VISCHER

Geltungsbereich des EU AI-Act.

Für KI-Entwickler, -Betreiber und Anwender

David Rosenthal, VISCHER AG
22. Januar 2025

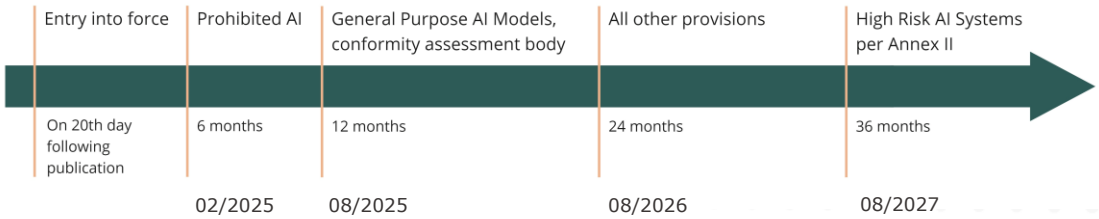
AI Act: Worum es geht

- Keine generelle Regulierung des Einsatzes von KI
- Stattdessen: **Produkteregulierung** mit Fokus Sicherheit
  - Verbotene KI-Praktiken
  - Vorgaben für "Hoch-Risiko"-KI-Systeme, Allzweck-KI-Modelle
  - Einzelne (Transparenz-)Vorgaben für weitere KI-Systeme
- **Ergänzt** bestehendes Recht (DSGVO, DSA, Vertragsrecht etc.)

