IHK Industrie- und Handelskammern in Bayern

Bayerisches Landesamt für Datenschutzaufsicht

Vortragsreihe: AI Act umsetzen

Künstliche Intelligenz und Datenschutz in der Praxis (Teil 2: Vertiefung)

Wer steht heute (virtuell) vor Ihnen?

Andreas Sachs

Informatiker

Bayerisches Landesamt für Datenschutzaufsicht (BayLDA)

Vize-Präsident
Bereichsleiter Cybersicherheit und Technischer Datenschutz
KI-Beauftragter

Carolin Loy

Juristin

Bayerisches Landesamt für Datenschutzaufsicht (BayLDA)

Bereichsleiterin „Digitalwirtschaft“
Rechtsfragen Künstlicher Intelligenz
Pressesprecherin

Das Webinar Vertiefung baut auf den Teil 1: Grundlagen auf (www.bihk.de/datenschutz.html)

Zielgruppe weiterhin: IHK Mitglieder und andere am Thema interessierte Personen

Kurzer Blick auf den Inhalt des **Webinars „Grundlagen“**:

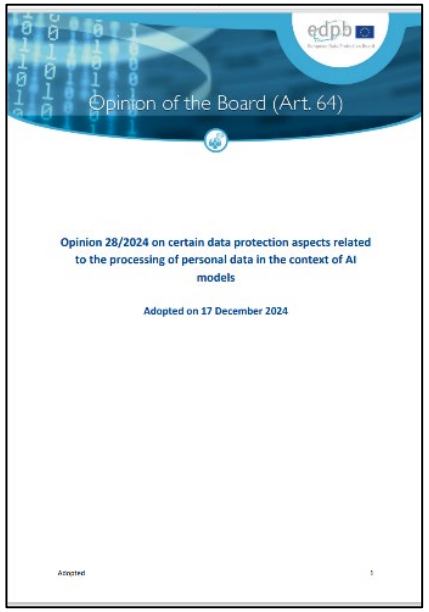
- **Zielsetzung** der **DS-GVO** und **KI-VO**: Primär Grundrechtsschutz
- **Anwendungsbereiche DS-GVO**: Verarbeitung **personenbezogener Daten** und KI
- **Anwendungsbereiche KI-VO**: **KI-System** im Sinne der Produktregulierung
- **Rollen und Verantwortlichkeiten**: **KI-Anbieter** und **KI-Betreiber**
- (Einige) **Rechtsgrundlagen** aus Sicht des Datenschutzes: Interessensabwägung und besondere personenbezogene Daten
- Kurzer Blick auf **Hochrisiko-KI** aus Sicht des Datenschutzes und der KI-VO
- **Herausforderung Betroffenenrechte** bei Recht auf Auskunft und Recht auf Löschung unter der DS-GVO
- Kurzer Blick auf **KI-as-a-Service** aus Sicht der DS-GVO

Heute werden wir einige im „Grundkurs“ vorgestellte Themengebiete vertieft betrachten

Ausgewählte Rechtsfragen bei KI

(Frau Loy)

EDPB Opinion 28/2024



https://www.edpb.europa.eu/system/files/2024-12/edpb_opinion_202428_ai-models_en.pdf

Verfahren

Request der irischen Aufsichtsbehörde vom 04.09.2024

Noch keine abgestimmte Position der nationalen Behörden vorhanden

Vollständig 13.09.2024

Frist 8 Wochen – Verlängerung + 6 Wochen = 23.12.2024

Beteiligung der KEYP und TECH Subgroup

Adressat: Datenschutzaufsichts-behörden in Europa

Frage 1:

***WANN** und **WIE** kann ein KI- Modell als „**anonym**“ angesehen werden?*

*“Is the final AI Model, which has been trained using personal data, in all cases, considered **not** to meet the definition of personal data (as set out in Article 4(1) GDPR)?*
(…)
*If the answer to question 1 is **“no”:** i. What are the **circumstances** in which that might arise?*
*a) If so, how can the steps that have been taken to ensure that the AI Model is not processing personal data be **demonstrated**?”*

Frage 1:

Wann kann von einer **Anonymisierung** ausgegangen werden?

- **Personenbezug** besteht auch bei Informationen die nur im **maschinenlesenbaren- Format** vorliegen -> keine automatische „Anonymisierung“
- **Extraktionsmöglichkeit aus dem Modell** direkt / Output
- Ausreichende **Beweise** (“sufficient evidence”), dass **keine Extraktion** pbD möglich

<u>Prüfschritte für Aufsichtsbehören:</u>

- **Schritt 1:** Verweis auf WP29 Opinion 05/2014 on Anonymisation Techniques u.a.

  Meist nicht anonym – **Risikobewertung** erforderlich!
- **Schritt 2**: Berücksichtigung **aller Mittel**, die nach **vernünftigem Ermessen** eingesetzt werden (“all the means reasonably likely to be used”)
- Schritt 3: Wurde eine **Risikobewertung** durchgeführt?

Frage 1:

Nachweis der Anonymisierung?

