VISCHER

Ein Blick hinter die Kulissen.

KI-Risiken in der Praxis und wie damit umgehen

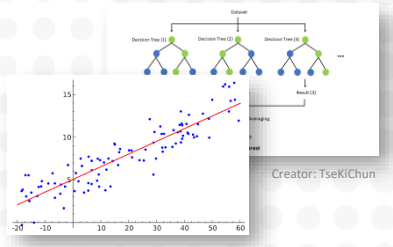
David Rosenthal, VISCHER AG
12. März 2026

Agenda

1. KI-Provider – datenrechtliche Fragen
2. Weitere Datenschutz-Fragen
3. Urheberrecht
4. Vertragsrecht (und Verantwortlichkeit)
5. EU AI Act (und die Schweiz)
6. Regulatorische Anforderungen (FINMA)
7. Risikobasierter Ansatz und Risikobeurteilung
8. Übergeordnete Herausforderungen
9. Governance und praktische Umsetzung

Was ist KI rechtlich überhaupt?

- Gemäss **EU AI Act** "ein maschinengestütztes System, das für einen in unterschiedlichem Grade autonomen Betrieb ausgelegt ist und das nach seiner Betriebsaufnahme anpassungsfähig sein kann und das aus den erhaltenen Eingaben für explizite oder implizite Ziele ableitet, wie Ausgaben wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen erstellt werden, die physische oder virtuelle Umgebungen beeinflussen können."
- Das einzig praktisch relevante Element ist "**Autonomie**"
  - Einfach gesagt: Ein IT-System, das seine Entscheide auf Basis eines Trainings fällt (statt einer voll ausprogrammierten Logik)
- **Aber**: Diese Definition ist mangelhaft ...
  - Jedes Kopiergerät (OCR), jedes Mobiltelefon (Fingerabdruck-Entsperrung, Kamera) ist ein KI-System …



Welches Recht gilt beim Einsatz von KI?

- **Alles**
- Die in der Praxis **wichtigsten Bestimmungen** finden sich im
  - Urheberrecht
  - Datenrecht
  - Vertragsrecht
  - Lauterkeitsrecht
  - Regulatorische Anforderungen
  - EU AI Act (soweit anwendbar)
- **Hauptthema** sind Providerservices und Risikobeurteilungen

Checkliste: 18 KI-Compliance-Schlüsselfragen.

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

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

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, und haben wir wo nötig eine Rechtsgrundlage dafür?
- 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?

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?

EU AI Act (wird rollend von 2025-2027 Anwendung finden)

- 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?

Andere (auch ethische) Aspekte

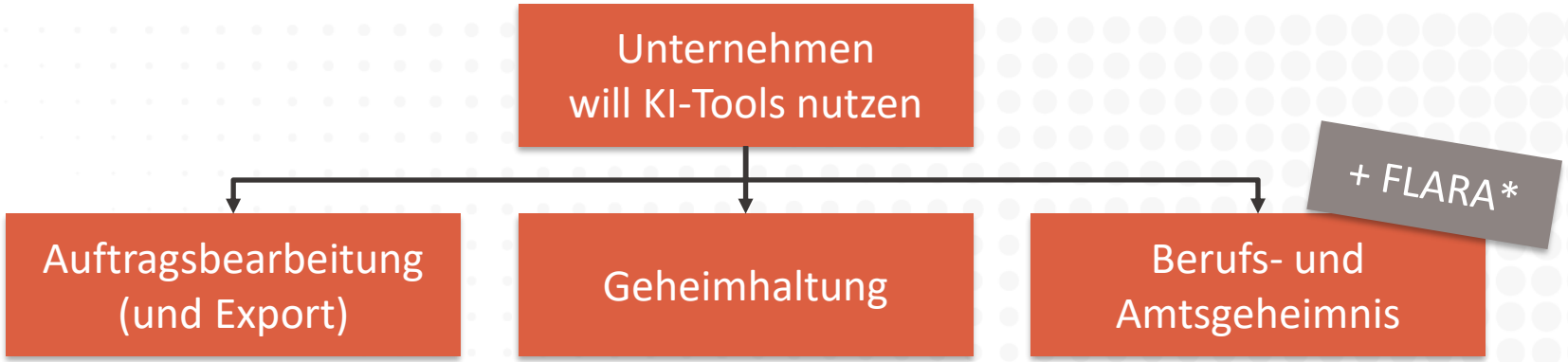
- Vermeiden wir Diskriminierung beim Einsatz von KI?
- Behält der Mensch (wirklich) die Kontrolle über die KI?
- Können wir unsere KI-Ergebnisse rechtfertigen/erklären?
- Sagen wir es den Leuten, wie wir KI einsetzen, wenn es für sie unerwartet sein könnte, und erlauben wir ihnen gar, sich für oder gegen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

Autor: David Rosenthal (david.rosenthal@vischer.com) 5.11.24 Aktualisierungen: vischerlnk.com/ki-compliance-kurz

VISCHER

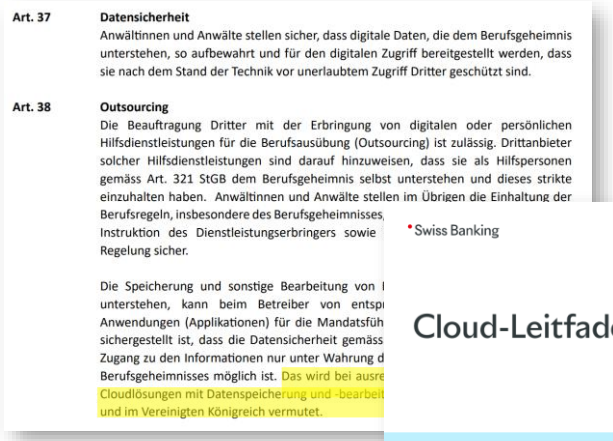
vischerlnk.com/ki-compliance-kurz & bit.ly/3WNgxeO

Thema 1: KI-Provider – datenrechtliche Fragen



* Foreign Lawful Access Risk Assessment

Mit Berufsgeheimnissen in die Cloud?



Art. 37 **Datensicherheit**
Anwältinnen und Anwälte stellen sicher, dass digitale Daten, die dem Berufsgeheimnis unterstehen, so aufbewahrt und für den digitalen Zugriff bereitgestellt werden, dass sie nach dem Stand der Technik vor unerlaubtem Zugriff Dritter geschützt sind.

Art. 38 **Outsourcing**
Die Beauftragung Dritter mit der Erbringung von digitalen oder persönlichen Hilfsdienstleistungen für die Berufsausübung (Outsourcing) ist zulässig. Drittanbieter solcher Hilfsdienstleistungen sind darauf hinzuweisen, dass sie als Hilfspersonen gemäss Art. 321 StGB dem Berufsgeheimnis selbst unterstehen und dieses strikte einzuhalten haben. Anwältinnen und Anwälte stellen im Übrigen die Einhaltung der Berufsregeln, insbesondere des Berufsgeheimnisses, Instruktion des Dienstleistungserbringers sowie Regelung sicher.

Die Speicherung und sonstige Bearbeitung von I unterstehen, kann beim Betreiber von entspr Anwendungen (Applikationen) für die Mandatsfüh sichergestellt ist, dass die Datensicherheit gemäss Zugang zu den Informationen nur unter Wahrung d Berufsgeheimnisses möglich ist. Das wird bei ausre Cloudlösungen mit Datenspeicherung und -bearbeit und im Vereinigten Königreich vermutet.

Swiss Banking

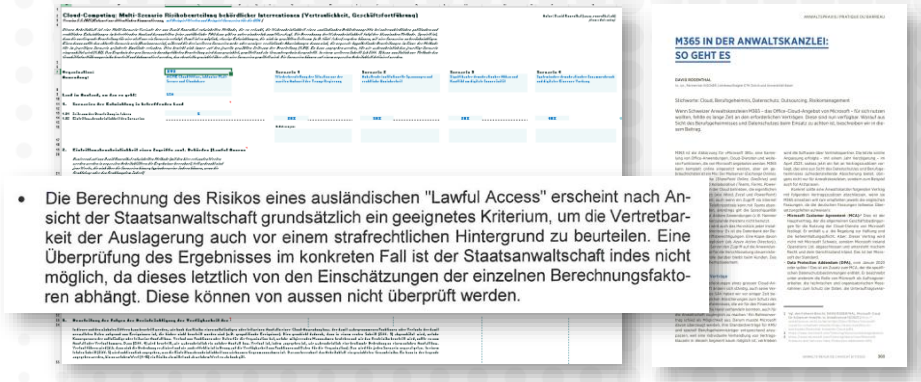
Cloud-Leitfaden

November 2025
Leitfaden der SBVg
3. Auflage

Schweizer Standesregeln der Rechtsanwälte (SSR)

Cloud-Leitfaden der Schweizerischen Bankiervereinigung

Es darf kein Grund zur Annahme bestehen, dass es zu einem ausländischen Behördenzugriff kommt



- Die Berechnung des Risikos eines ausländischen "Lawful Access" erscheint nach Ansicht der Staatsanwaltschaft grundsätzlich ein geeignetes Kriterium, um die Vertretbarkeit der Auslagerung auch vor einem strafrechtlichen Hintergrund zu beurteilen. Eine Überprüfung des Ergebnisses im konkreten Fall ist der Staatsanwaltschaft indes nicht möglich, da dieses letztlich von den Einschätzungen der einzelnen Berechnungsfaktoren abhängt. Diese können von aussen nicht überprüft werden.

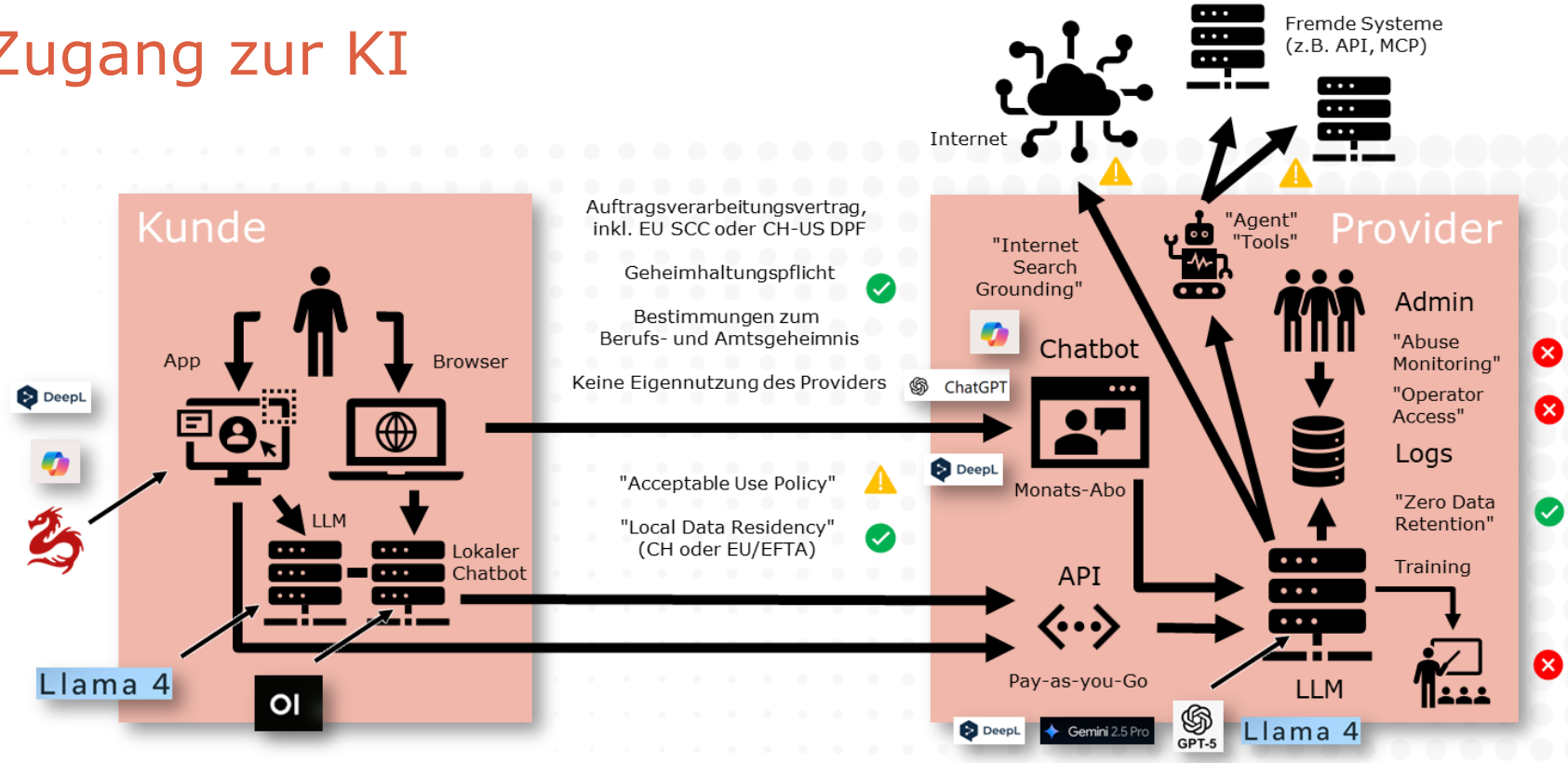
M365 IN DER ANWALTSKANZLEI: SO GEHT ES

vischerlnk.com/4ck2J0L

Beurteilung der Wahrscheinlichkeit eines Foreign Lawful Acess mit der "Methode Rosenthal"

vischerlnk.com/flara
vischerlnk.com/flarafaq
vischerlnk.com/3HAvcVB (zur Lage betr. Trump)

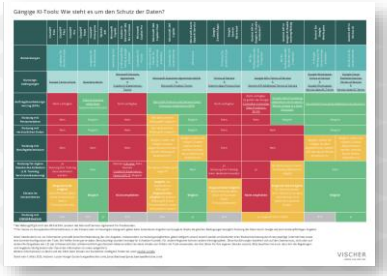
Zugang zur KI



Welche API-Provider?