vischerlnk.com/ai-act-uc

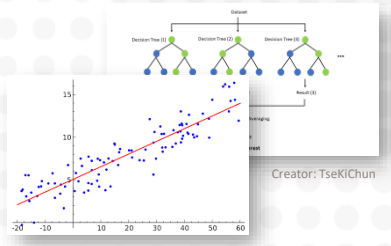
Text: vischerlnk.com/ai-act

Timeline EU AI Act

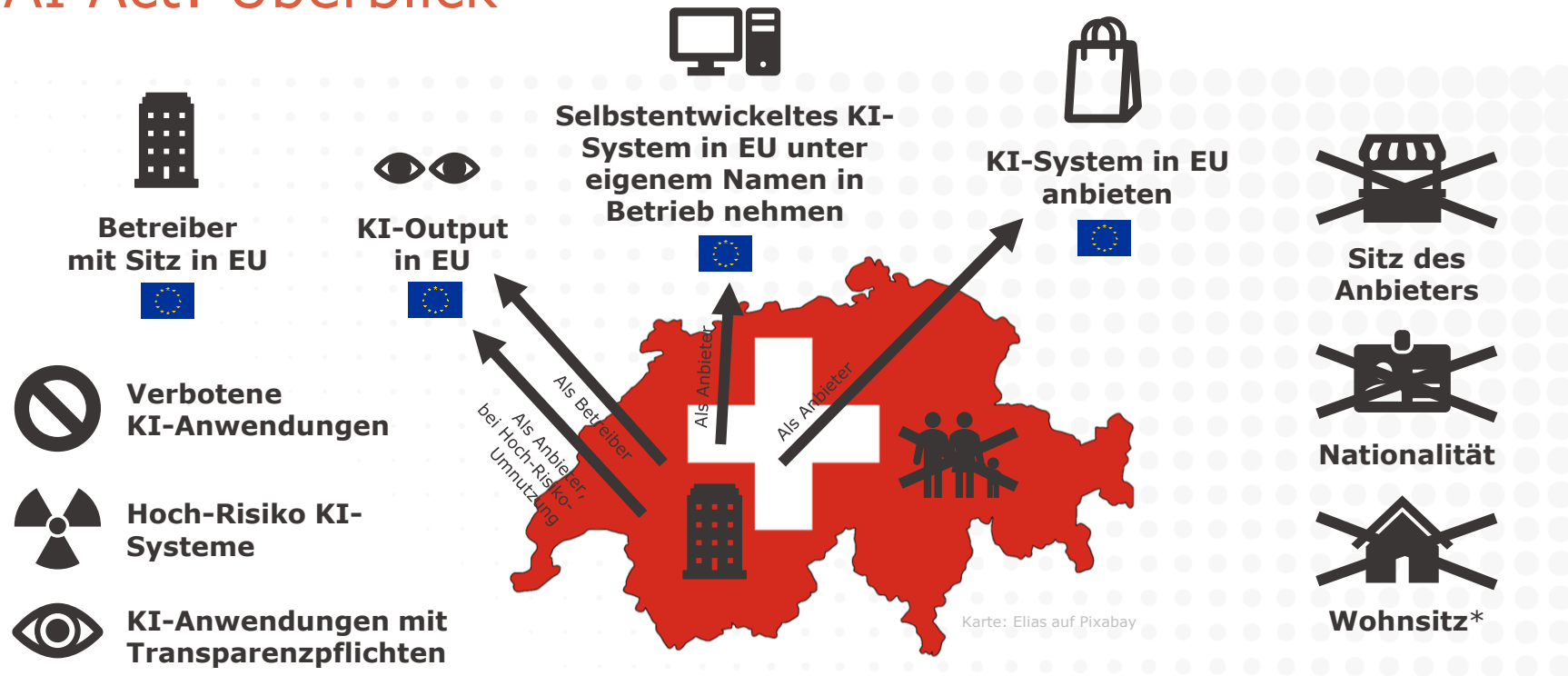


Aber was ist KI überhaupt?

- 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**: Die Definition ist mangelhaft ...
  - Jedes Kopiergerät (OCR), jedes Mobiltelefon (Fingerabdruck-Entsperrung, Kamera) ist KI; was ist mit linearer Regression?



AI Act: Überblick

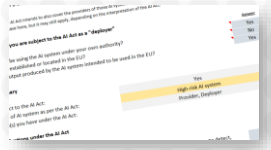


* Ort einer Person ist relevant für die Geltendmachung von Rechten

AI Act: Für wen er gilt

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

- **Deployer:** "using an AI system under its authority"
  - Falls Niederlassung in der EU *oder*
  - Falls KI-Output in der EU (bestimmungsgemäss) verwendet wird
- **Provider:** "develops an AI system … and places them on the [EU] market or puts the system into service [in the EU] under its own name or trademark"
  - Gemeinsame Entwicklung? Provider verpflichten + "powered by…"
  - Ggf. wer als Deployer ein System umfunktioniert (Art. 25)
- **Weitere Rollen:** Importer, Distributor, Product Manufacturer, EU-Vertreter



Siehe AI Act Check unter vischerlnk.com/gaira

KI-System umfunktionieren?

(1) In den folgenden Fällen gelten Händler, Einführer, Betreiber oder sonstige Dritte als Anbieter eines Hochrisiko-KI-Systems für die Zwecke dieser Verordnung und unterliegen den Anbieterpflichten gemäß Artikel 16:

a) wenn sie ein bereits in Verkehr gebrachtes oder in Betrieb genommenes Hochrisiko-KI-System mit ihrem Namen oder ihrer Handelsmarke versehen, unbeschadet vertraglicher Vereinbarungen, die eine andere Aufteilung der Pflichten vorsehen;

b) wenn sie eine wesentliche Veränderung eines Hochrisiko-KI-Systems, das bereits in Verkehr gebracht oder in Betrieb genommen wurde, so vornehmen, dass es weiterhin ein Hochrisiko-KI-System gemäß Artikel 6 bleibt;

c) wenn sie die Zweckbestimmung eines KI-Systems, einschließlich eines KI-Systems mit allgemeinem Verwendungszweck, das nicht als hochriskant eingestuft wurde und bereits in Verkehr gebracht oder in Betrieb genommen wurde, so verändern, dass das betreffende KI-System zu einem Hochrisiko-KI-System im Sinne von Artikel 6 wird.

Beispiel: ChatGPT für "hochriskante" Anwendungen nutzen

Was heisst "Entwicklung" eines KI-Systems?

- Eigenes Modell trainieren oder auch nur Finetuning betreiben?
- Programmierung eines Orchestrators bzw. Retrievers?
- Einrichten einer "Retrieval Augmented Generation"?
- Erstellen eines "benutzerdefinierten GPT"?
- Definieren der System-Prompts?
- Parametrisierung eines Chatbots für die Website?
- Installation eines quelloffenen LLM mit einem vorgegebenen, aber leicht angepassten Python-Skript?
- Beauftragen Sie einen Anbieter mit der Erstellung eines Chatbots für Ihre Website?

Weitere Fragen

- **Weitergabe von AI-Diensten innerhalb der Gruppe?**
  - Beispiel: Schweizer Muttergesellschaft kauft eine KI-Lösung und gibt sie an Konzerngesellschaften (auch) in der EU weiter → Risiko der Einstufung als Händlerin (es gibt Gegenargumente)
- **Weitergabe von KI-Funktionen an Kunden**
  - Beispiel: Eine Bank, die ihren Geschäftskunden über ihren Online-Zugang eine KI-gestützte Analyse ihrer Portfolios zur Verfügung stellt → Bank wird in der Regel zum Provider
- **Betrieb einer eigenentwickelten Lösung in der EU**
  - Beispiel: Eine Schweizer Privatschule entwickelt ein KI-System, welches Studierende während ihren Prüfungen überwachen kann und betreibt es in einer EU-Cloud → AI-Act kann anwendbar sein

Und KI-Modelle?

