VISCHER

Generative KI.

So klappt es mit dem Einsatz aus rechtlicher Sicht

David Rosenthal, Partner, VISCHER AG
28. November 2023

Ein Beispiel

Lebenslauf

Persönliche Daten:

Name:	Mustermann
Vorname:	Stephan
Adresse:	Musterstrasse 23, 3000 Musterstadt
Telefon:	077 XX XXX XX
E-Mail:	stephan.mustermann@musterstadt.ch
Geburtsdatum:	03.06.1978
Zivilstand:	ledig, keine Kinder

Berufliche Erfahrungen:

02/2004 – heute	Muster AG, Bern: Marketingkoordinator, Betreuung von Print- und Offlinekampagnen, Organisation der Messeauftritte, Vorbereitung und Durchführung von Kundenevents
02/2000 – 01/2004	Marketing Verlag Basel: Marketingfachmann, Anzeigenerstellung und –schaltung, Erstellen von Produktflyern und Prospekten, Direktwerbung
07/1998 – 01/2000	Marketing Zeitung Olten: Marketingassistent, Vorbereiten von Statistiken, Unterstützung bei der Marketingplanung und -umsetzung

Ausbildung:

05/1999 – 05/2000	HSO Schulen Thun Bern AG: «Abschluss als Marketingplaner»
08/1994 – 08/1997	Wirtschafts- und Kaderschule KV Bern: «Abschluss als Kaufmann E-Profil»

Sprachen:

Deutsch:	Muttersprache
Englisch:	schriftlich und mündlich sehr gut
Französisch:	schriftlich und mündlich gut

Spezielle Fähigkeiten:

- Erstellen von 3D-Animationen
- Erfahrung im Umgang mit Adobe Photoshop
- selbständige und strukturierte Arbeitsweise

Freizeit:

Asiatische Küche, Tennis, Basketball, Städtereisen

Referenzen:

Auf Anfrage

Was sagt die KI dazu?

Prompt in VISCHER GPT: "Du bist der weltbeste Personalchef und hast ein feines Gespür für Warnsignale in Lebensläufen von Bewerbern. Schaue Dir folgenden Lebenslauf sehr genau an und erkläre mir, welche Punkte auf mögliche Probleme bei diesem Kandidaten hinweisen:"

Quelle: https://www.jobscout24.ch/download/vorlagen/Lebenslauf_Marketing.pdf

Viel sagt sie, aber ist das ein Problem?

"Es gibt keine Hinweise auf kontinuierliche Weiterbildung oder berufsbezogene Zertifikate, Schulungen oder Kurse, die nach dem Abschluss als Marketingplaner absolviert wurden. In einem sich schnell verändernden Feld wie dem Marketing könnte dies bedeuten, dass der Bewerber in Bezug auf aktuelle Trends und Technologien nicht auf dem neuesten Stand ist."

"Während Kenntnisse in 3D-Animation und Adobe Photoshop wertvoll sein können, scheinen diese Fähigkeiten nicht direkt mit seiner Rolle als Marketingkoordinator in Verbindung zu stehen. Dies könnte darauf hinweisen, dass der Kandidat Interesse an einer Karriereänderung hat oder dass er über Qualifikationen verfügt, die er möglicherweise nicht vollständig nutzen konnte."

"Wenn jemand über einen längeren Zeitraum bei demselben Unternehmen beschäftigt ist, würde man normalerweise erwarten, dass er verschiedene Rollen durchlaufen oder Beförderungen erfahren hat. Es gibt keine Hinweise auf eine Beförderung oder zunehmende Verantwortung in seinem Lebenslauf, was darauf schliessen lässt, dass es möglicherweise an Ambition oder Erfolg gefehlt hat, um in seiner Karriere voranzukommen."

"Stephan Mustermann ist seit über 19 Jahren bei der Muster AG tätig. Während Kontinuität ein positives Zeichen sein kann, könnte dies auch bedeuten, dass der Bewerber Schwierigkeiten mit Veränderungen hat oder in seiner Entwicklung stagniert. Es könnte auch bedeuten, dass er wenig Anpassungsfähigkeit oder Motivation zur persönlichen Weiterentwicklung zeigt."

