VISCHER

KI-Projekte.

So meistern Sie die rechtlichen Aspekte

David Rosenthal, Partner, VISCHER AG
10., 12. und 13. Juni 2024

Streit über KI-Entwicklung

New York Times klagt gegen Micro
OpenAI

FORMEL-1-PILOTEN

Rechtsstreit um Fake-KI-Interview mit Michael Schumacher

TRANSPARENZ BEI KI

"Aktuell herrscht hier ziemliche Rechtsunsicherheit"

RECHT 30.10.2023

KI-verursachte Schäden: Wann haftet der Zahn-(Arzt)?

Getty Images und Adobe

KI-Training: Wie Getty Images und Adobe die Rechtsunsicherheit zu ihrem Vorteil nutzen

Quellen: Horizont.net, zwp-online.info, thepioneer.de, tagesschau.de, golem.de

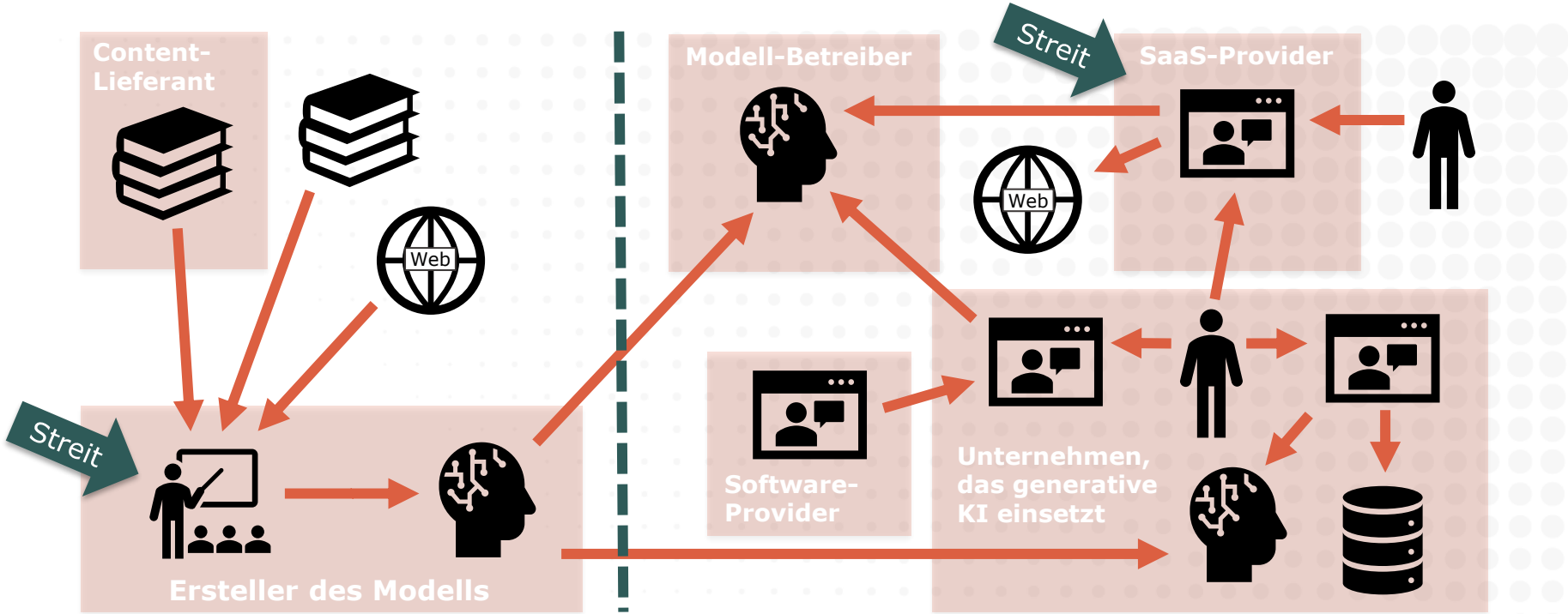
Rechtsunsicherheit?

- **Ja**, weil die Materie für uns unheimlich und ungewohnt ist
  - Wissen Sie, was in einem LLM steckt und warum es so gut ist?
  - Wir hatten beim Internet früher dieselbe Situation, und heute haben wir uns daran gewöhnt – es ist völlig normal geworden
- **Unbehagen** führt zum Ruf nach mehr Regulierung und "Ethik"
  - Beispiele: "Transparenz", "Nicht-Diskriminierung", "Erklärbarkeit"
  - EU AI Act als Reaktion (primär punktuelle Produkte-Regulierung)
  - Bundesrat will bis Ende Jahr Schweizer Regulierungsbedarf klären
- Das **bestehende Recht** regelt viele der Themen recht gut
  - Datenschutz, Urheberrecht, Lauterkeitsrecht, Geheimnisschutz
  - Viele 0815-Hausaufgaben und einzelne neue Herausforderungen

25. Juli 1994 (time.com, Titelseite: James Porto)

"KI" schon vielerorts im Einsatz: OCR, biometrische Authentifizierung, Stauwarner, Übersetzung, Malware-Erkennung, Sprachsteuerung etc.