- **DSFA** oder Begründung warum keine DSFA erforderlich
- Bewertung des **Datenschutzbeauftragten**
- **TOMs** die **Identifizierung verhindern** / Risikobewertung
- Tests, **Audits**, theoretische Widerstandsfähigkeit
- Dokumentationen des **Entwicklers** (insbesondere Restrisiken)

Frage 2+3:

*Wie kann die Angemessenheit des **berechtigten Interesses** als Rechtsgrundlage in der Entwicklungs- und Einführungsphase **nachgewiesen** werden?*

Frage 2+3:

Allgemeine Erwägungen:

- Einhaltung der **Datenschutzgrundsätze**

- **Rechenschaftspflicht** –> Rollen und Pflichten vorab prüfen
- Grundsatz der **Rechtmäßigkeit** -> Verarbeitungsphasen trennen
- **Transparenzgrundsatz** -> Informationspflichten (Art. 14 Abs. 5?)
- **Datenminimierung** und **Zweckbindung** -> Ermittlung der relevanten Verarbeitungstätigkeit / Kontext der Verarbeitung
- **Betroffenenrechte** -> Widerspruchsrecht durchsetzbar?

Frage 2+3:

Guidelines 1/2024 on processing of personal data based on Article 6(1)(f) GDPR

Drei- Stufen- Test:

1. **Berechtigtes Interesse** des Verantwortlichen / eines Dritten
2. **Notwendigkeit** der Verarbeitung (“necessary”)
3. **Interessensabwägung**

Frage 2+3:

Legitimer Zweck:

Rechtmäßig
Klar und präzise formuliert
Gegenwärtig und real

Beispiele:

Entwicklung eines Chatbots zur Unterstützung von Nutzern
Entwicklung eines KI-Systems zur Erkennung betrügerischer Inhalte oder Verhaltensweisen
Verbesserung der Bedrohungserkennung in einem Informationssystem

Frage 2+3:

Notwendigkeit:

Ermöglicht die Verarbeitung den **verfolgten Zweck**?

Gleich effektive, **weniger einschneidende Mittel möglich**? (Braucht es KI?)

Beispiele:

Benötigt das KI- Modell **überhaupt pbD**?

In welchem Maße benötigt das KI- Model pbD? (Datenminimierung!)

First- oder **Third Party- Data**?

Technische Maßnahmen zur Verringerung der Identifizierbarkeit

Frage 2+3:

Interessensabwägung:

- **Welche Interessen** können potenziell betroffen sein?
- Welche möglichen **Risiken** ergeben sich im Zusammenhang mit KI?
  - **Eintrittswahrscheinlichkeit** und **Schwere**
- Ist die Verarbeitung von den **vernünftigen Erwartungen** der Betroffenen umfasst?
- Unterscheidung **Entwicklungs-** und **Bereitstellungsphase**

Beispiele:

- Kontext
- Art und Umfang der Daten

Frage 4:

*Welche **Auswirkungen** hat eine unrechtmäßige Verarbeitung in der Entwicklungsphase auf die **spätere Verarbeitung** oder den **Betrieb des KI-Modells?***

Frage 4:

- **Rechtswidrig** im Kontext der Frage wird als **ohne Rechtsgrundlage** verstanden
- Verweis auf die **Rechenschaftspflicht** -> Verantwortlichkeiten
- **Aufsichtsbehörden** können **Maßnahmen** im Hinblick auf ursprüngliche Verarbeitung, auch in Bezug auf den Verantwortlichen treffen

  -> Angemessen, erforderlich und verhältnismäßig

Unterscheidung in drei Szenarien:

- Entwicklung und Einsatz bei demselben Verantwortlichen
- **Einsatz eines KI- Modells, dass von einem anderen Verantwortlichen entwickelt wurde**
- Beide Konstellation, aber das Modell wurde anonymisiert

Frage 4:

- Feststellung der datenschutzrechtlichen **Rollen**
- **Jeder Verantwortliche** muss die **Rechtmäßigkeit** der von ihm zu verantwortenden Datenverarbeitung sicherstellen und **nachweisen** können
- Hat der Verantwortliche **geprüft**, ob das **Modell rechtswidrig trainiert** wurde?
- **Offensichtlich rechtswidrig?** Von einem **Gericht** oder einer **Behörde** festgestellt
- Einbeziehung des Trainings in die **Interessensabwägung**