Schritt 1: Grundsätze

Ferner: Betroffenenrechte und Regelung für automatisierte Einzelentscheide

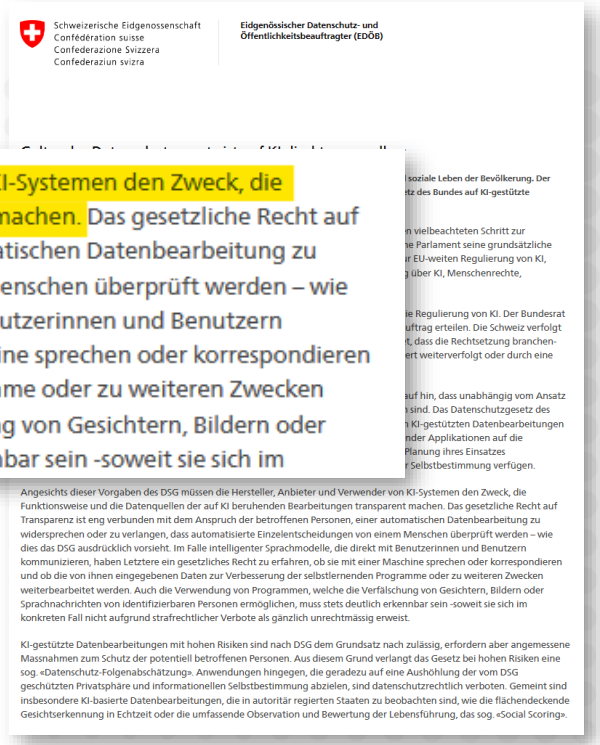
- Grundsatz der **Transparenz** eingehalten?
  - Ist das ein wesentlicher Aspekt der Datenbearbeitung, der für den Entscheid, mir die Daten zu geben, von Relevanz ist?
- Grundsatz der **Richtigkeit** eingehalten?
  - Ist gewährleistet, dass ich am Ende richtige Daten bearbeite?
- Grundsatz der **Zweckbindung** eingehalten?
  - Bearbeite ich die Daten nur für ersichtliche/erwartete Zwecke?
- Grundsatz der **Verhältnismässigkeit** eingehalten?
  - Erhebe ich nur Daten, die nötig sind? Bearbeite ich sie nur soweit nötig und geeignet? Ist die Bearbeitung für die Person zumutbar?
- Ist es trotz allem **fair**, was ich mache?

Transparenz?

- EDÖB: Absoluter Informationsanspruch

Angesichts dieser Vorgaben des DSG müssen die Hersteller, Anbieter und Verwender von KI-Systemen den Zweck, die Funktionsweise und die Datenquellen der auf KI beruhenden Bearbeitungen transparent machen. Das gesetzliche Recht auf Transparenz ist eng verbunden mit dem Anspruch der betroffenen Personen, einer automatischen Datenbearbeitung zu widersprechen oder zu verlangen, dass automatisierte Einzelentscheidungen von einem Menschen überprüft werden – wie dies das DSG ausdrücklich vorsieht. Im Falle intelligenter Sprachmodelle, die direkt mit Benutzerinnen und Benutzern kommunizieren, haben Letztere ein gesetzliches Recht zu erfahren, ob sie mit einer Maschine sprechen oder korrespondieren und ob die von ihnen eingegebenen Daten zur Verbesserung der selbstlernenden Programme oder zu weiteren Zwecken weiterbearbeitet werden. Auch die Verwendung von Programmen, welche die Verfälschung von Gesichtern, Bildern oder Sprachnachrichten von identifizierbaren Personen ermöglichen, muss stets deutlich erkennbar sein -soweit sie sich im

https://www.edoeb.admin.ch/edoeb/de/home/kurzmeldungen/2023/20231109_ki_dsg.html


Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössischer Datenschutz- und Öffentlichkeitsbeauftragter (EDÖB)