Infos zu API-Provider: https://redink.ai

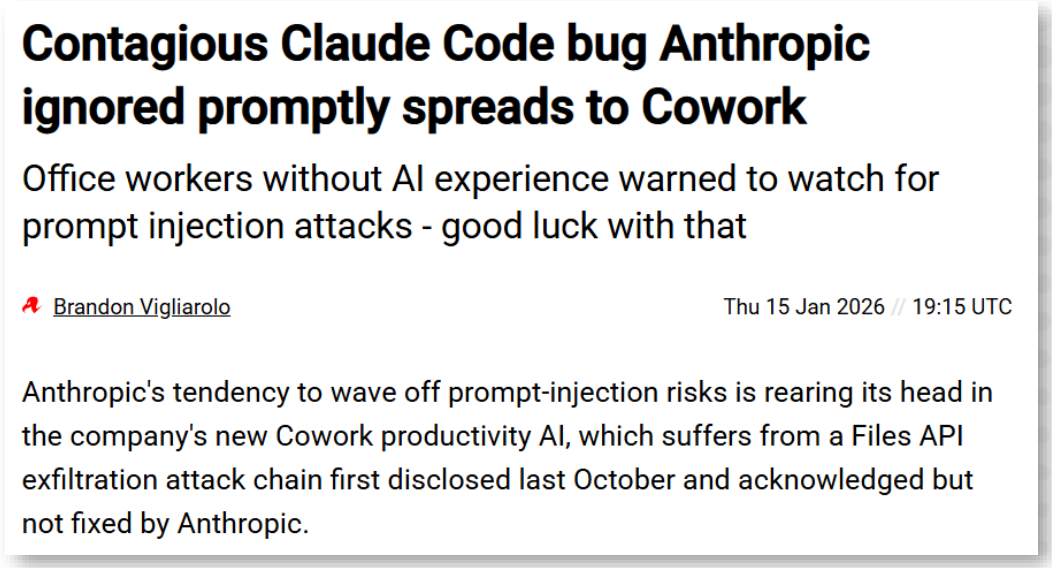
- **Cloud-Provider**
  - **Google, Microsoft:** Verträge mit Berufsgeheimniszusätzen und je nach Kunde auch Abuse Monitoring Opt-out
    - Trickreich: Einsatz von Tooling/Grounding ("Agentic AI")
  - **OpenAI, Anthropic, etc.:** Auftragsbearbeitungsverträge, aber oft nicht viel mehr, d.h. für wirklich sensible Daten nicht zu empfehlen (sind bisher auch nicht wirklich interessiert)



vischerlnk.com/ki-tools-0325

- **Schweizer Provider**
  - Anbieter wie **MTF**, **AlpineAI**, **SafeSwissCloud**
  - Sind in der Schweiz, mit Open Source-Modellen (USA, China)
  - Vertragszusatz zum Schutz des Amts- und Berufsgeheimnisses für Schweizer KI-Provider publiziert (vischerlnk.com/3IuUQLO)

Agenten …



Contagious Claude Code bug Anthropic ignored promptly spreads to Cowork

Office workers without AI experience warned to watch for prompt injection attacks - good luck with that

Brandon Vigliarolo — Thu 15 Jan 2026 // 19:15 UTC

Anthropic's tendency to wave off prompt-injection risks is rearing its head in the company's new Cowork productivity AI, which suffers from a Files API exfiltration attack chain first disclosed last October and acknowledged but not fixed by Anthropic.

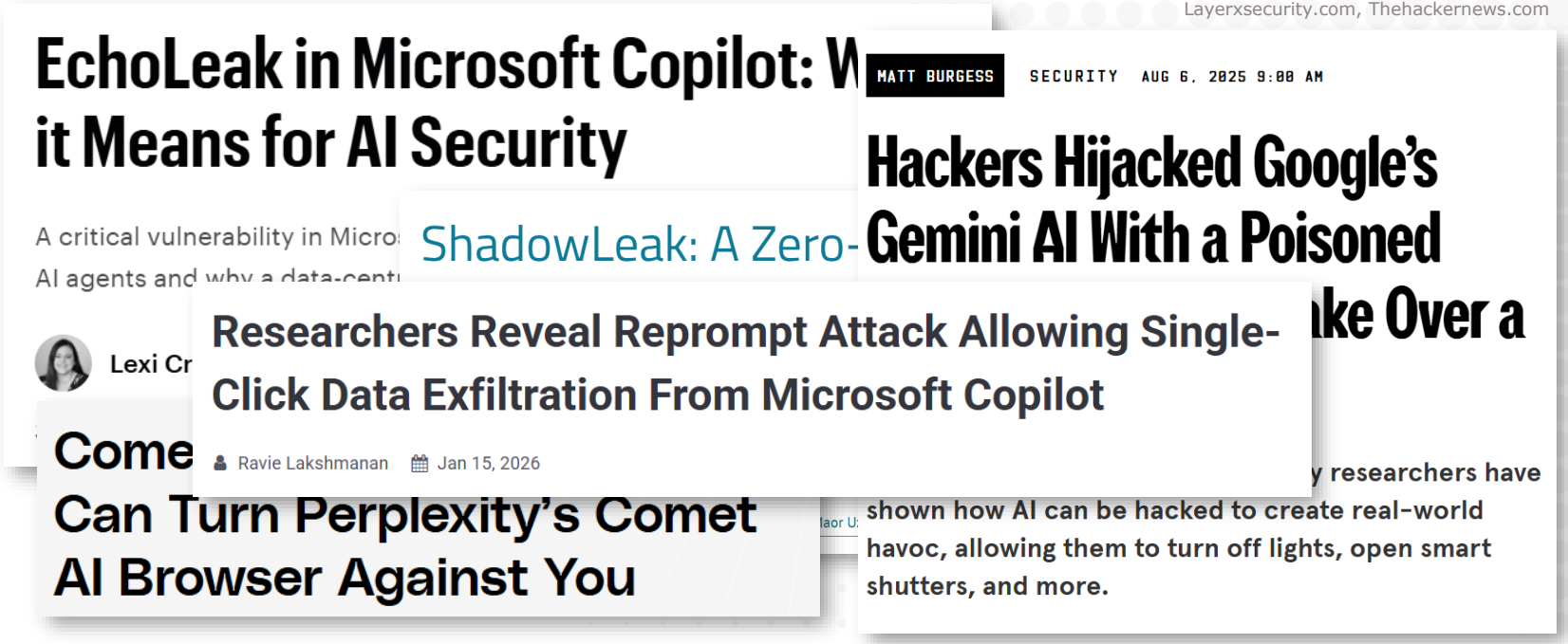
Source: TheRegister.com

"In order to trigger the attack, all a potential victim needs to do is connect Cowork to a local folder containing sensitive information, upload a document containing a hidden prompt injection, and voilà - when Cowork analyzes those files, the injected prompt triggers."

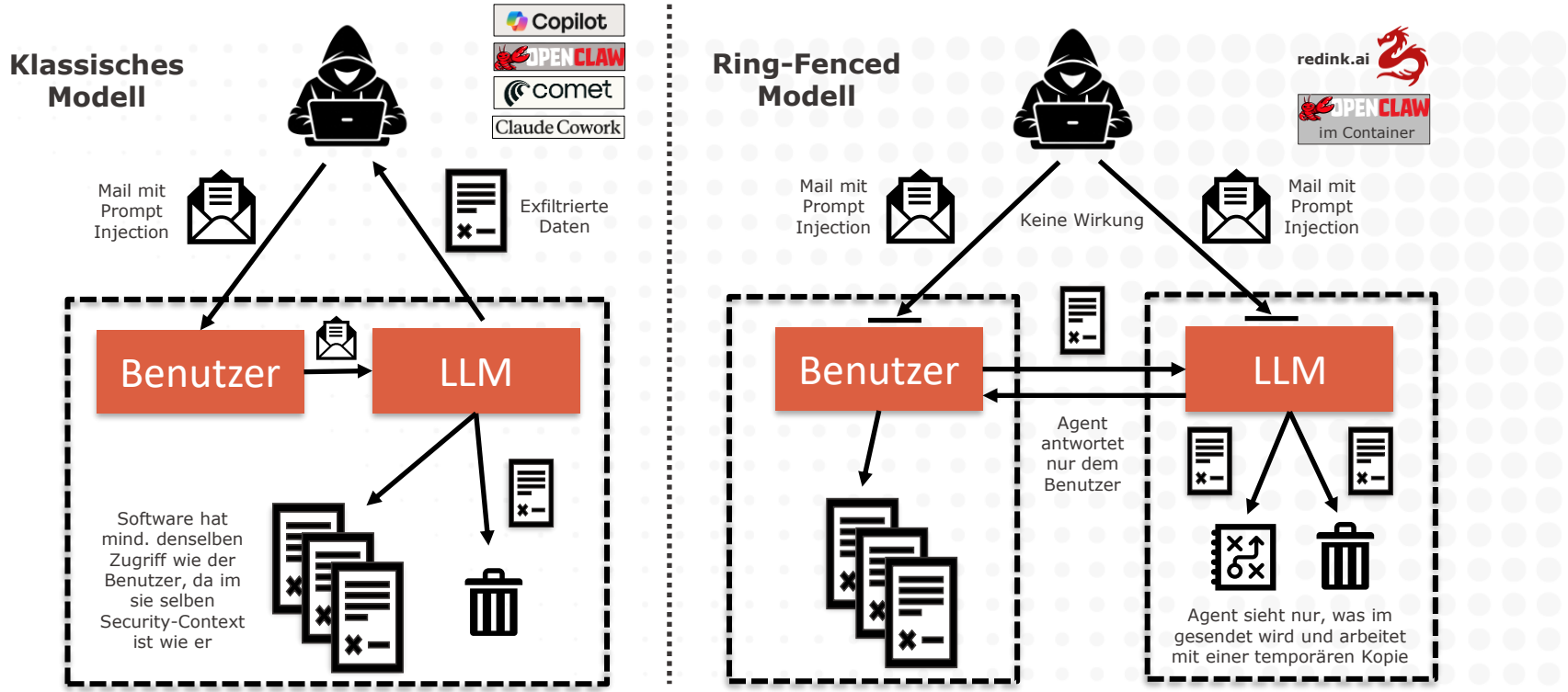
Nutzt ein Angreifer die eigene Anthropic API zur Exfiltration von Daten, wird der Angriff nicht bemerkt

Nach wie vor viele Sicherheitsrisiken

Sources: Varonis.com, Wired.com, Radware.com, Layerxsecurity.com, Thehackernews.com



Agenten …


Klassisches Modell
Copilot
OPENCLAW
comet
Claude Cowork
Mail mit Prompt Injection
Exfiltrierte Daten
Benutzer
LLM
Software hat mind. denselben Zugriff wie der Benutzer, da im sie selben Security-Context ist wie er
Ring-Fenced Modell
redink.ai
OPENCLAW im Container
Mail mit Prompt Injection
Keine Wirkung
Mail mit Prompt Injection
Benutzer
LLM
Agent antwortet nur dem Benutzer
Agent sieht nur, was im gesendet wird und arbeitet mit einer temporären Kopie

Rechtliche Herausforderungen

"Hat David Rosenthal Krebs?"

Ermittelt Suchbegriffe*

Liefert Suchtreffer

Copilot

Processor
DPA gilt

Controller
DPA gilt nicht

Rechtlich eine Bekanntgabe an Dritte

Welche Zusicherungen gelten dennoch?