Verantwortlichkeit für generative KI



Checkliste: 18 KI-Compliance-Schlüsselfragen.

KI = System, das Ergebnisse auf Basis eines Trainings und nicht nur einer Programmierung erzeugt

Unter vischer.com/ki finden Sie kostenlose Ressourcen zu diesen Themen sowie zu KI-Governance und Risikomanagement (keine Registrierung erforderlich)

vischerlnk.com/ki-compliance-kurz

bit.ly/3WNgxeO

Die auch sonst üblichen Fragen; haben Sie sie unter Kontrolle, so wie auch Ihre Provider

Datenschutz

- Haben wir einen angemessenen Vertrag mit den von uns genutzten Providern (z.B. einen ADV, EU SCC, Verbot der Eigennutzung unserer Daten)?
- Haben wir die Leute über die Zwecke informiert, zu denen wir Daten von ihnen bearbeiten oder erzeugen?
- Haben wir es im Griff, wenn die KI falsche oder anderweitig unzulässige Daten über sie produziert?
- Wenn eine KI wichtige Entscheidungen über sie trifft, können sie diese von einem Menschen prüfen lassen?
- Ist unsere KI vor Missbrauch und Angriffen geschützt und auch sonst sicher, insbesondere, wo wir Dritten die Nutzung erlauben (z.B. Chatbot)?
- Können wir Auskunfts- und Berichtigungsbegehren wie erforderlich umsetzen?
- Haben wir eine Risikobeurteilung für unser Vorhaben (inklusive einer DSFA) durchgeführt?

Wem geben Sie geheime Daten weiter? Halten Sie Ihre Verträge ein

Vertragspflichten, Geheimhaltung

- Kommen wir unseren Geheimhaltungspflichten nach (z.B. beim Einsatz von Providern, Verhinderung der unerwünschten Preisgabe von Daten)?
- Untersagen unsere Verträge die von uns ins Auge gefasste Anwendung (z.B. NDA, welches die Nutzung von Daten für unsere Zwecke einschränkt)?

Schutz von Inhalten Dritter

- Füttern wir KI-Systeme nur dann mit Inhalten Dritter, soweit unsere Lizenzen oder die gesetzlichen Schranken des Urheberrechts dies zulassen?
- Vermeiden wir die Erstellung von Inhalten, die bereits bestehenden Inhalten Dritter entsprechen?

Mit GMV in der Regel kein Problem für Anwender

EU AI Act (noch nicht in Kraft)

- Ist klar, dass wir entweder nicht unter den EU AI Act fallen oder unser Vorhaben keine verbotene Praktik ist und möglichst auch kein "Hoch-Risiko"-KI-System (und gehen wir ansonsten richtig damit um)?
- Wenn eine KI "Deep Fakes" erstellt oder mit Menschen interagiert oder sie beobachtet, werden sie dann darauf hingewiesen gemacht?

AI Act betrifft nur bestimmte KI-Systeme (kann aber in der CH gelten)

Andere (auch ethische) Aspekte

- Vermeiden wir Diskriminierung beim Einsatz von KI?
- Behält der Mensch (wirklich) die Kontrolle über die KI?
- Können wir unsere KI-Ergebnisse rechtfertigen/erklären?
- Sagen wir es den Leuten, wie wir KI einsetzen, wenn es für sie unerwartet sein könnte, und erlauben wir ihnen gar, sich für oder gegen deren Einsatz zu entscheiden?
- Haben wir ein angemessenes KI-Testing, angemessene Überwachung und ein angemessenes Risk-Management?

Was jedes Unternehmen für sich selbst ermitteln muss an weiteren Vorgaben

Use Case: Analyse eines Lebenslaufs

- Analyse eines **Stellenbewerber-CV** mit Hilfe eines Online-LLM
- Haben wir den KI-Provider **korrekt beauftragt** (AVV)? Ist eine Zweitverwertung der Personendaten durch ihn ausgeschlossen?
- Bearbeiten wir die Daten nur für den angedachten **Zweck**?
- Muss der Bewerber erwarten, dass wir das tun oder müssen wir es ihm sagen? Ist es in der **Datenschutzerklärung** erwähnt?
- Ist es ihm **zuzumuten**, dass sein CV so analysiert wird?
- Sind die (der KI gefütterten und die von ihr generierten) Personendaten im Hinblick auf den Zweck **korrekt**?
- Wird eine relevante automatisierte **Einzelentscheidung** getroffen?

