HÄRTING

KI und Datenschutz

Praktische Herausforderungen und (erste) Lösungsansätze

Stiftung Datenschutz | Datenschutz am Mittag | 13.06.2024

Sebastian Schulz | Rechtsanwalt

Agenda

1. Von hinten aufgezäumt ...
2. Ausgewählte Rechtsfragen & Usecases
3. Fazit

HÄRTING

1

Von hinten aufgezäumt ...

KI-Governance (nach AI-Act)

1. Evaluierung aller KI-Anwendungen im Unternehmen
2. Einteilung in Risikoklassen
3. Risiko- & Gap-Analyse
4. Umsetzung bestehender Transparenzpflichten
5. Einrichtung interner Kontrollsysteme
6. KI-Richtlinien
7. Schulung
8. Dokumentation

KI-Governance (nach AI-Act)

1. Evaluierung aller KI-Anwendungen im Unternehmen
2. Einteilung in Risikoklassen
3. Risiko- & Gap-Analyse
4. Umsetzung bestehender Transparenzpflichten
5. Einrichtung interner Kontrollsysteme
6. KI-Richtlinien
7. Schulung
8. Dokumentation

Praktisch alles mit Bezug zum Datenschutz

Typische Inhalte von KI-Richtlinien – **Fokus: Datenschutz**

1. Festlegung einzusetzender Tools
2. erlaubte und verbotene Einsatzfelder von KI
3. Festlegung von Verantwortlichkeiten (Wer darf einrichten? Wer gibt frei?)
4. (Beschränkung der) Eingabe personenbezogener Daten? (Training, Prompting)
5. Pflicht zur Prüfung der Ergebnisse (Personenbezug?)
6. Transparenz / Beschäftigten (P: Blackbox bei Drittanbieter-KI)
7. Dokumentation (Rechtsgrundlagen, VVT, DSFA, Verträge, etc.)
8. (...)

KI-Nutzung und datenschutzrechtliche Relevanz

1. **Datensammlung / Filterung**
   - regelmäßig: Scraping/Crawling/Harvesting auch von personenbezogenen Daten

2. **Training**
   - auch möglich: konfektionierte Texte, Bilder, Datenbanken, ...
   - strittig: Stellt Training eine Verarbeitung iSv Art. 4 Nr. 2 DSGVO dar?

3. **Input / Prompting**
   - kaum auszuschließen: personenbezogene Daten durch Upload von Daten oder Inhalten
   - strittig: ist das „statische LLM" datenschutzrechtlich relevant

4. **Output**
   - auch möglich: personenbezogene Daten durch generierten Text oder Bild

HÄRTING

2

Ausgewählte Rechtsfragen und Usecases

Fokus: Grundprinzipien

Durch den Verantwortlichen für alle Verarbeitungsphasen zu belegen

- **Rechtmäßigkeit**
  - Treu & Glauben? → AIA als Guidance?
- **Zweckbindung**
  - Ergebnisoffene Analyse eines Sachverhalts als hinreichend definierter Zweck?
  - nachträgliche Zweckkonkretisierung im Datenschutzrecht nicht vorgesehen
- **Datenminimierung**
  - Angemessenheit, Erheblichkeit, Notwendigkeit → Umdenken nötig
- **Richtigkeit**
  - Outputkontrolle, Fine-Tuning, Nachtrainieren, ...

Fokus: Rechtsgrundlagen

- differenziert nach Trainings-, Input-, Ausgabe- und Nutzungsdaten
- Einwilligung, Vertrag, öffentlicher Auftrag
  - Stellvertretung bei einwilligungsbasierter Verarbeitung? („Opa hat Demenz.“)
- **(Art. 6 Abs. 4 DSGVO iVm) Art. 6 Abs. 1 S. 1 lit. f DSGVO**
  - Art. 34 Abs. 1 lit. g EHDS-E; Weiterverwendung von elektr. Gesundheitsdaten für KI-Training
  - keine Interessenabwägung bei sensitiven Daten (weite Auslegung des EuGH!)
- **Anonymisierung / Aggregation / PETs als (derzeit) rechtssicherste Lösungen**
  - Vorsicht bei Quasi-Identifiern als Vehikel zur Herstellung von Personenbezug (pbD als Output) und zu kleinen Gruppengrößen
  - Anonymisierung rechtfertigungsbedürftig (Art. 6 Abs. 4 iVm Art. 6 Abs. 1 DSGVO)
- (Art. 25 DSGVO) vgl. EDSA, Leitlinien 4/2019