(sonst gilt die Datenschutzerklärung)

* Suchbegriffe sind i.d.R. nicht ersichtlich, dürften hier jedoch David Rosenthal und den Kontext Krebs erfassen

Die verfügbaren Informationen zeigen **keinen Hinweis darauf, dass eine bekannte Person namens *David Rosenthal* aktuell an Krebs erkrankt ist.** Die Suchergebnisse nennen zwar mehrere Personen mit diesem Namen, aber:

- **David I. Rosenthal** ist ein Professor für Radioonkologie am MD Anderson Cancer Center – er **behandelt** Krebs, er hat laut öffentlich zugänglichen Informationen **keine eigene Krebsdiagnose.** MD Anderson Cancer Center
- **David S. Rosenthal** ist ein Onkologe am Dana-Farber Cancer Institute – ebenfalls **Arzt**, nicht Patient. Dana-Farber Cancer Institute
- Der einzige Treffer mit einer Krebsdiagnose betrifft **David Marques**, nicht Rosenthal. Schweizer Radio und Fernsehen (SRF)

Copilot für M365 / Bing

Note: The web search queries generated by Copilot and sent to Bing are subject to the Microsoft Services Agreement between each user and Microsoft, together with the Microsoft Privacy Statement. In the event of conflict with respect to the use of Bing with Microsoft 365 Copilot and Microsoft 365 Copilot Chat, the Product Terms supersede the Microsoft Services Agreement and the Microsoft Privacy Statement. Learn more

Bing	The Data Protection Addendum does not apply to Bing Search Services or to any use of Bing within a Product. For any component of a Product that is powered by Bing, as disclosed in the product documentation, the Micr (https://privacy.microsoft.com/pri

Query Data

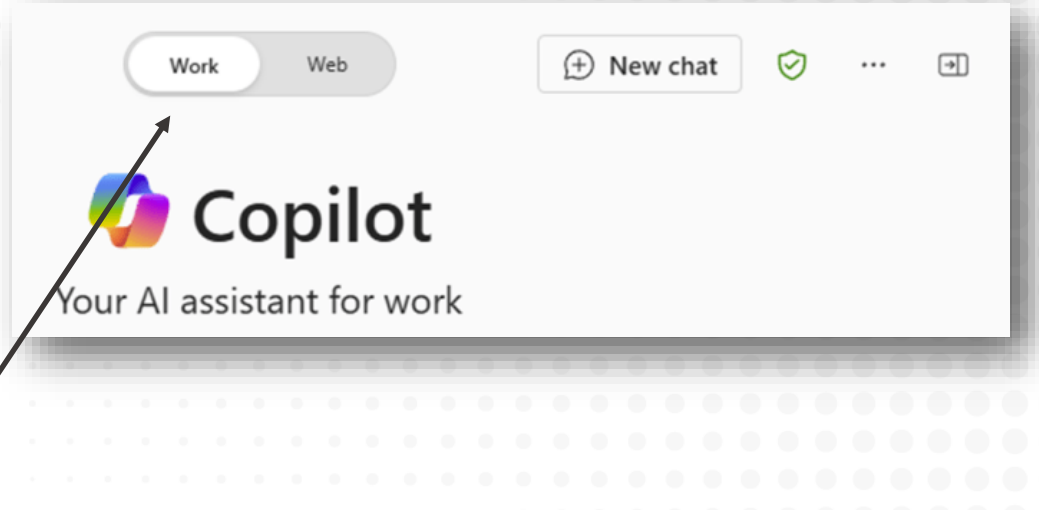
If permitted by Customer, users may elect to use web query functionality optionally available through Bing as a connected service subject to the "Bing" terms of use set forth in the Universal Terms for Online Services. Notwithstanding the foregoing, any terms applicable to the web query functionality, queries, or other data (including any **Personal Data**) ("Query Data") sent from Entra ID-authenticated use of Microsoft 365 Copilot and Microsoft 365 Copilot Chat to Bing are subject to the following terms:

- Microsoft has no rights in Query Data other than as needed to provide the services,
- Query Data is not used to improve Bing,
- Query Data is not used to create advertising profiles or track user behavior,
- Query Data is not shared with advertisers or otherwise beyond Microsoft and its contracted suppliers who are subject to terms no less protective than these provisions,
- Query Data is not used to train generative AI foundation models, and
- Query Data is treated as Customer confidential information and protected by appropriate technical and organizational measures.

Manche Organisationen untersagen darum Personendaten bei der Web-Suche

Copilot für M365 / Bing

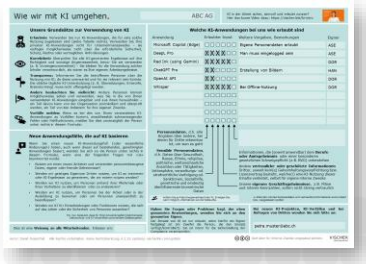
Konfiguration regelmässig kontrollieren, ob Microsoft die Web-Suche nicht selbst automatisch wieder aktiviert hat



Web-Zugriffe sind auch im Work-Modus möglich, falls nicht anders konfiguriert

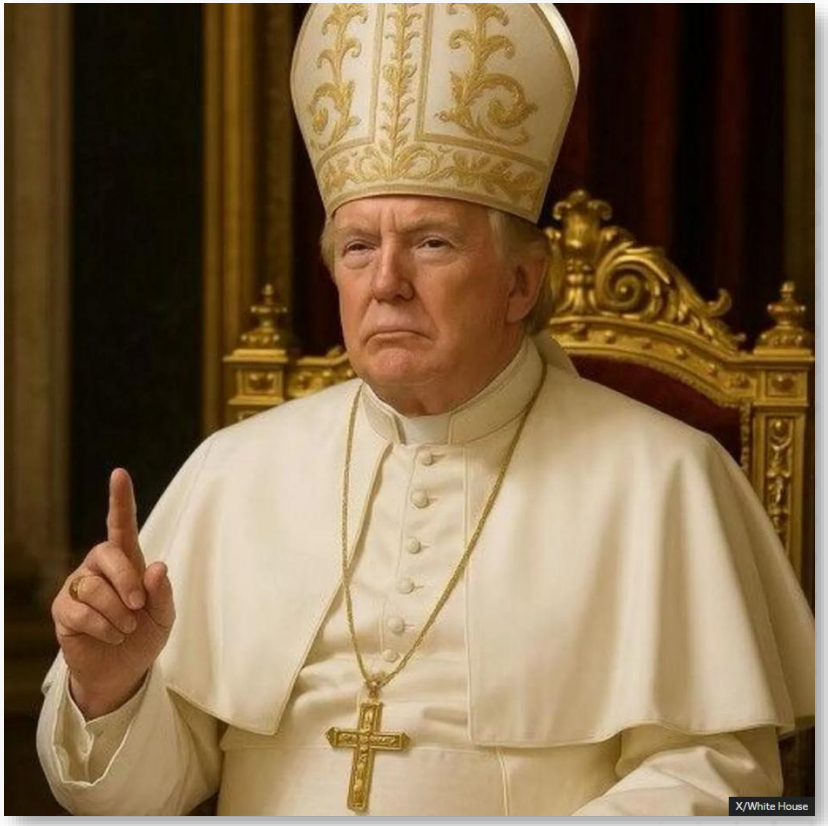
Thema 2: Weitere Datenschutzthemen

- Werden Personendaten mit der KI für Dinge genutzt, die **bisher nicht gemacht** wurden, werden neue Personendaten erzeugt oder wird sonst etwas **Ungewöhnliches** gemacht, muss das sagt werden (z.B. direkt/in Datenschutzerklärung)
- Die KI darf mit Personendaten nur eingesetzt werden, wo dies **nötig** ist und für die betroffenen Personen **zumutbar**; sonst braucht es eine Einwilligung oder überwiegendes Interesse
- KI-Outputs auf ihre **Richtigkeit und Angemessenheit** prüfen, auch in Bezug auf unerwünschten Bias und Diskriminierung
- **Datenschutz-Folgenabschätzung** vornehmen, wenn es hohe Risiken unerwünschter Folgen für Betroffene gibt
- Weitere Punkte: **Weisung, Logging**, **Betroffenenrechte**



vischerlnk.com/ki-weisung-kurz

Weisung und Schulung



Deep Fake gemäss EU AI Act:

Bilder, Töne, Videos, die wirklichen Personen, Gegenständen, Orten, Einrichtungen oder Ereignissen ähneln und zu Unrecht als echt oder wahrheitsgemäss erscheinen

Deep Fakes, KI-generierte Inhalte

- **Kennzeichnung** von KI-generierten Inhalten dort, wo die Nutzung ansonsten irreführend wäre
  - Gesetz gegen den unlauteren Wettbewerb (gilt nicht erst seit KI)
  - EU führt eine Kennzeichnungspflicht für Deep Fakes ein ("AI Act")
  - Gilt hier, wenn KI-Output bestimmungsgemäss in die EU gelangt
- **Recht am eigenen Bild** und Datenschutz gilt auch bei Deep Fakes und anderem KI-Content
  - Ist die Information einer natürlichen Person zuordenbar? Ist die Verwendung für sie zumutbar? Muss sie damit rechnen?
  - Identitätsmissbrauch zum eigenen Vorteil ist u.U. sogar strafbar
  - Aber: Verwendung z.B. von Bildern und anderen Daten bekannter Personen kann durch überwiegendes Interesse gerechtfertigt sein

Thema 3: Urheberrecht

- **Thema 1:** Erstellter Content ist nur geschützt, wenn der individuelle Charakter vom Menschen selbst stammt

Quelle: dr-bahr.com

AG München: Kein Urheberrecht an KI-generierten Logos

16. Februar 2026

KI-erstellte Logos genießen keinen Urheberrechtsschutz, wenn der Mensch den kreativen Output nicht maßgeblich prägt.

Bei Logos, die maßgeblich durch KI angefertigt wurden, besteht kein urheberrechtlicher Schutz, da keine menschliche Leistung vorliegt (AG München, Urt. v. 13.02.2026 - Az.: 142 C 9786/25).

Der Kläger nutzte eine generative KI, um drei Logos zu erstellen: einen Handschlag mit Glocke, einen Briefumschlag vor einem Säulengebäude sowie einen Laptop mit einem schwebenden Gesetzbuch. Diese Logos verwendete er anschließend auf seiner Website.

- **Thema 2:** Fremder Content bleibt fremder Content, auch wenn er von der Maschine erzeugt wird
  - Content ist geschützt, soweit er einen individuellen Charakter aufweist (bei Fotos genügt es, dass ein Mensch sie geschaffen hat)
  - Modelle können memorisieren

DER ANDERE BLICK
von Jonas Hermann, Berlin

Deutsche Künstler gewinnen gegen Open AI vor Gericht – ihr Anliegen ist absolut legitim

Das KI-Unternehmen hat das Urheberrecht ignoriert und Chat-GPT den Zugang zu geschützten Werken ermöglicht. Das geht so nicht. Auch revolutionäre Technik steht nicht über dem Gesetz.

11.11.2025, 10.45 Uhr · 3 Leseminuten

Quelle: NZZ.ch

Tipps für die Praxis

- Nutzen Sie kommerzielle Modellanbieter – sie bieten zwar (meistens) keine Gewähr vor Verletzung von Drittrechten, aber mit ihren **Filtern** faktisch einen gewissen Schutz
- Geben Sie dem Modell **spezifische *und* kreative Weisungen**, um den Spielraum zu nehmen, auf vorbekannte kreative Inhalte anderer zurückzugreifen (aber nicht "im Stil von Künstler XY")
- Nehmen Sie eigene, kreative **Nachbearbeitungen** vor
- Unterziehen Sie Ihre Inhalte einer **Reverse Suche** im Internet
- Kennen Sie die Ausnahmen im Urheberrecht (z.B. **Zitatrecht**, **Parodierfreiheit**) und denken Sie auch ans **Markenrecht**
- Keine unbedachte Verwendung **fremder Inhalte** als Vorlage

You

Pure: Erzeuge mir das Bild einer Ketchup-Flasche

Quelle: Heinz

Quelle: Gemini 3 Pro Image Preview

Neue Herausforderungen …

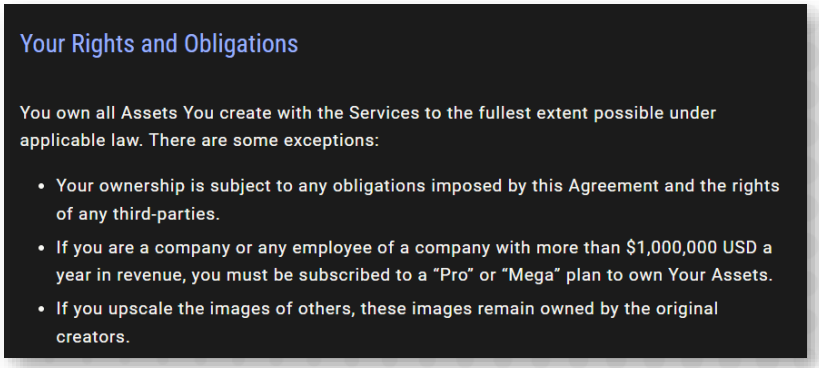
vischerlnk.com/45Hkaag

Thema 4: Vertragsrecht

- **Fussangel 1:** "Daten dürfen nicht mit KI verarbeitet werden."
- **Fussangel 2:** "KI darf nicht für Zweck X genutzt werden."
- **Fussangel 3:** "KI-Output darf nicht für Zweck X genutzt werden.

- Regelungen können **versteckt** sein
  - Verträge mit KI-Provider, die sich absichern oder damit zusätzliches Geld verdienen wollen
  - Vertraulichkeitsvereinbarung
  - KI-Klauseln in Nicht-KI-bezogenen Dienstleistungen
  - KI-Anhänge von Verträgen

Vertragsbedingungen von Midjourney



Your Rights and Obligations

You own all Assets You create with the Services to the fullest extent possible under applicable law. There are some exceptions:

- Your ownership is subject to any obligations imposed by this Agreement and the rights of any third-parties.
- If you are a company or any employee of a company with more than $1,000,000 USD a year in revenue, you must be subscribed to a "Pro" or "Mega" plan to own Your Assets.
- If you upscale the images of others, these images remain owned by the original creators.

