KI im Unternehmen – Grundlagen zum Datenschutz richtig gestalten

IHK-Webinar-Reihe „Daten in der Praxis“ am 28.02.2024

Rechtsanwältin, Fachanwältin IT-Recht Isabell Conrad (CSW Rechtsanwälte)

THEMEN

01 Anwendungsbeispiele

02 Datenschutzgrundlagen bei KI

03 Checklisten der Datenschutzbehörden

04 Datenschutz & KI-Verordnung

01

Anwendungsbeispiele

KI-Einsatz / KI-Projekte im Unternehmen

- **Machine Vision**: Verarbeitung und Analyse von Bildern, die mit Kameras aufgenommen wurden (Objekt-/Gesichtserkennung; Extraktion von Informationen; Ermittlung von Bewegungen)

- **Natural Language Processing** mit zeichen- und wortbasierter Mustererkennung sowie semantischer Analyse (Abgleich von Worthäufigkeiten; Erkennen von semantischen Beziehungen; Filtern von Metadaten)

- **Dokumentenanalyse**
- **Generative KI** (Text- und Bildgenerierung)

Form der Nutzung von KI

- **Backoffice-Applikation / Apps auf mobile Devices / Browser-basiert**: virtuelle Assistenten, Suchmaschinen, Transkription, Textkorrektur, automatische Übersetzung, Altersverifikation anhand von Fotos, medizinische Assistenz-Systeme (Erkennen von krankem Gewebe) …
- **„Embedded“** KI: Roboter, autonome Pkw, Drohnen, Anwendungen des „Internets der Dinge“ …

Aktuelle Beispiele

24 der besten KI-Chatbots für 2023

Quelle: https://www.zendesk.de/service/messaging/chatbot/#

Im Folgenden finden Sie weitere Informationen zu einigen der beliebtesten KI-Chatbots des Jahres 2023, einschließlich ihrer Funktionen und Preise. Diese KI-Chatbots wurden anhand von Nutzerbewertungen, ihren allgemeinen Merkmalen, Funktionen und ihrer Leistung bewertet.

- ChatGPT
- Zendesk
- Bard
- Bing
- Claude
- Perplexity AI
- Jasper Chat
- Character.AI
- ChatSonic
- Replika
- My AI
- Khanmigo
- Woebot Health
- Ideal
- Pi
- Copy.ai
- You.com
- ZenoChat
- Amazon CodeWhisperer
- HuggingChat
- Rose AI
- Socratic by Google
- ChatSpot
- Certainly

KI-Pioniere aus Köln und Heidelberg

Die deutschen KI-Unternehmen DeepL und Aleph Alpha wollen weltweit Maßstäbe setzen – bei Übersetzungen und beim digitalen Bürgerservice.

Quelle: Carsten Hauptmeier, 4.4.2023
www.deutschland.de

Quelle: https://www.blogmojo.de/ki-chatbots,

Platz	Tool	KI-Modelle	Kann coden	Deutsch	Internet -zugriff	Bilder-KI	Datei-Upload	Preis (netto, Monat)
1	Neuroflash*	GPT-3.5 & 4	✓	✓	ab Pro-Tarif	✓	✗	kostenlos, ab 25,21 €
2	Chatsonic*	GPT-3.5 & 4	✓	✓	✓	✓	Bilder, Dokumente	kostenlos, ab 16 $
3	ChatGPT Plus	GPT-3.5 & 4	✓	✓	✓	✓	Bilder, Dokumente	20 $
4	Microsoft Copilot	GPT-3.5 & 4	✓	✓	✓	✓	Bilder	kostenlos, ab 22 €
5	Claude 2	Claude 2.1	✓	✓	✗	✗	Dokumente	kostenlos
6	Google Gemini	Gemini Pro 1.0, Gemini Ultra 1.0	✓	✓	✓	✓	Bilder, Dokumente	kostenlos, 20 $
7	ChatGPT	GPT-3.5	✓	✓	✗	✗	✗	kostenlos
8	YouChat	C-A-L, GPT-4	✓	✓	✓	✓	✗	kostenlos, ab 9,99 $
9	Perplexity	GPT-3.5 & 4, Claude 2.1	✓	✓	✓	✗	Bilder, Dokumente	kostenlos, ab 16,70 $
10	Pi	Inflection-2	✓	✓	✓	✗	✗	kostenlos
11	OpenAI Playground	GPT-3.5 & 4	✓	✓	✗	✗	Bilder, Dokumente	kostenlos

Quelle: www.wiwo.de vom 8.7.23

KÜNSTLICHE INTELLIGENZ MADE IN GERMANY

Die 20 aussichtsreichsten KI-Start-ups in Deutschland*

Start-up	Sitz	Gründung	Finanzierung *(in Millionen Euro)*
Aleph Alpha	Heidelberg	2019	128,30
askUI	Karlsruhe	2021	1,80
Brighter AI**	Berlin	2017	
Celus	München	2018	30,40
Deepset	Berlin	2018	14,20
Frequenz	Berlin	2019	13,00
Helsing	München	2021	102,50
HQS Quantum Simulations	Karlsruhe	2017	14,30
Hyperganic	München	2014	6,40
Luminovo	München	2020	13,50
Navvis	München	2013	88,00
Parloa	Berlin	2017	24,00
Paretos	Heidelberg	2020	10,00
Qdrant	Berlin	2021	8,95
Scoutbee	Würzburg	2015	69,40
SPREAD	Berlin	2019	18,90
Taktile	Berlin	2020	22,50
Twaice	München	2018	71,50
Ultimate.ai	Berlin	2017	25,00
ZenML	München	2021	2,47

* laut Jurywertung (NVIDIA, Intel, Cherry Ventures, Earlybird Capital, UVC Partners, Digital+ Partners, Lakestar, High-Tech Founder Funds, eCAPITAL, La Famiglia, Asgard, Burda Principal Investments, HV Capital und Born2Grow); ** Finanzierung unbekannt ; **Quelle:** AppliedAI Institute for Europe, Crunchbase ; **Tabelle:** Gerd Weber