Usecase 1: Werbeprofiling mit Gradient Boost

- **Rechtsfrage**: Ist das Erstellen einer Werbe-Scorecard auf Grundlage umfangreicher historischer Kundendatensätze mittels eines Machine-Learning-Algorithmus einwilligungsfrei rechtlich zulässig?

- **Antwort**: Klar nein.
  - Werbeprofiling unterliegt strengen Voraussetzungen (vgl. DSK).
  - Die Scorecard ist nicht nur ein Merkmal, sondern die Summe vieler Merkmale.
  - Gilt auch für eine Scorecard, die mithilfe zuvor anonymisierter Daten erstellt wurde. (Die Anwendung auf Produktivdatensätze ist das Entscheidende.)

- Gilt erst recht für die Anwendung von ML auf Produktivdatensätze.

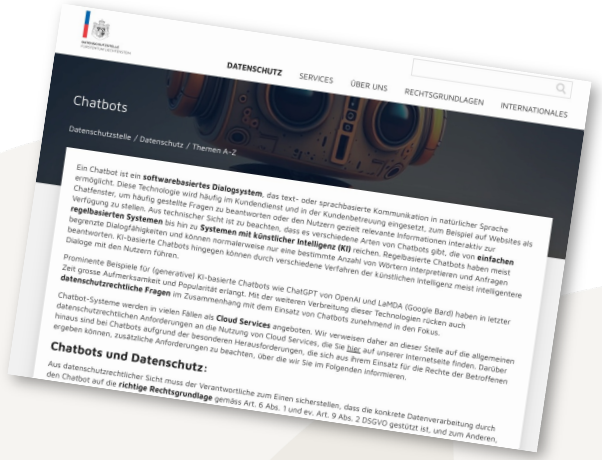

Niedersachsen. Klar.

900.000 Euro Bußgeld gegen Kreditinstitut wegen Profilbildung zu Werbezwecken

Die Landesbeauftragte für den Datenschutz (LfD) Niedersachsen hat gegen ein Kreditinstitut eine Geldbuße in Höhe von 900.000 Euro festgesetzt. Der Bußgeldbescheid ist noch nicht rechtskräftig.

Das Unternehmen hatte Daten aktiver sowie ehemaliger Kundinnen und Kunden ohne deren Einwilligung ausgewertet. Dazu analysierte es das digitale Nutzungsverhalten und wertete unter anderem das Gesamtvolumen von Einkäufen in App-Stores, die Häufigkeit der Nutzung von

Usecase 2: KI-gestützter Chatbot



- Nutzung von Chat-GPT 3.5 in privater Azure Cloud (Sandbox)
- Ausschluss von personenbezogenen Daten für das Training
  - Knowledge-Pool aus kuratierten Quellen
  - Anreicherung von Input durch Retrieval Augmented Generation (RAG)
  - Retriever aus der Knowledge-Base reduziert Halluzination
- Input- und Outputfilter verhindern ~~Eingabe~~ Verarbeitung personenbezogener Daten
- An § 25 TDDDG denken, wenn Daten lokal auf dem Endgerät verarbeitet werden.