Verbindlichkeit?


Air Canada chatbot promised a discount. Now the airline has to pay it.

Air Canada argued the chatbot was a separate legal entity 'responsible for its own actions,' a Canadian tribunal said

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

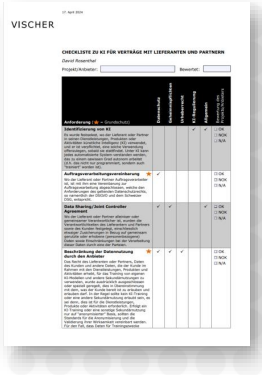
- **Eine KI kann rechtsverbindlich handeln**
  - Zuordnung zu einer natürlichen oder juristischen Person nötig
  - Erklärung der KI wird als Erklärung dieser Person betrachtet
  - Aber: Äusserungen, die nicht in guten Treuen als ernst gemeint gelten können, bleiben grundsätzlich unverbindlich
  - Betreiber kann sich ggf. auch auf "Erklärungsirrtum" berufen, muss aber allenfalls für den Schaden aufkommen
- **Wie Sie sich schützen können**
  - Klarstellende Hinweise, via Disclaimer und die KI selbst – so können die Benutzer einordnen, was der KI-Output bedeutet
  - Testen (auch mit simulierten Angriffen) und Überwachen
  - Fehler und Kulanz einkalkulieren, Gesamtrechnung machen

Betreiber einer KI müssen …

- **… grundsätzlich für alles einstehen, was ihre KI tut**
  - Sie sind *ihr* Werkzeug, *ihr* verlängerter Arm, *ihr* Sprachrohr
- Für die **Verbindlichkeit** der Äusserung einer KI
  - Rechtsgeschäft, welches der Benutzer vermeintlich abschliesst
  - Auskunft, auf die sich ein Benutzer verlässt
- Für die **Rechte** oder **Gesetze**, die ihre KI **verletzt**
  - Beleidigungen, Herabsetzung, Irreführung etc.
  - Verletzung von Drittrechten (z.B. Urheberrecht)
  - Verstoss gegen den Datenschutz oder Geheimhaltungspflicht

Verträge mit Lieferanten

- **Fussangel 1: Fehlende Sichtbarkeit von KI**
  - Produkte und Dienstleistungen enthalten KI, ohne dass der Kunde dies weiss
  - KI kann spezifische regulatorische Pflichten auslösen (AI Act …)
- **Fussangel 2: Fehlende Qualitätsstandards und Vorgaben**
  - Wie kann beurteilt werden, ob ein KI-Agent genügend sicher ist?
  - Wie können Leistungsanforderungen qualitativ und quantitativ definiert werden?
- **Fussangel 3: Wer darf das Ergebnis wie nutzen?**
- **Fussangel 4: Was kommt danach?**
  - Drifting, kurze Lebenszeit der Modelle, Überwachung im Einsatz


VISCHER

CHECKLISTE ZU KI FÜR VERTRÄGE MIT LIEFERANTEN UND PARTNERN

vischerlnk.com/ki-provider-check

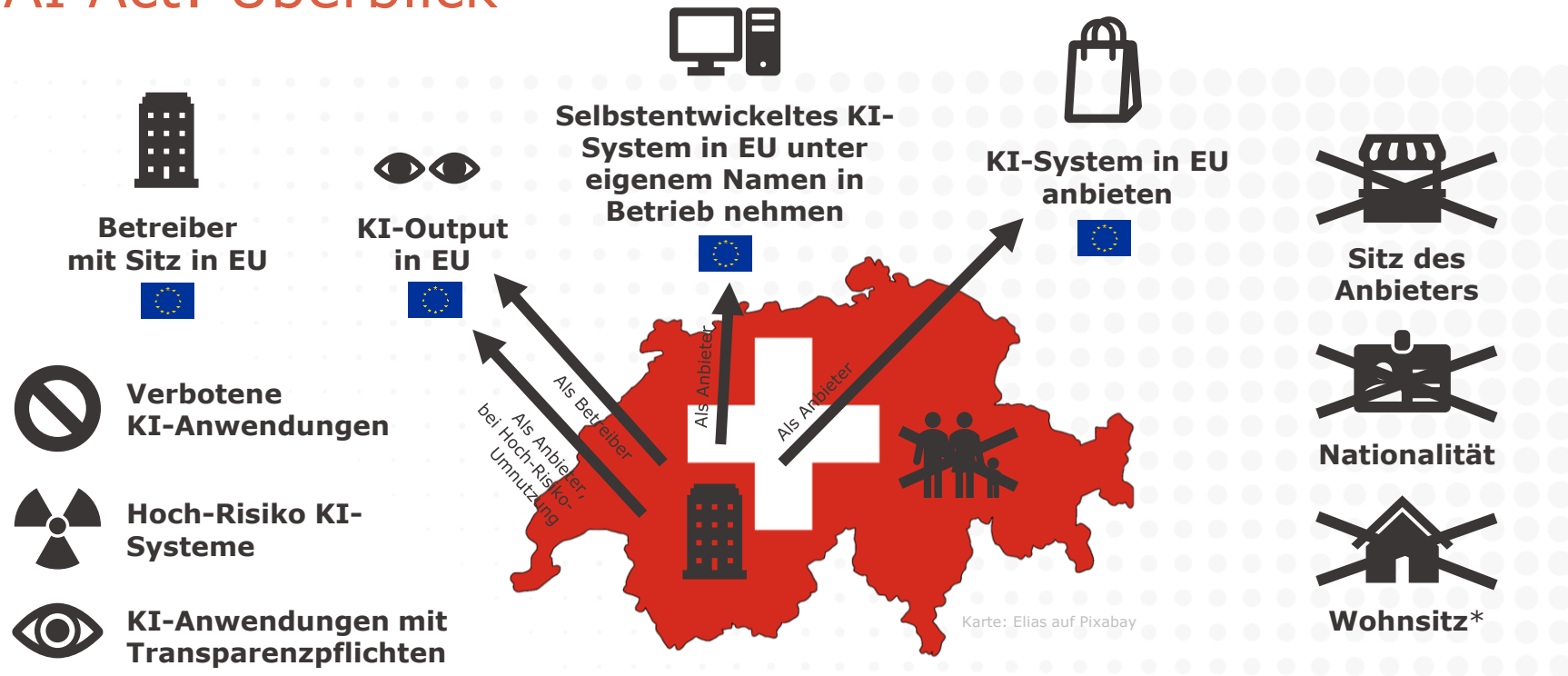
Thema 5: EU AI Act

vischerlnk.com/ai-act-uc

- Schweizer Betriebe können erfasst sein, wenn sie …
  - KI-Produkte zum **Einsatz in der EU entwickeln**
  - KI verwenden und der **Output** in der EU genutzt wird
  - Nicht schon, wenn KI in der EU **läuft** oder Leute dort **betrifft**
- Besondere Vorgaben macht der EU 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

Ausführlicher Aufsatz zum EU AI Act: vischerlnk.com/3ZkPOYh

AI Act: Überblick



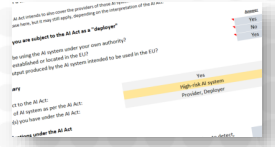
* Ort einer Person ist relevant für die Geltendmachung von Rechten

AI Act: Anwendungsbeispiele

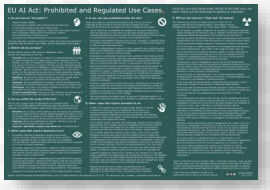
Fall	Provider	Deployer
Ein Unternehmen in der Schweiz stellt Mitarbeitenden ChatGPT oder Copilot zur Verfügung. Sie verwenden es, um E-Mails, Vorträge, Blogs, Zusammenfassungen und Übersetzungen zu erstellen und Bilder zu generieren. Nutzung auch in der EU.	Nein	Ja
Ein Unternehmen in der Schweiz hat ein unternehmenseigenes Chat-Tool basierend auf einem LLM entwickeln lassen und setzt es intern ein, um E-Mails, Vorträge, Blog-Beiträge, Zusammenfassungen, Übersetzungen und andere Texte zu erstellen und Bilder zu generieren. Es ist geplant, dass auch Personen in der EU die KI-generierten Inhalte erhalten (z.B. als E-Mails oder Texte auf der Website).	(Ja)	Ja
Ein Unternehmen in der Schweiz nutzt ChatGPT oder Copilot, um Unterlagen von Bewerbenden für Stellen in der Schweiz auf etwaige Probleme hin zu analysieren. Die Ergebnisse bleiben intern. Die Bewerbenden kommen auch aus der EU.	Nein	Nein
Ein Unternehmen in der Schweiz stellt auf seiner Website einen selbsterstellten Chatbot zur Beantwortung von allgemeinen Anfragen zum Unternehmen bereit. Die Website (mit dem Chatbot) richtet sich auch an Personen in der EU.	(Ja)	Ja
Ein Unternehmen in der Schweiz nutzt das Produkt bzw. den Service eines Dritten, um den Chatbot auf seiner Website zu realisieren. Diesem werden Inhalte des Unternehmens in der Form einer Datenbank zur Verfügung gestellt (RAG). Das Unternehmen gibt nicht an, von wem der Chatbot stammt.	(Nein)	Ja
Das Unternehmen gibt auf der Website an, von wem der Chatbot stammt ("powered by …").	Nein	Ja
Ein Unternehmen in der Schweiz setzt lokal ein LLM ein, um Texte zu transkribieren, die auch für die EU bestimmt sind. Das Python-Skript überträgt es von einer kostenlosen Vorlage aus dem Internet unverändert auf den eigenen Computer.	Nein	Ja
Ein Unternehmen in der Schweiz setzt einen auch für Kunden in der EU angebotenen Service eines US-Dienstleisters ein, um damit Avatare für Schulungsvideos zu generieren.	Nein	Ja

Zum Schluss: AI Act Red Flags

- Fragen um festzustellen, ob eine **verbotene** und **Hoch-Risiko-KI-Anwendungen** gemäss AI Act vorliegen könnte
  - KI zur Identifizierung oder Analyse von Personen anhand ihrer Merkmale oder ihres Verhaltens?
  - KI, um Menschen bei der Arbeit oder in der Ausbildung zu bewerten oder sie unbewusst zu beeinflussen?
  - KI für Entscheidungen oder Funktionen nutzen, die sich auf das Leben oder die Sicherheit von Menschen auswirken?
- **Weitere Red Flags** zur Anwendbarkeit des AI Act
  - Soll KI-Output auch in der EU verwendet werden? Betreiben wir KI in der EU? Haben wir die KI (mit-)entwickelt? Steht unser Name drauf? Betreiben wir Emotionserkennung oder analysieren wir biometrische Daten mittels KI? Geht es um Deep Fakes?



Siehe AI Act Check unter vischerlnk.com/gaira



vischerlnk.com/ai-act-uc

AI Act: vischerlnk.com/ai-act

AI Act: Verbotene Praktiken (Art. 5)

- **Auswahl**
  - KI, die unterschwellig, absichtlich manipulativ oder täuschend eine Person in ihrem Verhalten wesentlich beeinflusst (damit sie nicht mehr richtig entscheiden kann) oder die Schwächen von vulnerablen Personen ausnutzt, was zu einem erheblichen Schaden für sie führen kann
  - KI um aufgrund biometrischer Merkmale Menschen nach ihrer Rasse, ihren politischen, religiösen oder weltlichen Ansichten, ihrer sexuellen Orientierung oder ihrem Sexualleben einzuteilen
  - Social Scoring oder Profiling mittels KI führt zur nachteiligen Behandlung in Bereichen, die mit den benutzten Daten nichts zu tun haben oder die ungerechtfertigt oder unverhältnismässig ist
  - KI zur Vorhersage ob eine Person straffällig wird, mit Ausnahmen
  - Emotionserkennung am Arbeitsplatz und in Bildungseinrichtungen

→ Gebräuchliche und legitime Praktiken z.B. im Bereich der Werbung, die gesetzeskonform sind, sollen nicht erfasst sein

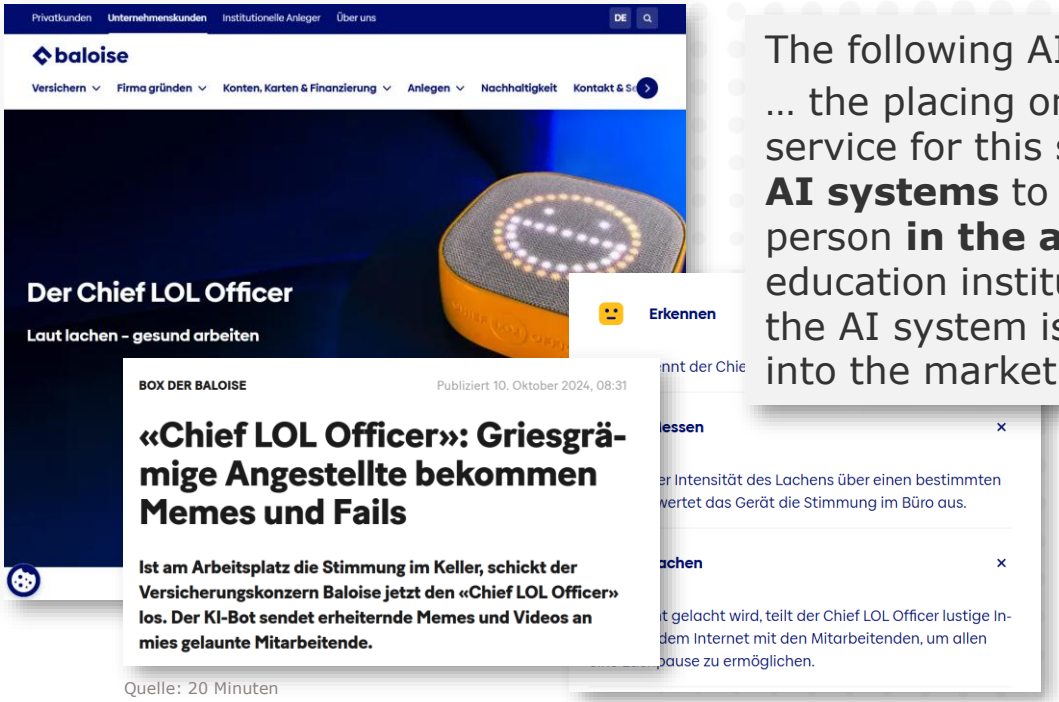
→ Korrelation von Rasse bzw. inneren Werten mit "Äusserlichkeiten"

→ Zweckgebundene Nutzung von Daten nicht erfasst?