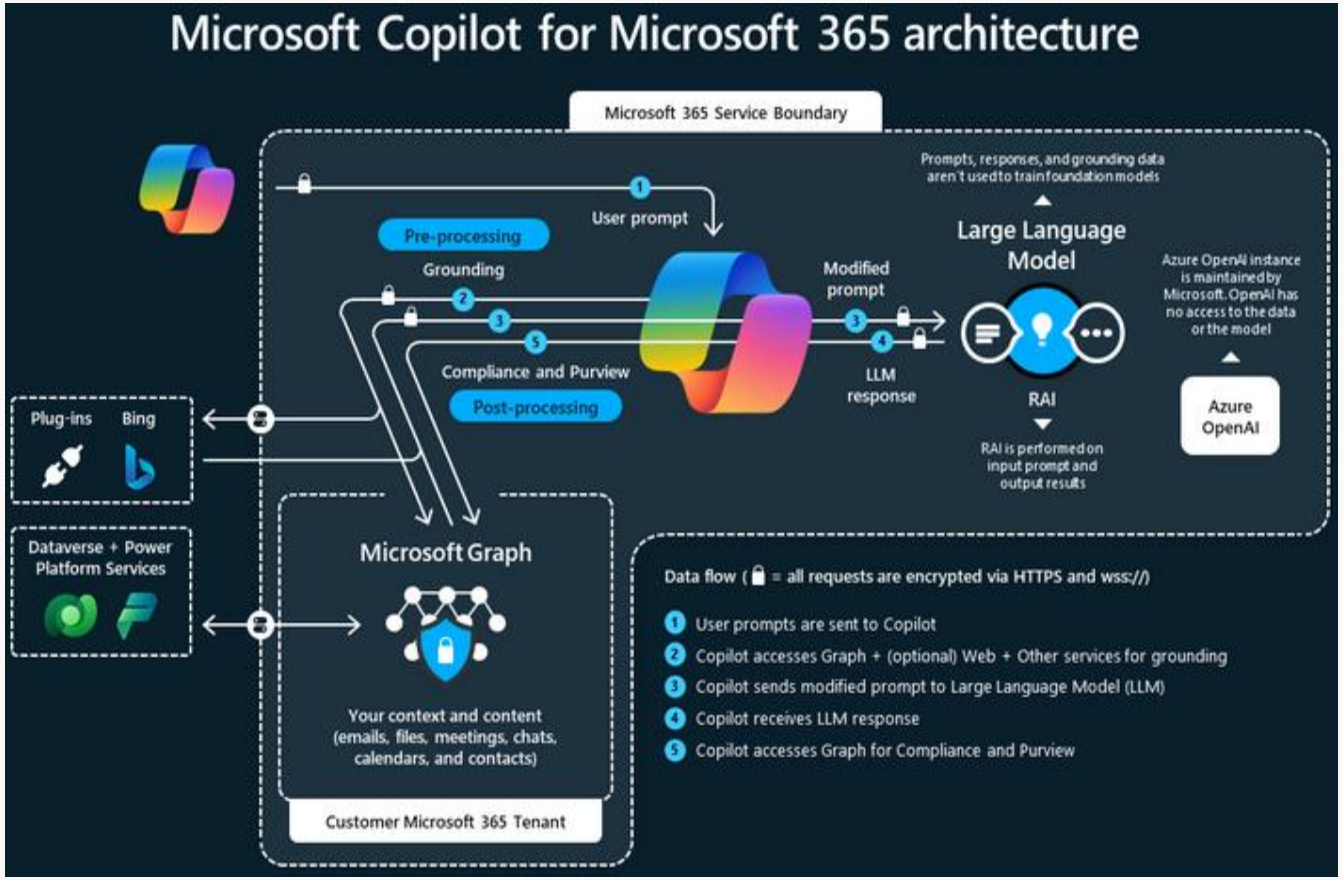
Wirtschafts Woche

Find the Azure service for your AI + machine learning needs

Quelle: AI and Machine Learning - Azure Services | Microsoft Azure

If you want to	Use this
Find an AI service that monitors metrics and diagnoses issues	Azure AI Metrics Advisor
Keep your content safer with better online experiences	Azure AI Content Safety
Easily extract meaningful insights from audio and video files using media AI	Azure AI Video Indexer
Create bots and connect them across channels	Azure AI Bot Service
Find insights using enterprise-scale search for app development	Azure AI Search
Design AI with Apache Spark™-based analytics	Azure Databricks
Use an enterprise-grade service for the end-to-end machine learning lifecycle	Azure Machine Learning
Add cognitive capabilities to apps with APIs and AI services	Azure AI Services
Empower users of all ages and abilities to read and comprehend text	Azure AI Immersive Reader
Easily add anomaly detection capabilities to your apps.	AI Anomaly Detector
Apply advanced coding and language models to a variety of use cases	Azure OpenAI Service PREVIEW
Automate content moderation for image, text, and video	Content Moderator

KI und BetrVG, GeschGehG, UrhG …



Quelle: https://learn.microsoft.com/de-de/microsoft-365-copilot/microsoft-365-copilot-overview

LAG Köln, Beschl. v. 21. Mai 2021 - 9 TaBV 28/20: unternehmensweite Einführung von Microsoft O365 ist nach § 87 Abs. 1 Nr. 6 BetrVG mitbestimmungspflichtig. Arbeitgeber hat den zentralen Zugriff auf die Administrationsrechte der 365 Cloud (im Azure Single Tenant) und damit auf die Log-Dateien der bearbeiteten Dokumente. Die Microsoft 365 Apps ermöglichen eine Leistungs- und Verhaltenskontrolle.

Microsoft Copilot nutzt für den Output den Prompt, den Kontext des Users und die Infrastruktur der 365 Cloud (zentraler Zugriff auf Logdaten, Emails, Vikos, Chats, Kalender etc. über Microsoft Graph). Microsoft Copilot wohl ebenfalls mitbestimmungspflichtig.

VG Hannover, Urt. v. 9.2.2023 – 10 A 6199/20: Umfang der Echtzeitdatenerfassung und -auswertung über Beschäftigte im Logistikzentrum

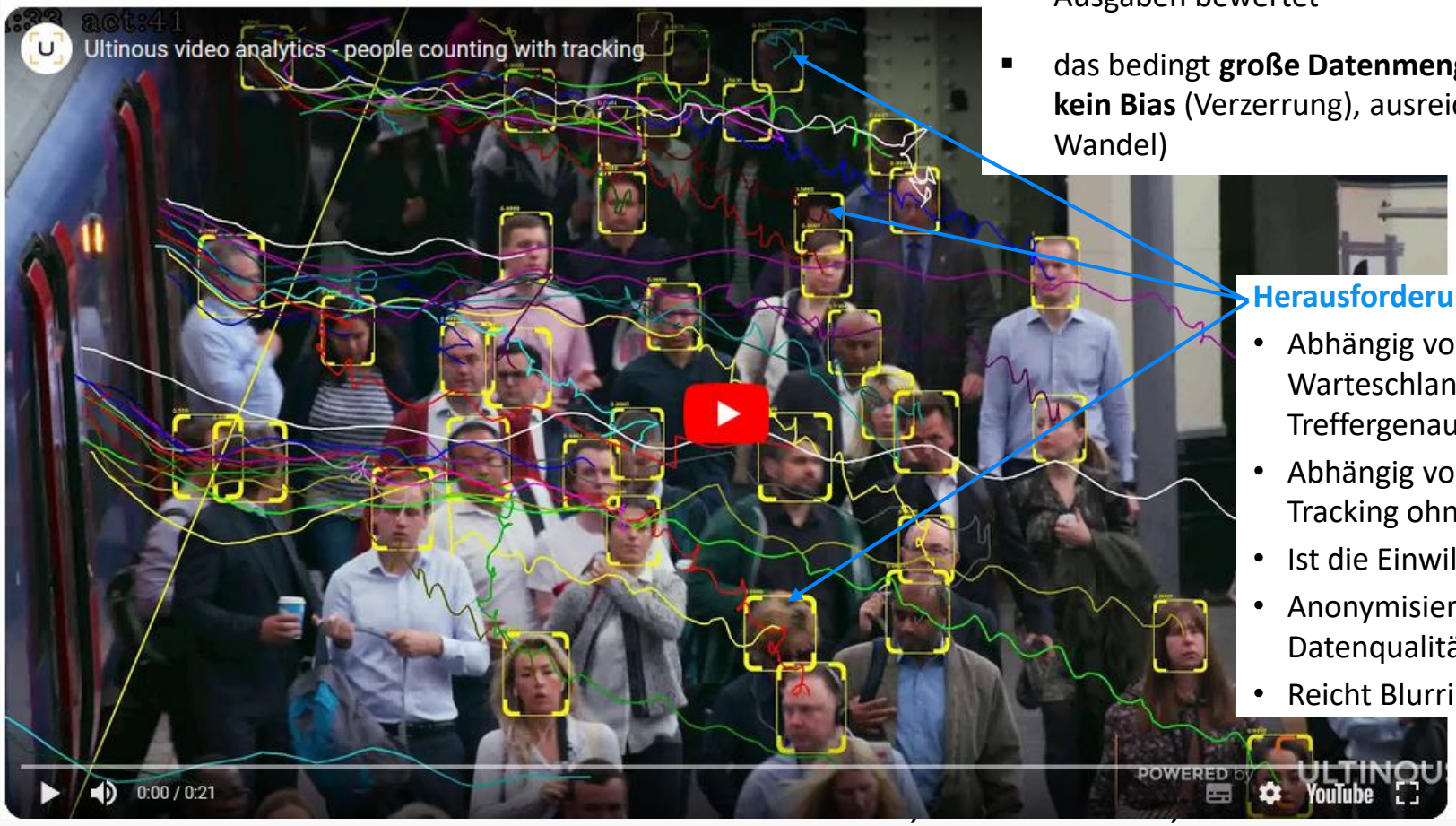
ArbG Hamburg, Beschl. v. 16.01.2024 - 24 BVGa 1/24:

Arbeitgeber erstellt **betriebliche Richtlinien (Guidelines, Handbook …)** mit Vorgaben an Arbeitnehmern, wenn diese bei der Arbeit KI-Tools nutzen. Im Intranet werden die Mitarbeiter über diese Vorgaben informiert und es heißt: *„Nutzen wir die generative KI als neues Werkzeug, um unsere Arbeit zu unterstützen.“* Die KI-Tools (wie ChatGPT) werden aber nicht on-premise beim Arbeitgeber betrieben und der Arbeitgeber legt auch keine dienstlichen Accounts an. Arbeitnehmer nutzen private Accounts auf eigene Kosten. Der Arbeitgeber erhält keine Daten, wann welcher Arbeitnehmer wie lange und mit welchem Anliegen ChatGPT genutzt hat.

Der KBR fordert im Wege der einstweiligen Verfügung, die Nutzung solche KI-Systeme zu untersagen, solange nicht eine Rahmen-KBV zu KI vereinbart ist. Das Gericht lehnt den Antrag des KBR ab.

Wie gut ist meine KI?

Beispiel Kopferkennung:



- sehr häufig wird beim Maschinellen Lernen die **statistische Korrektheit** der Ausgaben bewertet
- das bedingt **große Datenmengen** mit den Kriterien: **Verfügbarkeit**, **Datenqualität**, **kein Bias** (Verzerrung), ausreichend **Varianz**, Berücksichtigung von **Drift** (ständigem Wandel)

Herausforderungen für die KI (Beispiel Kopferkennung):

- Abhängig vom Einsatzzweck (z.B. Pflegeeinrichtung, Warteschlangen-Management im Supermarkt etc.) welche Treffergenauigkeit (Confidence) der Kopferkennung ist erforderlich?
- Abhängig vom Einsatzzweck: Sind Videos / Kopferkennung / Tracking ohne Einwilligung zulässig?
- Ist die Einwilligung praktikabel?
- Anonymisierung durch z.B. Blurring von Videos? = schlechte Datenqualität und beeinträchtigt die Confidence
- Reicht Blurring nur des Kopfes für eine Anonymisierung?

02

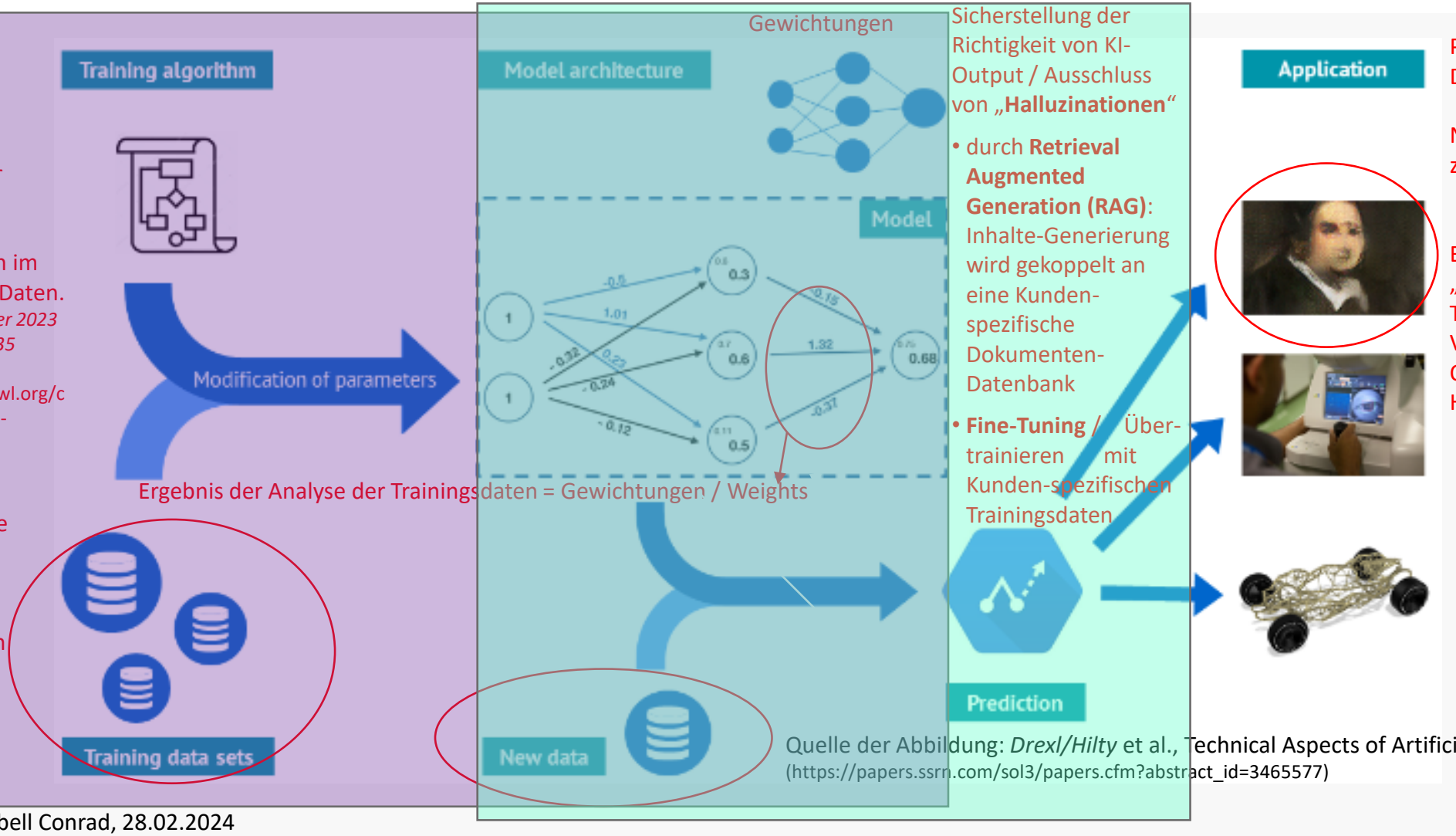
Datenschutzgrundlagen bei KI

Entwicklungsphase Projektphase Produktivphase

Transparenz hins. der Trainingsdaten?