Eine solche Nutzung von KI ist ein "hohes Risiko" im Sinne des AI Act

Leider wenig Transparenz bei den Anbietern

GÄNGIGE KI-TOOLS: WIE STEHT ES UM DEN DATENSCHUTZ?

Gesamtübersicht zu Teil 2 unserer KI-Blog-Serie

	ChatGPT Free	ChatGPT Plus	ChatGPT Team	ChatGPT Enterprise	OpenAI API	Microsoft Copilot	Microsoft Copilot Pro	Microsoft Copilot mit kommerziellem Datenschutz	Microsoft Copilot für Microsoft 365	Microsoft Azure OpenAI Service	Google Bard	Google Vertex AI
Nutzungs-bedingungen	Terms of use Ab 15.2.2024 Europe Terms of Use		Business terms			Microsoft-Servicevertrag und zusätzliche Bedingungen	Microsoft-Servicevertrag	Microsoft-Servicevertrag und zusätzliche Bedingungen	Microsoft Customer Agreement und Microsoft Product Terms		Bard Privacy Hub Nutzungsbedingungen	Nutzungsbedingungen der Google Cloud Platform
Auftragsbearbei-tungsvertrag (DPA)	Nicht verfügbar		OpenAI Data processing addendum			Nicht verfügbar	Nicht verfügbar	Nicht verfügbar	Microsoft Products and Services Data Protection Addendum (DPA)		Nicht verfügbar	Cloud Data Processing Addendum (Customers)
Nutzung mit Personendaten	Nein		Möglich (wenn DPA abgeschlossen)			Nein	Nein	Nicht empfohlen (DPA fehlt)	Möglich		Nein	Möglich
Nutzung mit vertraulichen Daten	Nein		Möglich (ausgeschlossen Amts- und Berufsgeheimnisse)			Nein	Nein	Nicht empfohlen (keine Vertraulich-keitsverpflichtung)	Möglich (gesetzliche Berufsgeheimnisse nur, wenn die richtigen Zusätze abgeschlossen sind)		Nein	Möglich (gesetzliche Berufsgeheimnisse nur, wenn zusätzliche vertragliche Zusicherungen von Google vorliegen)
Nutzung für eigene Zwecke des Anbieters (z.B. Training, Service-verbesserung)	Ja (Nutzung fürs Training kann aber deaktiviert werden)		Nein			Ja (Nutzung fürs Training und auch für weitere Zwecke - inklusive Veröffentlichung der Daten durch Microsoft)	Unklar (gemäss Daten-schutzerklärung wohl zumindest teilweise möglich)	Nein	Nein		Ja	Nein
Einsatz im Unternehmen	Eingeschränkt möglich (keine Personendaten / keine vertraulichen Daten / Training deaktiviert)		Möglich			Nicht empfohlen	Nicht empfohlen	Eingeschränkt möglich (keine Personendaten / keine vertraulichen Daten)	Möglich		Nicht empfohlen	Möglich

Diese Tabelle dient nur zur Information und stellt keine Rechtsberatung dar. Die Angaben, insbesondere zur Nutzungsmöglichkeit, geben lediglich unsere Ansicht wieder und bedürfen einer Risikoeinschätzung durch das jeweilige Unternehmen sowie eine korrekte Konfiguration der Tools. Wir helfen Ihnen gerne dabei. Diese Ausführungen beziehen sich auf den Datenschutz, nicht aber auf andere Rechtsgebiete wie z.B. das Urheberrecht (bei urheberrechtlich geschütztem Material sollten Sie keine Inhalte von Dritten mit Tools verwenden, die Ihre Daten für ihre eigenen Zwecke nutzen). Bitte beachten Sie auch, dass sich die Regelungen und Angebote häufig ändern (der Stand der Information ist unten aufgeführt).

Weitere Informationen zu Recht und der Ethik beim Einsatz von Künstlicher Intelligenz finden Sie unter vischer.com/ki.