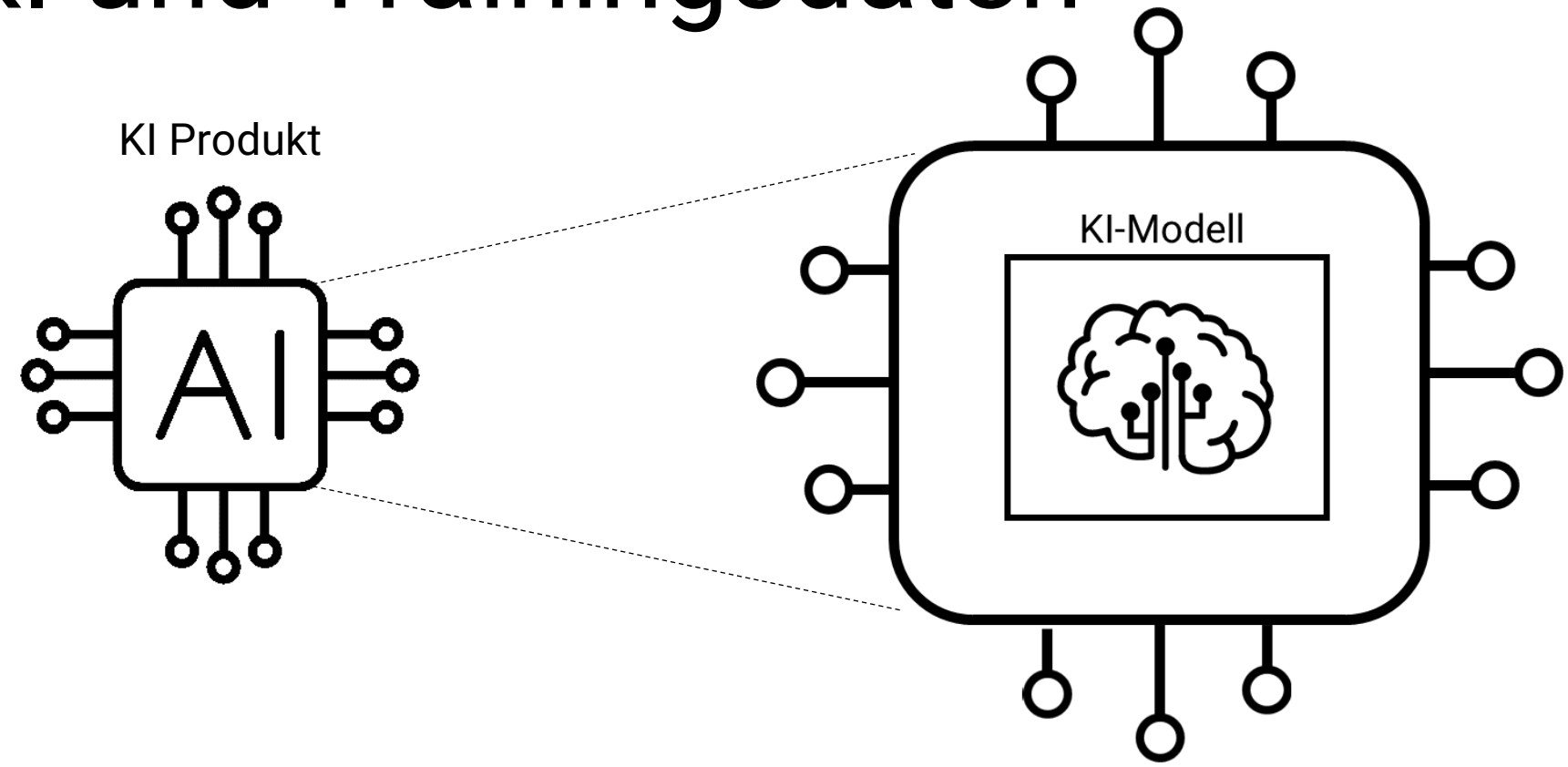
Was die Opinion nicht sagt.

- Umgang mit **Art. 9–Daten**
- **Automatisierte Entscheidungen** Art. 22 DS-GVO
- **Zweckänderung** Art. 6 Abs. 4 DS-GVO
- **DSFA**
- **Data Protection by design**

Anonymisierung von Trainingsdaten

(Herr Sachs)

KI und Trainingsdaten

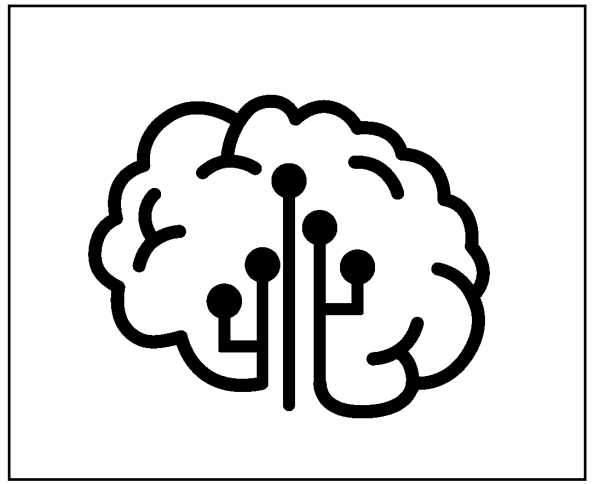


KI und Trainingsdaten

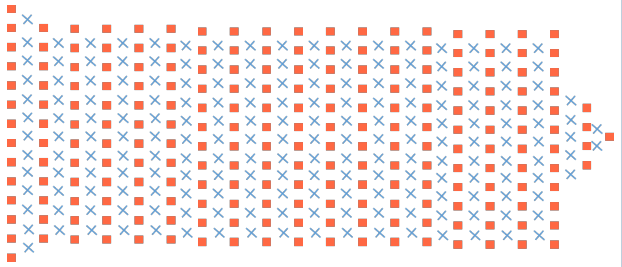
KI Training

(Personenbezogene) Trainingsdaten

KI Modell



z.B.