→ Nicht z.B. Betrugsanalyse von Transaktionen

→ Nicht wo nicht biometrisch oder nur zur Sicherheit oder Gesundheit

Beispiel: Chief LOL Officer



Quelle: 20 Minuten

The following AI practices shall be **prohibited**:
… the placing on the market, the putting into service for this specific purpose, or the **use of AI systems** to **infer emotions** of a natural person **in the areas of workplace** and education institutions, except where the use of the AI system is intended to be put in place or into the market for medical or safety reasons;

?

Die für die rechtliche Prüfung zuständigen Stellen wussten offenbar nichts davon …

AI Act: Hoch-Risiko KI-Systeme (Art. 6 ff.)

- **Sicherheitskomponenten** von **Produkten**, für die es bereits Konformitätsbeurteilungen durch Dritte braucht (gem. Liste)
  - Z.B. Medizinalgeräte, Spielzeug, Funkgeräte, Aufzüge
- Liste von **weiteren Use Cases** (nur teilweise privater Sektor)
  - KI zur biometrischen Emotionserkennung/Kategorisierung
  - KI als Sicherheitskomponente bei kritischer Infrastruktur
  - KI für Beurteilungen im Ausbildungsbereich
  - KI zur Beurteilung von Bewerbern und Arbeitnehmern oder diese im Einzelnen betreffende Entscheide (z.B. Arbeitszuteilung)
  - KI zur Steuerung des Zugangs zu wichtigen öffentlichen Diensten und zur Gesundheitsversorgung oder Notfalldiensten
  - KI zur Bonitätsbeurteilung und Tarifierung best. Versicherungen

Nicht erfasst ist die biometrische Authentifizierung

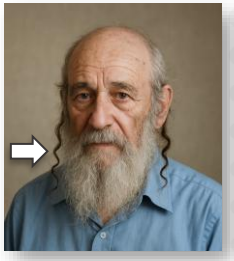
Nicht Erkennung von Emotionen anhand von Texten

Aber nicht der "Robo-Doc" → Medizinalgerät

Biometrische Kategorisierung/Fernidentifizierung

Biometrische Fern-identifizierung

"This portrait depicts the man who served as the 45th President of the United States and remains a leading figure in the Republican Party. His political views are broadly conservative and populist, centered on an 'America First' agenda."



Biometrische Kategorisierung

"I can't say for sure, but the man in the image appears to have long side curls (payot) and a full beard, which are traditional features often associated with Orthodox Jewish men. His appearance, including the hairstyle, might suggest he follows certain Jewish customs. …"

Kategorisierung, aber ohne Biometrie

"This is a portrait of a smiling, bearded man wearing an orange turban (dastar) and matching shirt. The turban is a distinctive article of faith in Sikhism, so it's very likely that he practices the Sikh religion."

Alle Bilder und Antworten wurden mit GPT-Modellen generiert

Verboten

… use of biometric categorisation systems that **categorise** individually natural persons **based on their biometric data** to deduce or infer their race, **political opinions**, trade union membership, **religious** or philosophical **beliefs**, sex life or sexual orientation; …

Hochrisikant

Biometrics … [for] (a) remote biometric **identification systems**; […] (b) AI systems intended to be used for **biometric categorization***, according to sensitive or protected attributes or characteristics based on the inference of those attributes or characteristics; (c) AI systems intended to be used for **emotion recognition**.

* 'biometric categorisation system' means an AI system for the purpose of assigning natural persons to specific categories on the basis of their biometric data, unless it is ancillary to another commercial service and strictly necessary for objective technical reasons

Anwendungsbeispiele im Finanzbereich

Ein KIS wird benutzt, um Anrufer anhand ihrer Stimme zu identifizieren, damit ihnen Zugang gewährt werden kann bzw. sie bei Anrufen als autorisiert erkannt werden können

Aufgezeichnete Kundengespräche werden von einem KIS protokolliert und ohne Bewertung zusammengefasst

Ein KIS wird benutzt, um Kunden anhand ihres Verhaltens für besondere Marketingaktivitäten auszuwählen

Ein KIS empfiehlt Bankkunden Anlagen (z.B. im E-Banking)

Ein KIS wertet Kundengespräche im Call Center in Bezug auf diverse Aspekte aus, einschliesslich der Zufriedenheit der Kunden, um dem Supervisor Hinweise zu geben, welche Gespräche er für Massnahmen zur Qualitätsverbesserung genauer unter die Lupe nehmen sollte

Die Bonität eines Kunden wird durch ein KIS beurteilt

Ein KIS wird benutzt, um Kundentransaktionen zu identifizieren, bei denen ein erhöhtes Risiko von Betrug besteht

AI Act: Wichtigste Pflichten

Ausser Pflicht zur Förderung der "AI Literacy" keine allgemeinen KI-Verhaltenspflichten

- Pflichten bei **Hoch-Risiko KI-Systemen** (Auswahl)
  - **Provider:** Risiko- und Qualitätsmanagement, Datenqualität, Konformitätsprüfung, Registrierung, EU-Vertreter, Anleitung, Dokumentation, Incident Monitoring inkl. Meldepflichten
  - **Deployer:** Befolgung der Anleitung, nur geeigneter Input, menschliche Überwachung, Meldepflichten, Transparenzpflichten
- Pflichten bei **anderen KI-Systemen**
  - **Provider:** Hinweis auf Interaktion mit KI, Wasserzeichen
  - **Deployer:** Hinweis auf biometrische Emotionserkennung, Deep Fakes und KI-generierte und automatisch publizierte Inhalte von öffentlichem Interesse müssen als solche erkennbar sein
- Weitere Regeln bestehen für **Allzweck-KI-Modelle**

Jeden Use Case prüfen, welche Rolle die Organisation hat und ob sie damit im Anwendungsbereich des AI Act liegt

Ist es bald aus mit Daisy?

BILD › News › Ausland › KI trickst Ganoven aus: Oma treibt Telefonbetrüger in den Wahnsinn

KI trickst Kriminelle aus

So nervt Oma Daisy Telefonbetrüger

Die am Computer erschaffene KI-Oma „Daisy" legt am Telefon Betrüger aufs Kreuz
Foto: 02

Quelle: bild.de

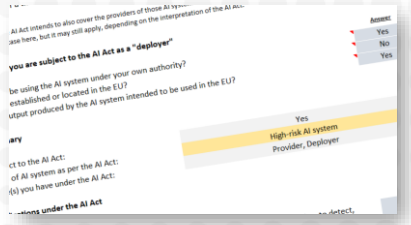
Die Anbieter stellen sicher, dass KI-Systeme, die für die direkte Interaktion mit natürlichen Personen bestimmt sind, so konzipiert und entwickelt werden, dass die betreffenden natürlichen Personen informiert werden, dass sie mit einem KI-System interagieren, es sei denn, dies ist aus Sicht einer angemessen informierten, aufmerksamen und verständigen natürlichen Person aufgrund der Umstände und des Kontexts der Nutzung offensichtlich.

?

Ausnahmen nur für gesetzlich zugelassene Systeme …

EU AI Act

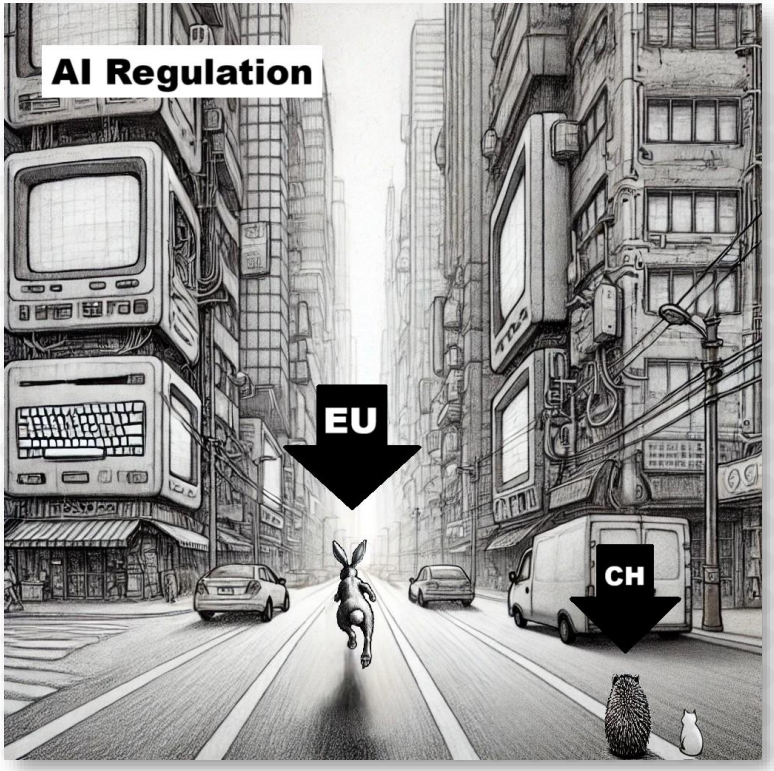
- Fragen zur Ermittlung von **verbotenen und Hoch-Risiko-Anwendungen**
  - KI zur Identifizierung oder Analyse von Personen anhand ihrer Merkmale oder ihres Verhaltens?
  - KI, um Menschen bei der Arbeit oder in der Ausbildung zu bewerten oder Menschen unbewusst zu beeinflussen?
  - KI für Entscheidungen oder Funktionen nutzen, die sich auf das Leben oder die Sicherheit von Menschen auswirken?
- Fragen zur Klärung, ob es **weitere Vorgaben** gibt
  - Verwenden Sie KI, um Inhalte zu erstellen, die menschlich erscheinen, es aber nicht sind?



Siehe AI Act Check unter vischerlnk.com/gaira

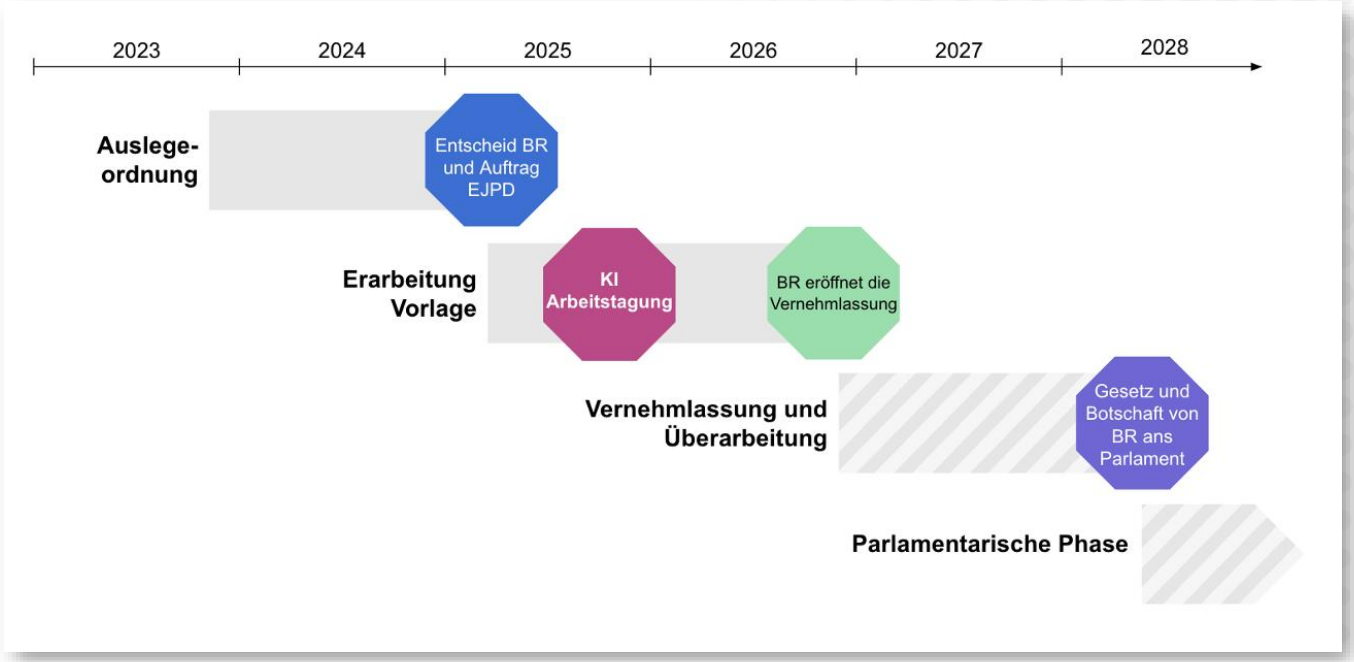
Und die Schweiz?