Z.B. Crawler sammeln im Internet zugängliche Daten.
„The November/December 2023 crawl archive contains 3.35 billion pages“
https://data.commoncrawl.org/crawl-data/CC-MAIN-2023-50/index.html

- Personenbezogene Daten?
- Art. 9 DSGVO?
- Zweckänderung?
- Info-Pflichten nach Art. 14 DSGVO?



Sicherstellung der Richtigkeit von KI-Output / Ausschluss von „**Halluzinationen**“

- durch **Retrieval Augmented Generation (RAG)**: Inhalte-Generierung wird gekoppelt an eine Kunden-spezifische Dokumenten-Datenbank
- **Fine-Tuning** / Über-trainieren mit Kunden-spezifischen Trainingsdaten

Personenbezogene Daten in Prompts?

Nutzung der Prompts zum Weitertraining?

Bewertung der „Memorization“ von Trainingsdaten als Verarbeitung beim Output oder schon bei Hosting des Modells

Quelle der Abbildung: *Drexl/Hilty* et al., Technical Aspects of Artificial Intelligence, S. 5-6 (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3465577)

Verarbeitung personenbezogener Daten bei LLM

Memorization von Trainingsdaten:

(1) Training eines LLMs regelmäßig auch mit personenbezogenen Daten; Text- und Data-Mining ist nach § 44b UrhG zulässig; aber die Daten sind nach der Analyse (Training) zu löschen

(2) Sehr strittig, ob allein durch Speicherung (Hosting) eines trainiertes LLM (ohne Prompts und Output) personenbezogene Daten verarbeitet werden
- **Ansicht 1)** nimmt Personenbezug bei Reproduzierbarkeit von Trainingsdaten im Output an (ggf. vergleichbar mit verschlüsselten Daten; komprimierten Daten)
- **Ansicht 2)** verneint den Personenbezug der Weights; Reproduzierbarkeit nur bei Brut Force Attacken; ggf. pb Output ist Ergebnis einer Ad-hoc Errechnung der Daten auf Basis von Wahrscheinlichkeitswerten

(3) Möglicherweise kann der Einsatz von Output-Filtern, RAG oder Fine-Tuning o.a., soweit sie eine Reproduzierbarkeit ausschließen, eine Anonymisierungs- oder Löschungsmaßnahme sein.
Datenschutzrechtl. sicherer: Bereinigung der Trainingsdaten vor dem Training

Personenbezogene Daten
Informationen mit Bezug zu identifizierten oder identifizierbaren natürlichen Person (**EuGH Rs. C-413/23 anhängig**)

Nutzungsdaten der KI-Anwendung
z.B.: welcher Beschäftigte das LLM wann nutzt

Trainingsdaten

Google DeepMind

Language Modeling Is Compression

Grégoire Delétang[*1], Anian Ruoss[*1], Paul-Ambroise Duquenne[2], Elliot Catt[1], Tim Genewein[1], Christopher Mattern[1], Jordi Grau-Moya[1], Li Kevin Wenliang[1], Matthew Aitchison[1], Laurent Orseau[1], Marcus Hutter[1] and Joel Veness[1]
[*]Equal contributions, [1]Google DeepMind, [2]Meta AI & Inria https://arxiv.org/pdf/2309.10668.pdf

It has long been established that predictive models can be transformed into lossless compressors and vice versa. Incidentally, in recent years, the machine learning community has focused on training increasingly large and powerful self-supervised (language) models. Since these large language models exhibit impressive predictive capabilities, they are well-positioned to be strong compressors. In this work, we advocate for viewing the prediction problem through the lens of compression and evaluate the compression capabilities of large (foundation) models. We show that large language models are

LLM als Solches?
Speicherung der Gewichtungen (Weights) im Modell als Verarbeitung pb Daten?

Ausgabedaten (Output)
Bei Personenbezug des Prompts immer gegeben

Eingabedaten (Prompts)
z.B.: Eingabe eines Namens in das LLM

Bei Microsoft Copiloten **„Grounding" der Prompts**
Berücksichtigung von weiteren Daten (wie Chats, Emails, Kalendereinträge, Dokumente etc. via Microsoft Graph), Kontext des Nutzers und ggf. Bing für den Output

Zu prüfen beispielsweise:
- Mandantentrennung im Azure Single Tenant?
- Datenbasis bei Nutzern mit Admin-Rechten?

Datenschutzrechtliche Verantwortlichkeit

Verantwortlicher

entscheidet allein oder gemeinsam mit anderen über Zweck und Mittel der Verarbeitung

Pflichten, insbes.

- Erlaubnis der Verarbeitung (zu differenzieren zwischen Phasen und Zwecken; bei Training relevant: Interessen-abwägung nach Art. 6 Abs. 1 lit. f DSGVO dokumentieren)
- Art. 9 DSGVO sehr weit ! (**EuGH Rs. C-184/20; EuGH Rs C-252/21**); relevant: Art. 9 Abs. 2 lit. a) oder e) oder g) DSGVO
- Daten von Minderjährigen?!
- Zulässigkeit/Dokumentation der Zweckänderung (Art. 6 Abs. 4 DSGVO)
- **Betroffenenrechte** (v.a. Auskunft, Löschung, Datenrichtigkeit)
- Beschreibung der Verarbeitungsvorgänge (VVT, ggf. DSFA Art. 30/35 DSGVO)
- Drittlandübermittlung (Kap. V. DSGVO)

- Art. 14 Abs. 2 lit. f DSGVO Information über (öffentl. zugängl.) Quellen
- Art. 14 Abs. 2 lit. g DSGVO Information über involvierte Logik, Tragweite und Auswirkung bei Profiling
- Art. 14 Abs. 4 DS-GVO Information über Zweckänderung
- ⇨Art. 10 Abs. 2 KI-VO (E) Information über Quellen und urspgl. Zweck der Trainingsdaten

KI-Entwickler

KI-Trainer

Provider der KI-Anwendung

„AI-as-a-Service“
oder
„On permise“

Kunde/KI-Anwender

Beschäftigte / Nutzer

Auftragsverarbeitung (AVV) auf Anbieterseite (Art. 28 DSGVO)

