soll[1]

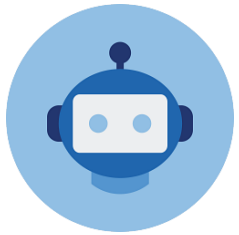
EDÖB

- **Geltendes** und technologieneutrales **Datenschutzgesetz** ist direkt **auf KI anwendbar**
- **Transparenzpflichten** beim Einsatz von KI abgeleitet aus Art. 19 DSG (Informations-pflichten)
- Zusätzliche Informationspflichten und «menschliches Gehör» bei **automatisierten Einzelentscheidungen**
- **Datenschutz Folgenabschätzungen** (DSFA) bei potenziell hohem Risiko für die betroffenen Personen

[1] **FINMA-Medienmitteilung 2025** https://www.finma.ch/de/news/2025/04/20250424-mm-umfrage-ki/ I **FINMA-Risikomonitor 2023:** Längerfristige Trends und Risiken: Künstliche Intelligenz im Schweizer Finanzmarkt I **FINMA-Jahresbericht 2023:** Künstliche Intelligenz: Die FINMA formuliert ihre Aufsichtserwartungen I **FINMA-Jahresbericht 2021:** Künstliche Intelligenz im Schweizer Finanzmarkt

Ohne die folgende Basis kann nicht bewusst gesteuert werden

Erster Schritt: Eine für alle Stakeholder passende Definition für «AI» und «Wesentlichkeit»

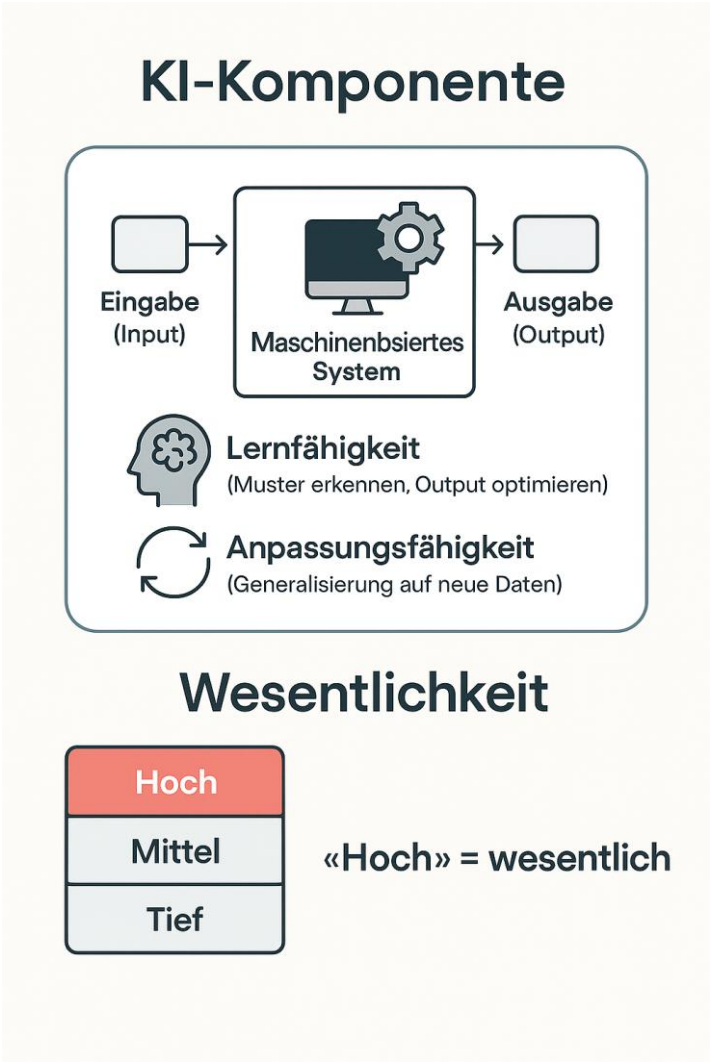


KI-Komponente

- Eine KI-Komponente ist ein intern oder extern entwickeltes maschinenbasiertes System, welches Outputs anhand von Inputs generiert. Sie zeichnet sich durch mindestens folgende Eigenschaften aus:
  - Lernfähigkeit: Die KI-Komponente optimiert ihren Output für eine vorgegebene Zielsetzung, indem sie (komplexe) Muster und Zusammenhänge in zugrunde liegenden Daten identifiziert.
  - Anpassungsfähigkeit: Die KI-Komponente ist darauf ausgelegt anhand von zuvor ungesehenen Daten generalisierte Outputs zu generieren.