- **Modelle** können zwar erfasst sein, kommen aber nie alleine zum Einsatz, sondern immer nur als Teil eines KI-Systems
- Zwei Kategorien
  - General Purpose AI Model (GPAIM) – Allzweck-Modelle, die mit grossen Datenmengen trainiert worden sind (z.B. GPT-4o)
  - GPAIM mit systemischen Risiken – sehr grosse Allzweck-Modelle (gibt es bisher möglicherweise noch nicht) oder solche, die wesentliche negative Auswirkungen auf den EU-Markt haben können
- Insbesondere **Informationspflichten** und ein (untauglicher?) Versuch, die Einhaltung des **EU-Urheberrechts** zu erzwingen
  - Wenige Ausnahmen für Open-Source-Modelle

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

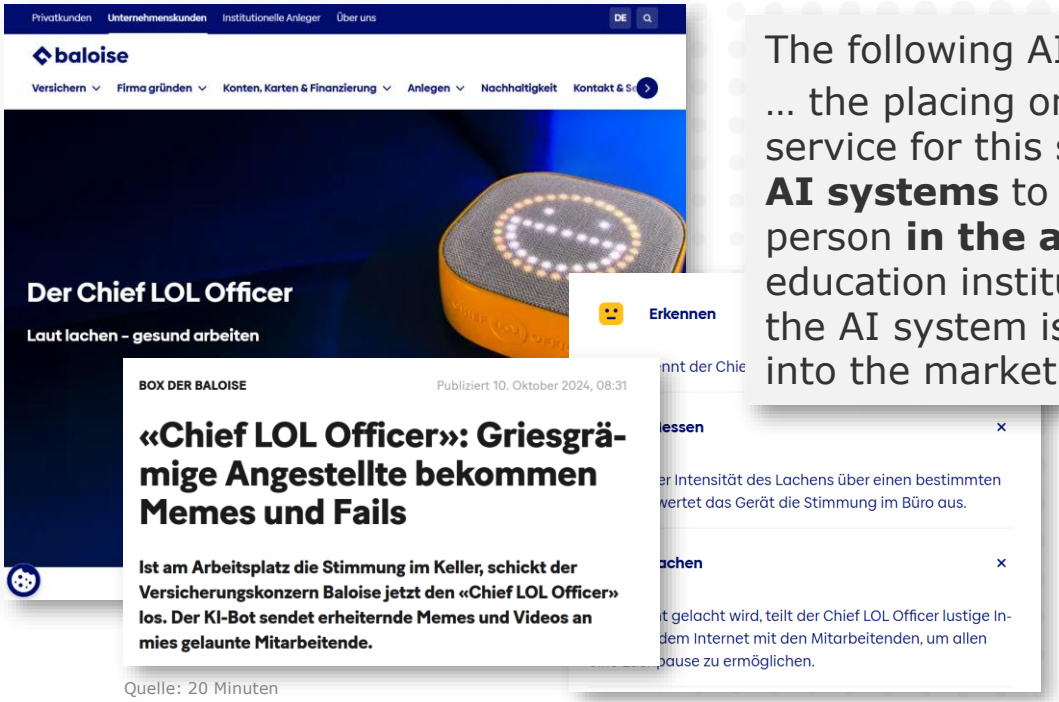
→ Korrelation von Rasse bzw. inneren Werten mit "Äusserlichkeiten"

→ Zweckgebundene Nutzung von Daten nicht erfasst?

→ Nicht z.B. Betrugsanalyse von Transaktionen

→ Nicht falls 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

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 …

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

AI Act: Anwendungsbeispiele einer Bank

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 Kunden Anlagen (z.B. im E-Banking)

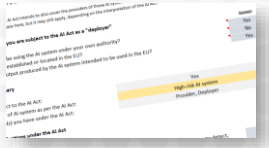
Ein KIS wertet Kundengespräche im Call Center in Bezug auf diverse Aspekte aus, einschliesslich der Zufriedenheit der Kunden, um dem Supervisor Hinweise zu geben, welche Gespräche er für Massnahmen zur Qualitätsverbesserung genauer unter die Lupe nehmen sollte

Die Bonität eines Kunden wird durch ein KIS beurteilt

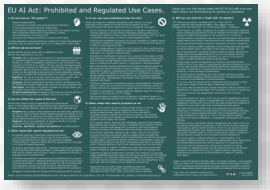
Ein KIS wird benutzt, um Kundentransaktionen zu identifizieren, bei denen ein erhöhtes Risiko von Geldwäscherei besteht

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

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