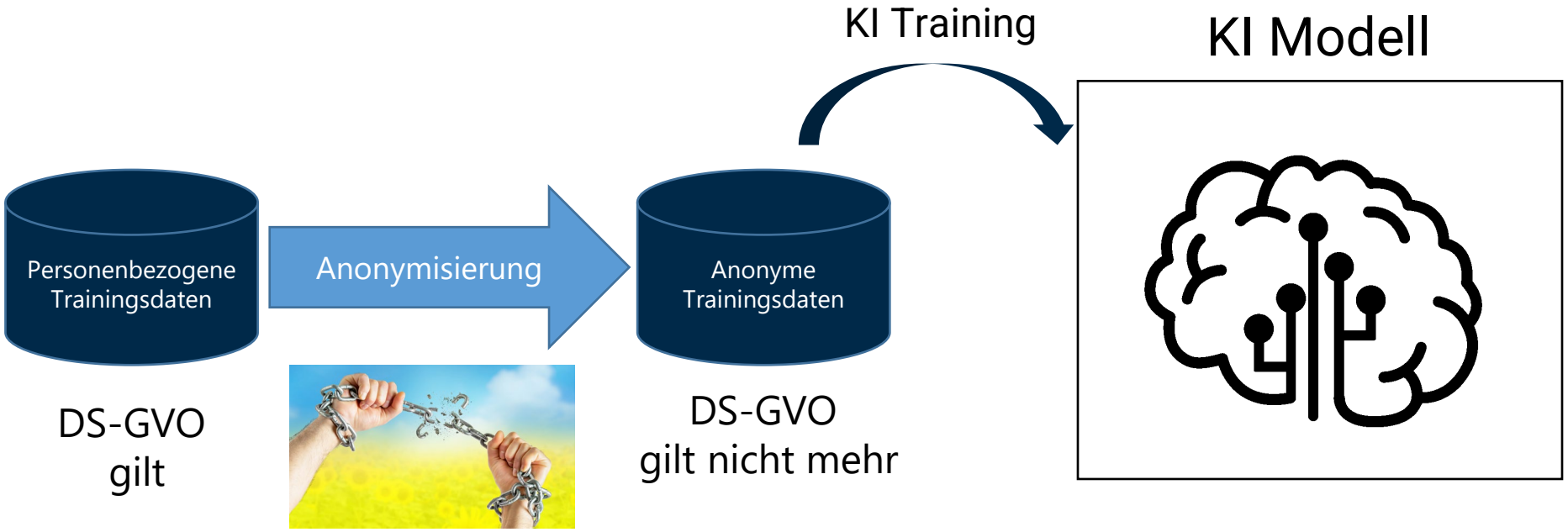


Tiefes Neuronales Netzwerk

Erforderlichkeit großer Mengen an Trainingsdaten

- Die **Großen Sprachmodellen** legen folgende Aussage nahe: „**Je größer, desto besser**“
- Die **Anzahl der Modellparameter** hat mitunter den Bereich von **mehreren Milliarden** bis über eine **Billion** erreicht
- Es gibt einen **Zusammenhang** zwischen der **Anzahl der Parameter** und der **Menge an Trainingsdaten. Grobe Daumenregel: Pro Parameter** ungefähr **10-100 Trainingstoken**
- Auch die DS-GVO stellt grundsätzlich bei der **Erforderlichkeit** von vielen und guten Daten für ein KI-Training den (alleinigen) Grundsatz der **Datenminimierung** hinten an.
- Beispiele:
  - **GPT4** geschätzt **10 Billionen Token**
  - **DALL·E 2** geschätzt **650 Mio. Bild-Text-Paare**
  - **ResNet** ca. 14 Millionen Fotos

KI und Trainingsdaten



Bewertungsmaßstab Anonymisierung

COUR DE JUSTICE
DE L'UNION EUROPÉENNE

Risikoorientierer Ansatz

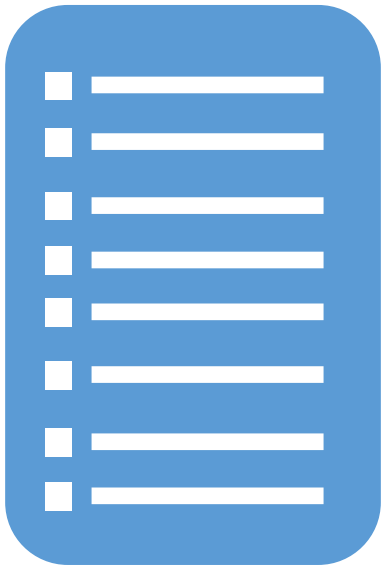
Relativer Personenbezug

„Nur das eigene Wissen zur Re-Identifizierung zählt“

Absoluter Personenbezug

„Das Weltwissen zur Re-Identifizierung zählt“

Anonymisierung von Trainingsdaten



Strukturierte Daten

<u>Etablierte Methoden:</u>

- K-Anonimität
- L-Diversität
- T-Geschlossenheit
- Differential Privacy

Anonymisierung ist mit Verfahren des technischen Datenschutzes gut erreichbar

Anonymisierung von Trainingsdaten

Anonymisierung

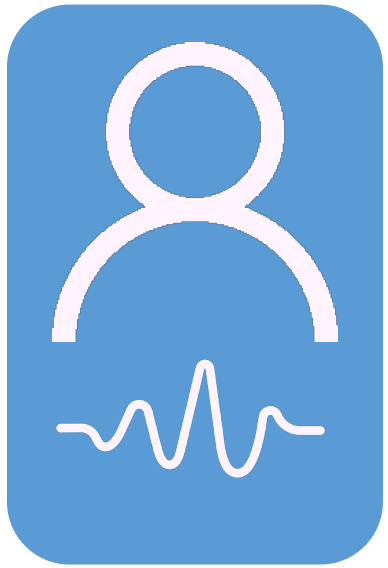
- Unmittelbar identifizierende Merkmale
- Beispiel: Name, Versicherungsnummer, Mailadresse