Fokus: Datenschutzrechtliche Rollenkonstellation

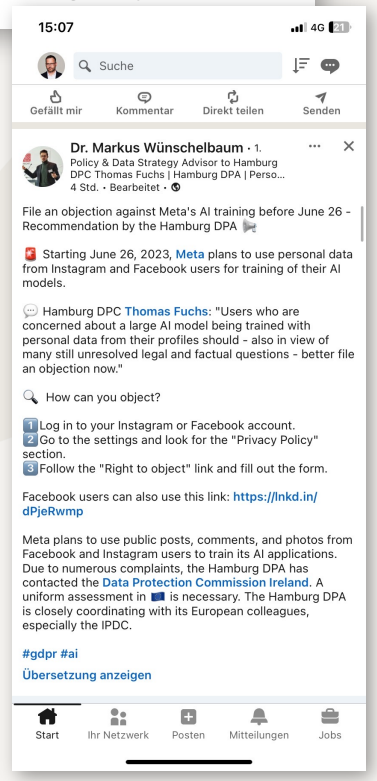
- Wesentliche Weggabelung: **on prem oder cloud?**
  - ChatGPT Enterprise API; private Cloud (z.B. Azure)
- **Was gilt nach Maßgabe des Tatsächlichen?** (nicht nach Vertragslage)
  - bei Bereitstellung von Input zu Trainingszwecken Art. 26 DSGVO naheliegend (Opt-Out nutzen)
- bei AV-Konstellationen – **Art. 28 Abs. 1 DSGVO**!
  - Speicherorte: Inputverarbeitung nur auf EU-Tenants? Belastbarkeit vertragl. Zusicherungen?
  - jedenfalls: keine privaten, nur Business-Accounts mit AV (naja...)
- **Prompter als Verantwortlicher?**
  - jedenfalls: bei dienstlich veranlasster Nutzung (regelmäßig) nur der AG!
  - EDSA: keine Verantwortungsverlagerung auf Prompter bei Nutzung für Training

Fokus: Betroffenenrechte

- Transparenz
  - Problem: Blackbox; explainable AI vor dem Durchbruch?

- **Umsetzung von Lösch-, Berichtigungs-, Widerspruchsrechten?**
  - Liegen die gesetzlichen Voraussetzungen vor?
  - neues Training unverhältnismäßig und nicht zielführend (Halluzination bleibt)
  - **Was gilt eigentlich bei Halluzination?** (LLMs sind probalistisch)
  - Umgang mit durch die KI kontextual „generierten" Daten? Personenbezug?

- Umsetzung von Benachrichtigungen nach Art. 34 DSGVO?
  - Wer ist betroffen? Und wie erreiche ich die betroffene Person?



Fokus: Automatisierte Entscheidungen

- **EuGH, Urt. v. 14.12.2023 – C-634/21**
  - Ein durch eine Auskunftei erstellter Scorewert unterfällt Art. 22 DSGVO, wenn ein Dritter diesen Scorewert maßgeblich in seine unternehmerische Entscheidung einfließen lässt.

- **Übertragbarkeit auf KI-Output?**
  - Voraussetzung: nicht nur geringe Beeinträchtigung
  - nicht bei menschlicher Letztentscheidung
  - Outputkontrolle allein genügt nicht

Der Hamburgische Beauftragte für Datenschutz und Informationsfreiheit

Datenschutz Rechtsgrundlagen Informationsfreiheit

Bild von Gerd Altmann von Pixabay

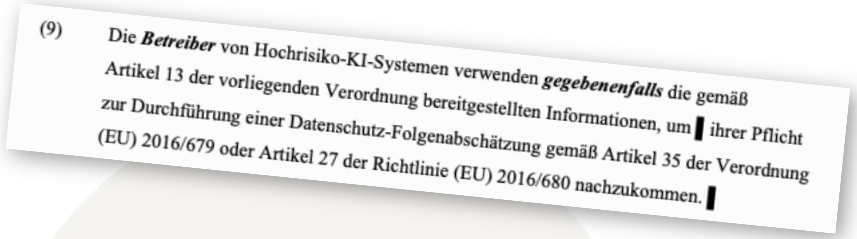
Automatisierte Entscheidungen dürfen keine maßgebliche Rolle spielen