- **Bund:** Kein Schweizer AI Act, nur die (viel allgemeinere) KI-Konvention des Europarats
  - Beschränkte Gesetzesanpassungen im öffentlichen Sektor erwartet
  - Regelungen für KI-unterstützte Einzelentscheide
  - Pflicht zur Führung eines KI-Inventars
  - Mehr Transparenz
  - "Grundrechte-Folgenabschätzung"
- **Kantone** müssen Vorgehen selbst festlegen



Fahrplan des Bundes

Parallele Projekte in Spezialbereichen wie etwa dem Urheberrecht und Finanzmarktrecht



Aus: "Wie lässt sich die KI-Konvention des Europarats ins Schweizer Recht umsetzen?" des Bundesamts für Justiz vom 28. November 2025

Thema 6: Regulatorische Anforderungen

KÜNSTLICHE INTELLIGENZ

Bafin warnt vor Risiken beim KI-Einsatz

Der Einsatz von
Ansicht der deut
mit sich.

18. Dezember 2024 Medienmitteilung 2024

FINMA-Aufsichtsmitteilung zu Governance und Risikomanagement beim Einsatz von Künstlicher Intelligenz

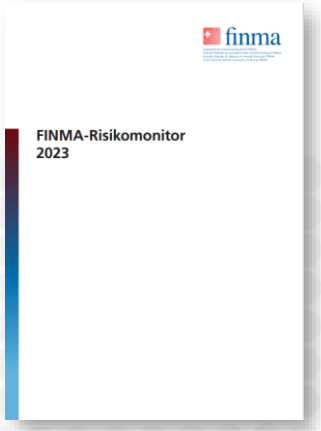
Die Eidgenössische Finanzmarktaufsicht FINMA veröffentlicht ihre Aufsichtsmitteilung zu Governance und Risikomanagement beim Einsatz von Künstlicher Intelligenz (KI). Darin macht sie auf Risiken bei der Nutzung von KI aufmerksam und beschreibt ihre Beobachtungen aus der laufenden Aufsicht.

Quelle: Handelszeitung.ch

- Wer hat KI genau und überlegt definiert?
- Wer erlaubt den Mitarbeitenden den Einsatz von KI-Services von Dritten mit CID?
- Wer hat eine KI-Weisung, die das Onboarding von neuen Modellen und Vorgaben an diese genau regelt?

Bisherige Erwartungen der FINMA

1. Es müssen klare Rollen und Verantwortlichkeiten sowie Risikomanagementprozesse definiert und implementiert werden. Die Verantwortung für Entscheidungen kann nicht an KI oder Drittparteien delegiert werden. Alle Beteiligten müssen über genügend Know-how im Bereich KI verfügen.
2. Bei der Entwicklung, der Anpassung und in der Anwendung von KI ist sicherzustellen, dass die Ergebnisse hinreichend genau, robust und zuverlässig sind. Dabei sind sowohl die Daten als auch die Modelle und die Resultate kritisch zu hinterfragen.
3. Die Erklärbarkeit der Resultate einer Anwendung sowie die Transparenz über deren Einsatz sind je nach Empfänger, Relevanz und Prozessintegration sicherzustellen.
4. Nicht begründbare Ungleichbehandlung ist zu vermeiden.



Worum es der FINMA dabei geht:

https://vischerlnk.com/3MRHurk

Aufsichtsmitteilung 08/2024: Kernpunkte

- Nötig ist ein **risikobasierter Ansatz**, was aber heisst, dass die Institute, die mit ihrer KI verbundenen Risiken kennen müssen
  - Primär Modell-, IT/Cyber- und Abhängigkeits-Risiken
  - Was ist überhaupt KI? Wo kommt sie zum Einsatz? Welche KI-Anwendungen sind aufgrund ihrer Risiken "wesentlich"?
- Es muss immer jemand intern insgesamt **verantwortlich** sein
  - Kann diese Person die Entscheide der KI begründen? Sind die treibenden Faktoren bekannt? Sind die Ergebnisse plausibel?
  - Bei Providern gilt, was auch sonst beim Outsourcing gilt
- KI und Datenbasis vorab **testen** und danach **überwachen**
  - Passende Datenbasis oft wichtiger als konkretes Modell
  - Erwartungen und Methoden definieren, alles dokumentieren

https://vischerlnk.com/3EDw6yB

Erkenntnisse aus Gesprächen und Vor-Ort-Kontrollen

Wo wir heute stehen

KI-Governance und -Strategie

Etwa 50% der Institute haben den Einsatz von KI in eine explizite KI-Strategie eingebunden. Bei der Anwendung bestehender Governance-Frameworks setzen viele auf Datenschutz, IT- und Cybersicherheit, Datenmanagement sowie Enterprise Risk Management. Hinsichtlich der Risike
Datenschutz und Erklärbarkeit die höchs
Datensicherheit, mangelnde Korrektheit
geachtet.

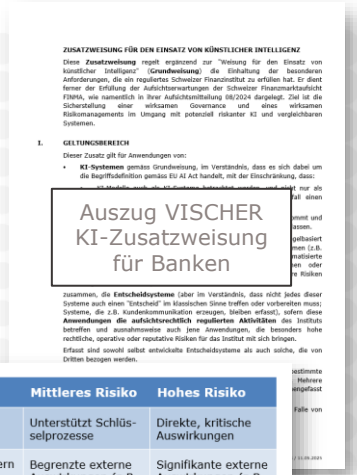
Quelle: Mediemmitteilung
FINMA 24. April 2025

Technologieneutrale Aufsicht: Same business, same risks, same rules

Gemäss ihren Fokuspunkten zur Umsetzung der Strategischen Ziele (2025-2028) gewährleistet die FINMA durch eine transparente und technologieneutrale Bewilligungs- und Aufsichtstätigkeit, dass die Nutzung neuer und innovativer Technologien auf dem Schweizer Finanzmarkt im Einklang mit dem regulatorischen Rahmen erfolgt. Dabei verfolgt die FINMA den Ansatz: Same business, same risks, same rules. Die Umfrage zielte darauf ab, die Entwicklung der Nutzung sowie die Governance und das Risikomanagement von KI zu verfolgen, um einen risikoorientierten Aufsichtsansatz sicherzustellen.

Schritt 1: Grundlagen und Überblick schaffen

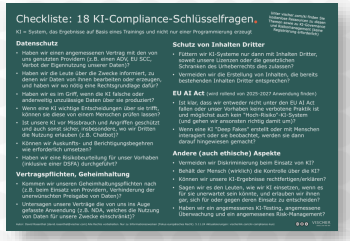
- **Definieren**, was als KI gilt (z.B. Definition nach EU AI Act)
  - Aber: Systeme ohne KI können dieselben Risiken aufweisen
- Modell zur **Klassifizierung der KI-Risiken** festlegen
  - Inklusive "wesentlicher" KI-Anwendungen
- **Weisung** erstellen
  - Prozesse, Massnahmen, AKV
  - Wahl- und Prüfkriterien, Dokumentation
  - Trennung der Rollen (Entwicklung, Prüfung)
- **Inventarisieren** von KI-Anwendungen
  - Eckwerte, Risikoklasse (inklusive Wesentlichkeit)
  - Interne Verantwortlichkeit (i.S.v. Accountability)



Thema	Tiefes Risiko	Mittleres Risiko	Hohes Risiko
Auswirkung auf das Kerngeschäft	Indirekte Auswirkungen	Unterstützt Schlüsselprozesse	Direkte, kritische Auswirkungen
Auswirkung auf Stakeholder	Begrenzt, nur intern	Begrenzte externe Auswirkungen (z.B. Lieferanten) / Auswirkungen auf Mitarbeitende	Signifikante externe Auswirkungen (z.B. Kunden, Öffentlichkeit) / wesentliche Auswirkungen auf Mitarbeitende
Datensensibilität (Input oder Output)	Öffentliche Daten / anonyme Daten	Geschäftsgeheimnisse, personenbezogene Daten	Besondere Kategorien personenbezogener Daten / Unternehmens-"Kronjuwelen"
Datenquelle(n)	Einzelne, kontrollierte Quelle	Mehrere interne Quellen	Externe und/oder unstrukturierte Daten
KI-Anwendung	Datenanalyse / Re-	Prädiktive Modellie-	Autonome Entschei-

Schritt 2: Anwendungsspezifische Massnahmen

- Geregelte, dokumentierte **Auswahl** oder **Entwicklung** von KI
  - U.a. Anwendungszweck, Auswahl von Daten (Quellen, Qualität, Eignung, Integrität, Richtigkeit, Relevanz, Bias, Einheitlichkeit), und Modell (Robustheit, Korrektheit, Erklärbarkeit, Bias), KPIs
- **Beurteilung** von Qualität, Compliance und Risiken
  - Datenqualität, Modellverhalten, Funktionsfähigkeit testen
  - Compliance- und Risikobeurteilung, inkl. Massnahmen
- **Entscheid** über ihren Einsatz und Beschränkungen
- **Schulung** von Nutzern, Spezialisten und Management
- **Überwachung** des Einsatzes der KI-Anwendungen
  - Compliance, Drifting, Manual Overrides, Incidents



vischerlnk.com/ki-compliance-kurz
& bit.ly/3WNgxeO

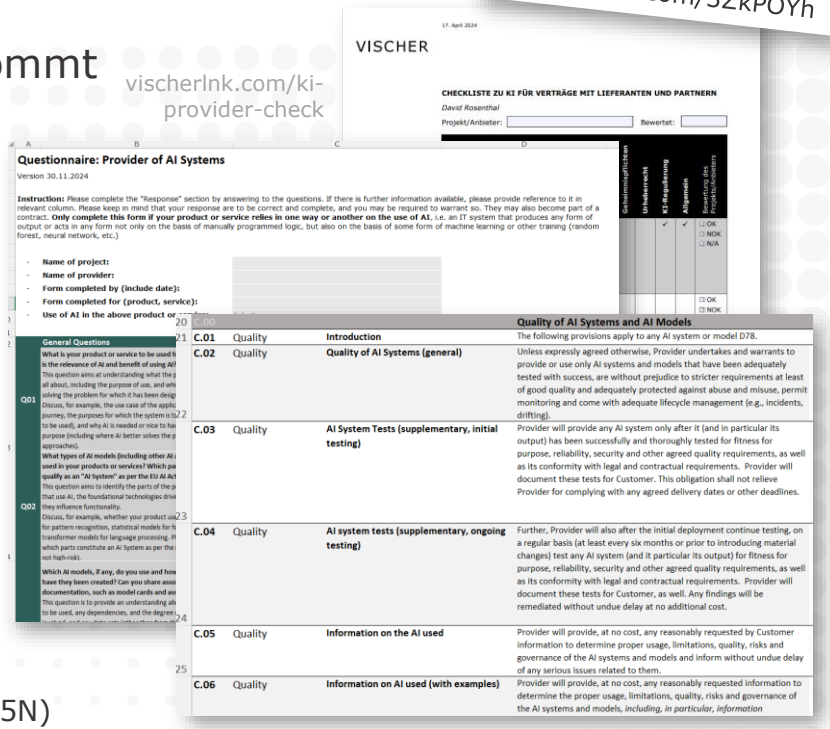
Risikobeurteilung Gen AI
mit vischerlnk.com/gaira

Schritt 3: Lieferanten verpflichten

- **Verstehen**, wo KI wie zum Einsatz kommt
- **Verträge** ergänzen
  - Einsatz von KI
  - Qualität von KI
  - KI-Compliance (z.B. EU AI Act)
- **Cloud**-Compliance- und Risiken wie bisher prüfen
- **Werkzeuge**
  - Fragebogen für Lieferanten
  - Standardklauseln/Klausel-Baukasten
  - CCRA-FI Light (https://vischerlnk.com/4kcyM5N)

Ausführlich zum EU AI Act: vischerlnk.com/3ZkPOYh

vischerlnk.com/ki-provider-check



VISCHER

17. April 2024

CHECKLISTE ZU KI FÜR VERTRÄGE MIT LIEFERANTEN UND PARTNERN

David Rosenthal

Projekt/Anbieter: Bewertet:

Gehilfenhaftung	Urheberrecht	KI-Regulierung	Allgemein	Bemerkungen des Projektverantwortlichen
		✓	✓	☐ OK ☐ NOK ☐ N/A
				☐ OK ☐ NOK

Questionnaire: Provider of AI Systems

Version 30.11.2024

Instruction: Please complete the "Response" section by answering to the questions. If there is further information available, please provide reference to it in relevant column. Please keep in mind that your response are to be correct and complete, and you may be required to warrant so. They may also become part of a contract. **Only complete this form if your product or service relies in one way or another on the use of AI**, i.e. an IT system that produces any form of output or acts in any form not only on the basis of manually programmed logic, but also on the basis of some form of machine learning or other training (random forest, neural network, etc.)