- Kontextinformationen
- Beispiel: Bestimmte Gesundheitsinformationen aus Entlassbrief aus dem Krankenhaus

Semistrukturierte Daten
(z.B. Texte)

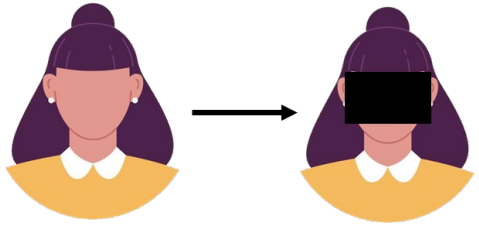
Bei komplexeren Informationsstrukturen in längeren Texten wird eine Anonymisierung sehr aufwändig

Anonymisierung von Trainingsdaten



Anonymisierung

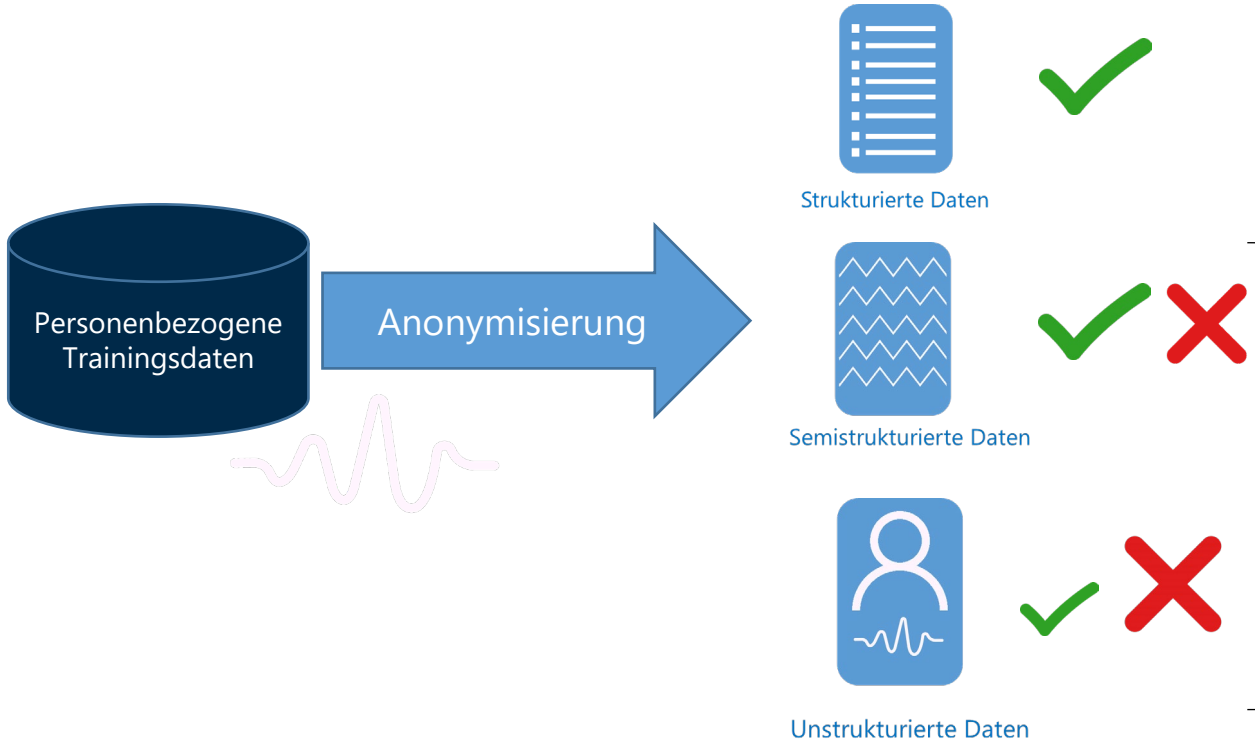
Anonymisierung häufig nur zu Lasten eines Qualitätsverlusts möglich



Unstrukturierte Daten
(z.B. Fotos, Audiosignale)

Bei unstrukturierten Daten ist eine Anonymisierung unter Beibehaltung einer guten Datenqualität nur in Ausnahmefällen möglich

Anonymisierung von Trainingsdaten



Plan B zur Anonymisierung:
In der **DS-GVO bleiben**

Bausteine:

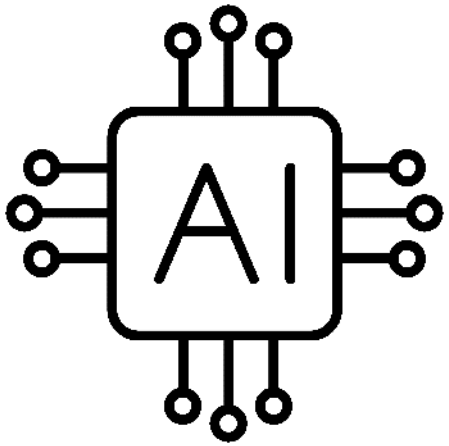
- Privilegierung Forschung
- Gesetzliche Regelungen für KI
- Interessensabwägung geht i.d.R. gut

Was sind denn eigentlich KI-Risiken?