Fuchs: „Urteil von wegweisender Bedeutung auch für KI-basierte Entscheidungen“

Mit seinem Urteil vom 07.12.2023 hat der Europäische Gerichtshof (EuGH) die Rechte Betroffener gestärkt. Es schützt Bürger:innen davor, zum Objekt undurchsichtiger Entscheidungsvorschläge zu werden. Konkret hat der EuGH die Schuldnerbewertung der Schufa als automatisierte Entscheidung eingestuft, die sich Betroffene nur in Ausnahmefällen und unter Überwindung hoher Hürden gefallen lassen müssen. Denn gemäß Artikel 22 der Datenschutzgrundverordnung (DSGVO) hat jeder Mensch das Recht, nicht durch rein automatisierte Entscheidungen benachteiligt zu werden. Was nun für algorithmisch errechnete Schufa-Scores gilt, hat Folgen weit über den Wirkungsbereich von Auskunfteien hinaus – denn es ist auf den Einsatz vieler KI-Systeme übertragbar.

Bereits die abstrakte Bewertung des Risikos eines Zahlungsausfalls durch die Schufa stufte der EuGH nun als Entscheidung mit Rechtswirkung ein. Diese Haltung ist nachvollziehbar, bedenkt man, wie stark sich Banken und Unternehmen in aller Regel maßgeblich auf diese Vorbewertung stützen, ohne den Score überprüfen zu können. Im Gegenteil sind Letztentscheider:innen faktisch gezwungen, der Schufa-Bewertung blind zu vertrauen. Ebenfalls intransparent und damit für das EuGH-Urteil gleichermaßen relevant sind viele KI-Anwendungen.

Fokus: Datenschutz-Folgenabschätzung



(9) Die ***Betreiber*** von Hochrisiko-KI-Systemen verwenden ***gegebenenfalls*** die gemäß Artikel 13 der vorliegenden Verordnung bereitgestellten Informationen, um ihrer Pflicht zur Durchführung einer Datenschutz-Folgenabschätzung gemäß Artikel 35 der Verordnung (EU) 2016/679 oder Artikel 27 der Richtlinie (EU) 2016/680 nachzukommen.

- **Spätestens hier: Warum das Ganze?**
  - Welchen Mehrwert bringt die einzelne KI-Anwendung als Teil des Verarbeitungsvorgangs?
  - Erforderlichkeit? Einsatz milderer Mittel zur Zweckerreichung?
- **Höhere Risiken** durch Data Linkage / Profilbildung durch Kontextwissen
- **Neue Risiken** für Vertraulichkeit, u.a.
  - Schutz (der KI) vor Modellextraktion / Modellinversion
  - Lesenswert: BSI, Generative KI-Modelle – Chancen und Risiken ... vom 27.03.24
- insgesamt: **Konfidenzniveau** (nicht nur ein DS-Thema!)
  - Algorithmic Impact Assessement (AIA; Tools der OECD, Canada, ...)
  - Trifft Risikoeinordnung nach AIA Aussage zur Notwendigkeit einer DSFA? wohl (-)

Usecase 3: DSFA zu Microsoft Copilot

- **Fokus**: KI-spezifische Datenschutzrisiken
  - u.a. ergebnisoffene Nutzung von MS Graph
  - Sachverhalt: Kernerarbeit

- **Bewertungsphase** (Auswahl)
  - kein Training mit User-Input
  - AI Safety Mechanism (u.a. Schutz vor Prompt-Injection)

- **Maßnahmenphase** (Auswahl)
  - DKE für Outlook, Word, Excel und Powerpoint (Purview Complianceportal)
  - Sekundärnutzung von Daten (Logfiles) ausschließen
  - Pilotphase mit wenigen Usern, (nicht zu eng) definierten Usecases, ...