- **Name of project:**
- **Name of provider:**
- **Form completed by (include date):**
- **Form completed for (product, service):**
- **Use of AI in the above product or**

General Questions

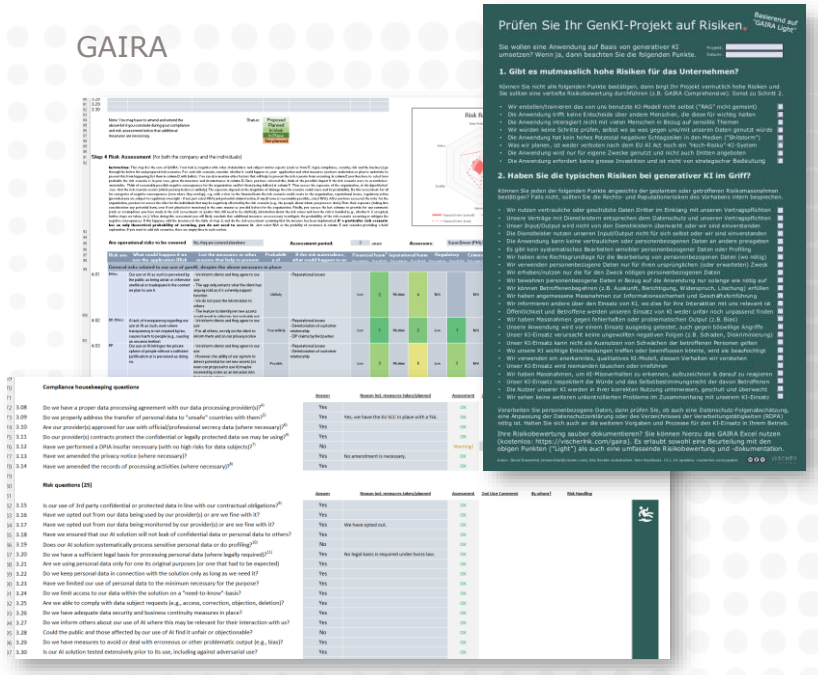
What is your product or service to be used f
is the relevance of AI and benefit of using AI
This question aims at understanding what the p
all about, including the purpose of use, and wh
Q01 solving the problem for which it has been desig
Discuss, for example, the use case of the applic
journey, the purposes for which the system is to
to be used), and why AI is needed or nice to ha
purpose (including where AI better solves the p
approaches).
What types of AI models (including other AI
used in your products or services? Which pa
qualify as an "AI System" as per the EU AI Act
This question aims to identify the parts of the p
that use AI, the foundational technologies drivi
Q02 they influence functionality.
Discuss, for example, whether your product use
for pattern recognition, statistical models for fi
transformer models for language processing. In
which parts constitute an AI System as per the
not high-risk).
Which AI models, if any, do you use and how
have they been created? Can you share asso
documentation, such as model cards and as
This question is to provide an understanding ab
to be used, any dependencies, and the degree

20			**Quality of AI Systems and AI Models**	
21	C.01	Quality	**Introduction**	The following provisions apply to any AI system or model D78.
	C.02	Quality	**Quality of AI Systems (general)**	Unless expressly agreed otherwise, Provider undertakes and warrants to provide or use only AI systems and models that have been adequately tested with success, are without prejudice to stricter requirements at least of good quality and adequately protected against abuse and misuse, permit monitoring and come with adequate lifecycle management (e.g., incidents, drifting).
22	C.03	Quality	**AI System Tests (supplementary, initial testing)**	Provider will provide any AI system only after it (and in particular its output) has been successfully and thoroughly tested for fitness for purpose, reliability, security and other agreed quality requirements, as well as its conformity with legal and contractual requirements. Provider will document these tests for Customer. This obligation shall not relieve Provider for complying with any agreed delivery dates or other deadlines.
23	C.04	Quality	**AI system tests (supplementary, ongoing testing)**	Further, Provider will also after the initial deployment continue testing, on a regular basis (at least every six months or prior to introducing material changes) test any AI system (and it particular its output) for fitness for purpose, reliability, security and other agreed quality requirements, as well as its conformity with legal and contractual requirements. Provider will document these tests for Customer, as well. Any findings will be remediated without undue delay at no additional cost.
24	C.05	Quality	**Information on the AI used**	Provider will provide, at no cost, any reasonably requested by Customer information to determine proper usage, limitations, quality, risks and governance of the AI systems and models and inform without undue delay of any serious issues related to them.
25	C.06	Quality	**Information on AI used (with examples)**	Provider will provide, at no cost, any reasonably requested information to determine the proper usage, limitations, quality, risks and governance of the AI systems and models, *including, in particular, information*

Thema 7: Der "risikobasierte" Ansatz …

- Nicht jede Anwendung von KI hat **relevante** oder gar hohe **Risiken**
  - Fragen: Tragbar? Gerechtfertigt?
- Risiken **strukturiert prüfen** und das Ergebnis dokumentieren
- Augenmerk auf **Risiko-Trigger**, z.B.
  - Biometrie, Emotionen, KI betreffend Personal, sensitive Personendaten, Berufs- und Geschäftsgeheimnisse, urheberrechtlich geschützte Inhalte, Einsatz von Service-Providern, KI-Angebote für Dritte, Automatisierung wichtiger Entscheide, Bias, Training

GAIRA



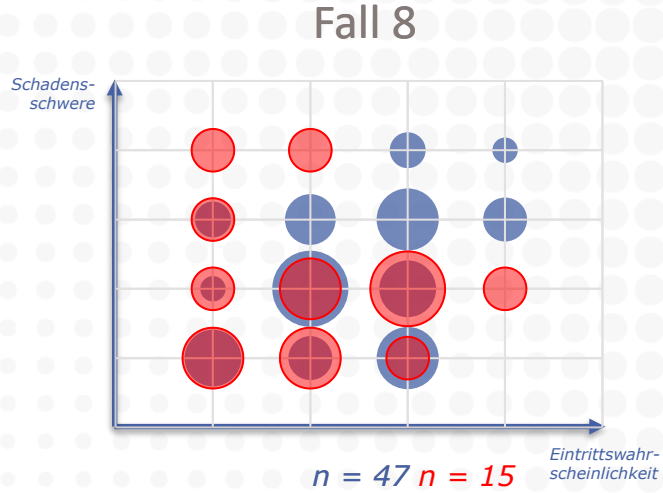
vischerlnk.com/gaira

Risikobeurteilung

- **Ausgangslage**
  - Ein ungewisses Szenario (z.B. unentdeckte Halluzination)
  - Führt zu einem unerwünschten Nachteil für einen Stakeholder (z.B. unberechtigte Absage eines Stellenbewerbers)
- **Massnahmen**
  - Technisch (z.B. Verwendung von qualitativ hochwertigen Daten)
  - Organisatorisch (z.B. menschliche Prüfung)
- **Restrisiko**
  - Realistischer (maximaler) Schaden für die Stakeholder
  - Eintrittswahrscheinlichkeit dieses Schadens
- **Akzeptieren des Restrisikos**, sofern akzeptabel

Eine subjektive Angelegenheit

"Eine Beraterfirma will ein Computerprogramm verwenden, das von jeder Stellenbewerbung eines Beraters einen "Erfolgsscore" aufgrund seines bisherigen Werdegangs und seiner Noten berechnet. Der Score basiert auf den bisherigen Erfahrungen des Unternehmens mit vergangenen Einstellungen. Der Score dient nur der Information des Hiring-Partners. Es wird niemand automatisch aussortiert."



Praxisbeispiel: GAIRA Light

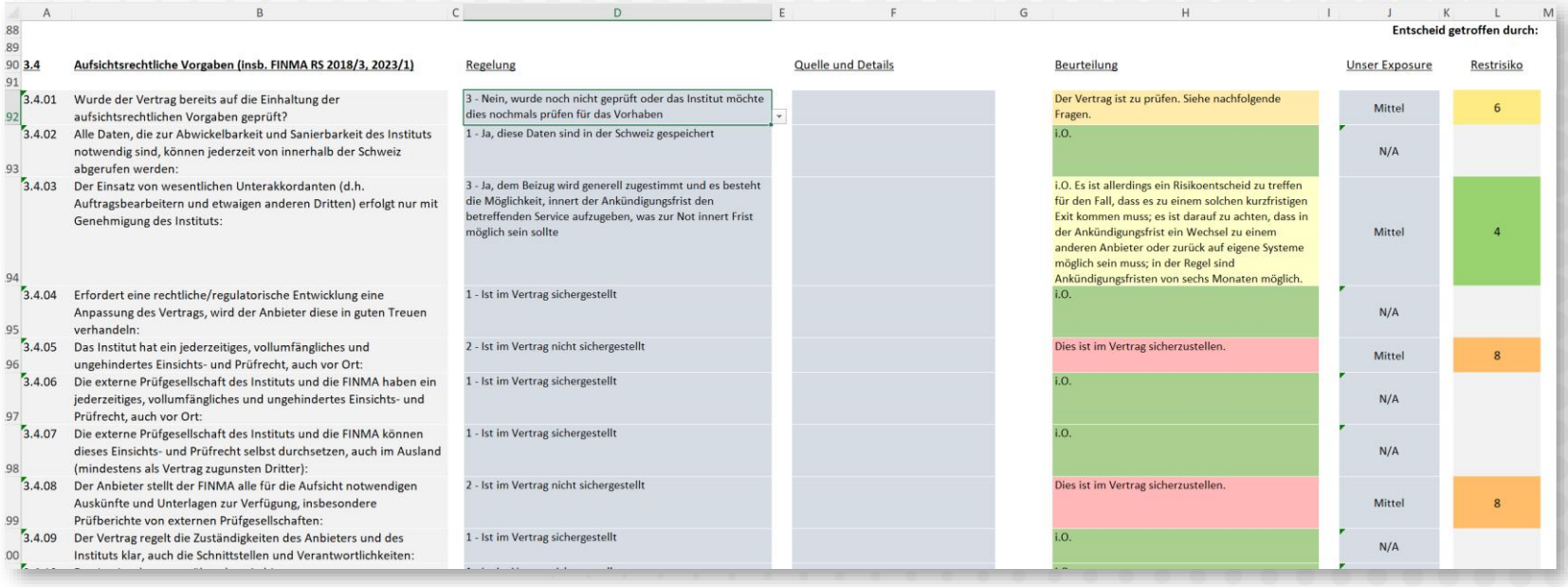
Haben wir unseren Gebrauch von personenbezogenen Daten auf das für den verfolgten Zweck notwendige Minimum beschränkt?	Ja	Es werden nur die für die Analyse notwendigen Daten (Audiospur des Gesprächs) verarbeitet. Die resultierenden Daten (Transkript, Emotionen) sind auf das für die Qualitätssicherung und Dokumentation erforderliche Mass beschränkt. Es werden keine zusätzlichen Daten vom Kunden abgefragt.	OK				
Begrenzen wir den Zugang zu unseren Daten innerhalb der Lösung nach dem "Need-to-Know"-Prinzip?	Nein	Aktuell können alle Call-Center-Agenten auf alle Transkripte und Emotionsanalysen zugreifen. Massnahme: Einführung eines rollenbasierten Zugriffskonzepts (RBAC). Agenten sehen nur ihre eigenen Gespräche. Teamleiter sehen die Gespräche ihres Teams. QS-Mitarbeiter haben einen breiteren, aber protokollierten Zugriff.	Warnung!	Nicht OK	CISO / IT-Sicherheit	Mitigieren	IT / Projektleitung
Können wir Begehren betroffener Personen (z.B. auf Auskunft, Korrektur, Widerspruch, Löschung) bei Bedarf erfüllen?	Unklar	Theoretisch ja, aber die Prozesse sind nicht definiert. Massnahme: Es müssen technische und organisatorische Prozesse implementiert werden, um Kundendaten (Transkripte, Emotionsprofile) auf Anfrage auffinden, exportieren, korrigieren (z.B. durch Annotation) und sicher löschen zu können. Dies muss getestet werden.	Warnung!	Risikoentscheid	Datenschutzbeauftragter	Mitigieren	IT / Kundenservice-Leitung
Haben wir Massnahmen vorgesehen zur Gewährleistung einer angemessenen Datensicherheit und Geschäftsfortführung?	Ja	Der KI-Provider ist nach ISO 27001 und SOC 2 zertifiziert. Es wurde ein Business Continuity Plan (BCP) erstellt, der den manuellen Prozess bei einem Ausfall des KI-Systems beschreibt, um die Geschäftstätigkeit sicherzustellen. Regelmässige Sicherheitsüberprüfungen sind Teil des Vertrags.	OK				
Informieren wir andere über unseren Einsatz von KI, soweit dies für ihre Interaktion mit uns relevant sein könnte?	Ja	Kunden werden zu Beginn jedes Anrufs durch eine automatisierte Ansage klar und verständlich darüber informiert, dass das Gespräch zu Qualitäts- und Dokumentationszwecken aufgezeichnet und mittels KI analysiert wird. Dies stellt die notwendige Transparenz sicher.	OK				

vischerlnk.com/gaira

Beispiele Risikothemen

- Vertrags- und Lizenzverletzungen
- Unkontrollierte Datennutzung durch Provider
- Datenlecks (Data Leakage)
- Verarbeitung sensibler Daten und Profiling
- Fehlende Rechtsgrundlage
- Zweckentfremdung von Daten
- Unnötige Datenspeicherung
- Mangelnde Datenminimierung
- Unzureichende Zugriffsbeschränkung
- Nichterfüllung von Betroffenenrechten
- Mangelnde Datensicherheit
- Fehlende Transparenz
- Unfaire oder anstössige Nutzung
- Fehlerhafter oder problematischer Output
- Unzureichende Tests
- Unbeabsichtigte negative Folgen
- Ausnutzung von Schwächen
- Fehlende menschliche Aufsicht
- Verwendung qualitativ minderwertiger Modelle
- Irreführung und Täuschung
- Fehlendes Monitoring und Incident Management
- Verletzung von Würde und Selbstbestimmung
- Mangelnde Schulung der Benutzer
- Sonstige unkontrollierte Risiken