Angesichts dieser Vorgaben des DSG müssen die Hersteller, Anbieter und Verwender von KI-Systemen den Zweck, die Funktionsweise und die Datenquellen der auf KI beruhenden Bearbeitungen transparent machen. Das gesetzliche Recht auf Transparenz ist eng verbunden mit dem Anspruch der betroffenen Personen, einer automatischen Datenbearbeitung zu widersprechen oder zu verlangen, dass automatisierte Einzelentscheidungen von einem Menschen überprüft werden – wie dies das DSG ausdrücklich vorsieht. Im Falle intelligenter Sprachmodelle, die direkt mit Benutzerinnen und Benutzern kommunizieren, haben Letztere ein gesetzliches Recht zu erfahren, ob sie mit einer Maschine sprechen oder korrespondieren und ob die von ihnen eingegebenen Daten zur Verbesserung der selbstlernenden Programme oder zu weiteren Zwecken weiterbearbeitet werden. Auch die Verwendung von Programmen, welche die Verfälschung von Gesichtern, Bildern oder Sprachnachrichten von identifizierbaren Personen ermöglichen, muss stets deutlich erkennbar sein -soweit sie sich im konkreten Fall nicht aufgrund strafrechtlicher Verbote als gänzlich unrechtmässig erweist.

KI-gestützte Datenbearbeitungen mit hohen Risiken sind nach DSG dem Grundsatz nach zulässig, erfordern aber angemessene Massnahmen zum Schutz der potentiell betroffenen Personen. Aus diesem Grund verlangt das Gesetz bei hohen Risiken eine sog. «Datenschutz-Folgenabschätzung». Anwendungen hingegen, die geradezu auf eine Aushöhlung der vom DSG geschützten Privatsphäre und informationellen Selbstbestimmung abzielen, sind datenschutzrechtlich verboten. Gemeint sind insbesondere KI-basierte Datenbearbeitungen, die in autoritär regierten Staaten zu beobachten sind, wie die flächendeckende Gesichtserkennung in Echtzeit oder die umfassende Observation und Bewertung der Lebensführung, das sog. «Social Scoring».

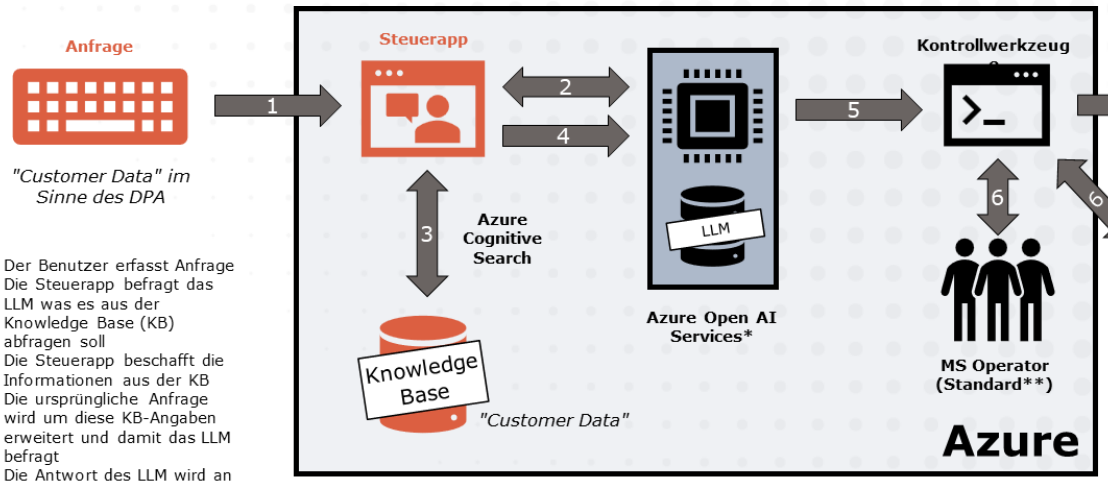
Hier werden jedoch gesetzliche Vorgaben und "ethische" Erwartungen vermengt; eine solche absolute Pflicht gibt es in der Schweiz nicht