(Herr Sachs)

KI-Risiken

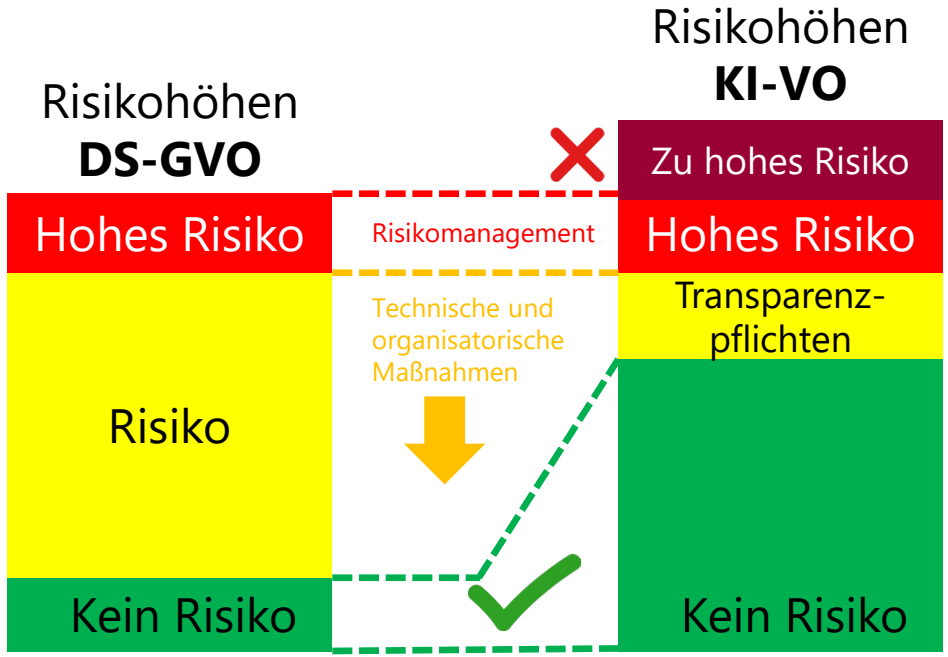
KI-Systeme



Einsatzzweck

DS-GVO und KI-VO: Der **Einsatz von KI-Technologie** zu konkreten **Zwecken** kann zu Risiken führen

(Ausnahme KI-VO: KI-Modelle mit systemischen Risiken)



KI-Risiken

- Es etablieren sich (international) zunehmend Risikoblickwinkel aus dem Bereich „**Vertrauenswürdige KI**“
  - Mit Blick auf die **KI-VO** dürften die sog. „***Ethik-Leitlinien für eine vertrauenswürdige KI***“ bei einer Bestimmung von KI-Risiken **eine zentrale Rolle** einnehmen

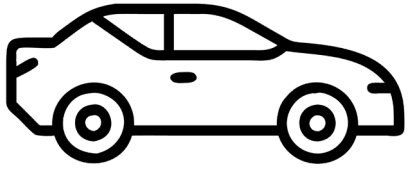
  - Daraus lassen sich sog. „**Schutzziele**“ ableiten, bei deren Nichterreichung Risiken auftreten können

https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

KI-Risiken (Ein paar Beispiele)

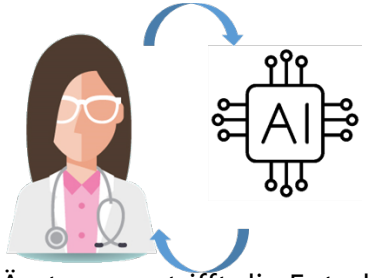
Risiko „Mangelnde menschliche Kontrolle“

- Anwendbar in DS-GVO und KI-VO: Hohes Risiko
- Implementierung von **menschlicher Kontrollmöglichkeit** in bestimmten Hochrisiko-KI-Systemen
- Beispiele:



Autonomes Fahren (in sog. Level 3)

Automatisierte Kreditentscheidung

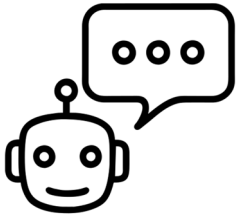


KI und Ärzte – wer trifft die Entscheidung? (sog. Automation Bias)

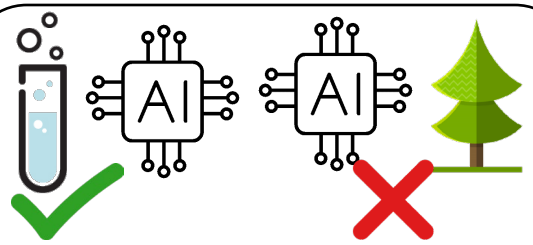
KI-Risiken (Ein paar Beispiele)