Eine Nutzung kann auch anonym sinnvoll möglich sein

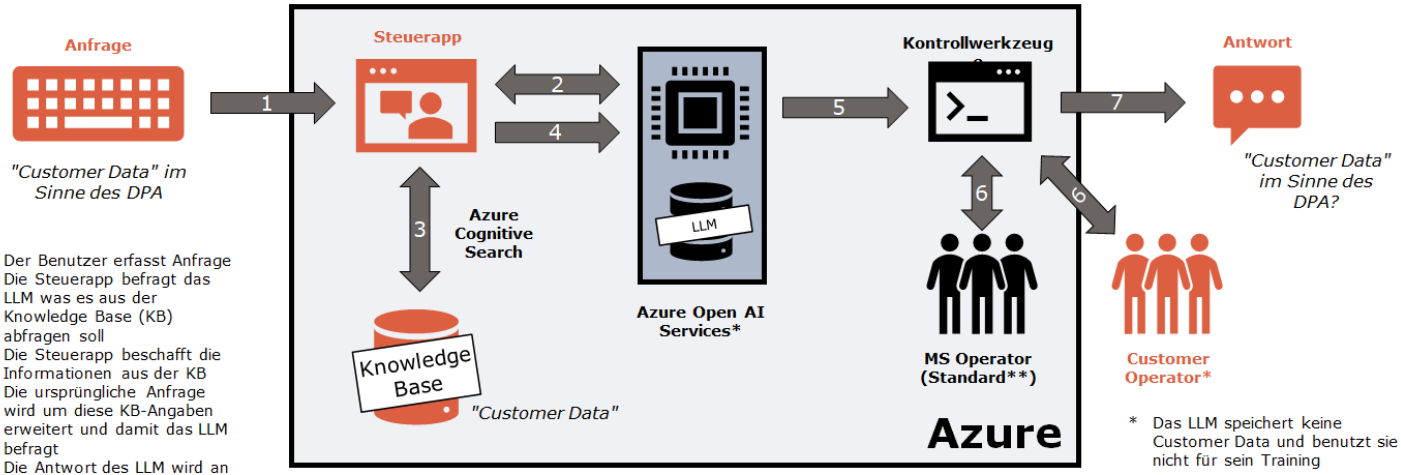
Nicht vergessen:

- Abuse Monitoring
- Nutzungsbeschränkungen

vischerlnk.com/ki-tools

LLM: 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

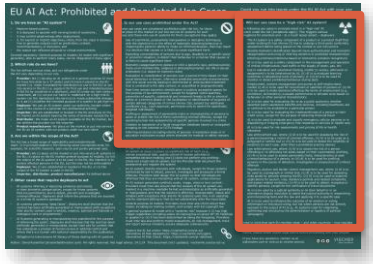
* Das LLM speichert keine Customer Data und benutzt sie nicht für sein Training

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

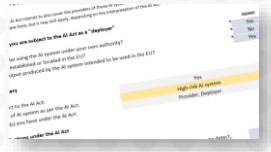
vischerink.com/3PnEqF4

Und der AI Act?

- Schweizer Unternehmen können erfasst sein, wenn sie …
  - KI-Produkte zum **Einsatz in der EU entwickeln**
  - KI verwenden und der **Output** in der EU benutzt wird
  - Nicht schon, wenn KI in der EU **läuft** oder Leute dort **betrifft**
- Besondere Vorgaben macht der AI Act für …
  - **Verbotene** KI-Anwendungen (z.B. KI-Emotionserkennung am Arbeitsplatz und in der Schule, Manipulation durch KI-Einsatz)
  - **Hoch-Risiko** KI-Anwendungen (z.B. regulierte Produkte, KI-Beurteilung von Mitarbeitenden/Schülern, KI-Bonitätsbewertung)
- Darüber hinaus: Nur sehr begrenzte Pflichten zur **Transparenz**
  - Z.B. Emotionserkennung, Wasserzeichen, Deep Fakes, Chatbots



vischerlnk.com/ai-act-uc



Siehe AI Act Check unter vischerlnk.com/gaira

AI Act: "Hoch-Risiko"-KI-Systeme vermeiden …

Anbieter sind unter anderem verpflichtet, (i) ein Risiko- und Qualitätsmanagement zu betreiben, (ii) eine Konformitätsbewertung durchzuführen und eine CE-Kennzeichnung mit ihren Kontaktdaten anzubringen, (iii) bestimmte Qualitätsniveaus für Schulungs-, Validierungs- und Testdaten zu gewährleisten, (iv) eine detaillierte technische Dokumentation bereitzustellen, (v) automatisches Protokollieren vorzusehen und Protokolle aufzubewahren, (vi) Anweisungen für Betreiber bereitzustellen, (vii) das System so zu gestalten, dass menschliche Aufsicht möglich ist, es robust, zuverlässig, gegen Sicherheitsbedrohungen (einschliesslich KI-Angriffe) geschützt und fehlertolerant ist, (vii) das KI-System behördlich zu registrieren, (viii) eine Überwachung des Systems nach seiner Markteinführung zu betreiben, (ix) Vorfälle den Behörden zu melden und Korrekturmassnahmen zu ergreifen, (x) mit den Behörden zusammenzuarbeiten, (xi) die Einhaltung der vorstehenden Anforderungen zu dokumentieren und (xii) einen Vertreter in der EU zu haben, falls der Anbieter selbst nicht in der EU ansässig ist, aber dem AI Act unterliegt.