Schritt 2: Tools und Provider

- Arbeitet der Provider **nur für mich** oder auch **für sich**?
  - Für sich: Training seines Modells, Werbung, Inhaltskontrollen
- **Variante A:** Er ist Auftragsbearbeiter
  - An sich unproblematisch – wie jedes "Outsourcing"
  - Datensicherheit gewährleistet? Vertrauenswürdig?
  - Braucht Auftragsverarbeitungsvertrag
  - Falls nicht im EWR: Zusätzliche Vereinbarung oder Zertifizierung
- **Variante B:** Er ist eigener Verantwortlicher
  - Problematisch → näher prüfen
  - Erfordert bei Personendaten i.d.R., dass dies vorher transparent gemacht worden ist und dem nicht widersprochen wurde

Retrieval-Augmented Generation

Beispiel Microsoft mit OpenAI



1. Der Benutzer erfasst Anfrage
2. Die Steuerapp befragt das LLM was es aus der Knowledge Base (KB) abfragen soll
3. Die Steuerapp beschafft die Informationen aus der KB
4. Die ursprüngliche Anfrage wird um diese KB-Angaben erweitert und damit das LLM befragt
5. Die Antwort des LLM wird an Apps von Microsoft zur Prüfung der Einhaltung der Vorgaben gemeldet
6. Bei Bedarf wird die Antwort manuell überprüft
7. Die Antwort wird ausgespielt

Von/für Kunde

Von Microsoft

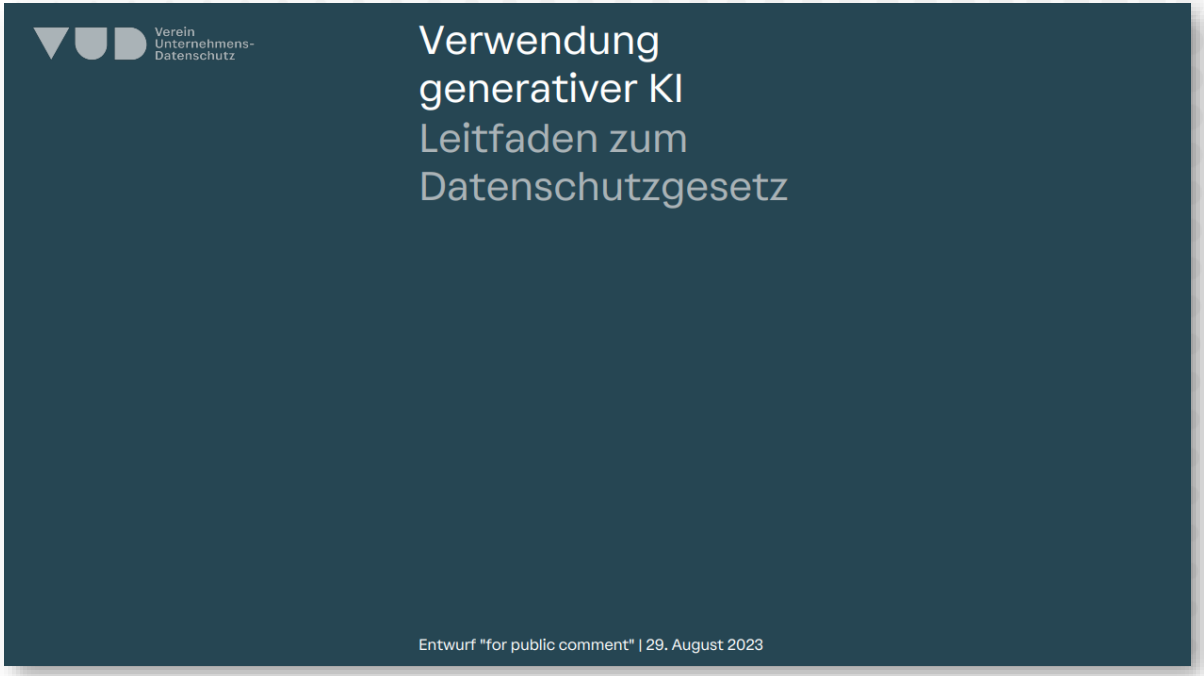
→ Einsatz von Azure im Rahmen der bestehenden Vertragsbedingungen

Infos: https://learn.microsoft.com/en-us/legal/cognitive-services/openai/data-privacy

* Das LLM speichert keine Customer Data und benutzt sie nicht für sein Training

** Der Kunde kann unter Umständen die Prüfung der Antwort durch Microsoft abwählen (und sie mit den Werkzeugen von Microsoft ggf. selbst vornehmen)