Risiko „Mangelnde Robustheit und Genauigkeit“

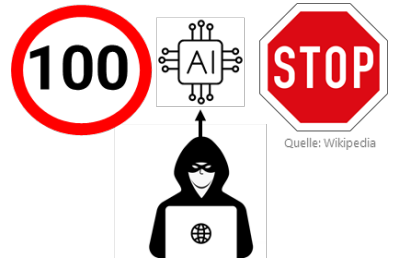
- Anwendbar in DS-GVO: Hohes Risiko | Risiko
- Anwendbar in KI-VO: Hohes Risiko
- KI macht (wie wir Menschen) **Fehler** und kann auch „**ausgetrickst**“ werden
- Beispiele:



Systembedingte „Halluzinationen“ bei Großen Sprachmodellen



Schlechte Datenqualität führt zur Unzuverlässigkeit im Echtbetrieb



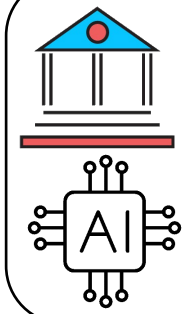
Adversiale Angriffe

KI-Risiken (Ein paar Beispiele)

Risiko „Mangelnde Fairness“

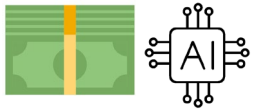
- Anwendbar in DS-GVO: Hohes Risiko | Risiko
- Anwendbar in KI-VO: Hohes Risiko
- KI hat (wie wir Menschen) **eine Art Charakter** – dieser kommt aus den **Trainingsdaten**
- Beispiele:

Verzerrung(Bias) in Trainingsdaten: Eine **Bewerbungs-KI bevorzugt** systematisch **Männer**



Diskriminierende KI: **Kreditentscheidungen** sind vom **Wohnviertel** eines Antragstellers abhängig

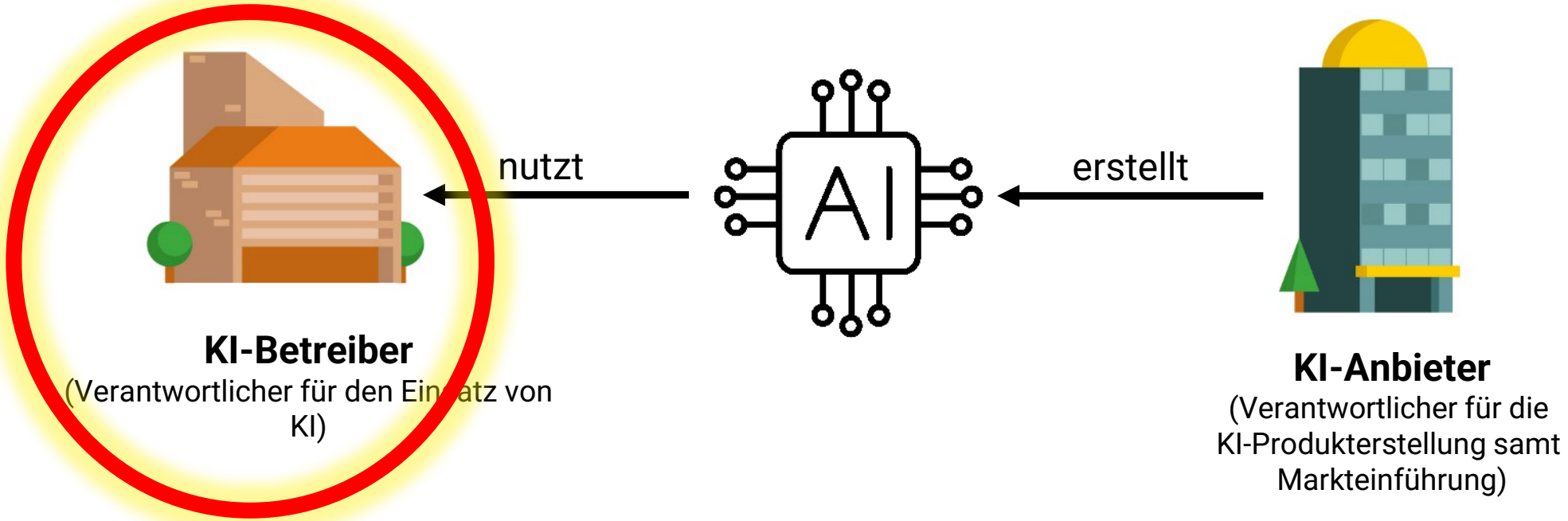
Diskriminierende KI: **Betrugserkennungs-KI** bei Sozialleistungen **klassifiziert** Menschen mit **Migrationshintergrund** häufig als potentielle **Betrüger**



KI-VO aus Sicht des Datenschutzes denken – Synergien nutzen und Aufwände reduzieren

(Herr Sachs)

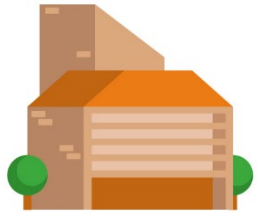
Aus Grundlagen-Webinar: Rollen und Verantwortlichkeiten