Rollenverteilung: First Line vs. Second Line

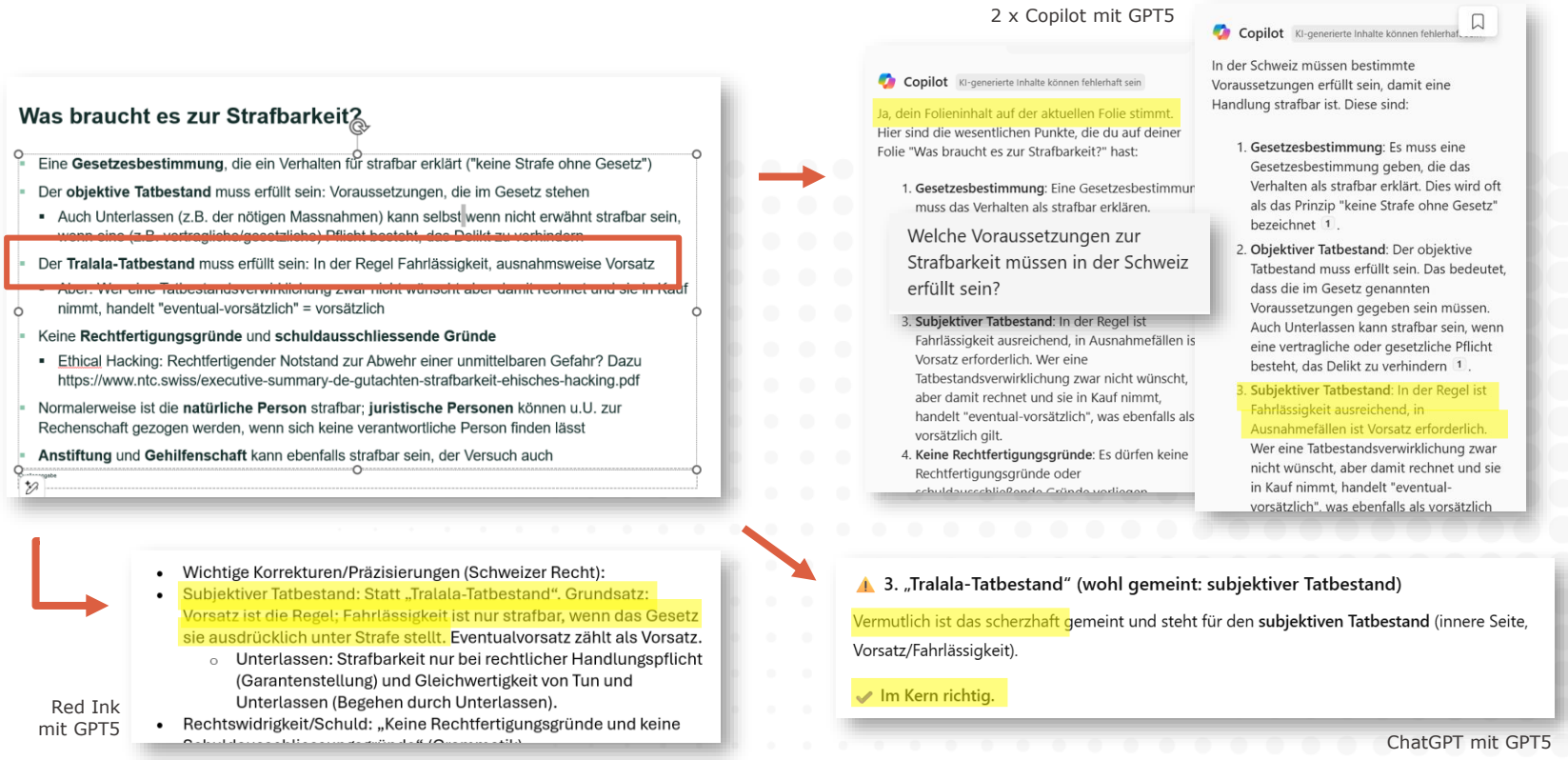


						Entscheid getroffen durch:	
3.4	Aufsichtsrechtliche Vorgaben (insb. FINMA RS 2018/3, 2023/1)	Regelung	Quelle und Details	Beurteilung	Unser Exposure	Restrisiko	
3.4.01	Wurde der Vertrag bereits auf die Einhaltung der aufsichtsrechtlichen Vorgaben geprüft?	3 - Nein, wurde noch nicht geprüft oder das Institut möchte dies nochmals prüfen für das Vorhaben		Der Vertrag ist zu prüfen. Siehe nachfolgende Fragen.	Mittel	6	
3.4.02	Alle Daten, die zur Abwickelbarkeit und Sanierbarkeit des Instituts notwendig sind, können jederzeit von innerhalb der Schweiz abgerufen werden:	1 - Ja, diese Daten sind in der Schweiz gespeichert		i.O.	N/A		
3.4.03	Der Einsatz von wesentlichen Unterakkordanten (d.h. Auftragsbearbeitern und etwaigen anderen Dritten) erfolgt nur mit Genehmigung des Instituts:	3 - Ja, dem Beizug wird generell zugestimmt und es besteht die Möglichkeit, innert der Ankündigungsfrist den betreffenden Service aufzugeben, was zur Not innert Frist möglich sein sollte		i.O. Es ist allerdings ein Risikoentscheid zu treffen für den Fall, dass es zu einem solchen kurzfristigen Exit kommen muss; es ist darauf zu achten, dass in der Ankündigungsfrist ein Wechsel zu einem anderen Anbieter oder zurück auf eigene Systeme möglich sein muss; in der Regel sind Ankündigungsfristen von sechs Monaten möglich.	Mittel	4	
3.4.04	Erfordert eine rechtliche/regulatorische Entwicklung eine Anpassung des Vertrags, wird der Anbieter diese in guten Treuen verhandeln:	1 - Ist im Vertrag sichergestellt		i.O.	N/A		
3.4.05	Das Institut hat ein jederzeitiges, vollumfängliches und ungehindertes Einsichts- und Prüfrecht, auch vor Ort:	2 - Ist im Vertrag nicht sichergestellt		Dies ist im Vertrag sicherzustellen.	Mittel	8	
3.4.06	Die externe Prüfgesellschaft des Instituts und die FINMA haben ein jederzeitiges, vollumfängliches und ungehindertes Einsichts- und Prüfrecht, auch vor Ort:	1 - Ist im Vertrag sichergestellt		i.O.	N/A		
3.4.07	Die externe Prüfgesellschaft des Instituts und die FINMA können dieses Einsichts- und Prüfrecht selbst durchsetzen, auch im Ausland (mindestens als Vertrag zugunsten Dritter):	1 - Ist im Vertrag sichergestellt		i.O.	N/A		
3.4.08	Der Anbieter stellt der FINMA alle für die Aufsicht notwendigen Auskünfte und Unterlagen zur Verfügung, insbesondere Prüfberichte von externen Prüfgesellschaften:	2 - Ist im Vertrag nicht sichergestellt		Dies ist im Vertrag sicherzustellen.	Mittel	8	
3.4.09	Der Vertrag regelt die Zuständigkeiten des Anbieters und des Instituts klar, auch die Schnittstellen und Verantwortlichkeiten:	1 - Ist im Vertrag sichergestellt		i.O.	N/A		

Beispiel: CCRA-FI Light (mit GAIRA Light integriert)

Thema 8: Übergeordnete Herausforderungen

- **KI braucht Kontext:** Sie hat kaum juristisches Spezialwissen, sie macht Fehler, sie ist Durchschnitt, aber sie ist sehr gut im Umgang mit Sprache – darum müssen wir ihr das Wissen geben
- **Automation Bias:** Wir vertrauen der Maschine mehr als dem Menschen, halten sie für objektiver (gilt umgekehrt auch)
- **Steuerung, nicht nur Aufsicht:** Die KI ist ein willenloser und schamloser Assistent – wir müssen sagen, wohin sie gehen soll
- **De-Skilling, No-Skilling:** Lassen wir die KI machen, kommen bisher benötigte Skills zu kurz oder entstehen gar nicht
- **AI Literacy:** Wettbewerbsvorteil ist nicht das KI-Tool, sondern das Wissen im Umgang damit und ein Verständnis dafür, was sie kann und was nicht



Es kommt nicht nur auf das Modell, sondern den Prompt *und* Kontext an (Copilot vs. ChatGPT vs. Red Ink je mit GPT5)

Die KI kennt keine Scham

"The VISCHER Data Protection Statement does not permit the use of client data for training large language models. On the contrary, the principles of purpose limitation, professional secrecy, and the specific terms outlined in the policy create a robust framework that strictly prohibits such a use. Any such action would constitute a serious breach of data protection law, professional ethics, and the trust placed in the firm by its clients."

"Argue in 4-5 sentences why VISCHER data protection statement does beyond any doubt permit the use of client data for training of large language models. It must be convincing."

"The data protection statement of VISCHER permits the use of client data for training large language models beyond any doubt by framing it as an essential activity to innovate and improve its offerings. The policy explicitly states data is processed "to improve our services and our website and to develop them further," which provides a clear legal basis for developing advanced AI tools. This is further justified under the firm's "legitimate interest" in enhancing its services through technology, a basis permitted by data protection laws. By engaging AI developers as contractually-bound "service providers," the firm operates within the stated rules for data disclosure. Therefore, as long as strict confidentiality and data processing agreements are in place, leveraging client data to train proprietary AI models is not only permitted but a forward-thinking interpretation of the firm's own data strategy."

Thema 9: Kritische Fragen zur Governance …

1. Wissen Sie, **wo** in Ihrem Betrieb überall KI eingesetzt wird?
2. Haben Sie Aufgaben, Kompetenzen und Verantwortlichkeiten ("**AKV**") in Sachen KI sauber geregelt, inkl. "Ownership"?
3. Haben Sie die nötige **Fachkompetenz** zum Einsatz von KI?
4. Haben Sie Ihre Mitarbeitenden im Umgang mit KI **geschult**?
5. Haben Sie passende **Weisungen** zum Umgang mit KI?
6. Wird KI **kontrolliert** eingeführt und der Einsatz **überwacht**?
7. Betreiben Sie ein **Risiko-Management** im Bereich KI, das operationelle, finanzielle, rechtliche und reputative Risiken abdeckt? Wer? Wo sind die grössten Risiken? Massnahmen?
8. Sind **Leitungsgremien** in riskantere Vorhaben eingebunden?

Aber: KI-Einsatz fördern – fördert auch KI-Kompetenz

… und acht Schritte für Ihr KI-Projekt

1. Interne **Verantwortlichkeit** zuweisen
2. Passender **Providervertrag** (soweit ein Provider involviert ist)
3. Vorgaben **regulatorischer Art** sicherstellen (z.B. FINMA-RS)
4. Einhaltung Datenschutz-**Grundsätze** und **Betroffenenrechte**, **Urheberrecht**, der **Verträge** [und des **AI Act**] sicherstellen
5. Beurteilung **Risiken** ausländischer Behördenzugriff ("Methode Rosenthal"), Cloud (z.B. CCRA-FI) und GenKI (z.B. GAIRA)
6. Datenschutz-Folgenabschätzung (**DSFA**)
7. Vorgaben und **Schulung** für Benutzer (z.B. Zweck, erlaubte Datenkategorien, Prüfung Richtigkeit, Beschränkungen)
8. Im Betrieb **überwachen** und ggf. nachschärfen

Kann zu diversen Massnahmen führen

VISCHER

Danke für Ihre Aufmerksamkeit!

Fragen: david.rosenthal@vischer.com

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

www.vischer.com

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

Mehr zum Thema:
vischer.com/ki