Weitere Informationen



www.vud.ch

Schritt 3: Andere Rechtsgebiete (Auswahl)

- **(Urheber-)Rechte an Inhalten**
  - Habe ich die Rechte der KI zu füttern, was ich ihr füttern will?
  - Wer hat die Rechte am Output? Stecken fremde Werke drin?
  - Was sagt der Vertrag mit dem Anbieter dazu?
- **Geheimnispflichten**
  - Enthält der Input oder Output Geheimnisse Dritter (planmässig oder fehlerbedingt), die so nicht bekanntgegeben werden dürfen?
  - Betr. Provider: Passende(r) Vertrag & "TOMS" genügt i.d.R.
- **Lauterkeitsrecht**
  - Wird das Publikum irgendwie irregeführt oder getäuscht?
  - Werden fremde Arbeitsergebnisse schmarotzerisch verwertet?

Schritt 4: Interne Hausaufgaben

- **Nutzung der KI regeln und schulen**
  - Vorgaben und Schulung für Projekte & Mitarbeitende
- **Einsatz der KI intern und extern dokumentieren**
  - "ROPA" (Art. 12 DSG) und "ROAIA" ("Records of AI Activities")
  - Datenschutz-Erklärung wo nötig anpassen (Zwecke, Empfänger)
- **Risikobeurteilung vornehmen und dokumentieren**
  - Datenschutz-Folgenabschätzung (DSFA, Art. 22 DSG)
  - Für grössere Vorhaben: Generative AI Risk Assessment (GAIRA)
  - Ziel: Mögliche negative Auswirkungen für Betrieb und Betroffene und Lücken bei den Schutzvorkehrungen identifizieren
  - Auch "ethische" Vorgaben berücksichtigen, ggf. Gremium bilden

Tool für eine Datenschutz-Folgenabschätzung

Datenschutz-Folgenabschätzung (DSFA)
Version 25.9.2023 for public comment - Private CH-DSG/DSGV

Hinweis: Eine Anleitung zum Ausfüllen dieser DSFA und zur KI-gestützten Ausfüllhilfe (optional, nur in der Version des Excels mit Makros) findet sich am Ende dieses Arbeitsblatt

Unternehmen (Verantwortlicher): Musterfirma AG

Abteilung:
Verantwortlich intern:
Status der DSFA:
Name des Vorhabens:
Aktivität gemäss Bearbeitu

1. Beschreibung de
1.01 In welchem Bereic bzw. welcher Gesc
1.02 Was vorgesehen i
1.03 Welche Interesser
1.04 Welche Mittel unc
1.05 Welche Dritten an
1.06 Welche Daten bea
1.07 Wessen Daten bei
1.08 Wo überall Daten
1.09 Wann die Datenb
1.10 Weitere Besonder

2. Erforderlichkeit /
2.01 Warum die Daten Sicht der betroffer eingreifende Vorg
2.02 Warum die Datenbearbeitung datensparsam, zeitlich auf das nötige begrenzt und auch sonst verhältnismässig ist:

4. Risiken von negativen Folgen für die betroffenen Personen, die trotz der obigen Massnahmen verbleiben

10 Risiken vorschlagen (überschreibt bisherige Werte)

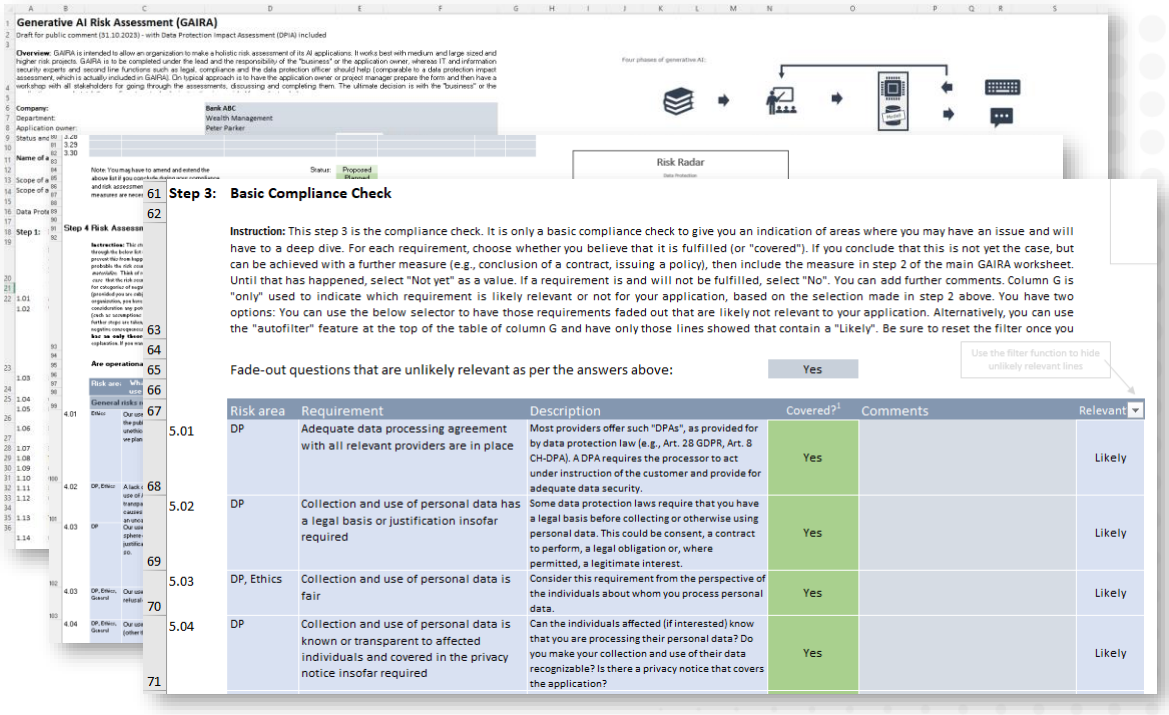
Hinweis: Falls die ermittelten Risiken als zu hoch erscheinen oder sich zeigt, dass es noch weitere Massnahmen zur Minimierung gibt, sollten diese oben unter Ziff. 3 eingetragen werden und bei der Risikobeurteilung hier berücksichtigt werden.

	Mögliche unerwünschte negative Folgen Weiteres Risiko vorschlagen*	Was wir dagegen tun Massnahmen vorschlagen* / Aus obigen formulieren*	Wie wir das Restrisiko einschätzen Risikobeurteilung vorschlagen*	Mögliche Folgen für die Person	Eintrittswahrscheinlichkeit (alles in allem)	Risiko (1-16)
4.01	Personendaten des Vorhabens gelangen wegen eines Fehlers oder absichtlich an **unbefugte Dritte**. Diese missbrauchen sie zum Schaden der betroffenen Personen.	- Berechtigungskonzept: Da wir nur autorisierten Personen Zugriff auf die Personendaten geben, wird das Risiko von unbefugtem Zugriff und Missbrauch reduziert. - Schulung: Durch Schulungen stellen wir sicher, dass die Mitarbeitenden die Lösung korrekt und sicher nutzen, was das Risiko von Fehlern und Missbrauch verringert. - Zugriffskontrolle: Durch die Beschränkung des Zugriffs auf autorisierte Personen können wir Missbräuche und unbefugte Nutzung der Personendaten verhindern. - Verschlüsselung "at rest": Die Verschlüsselung der Personendaten in unserem System schützt vor unautorisiertem Zugriff, falls jemand physischen Zugriff auf die Speichermedien erhält. - Datenlöschungsfunktionen: Durch die Möglichkeit, nicht mehr benötigte Personendaten zu löschen oder zu anonymisieren, minimieren wir das Risiko eines unbefugten Zugriffs auf diese Daten.	Das konkrete Restrisiko für die betroffene Person besteht darin, dass ihre Personendaten aufgrund eines Fehlers oder absichtlich an unbefugte Dritte gelangen könnten. Diese könnten die Daten dann zum Schaden der betroffenen Person nutzen, beispielsweise für Identitätsdiebstahl oder Missbrauch in sozialen Medien. Die Wahrscheinlichkeit dieses Szenarios ist jedoch insgesamt gering, da strenge Sicherheitsmassnahmen wie Zugriffskontrollen und Verschlüsselung implementiert wurden.	Substanziell	Tief	Mittel (6)
4.02	Personendaten des Vorhabens gelangen wegen eines Fehlers oder absichtlich an eine **unbefugte interne Person**.					