- Denkbar, soweit z.B. Trainingsdaten vom Kunden bereitgestellt nur im Auftrag und auf Weisung eines spezifischen Kunden erfolgt
- Weitertraining mit Prompts/Output: **AVV nein**, wenn das Modell so auch für andere Kunden weitertrainiert wird (Produktverbesserung etc.); dann ggf. gemeinsam Verantwortlichkeit mit dem Kunden
- **AVV nein**, soweit der Anbieter z.B. bei einem Telemediendienst (Login, Cookies) selbst Vertragspartner der Endnutzer wird (hins. der persönlichen Accounts der Nutzer)

Gemeinsame Verantwortlichkeit (JC) des Kunden (Art. 26 DSGVO: JC-Vertrag) strittig hins.:
Vor-Training des Modells (Zugriffsmöglichkeit des Kunden auf Trainingsdaten nicht erforderlich, wenn Training im Interesse des Kunden); aA.: phasenweise differenzierte Verantwortlichkeit

Art. 22 DSGVO: **Verbot der automatisierten Entscheidungsfindung sehr weit**: **EuGH v. 7.12.23 zum Schufa-Score**! Auch der Ersteller eines Scores, der ihn nicht verwendet, fällt unter Art. 22 DSGVO; große Relevanz für KI

https://datenschutz-hamburg.de/news/auswirkungen-des-schufa-urteil-auf-ki-anwendungen

Datenschutzgrundlagen gestalten

Auswahlkriterien:

- Eignung des Modells (Qualität / Reifegrad des Modells; verfügbare Sprachen; ist generative KI / LLM immer die geeignetste Wahl für den Einsatzzweck?)
- Vendor-Lock-in vermeiden / Kostenkontrolle (kostenpflichtige proprietärer Lizenz oder Open Source; Vertragsbedingungen des Anbieters etc.)
- Transparenz und Compliance im Hinblick auf das Vortraining (neue Abhängigkeit der Unternehmensprozesse von externen (Trainings-)Daten?)
- Serverstandorte (gehostet oder on premise); Zugriffsmöglichkeiten für Anbieter aus Drittländern?
- Richtigkeit der Ergebnisse der KI (Fehlertoleranz; Fine-Tuning; RAG)
- Transparenz und Compliance beim Umgang mit Prompts, Output und anderen Unternehmensdaten („Grounding“; Fine-Tuning etc.)
- Erfüllbarkeit von Betroffenenrechten (v.a. Art. 13, 14, 15, 17, 22 DSGVO)
- Unterstützung durch den Anbieter bei Verarbeitungsverzeichnis (Art. 30 DSGVO) und Datenschutzfolgenabschätzung (Art. 35 DSGVO; soweit für eine Verarbeitung keine Rechtsgrundlage (Art. 6, 9 DSGVO) vorliegt, muss keine DSFA durchgeführt werden)

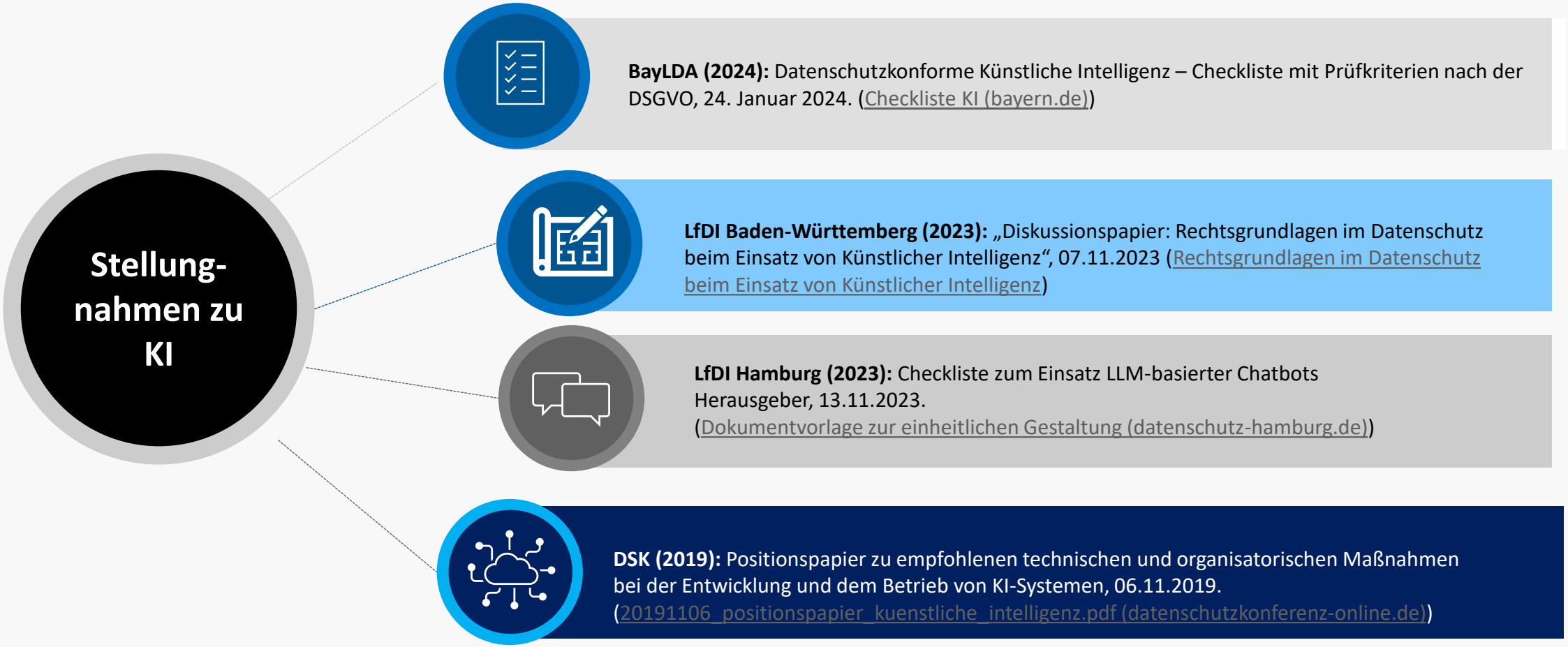
⇨ Sehr viele verschiedenen Einsatzmöglichkeiten / KI-Angebote / Risikoszenarien; one size fits all-Ansatz nicht möglich

Faustregel: je weniger Funktionalitäten / Datenquellen / Datenarten / Betroffenengruppen / Datenempfänger – desto schneller sind die Datenschutzaufgaben erledigt

03
Checklisten der Datenschutzbehörden

DSK, Hamburg, Baden-Württemberg, Frankreich (CNIL), BayLDA

Datenschutzbehörden Deutschland



BayLDA – Checkliste datenschutzkonforme KI

Schutzziele: Fairness, Autonomie, Transparenz, Verlässlichkeit, Sicherheit (nicht abschließend)

Datenschutz: Rechtsgrundlage für die Erzeugung und Nutzung von KI-Modellen, Erfüllung von Betroffenenrechten und weiterer Compliance-Anforderungen.

Zwei Checklisten: Prüfpunkte für Training und Einsatz von KI-Modellen.