Betreiber sind unter anderem verpflichtet, (i) den Anweisungen des Anbieters zu folgen, (ii) angemessene menschliche Aufsicht zu gewährleisten, (iii) automatisch generierte Protokolle mindestens sechs Monate lang aufzubewahren, (iv) angemessenen Input zu gewährleisten, (v) an der Überwachung des KI-Systems nach seiner Einführung durch den Anbieter teilzunehmen, (vi) schwere Vorfälle und bestimmte Risiken den Behörden und dem Anbieter zu melden, (vii) Mitarbeiter zu informieren, falls das KI-System sie betrifft, (viii) betroffene Personen über Entscheidungen zu informieren, die durch oder mit Hilfe des KI-Systems getroffen wurden, und (ix) Anfragen betroffener Personen bezüglich solcher Entscheidungen zu befolgen.

Offizielle Schätzung: Max. 5-10% der KI-Systeme

Quelle: vischerlnk.com/gaira

KI-Governance: Sechs Schritte

- Voraussetzungen schaffen: Robustes **Data Management**
- Aufgaben, Kompetenzen und Verantwortlichkeiten (**AKV**) regeln
- **Richtlinie** mit Vorgaben zum Umgang mit KI um Mitarbeitende "sicher" zu machen und einen KI-Einsatz zu ermöglichen
- **Schulung** im sicheren und verantwortungsvollen Umgang mit KI und Vermittlung von KI-Kenntnissen – bis zur GL und zum VR, damit die Risiken bekannt sind und übernommen werden können
- **Map & Track** von (relevanter) KI im Unternehmen
- **Risiko-Management** für KI-Vorhaben und Tools (heisst: die wichtigsten Risiken beurteilen und Massnahmen dazu treffen)

KI-Projekte: Risikobeurteilung

- Differenzieren nach **Risiko für das Unternehmen**
  - Nur wenige Vorhaben werden hohe Risiken haben
  - Tipp: Risikoschwellen für Prüftiefe definieren
- Risiko eines KI-Projekts **strukturiert beurteilen**
  - Rechtliche, operationelle, reputative und "ethische" Risiken
  - Finden und Ausmerzen von Schwachstellen
  - Nicht durch Fachleute, sondern durch das "Business", welches auch entscheiden (und verantwortlich sein) muss
  - Alles dokumentieren, inklusive die getroffenen Massnahmen
  - Tipp: https://vischerlnk.com/gaira (Light & Comprehensive)
  - Überwachung im Betrieb nicht vergessen

GAIRA Light

vischerlnk.com/
ki-risikocheck

Use Case: Chatbot auf der Website

Source: WashingtonPost.com

Wieso sollte sie für fehlerhafter Auskünfte nicht bezahlen? Wo ist der Unterschied zum Call Center? Und lohnt es sich nicht trotzdem?

- Wie riskant ist der Bereich, um den es geht? Wie wird die **Thementreue** sichergestellt? Wie gut wurde der Bot getestet?
- Ist der Chatbot auf eigene, "gute" Daten begrenzt (sog. **RAG**)?
- Wie wird kommuniziert, wie verlässlich ist die Auskunft, die der Chatbot erteilt? Pauschalvorbehalt auf der Website (schwächer) oder im Output selbst integrierter **Vorbehalt** und Vermeidung von Einzelfallauskünften via *Alignment* (besser/stärker)?
- Ist eine **Eskalation** an den Menschen vorgesehen? Mittels Themen und Konfidenzschwellen? Über Standardhinweise?
- Welche Massnahmen gibt es gegen (unerwünschten) **Bias**?
- Wird geloggt? Werden **Logs** und User-**Feedback** ausgewertet?

1. **Haben wir ein vernünftiges Set an Massnahmen getroffen?**
2. **Was sind die Restrisiken?**
3. **Sind sie akzeptabel und lohnt sich das ganze wirklich?**

VISCHER

Danke für Ihre Aufmerksamkeit!

Fragen: david.rosenthal@vischer.com

Mehr Unterlagen:
www.vischer.com/ki
www.rosenthal.ch

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