Zugriffen geschützt werden.
Die Datenbearbeitung ist datensparsam, da nur der Stimmabdruck, die ID der Person und Tonaufnahmen gespeichert werden, die für die Identitätsverifizierung notwendig sind. Die Datenbearbeitung ist zeitlich begrenzt, da der Stimmabdruck bei jedem Anruf neu erstellt und nicht länger als nötig gespeichert wird. Die Datenbearbeitung ist verhältnismässig, da sie zur Sicherheit der Anrufer im Call-Center beiträgt und die einzigen Daten bearbeitet werden, die dafür erforderlich sind.

https://vud.ch/dsfa

Tool für ein GenAI Risk Assessment (GAIRA)

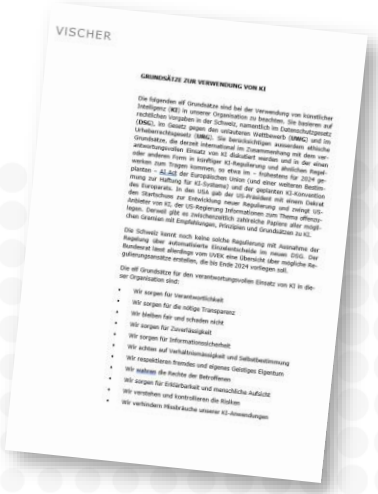


- Risiko-Assessment beinhaltet zugleich eine Datenschutz-Folgenabschätzung (DSFA)
- Folgt hinsichtlich der Methodik dem Prinzip einer DSFA
- Separate Compliance-Checkliste
- ROAIA-Vorlage

Kostenlos abrufbar unter https://bit.ly/gaira

11 Vorgaben für KI-Projekte

- Wir sorgen für Verantwortlichkeit
- Wir sorgen für die nötige Transparenz
- Wir bleiben fair und schaden nicht
- Wir sorgen für Zuverlässigkeit
- Wir sorgen für Informationssicherheit
- Wir achten auf Verhältnismässigkeit und Selbstbestimmung
- Wir respektieren fremdes und eigenes Geistiges Eigentum
- Wir wahren die Rechte der Betroffenen
- Wir sorgen für Erklärbarkeit und menschliche Aufsicht
- Wir verstehen und kontrollieren die Risiken
- Wir verhindern Missbräuche unserer KI-Anwendungen


VISCHER

GRUNDSÄTZE ZUR VERWENDUNG VON KI

Grundsätze für die Verwendung von KI im Unternehmen (10 Seiten)

Wie Betriebe mit KI umgehen (sollten)

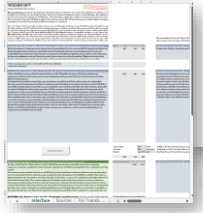
Rechtliche und ethische Vorgaben für KI-Anwendungen in Betrieben (hier: Die 11 Grundsätze für KI für Organisationen von VISCHER)

Vorgaben für einzelne Anwendungen

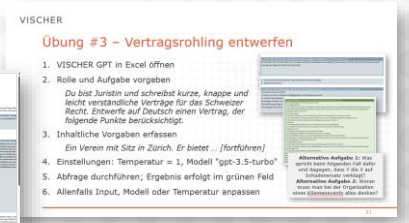
Hilfestellungen für Mitarbeitende (hier: VISCHER GenKI Cheat-Sheet)

GAIRA zur Risikobeurteilung vor KI-Projekten

Verzeichnis: Records of AI Activities (ROAIA)