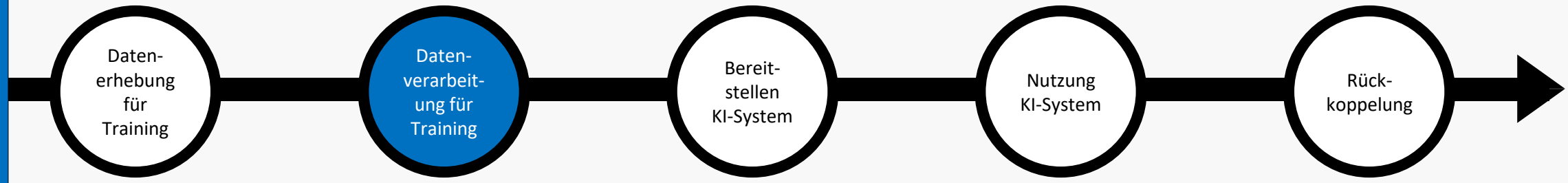
LfDI BaWü - Rechtsgrundlagen beim Einsatz von KI

RECHTS GRUNDLAGEN (hier nur Auswahl)

Einwilligung	Erfüllung eines Vertrages	Schutz lebenswichtiger Interessen	Berechtigte Interessen	Im Beschäftigtenkontext
Gewährleistung von Transparenz und Verständlichkeit	Relevant für KI-basierte Produkte und Dienstleistungen	Relevant in regulierten Bereichen (Gesundheit, Lebensschutz)	**Besonders geeignet, weil offener Tatbestand!** (aber grds. nicht für öffentliche Stellen)	**§ 26 Abs. 1 S. 1 BDSG (-), aber Art. 88 DSGVO mit Betriebsvereinbarung**

PHASEN

Wann werden Personenbezogene Daten von KI-Systemen verarbeitet?



Entwicklung **Realeinsatz**

VERANTWORT-LICHKEIT

Alleinige Verantwortlichkeit
- Entscheidung über Zweck und Mittel der Verarbeitung

Gemeinsame Verantwortlichkeit
- Verarbeitung ohne andere Partei möglich?
- Beiträge untrennbar miteinander verbunden?

Auftragsverarbeitung
- Denkbar z.B. bei KI-Entwicklung-, Cloud-, Analysedienste

Datenschutzbehörde Hamburg - Checkliste zum Einsatz LLM-basierter Chatbots

Konkrete Anleitung für den datenschutzkonformen Einsatz von (Cloud-)Chatbots (u.a. Bereitstellung beruflicher Funktionsaccounts; sichere Authentifizierung; Opt-Out-Settings für KI-Training; Opt-out-History).

Wesentliche **Herausforderungen** hinsichtlich des Datenschutzes und der Informationssicherheit.

Primäres Datenschutzrisiko: Kompromittierung von Geschäftsgeheimnissen und persönlichen Daten.

Maßnahmen: klare Nutzungsrichtlinien, Gestaltung im Einklang mit Datenschutz, kontinuierliche Überwachung der Datenverarbeitungsvorgänge.

Datenschutzkonferenz - empfohlenen TOM bei Entwicklung und Betrieb von KI-Systemen

Hintergrund: Konkretisierung der Hambacher Erklärung (vom 03.04.2019)

Datenschutzkonforme Gestaltung über den gesamten **Lebenszyklus** von KI

TOMs für

1) Design des KI-Systems

2) Rückkoppelung bei Nutzung des KI-Systems

3) Selbstveränderung des Systems

Risiken nach Einsatzszenario und Art der KI-Komponenten.

Zuordnung der TOMs zu Gewährleistungszielen (Transparenz, Datenminimierung, Nichtverkettung, Intervenierbarkeit, Verfügbarkeit, Integrität, Vertraulichkeit)

Zuordnung der Maßnahmen zu Verarbeitungsschritten und Gewährleistungszielen

CNIL (2022) - Leitfaden zur Selbsteinschätzung

Für wen? Für Anbieter und Betreiber - Handreichung zur datenschutzkonformen Entwicklung von KI-Systemen

Ziel: Selbstbewertung aller relevanten Aspekte in Bezug auf Datenschutz und Ethik für Verarbeitungen durch KI ermöglichen.

Umsetzung: 7 Faktenblätter zu spezifischen Themen

1) Vorfragen zum verhältnismäßigen KI-Einsatz
2) Sammeln und Qualifizieren von Trainingsdaten (Aufbau einer hochwertigen Datenbank)
3) Entwicklung und Training eines Algorithmus mit bewährten Verfahren
4) Gewährleistung von Qualität und Transparenz im Produktivbetrieb
5) Analysieren von Risiken und Verhindern von Schwachstellen und Angriffen
6) Förderung von Transparenz und Ermöglichung der Betroffenenrechte
7) Verantwortlichkeiten zuweisen und Verarbeitung dokumentieren

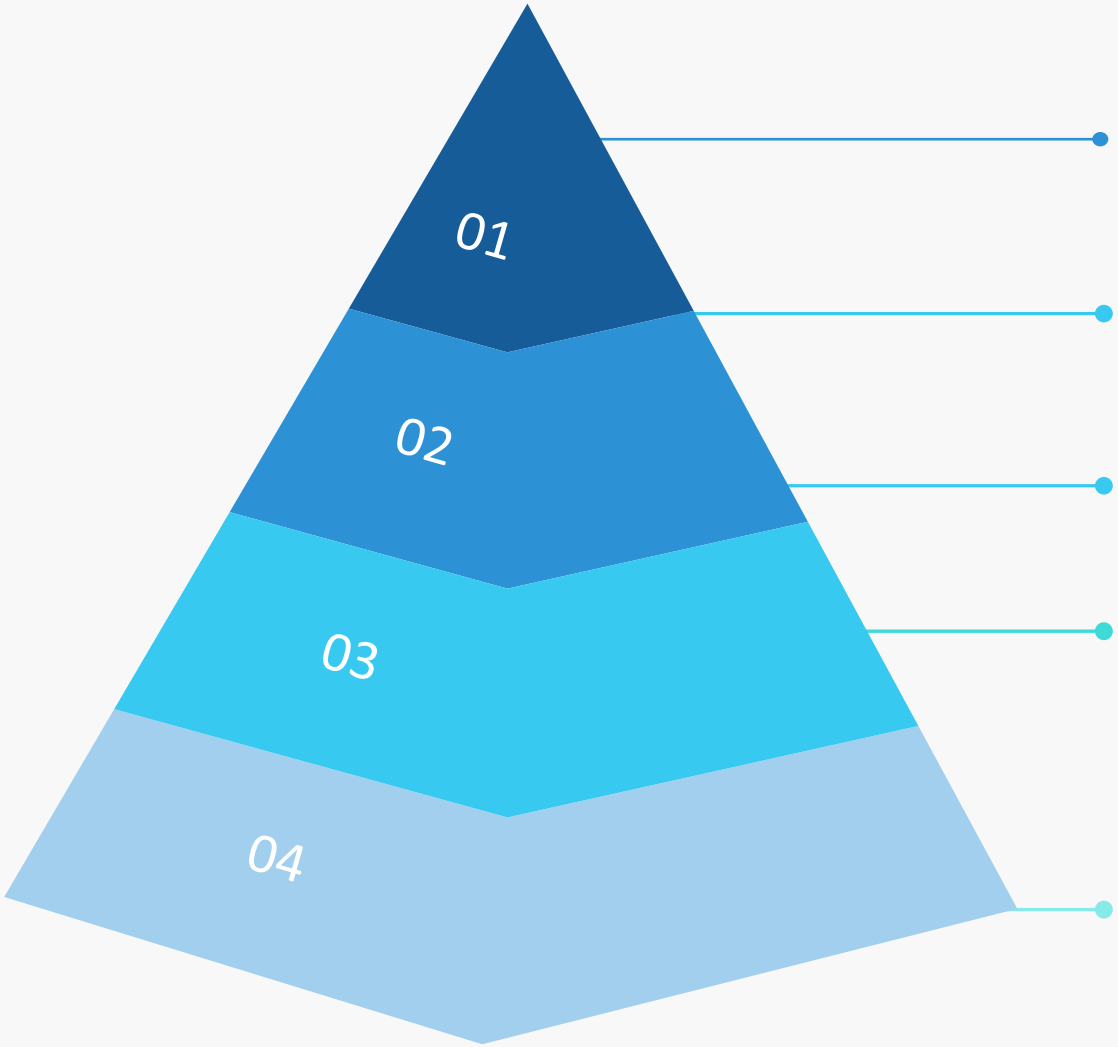
https://www.cnil.fr/fr/intelligence-artificielle/guide/conformite-des-systemes-dia-les-autres-guides-outils-et-bonnes-pratiques