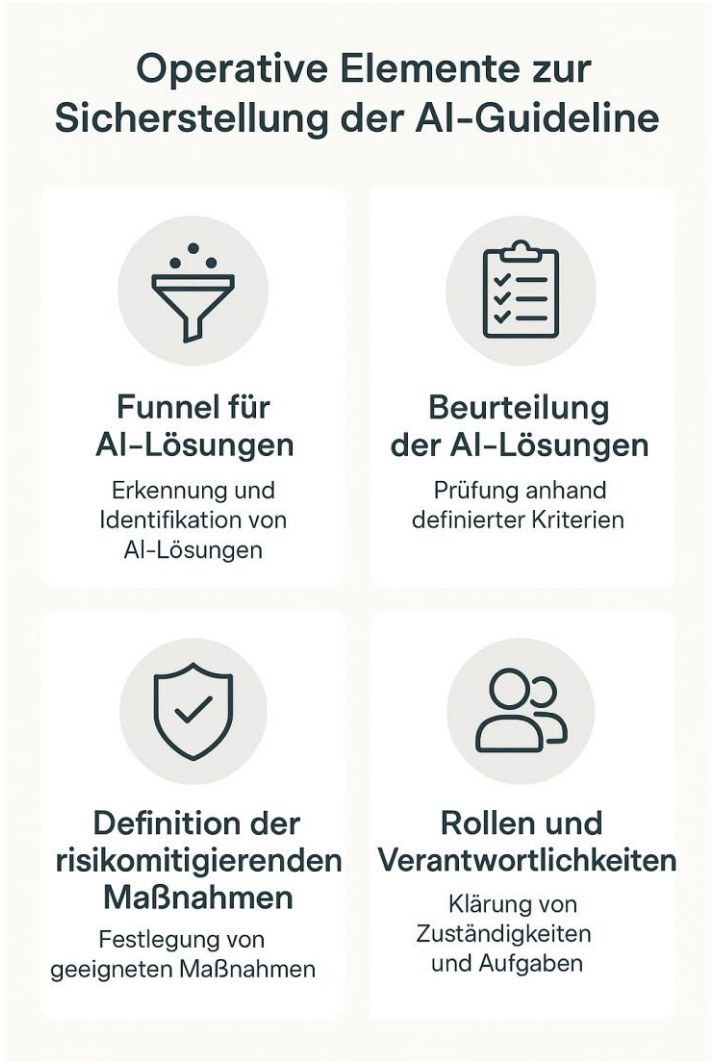
Wesentlichkeit

- KI-Komponenten, welche per Risikokategorisierung als «hoch» eingestuft werden, gelten als wesentlich.



Die wichtigsten «Zutaten» unseres AI-Steering

Funnel ➔ Beurteilung ➔ risikomindernde Massnahmen ➔ Rollen und Verantwortlichkeiten



1. Funnel für AI-Lösungen (Erkennung & Identifikation)

Ziel - Sicherstellen, dass jede neue oder geänderte KI-Komponente identifiziert wird.

Kernpunkte:

- Initiierung des AI-Steering Prozesses.
- Zuweisung eines KI-Komponenten-Owners.
- Inventarisierung und Dokumentation der Komponente.

2. Beurteilung der AI-Lösungen (Risikokategorisierung & Risk Assessment)

Ziel - Bewertung der Risiken und Festlegung der Risikoklasse (tief, mittel, hoch).

Kernpunkte:

- Durchführung der Risikokategorisierung.
- Validierung
- Risk Assessment inkl. Definition wesentlicher Änderungen und risikomindernder Faktoren.

3. Definition der risikomindernden Massnahmen

Ziel - Festlegung geeigneter Massnahmen zur Risikominderung.

Kernpunkte:

- Massnahmen im Risk Assessment dokumentieren.
- Monitoring und Kontrollvorgaben (z. B. Bias-Checks, Logging, Performance-Überwachung).
- Fallback-Szenarien und Business-Continuity-Pläne.

4. Rollen und Verantwortlichkeiten

Ziel - Klare Zuständigkeiten für alle Schritte.

Kernpunkte:

- Owner der KI-Komponente: Umsetzung der Vorgaben, Durchführung von Risikoanalysen.
- IT-Owner: Technische Implementierung.
- AI Experte: Unterstützung, Inventarführung, Validierung.
- AI Gremien: Empfehlungen und Freigaben.
- Risk Management: Moderation des Risk Assessments und Validierung der Maßnahmen.