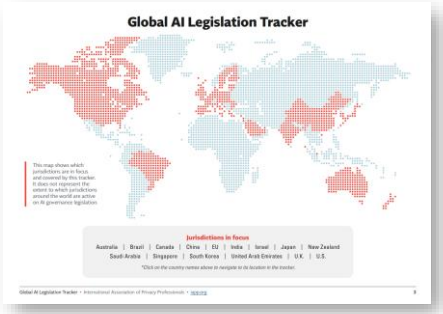


Geprüfte Tools



Schulung der Mitarbeitenden im sicheren und nutzbringenden Einsatz von GenKI

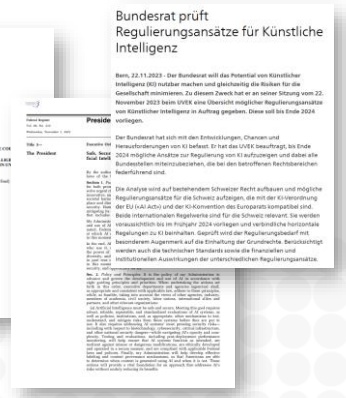
Und der Gesetzgeber?



IAPP Global AI Legislation Tracker (https://iapp.org/resources/article/global-ai-legislation-tracker/)

- Weltweit im Lead: **EU AI Act**
  - Noch in Diskussion
  - Produkteregulierung, nicht primär GenKI bzw. generische Modelle
  - Verbietet Aktivitäten (z.B. Emotionserkennung am Arbeitsplatz oder in der Schule)
  - Definiert "Hoch-Risiko"-KI (z.B. biometrische Identifikation, Produkte für Arbeitgeber, Betrieb kritischer Infrastrukturen) und Pflichten für Anbieter solcher (z.B. Risikomanagement, Selbstzertifizierung, Meldepflicht, Überwachung)
  - Einige wenige weitere Pflichten (z.B. Transparenz bei Interaktion mit KI und Deep Fakes)

AI Act (Draft) (EU)



Executive Order 14110 (USA)

www.sbfi.admin.ch/ki

Schlussbemerkungen

- **Bisherige Regeln** gelten und passen oft auch bei GenKI
- Rechtliche Wogen werden sich **glätten**, und die Unsicherheit beim Einsatz von GenKI wird schwinden
- Hautprobleme liegen beim **Output** und dessen Verwendung, und dort mehr in der **Richtigkeit** als der Transparenz
- GenKI-Projekte sind oft auch **Cloud-Projekte** – und fehlende **Transparenz** seitens der Provider die Herausforderung
- Eine **strukturierte Risiko-Beurteilung** hilft, die Sache in den Griff zu bekommen
- Gefahr der Überregulierung und **Vermengung Ethik/Gesetz**
- **Keine Angst** haben vor GenKI, auch nicht rechtlich

VISCHER

Danke für Ihre Aufmerksamkeit!

Fragen: drosenthal@vischer.com

Zürich
Schützengasse 1
Postfach
8021 Zürich, Schweiz
T +41 58 211 34 00

Basel
Aeschenvorstadt 4
Postfach
4010 Basel, Schweiz
T +41 58 211 33 00

Genf
Rue du Cloître 2-4
Postfach
1211 Genf 3, Schweiz
T +41 58 211 35 00

www.vischer.com


www.jusletter.ch

David Rosenthal

Datenschutz beim Einsatz generativer künstlicher Intelligenz

Ist der datenschutzkonforme Einsatz generativer künstlicher Intelligenz möglich? Eine Analyse zeigt: Das Schweizer Datenschutzgesetz kommt auch mit dieser technischen Entwicklung gut zurecht. Es braucht nicht weitere Grundsätze und Pflichten, sondern neue technische und organisatorische Massnahmen, mit denen das bestehende Recht sinnvoll umgesetzt werden kann. Dabei ist klar zwischen dem zu unterscheiden, was der Gesetzgeber vorschreibt, und dem, was aus «ethischen» Überlegungen wünschenswert sein mag. Dieser Beitrag beschäftigt sich bewusst nur mit Ersterem. Er setzt einen früheren Beitrag des Autors zum Datenschutz bei KI-Systemen fort.

Beitragsart: Wissenschaftliche Beiträge
Rechtsgebiete: Datenschutz, Medienrecht

Zitiervorschlag: David Rosenthal, Datenschutz beim Einsatz generativer künstlicher Intelligenz, in: Jusletter 6. November 2023

https://www.rosenthal.ch/downloads/Rosenthal-Jusletter-GenKI-Datenschutz.pdf