04
Datenschutz & KI-Verordnung

Anwendungsbereich, Geltung,
Regulierungsansatz, Datenschutzvorschriften

Anwendungsbereich, Art. 2

Geltung für wen?	Anbieter, Betreiber, Importeure und Händler, Produkthersteller, EU-Vertreter Ausnahmen	Bereichsausnahmen	Nationale Sicherheit
internationale Zusammenarbeit bei Strafverfolgung	Forschung	ausschließlich private Nutzung	Open Source (mit Ausnahmen)



UNANNEHMBARES RISIKO

z.B. Social scoring, Emotionserkennung am Arbeitsplatz ….

Verboten

HOHES RISIKO

z.B. Medizinprodukte

Zulässig, vorbehaltlich der Erfüllung der Hochrisiko-KI-Anforderungen und der Ex-ante Konformitätsbewertung

GPAI mit systemischem Risiko

Zulässig, vorbehaltlich von Informations-/Transparenzpflichten, Modellevaluierung, Verhaltenskodex etc., Kapitel III Art. 52d-e KI-VO (E)

TRANSPARENZ / GPAI

z.B. Chatbots

Zulässig, vorbehaltlich von Informations-/Transparenzpflichten, Kapitel II Art. 52c KI-VO (E)

MINIMALES RISIKO

z.B. Spamfilter

Zulässig ohne Einschränkungen durch KI-VO; aber DSGVO, UrhG …

CSW

Risikoklassifizierung in der Praxis

Impact Assessment der Kommission zur KI-Verordnung, 21. April 2021:

https://digital-strategy.ec.europa.eu/en/library/impact-assessment-regulation-artificial-intelligence, S. 68

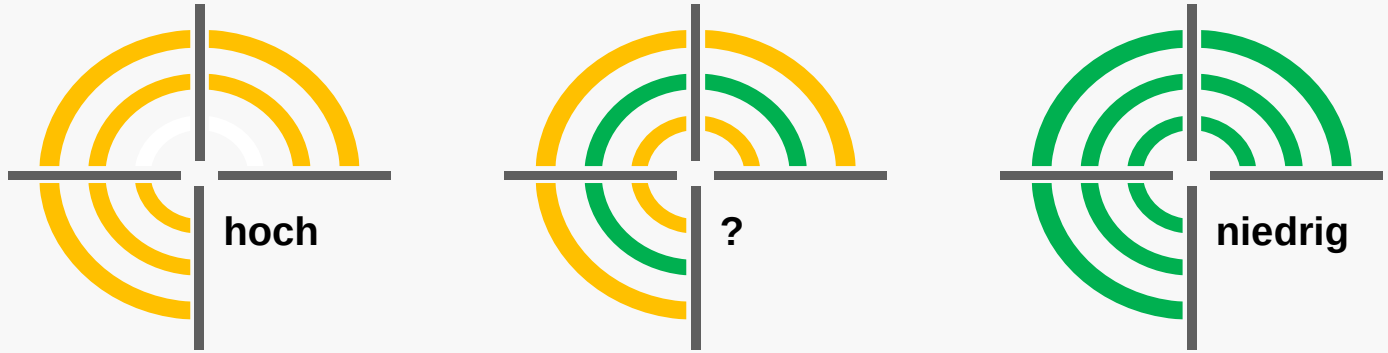
5-15 %

85-95 %

nicht bewertet

Studie 2023: Potenzieller Hochrisikoanteil zwischen 18% und 58%:

AI Act: Risikoklassifizierung von KI-Anwendungen aus der Praxisperspektive (appliedai.de)