Erste Zutat: Funnel & AI Steering / Zweite Zutat: Beurteilung

Klarer Ablauf von der Identifikation einer KI-Komponente bis zur Re-Validierung

AI Steering

Identifikation
Neue KI-Komponente erkennen,
Owner festlegen
Mitarbeitende, IT, Legal, Risk, Compliance, Audit

Risikokategorisierung
Einstufung in »hoch«, »mittel«oder tief«
KI-Komponenten Owner

Validierung
Überprüfung der Einstufung

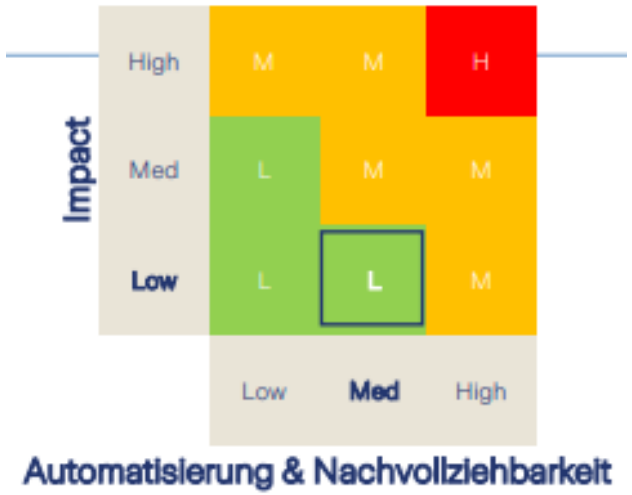
Risk Assessment
Risiken bewerten,
Maßnahmen definieren
Moderation: Risk Management
Unterstützung: AI-Champion

Review & Freigabe
Präsentation an AI-Review Pane
oder AI-Oversight Board,
Empfehiung und Entscheidung
Owner, AI-Champion

Freigabe
durch Risikoträger
(Management)
Abweichung:

Inventarisierung& Dokumentation
Dokumentation und Inventar pflegen

Jährliche Validierung
Überprüfung und ggf. Aktualisierung



Dritte Zutat: Definierte risikomindernde Massnahmen

Um KI-Komponenten ähnlich zu schützen, wurde ein Standard-Set an Massnahmen formuliert

Minimal-Set an Massnahmen für alle KI-Komponenten (unabhängig von Risikoeinschätzung

- Inventarisierung
- Minimum jährliche Prüfung der Aktualität und Richtigkeit der Risikobewertung / Risikokategorisierung und update des Inventars im Falle von Änderungen.
- Dokumentation (Zweck, Umfang, Rollen und Verantwortlichkeiten, Training, Validation und Testkonzept, BCP, Änderungen, Versionskontrolle etc.)

Für KI-Komponenten mit Risiko mittel / hoch zusätzlich verpflichtend

Inputs

– Überprüfung der Datenquellen und der zugrunde liegenden Daten auf Eignung und Bias

Methodologie

– Auswahl der KI-Komponente und Kalibrierungsprozess und Performance Evaluation anhand von quantitativen und qualitativen Bewertungsmetriken

Implementation

– Versionskontrolle / Definition SLAs des Services (Drittanbieter / Intragroup) sowie Überwachung und Logging

Outputs

– Überprüfung des Outputs auf Bias und Fairness, Erklärbarkeit, Fortlaufende Überwachung der Leistung auf vordefinierten Metriken und Handlungstrigger

– Etablierung eines Feedbackloops (Nutzerfeedback) mit Feedbackloop-Monitoring.

Vierte Zutat: Rollen und Verantwortlichkeiten

Umfassende Definition von (neuen) Rollen / Verantwortlichkeiten über alle Hierarchiestufen

CEO

Umsetzung der Guideline

Mitarbeitende

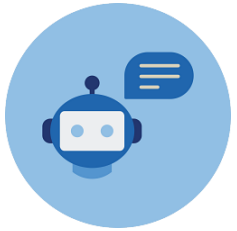
Sicherstellung, dass ein Owner für jede neue oder ergänzte KI-Komponente bestimmt wird

Owner einer KI-Komponente

Einhaltung der Vorgaben über den gesamten Lebenszyklus; Identifikation, Risikokategorisierung, Risk Assessment, Dokumentation, Inventarisierung

IT Owner

Technische, produktive Implementation der KI-Komponente; Unterstützung des Owners bei technischen Anforderungen.



AI Experte

Unterstützung der Owner beim AI-Steering Prozess; Führung des Inventars; Sicherstellung der Risikokategorisierung und Dokumentation

AI Panel

Empfehlung an Risikoträger zur Freigabeentscheidung; Validierung der Risikokategorisierung; Dokumentation der Empfehlungen

AI Board

Validierung der Empfehlungen des AI-Review Panels bei „hoch“ eingestuften Risiken; Empfehlung an Risikoträger auf Geschäftsleitungsebene; Bestätigung, Anpassung oder Rückweisung der Empfehlungen

Besten Dank und weiterhin lehrreiche Momente.....

HEUTE IN BERUFE DER ZUKUNFT:
DER KI-FLÜSTERER

Quelle: cloud-science.de