KI-Betreiber
(Verantwortlicher für den Einsatz von KI)

KI-Anbieter
(Verantwortlicher für die KI-Produkterstellung samt Markteinführung)

KI-VO aus Sicht des **Datenschutzes** gedacht: **Synergien** nutzen



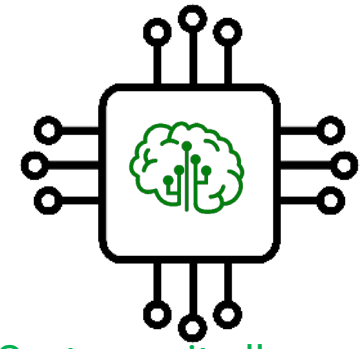
KI-Betreiber

DS-GVO	**Voraus-setzungen** prüfen	**Informationspflichten** Zielgruppenspezifisch umsetzen	Betrieblicher DSB führt Beratung/**Schulung** durch (sofern bestellt)	Auf **Hochrisiko-KI prüfen** und ggf. DSFA umsetzen	**Technische und organisatorische Maßnahmen** für KI-Betrieb anwenden	**Menschliche Kontrolle** (Art. 22 DS-GVO)
KI-VO	**Voraus-setzungen** prüfen	**Transparenz-**anforderungen (Art. 50 KI-VO)	**KI-Kompetenz** (Art. 4 KI-VO), ggf. KI-Beauftragter	**Risikoklassifik-ation** nach KI-VO	**Betreiberpflichten** nach **Anleitung** eines KI-Systems (Art. 26 KI-VO)	**Menschliche Kontrolle** (Art. 14 KI-VO)

KI-VO vs. DS-GVO: Herausforderung KI-Betreiber bei Hochrisiko-KI

KI Modell mit allgemeinem Verwendungszweck

z.B. OpenAI GPT4



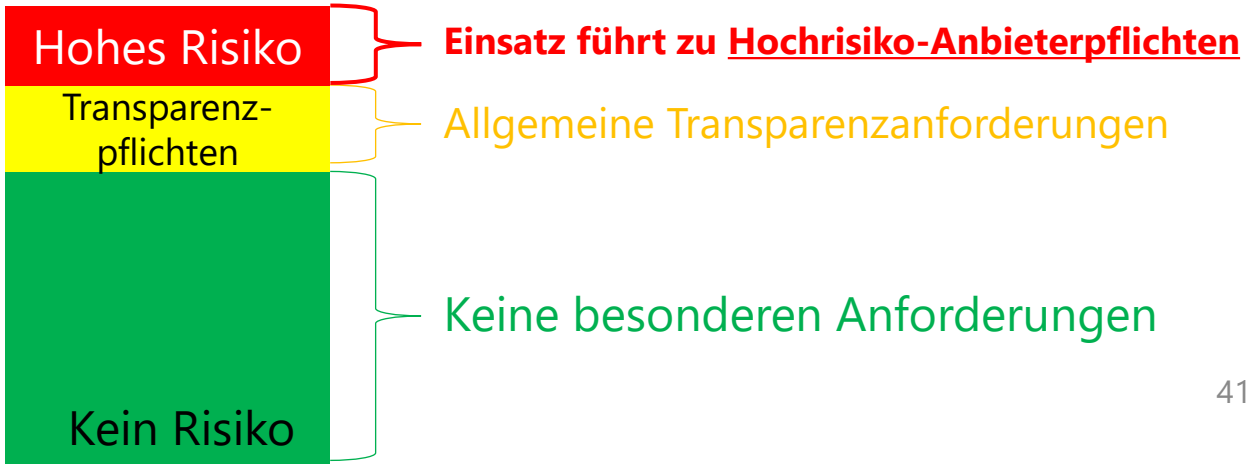
z.B. MS 365 Copilot

KI System mit allgemeinem Verwendungszweck

DS-GVO: „Normaler“ Verantwortlicher, z.B.

- Vertrag zur Auftragsverarbeitung
- Zweckbindung der Inhaltsdaten sicherstellen
- Rollen-/Rechtekonzepte bei größeren Einheiten

KI-VO: Zwecksetzung führt zur Rolle nach KI-VO



Zusammenfassung

Zusammenfassung DS-GVO vs. KI-VO (Vertiefung)

- ☑ EDPB Opinion: **Personenbezug** von **KI-Modellen**
- ☑ EDPB Opinion: **3-Stufen-Prüfung** Interessensabwägung
- ☑ EDPB Opinion: Auswirkung **Unrechtmäßigkeit** von **Trainingsdaten** auf KI-Betrieb
- ☑ **Erforderlichkeit** von **vielen und guten Daten** für **KI-Training** erkennt auch die DS-GVO an
- ☑ **Anonymisierung**: Bei **strukturierten Daten** sehr gut umsetzbar
- ☑ **Anonymisierung:** Bei **semi-/unstrukturierten Daten** eine Herausforderung
- ☑ **KI-Risiken:** Ansätze aus der „ethischen KI“ auch für Datenschutz nutzbar
- ☑ **Synergien** aus Sicht der DS-GVO bei KI-Betreibern bei der KI-VO nutzen

Vielen Dank für Ihre Aufmerksamkeit