18%

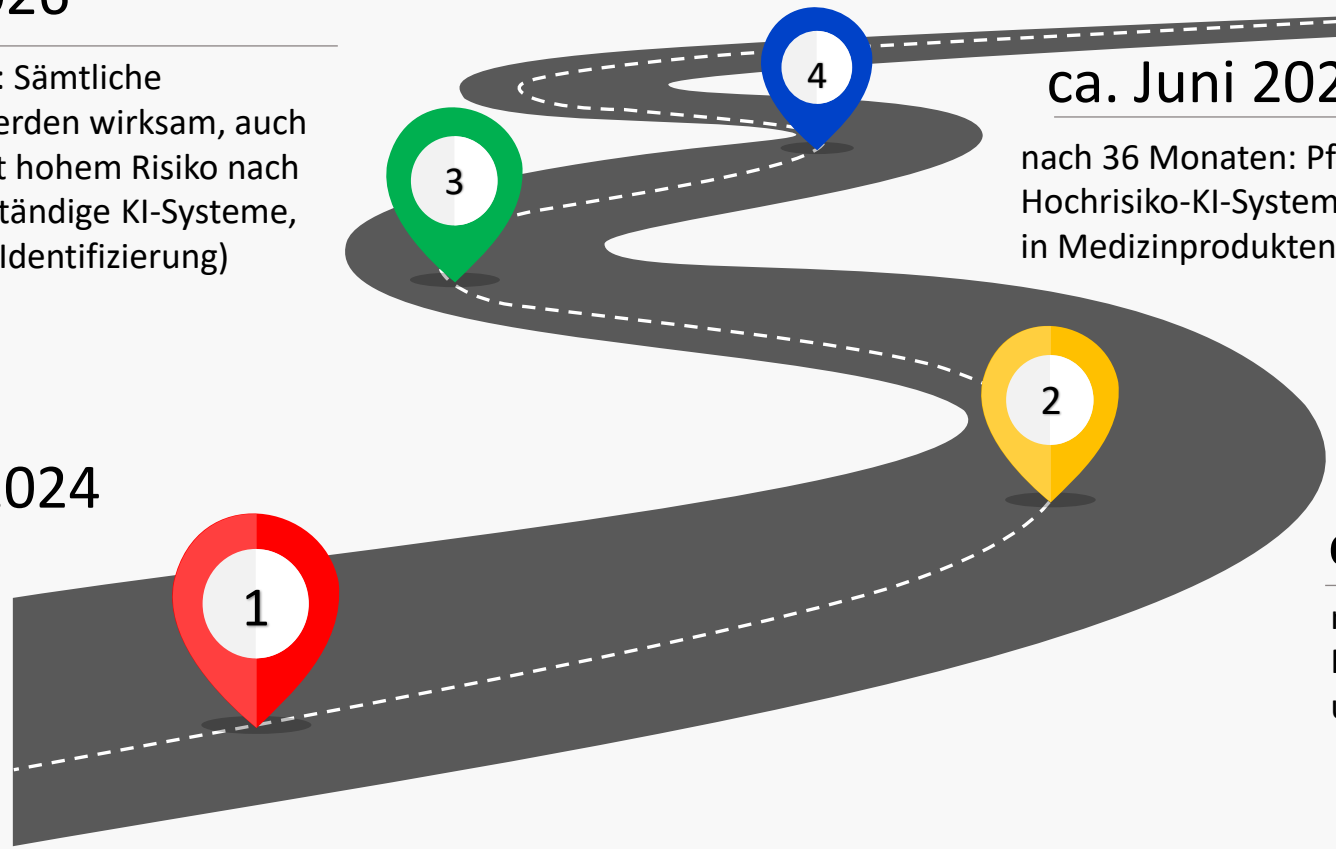
40% unklar

42%

Gestufte Anwendbarkeit

ca. Juni 2026

nach 24 Monaten: Sämtliche Bestimmungen werden wirksam, auch für KI-Systeme mit hohem Risiko nach Anhang III (eigenständige KI-Systeme, z.B. Biometrische Identifizierung)



ca. Juni 2027

nach 36 Monaten: Pflichten für Hochrisiko-KI-Systeme nach Anhang II (KI in Medizinprodukten, Maschinen etc.)

ca. Dezember 2024

nach 6 Monaten: Verbote greifen

ca. Juni 2025

nach 12 Monaten: Regeln zu Governance und GPAI

Die Regulierung von KI mit allgemeinem Verwendungszweck (GPAI)

Einheitlicher **Begriff**: "GPAI" wird als Standardbegriff verwendet, "Foundation Model" entfällt.

- **Abgestufte Regulierung** für GPAI:
  - Alle GPAI-Anbieter müssen technische Dokumentationen erstellen.
  - GPAI mit systemischen Risiken unterliegen zusätzlichen Anforderungen wie Modellbewertung und Risikominderung.
- **Schwellenwert** für systemische Risiken: Bestimmt durch hochwirksame Fähigkeiten, basierend auf Benchmarks oder einem Rechenaufwand von über 10^25 FLOPs.
- **Klassifizierungsverfahren:** Meldung an die Kommission, Argumentationsmöglichkeit für Anbieter, öffentliche Liste der GPAI-Modelle mit systemischen Risiken.

Datenschutzrecht bleibt unberührt?

Art. 2 Abs. 5a KI-VO: Datenschutzrecht bleibt unberührt:

Datenschutz-Grundverordnung (DSGVO)

Verordnung bei der Verarbeitung personenbezogener Daten durch die Organe und Einrichtungen der Union (EU) 2018/1725

ePrivacy-Richtlinie - Richtlinie 2002/58/EG

Richtlinie für den Datenschutz bei Polizei und Justiz (EU) 2016/680

Aber: Data Governance-Vorschriften in der KI-Verordnung über Trainieren, Testen und Validieren von KI, etwa …

Artikel 52c Abs. 1 KI-VO (E)
Anbieter von AI-Modellen für allgemeine Zwecke müssen:
(…)
(d) eine hinreichend detaillierte Zusammenfassung der für das Training des allgemeinen KI-Modells verwendeten Inhalte nach einer vom Amt für künstliche Intelligenz bereitgestellten Vorlage zu erstellen und öffentlich zugänglich zu machen.

Artikel 10 Abs. 2 KI-VO (E)
2. Für die Trainings-, Validierungs- und Testdatensätze gelten geeignete Datenverwaltungs- und -management-praktiken, die für den beabsichtigten Zweck des KI-Systems geeignet sind. Diese Praktiken betreffen insbesondere Folgendes:
(a) die relevanten Designentscheidungen;
(aa) Verfahren der Datenerhebung und Herkunft der Daten sowie bei personenbezogenen Daten den ursprünglichen Zweck der Datenerhebung

Datenschutzerlaubnis in der KI-Verordnung

Rechtsgrundlage nach Art. 9 Abs. 2 lit. g) DSGVO: *„erhebliches öffentliches Interesse“*?

Hilft vermutlich nicht in frühen Phasen der KI-Entwicklung, weil da Zweck oft (noch) nicht Bias-Detection

ERKENNUNG UND KORREKTUR VON BIAS/VERZERRUNGEN IN HOCHRISIKO-KI (ART. 10 V)

- Gilt zusätzlich zu **Art. 9 DSGVO**
- **Keine andere Daten verfügbar,** wie synthetische oder anonymisierte Daten.
- **Starke Sicherheitsmaßnahmen,** wie Pseudonymisierung und technische Beschränkungen.
- **Strenge Kontrollen und Dokumentation des Datenzugangs** vorhanden, um Missbrauch zu verhindern.
- **Keine Weitergabe** oder Übermittlung der Daten an Dritte.
- **Datenlöschung,** sobald die Verzerrung korrigiert ist oder die Aufbewahrungsfrist endet.
- **Dokumentierte Nachweise für** die Notwendigkeit der Verarbeitung.

+

ENTWICKLUNG VON KI-SYSTEMEN IM ÖFFENTL INTERESSE IN KI-REGULIERUNGSSANDBOX (ART. 54)

- **Entwicklungsziel im öffentlichen Interesse** (Gesundheit, Umweltschutz, Energie, etc.).
- **Keine effektive Alternative** zu personenbezogenen Daten vorhanden.
- **Effektive Überwachung** und Reaktionsmechanismen gegen Risiken implementiert.
- **Isolierte, geschützte Umgebung** für Daten.
- **Keine Datenweitergabe außerhalb der Sandbox.**
- **Persönliche Entscheidungen** oder Rechte der Betroffenen werden nicht beeinflusst.

…

=

SEHR ENGE AUSNAHMEN

Verpasste Chance?

Datenschutz spielt herausragende Rolle insb. bei KI-Entwicklung

Verhältnis bleibt ungeklärt

Hilfestellung durch Datenschutzbehörden?

Zeit für Ihre Fragen!

CSW
Rechtsanwälte

Beethovenstraße 6
80336 München
Tel. 089/54349-120
https://csw.partners/