Zusammenfassung: Checkliste Datenschutz und KI

1. Rechtsgrundlage (Art. 6 DSGVO); Mitbestimmung des BRat? (§§ 87, 80, 75 BetrVG)
2. Datenschutzrechtliche Rollenverteilung
   - (datenschutzrechtlicher) Vertrag mit KI-Anbieter (Art. 26 vs. 28 DSGVO)
3. VVT ☺
4. Datenschutzfolgenabschätzung (Art. 35 DSGVO; DSK-Positivliste Nr. 11); P: Blackbox
5. (zusätzliche) Maßnahmen zur Datenminimierung und Datensicherheit
   - Pseudonymisierung, Anonymisierung; **Input- und/oder Outputfilter**
   - **Historie deaktivieren** wenn mehrere Personen zur Nutzung befugt
6. Transparenz (insb. Art 13, 14 DSGVO) und Betroffenenrechte; P: Blackbox
7. Drittstaatentransfer → DPF, TIA & andere Garantien (Art. 44 ff DSGVO)
8. TOMs; vgl. DSK, Positionspapier vom 6.11.2019
9. Schulung / Qualifikation / Sensibilisierung


DSK

Orientierungshilfe
der Konferenz der unabhängigen Datenschutzaufsichtsbehörden
des Bundes und der Länder vom 6. Mai 2024

Künstliche Intelligenz und Datenschutz

Version 1.0

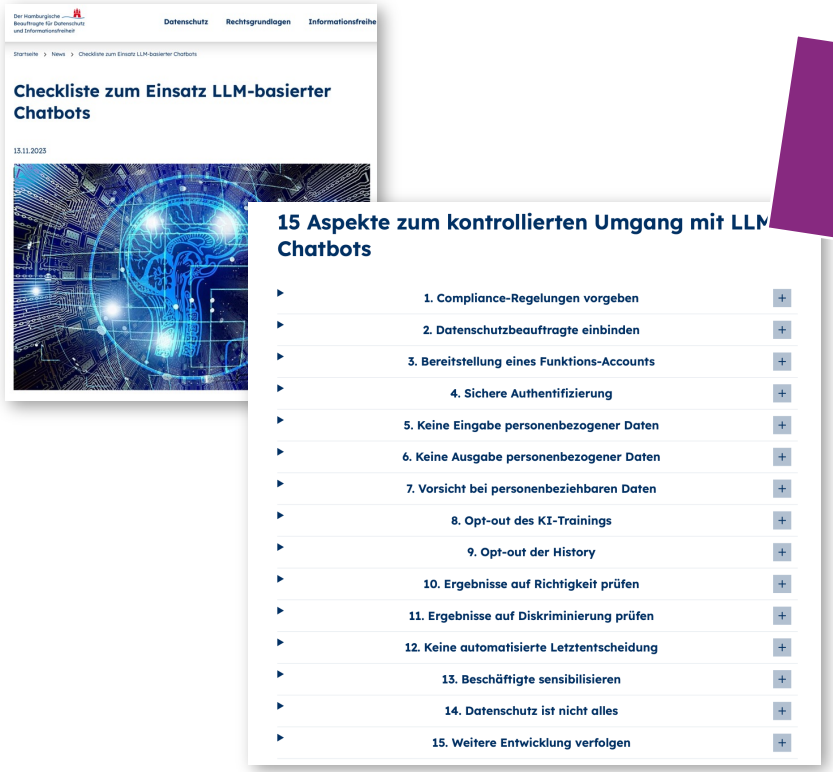
Viele Unternehmen, Behörden und andere Organisationen fragen sich derzeit, unter welchen Voraussetzungen sie KI-Anwendungen datenschutzkonform einsetzen können. Dabei lag der Schwerpunkt ab dem Jahr 2023 auf sogenannten Large Language Modelli (LLM), die häufig als Chatbots angeboten werden, aber auch als Grundlage für andere Anwendungen dienen können. Der Schwerpunkt der folgenden Orientierungshilfe liegt daher derzeit auf diesen KI-Anwendungen. Über die LLM hinaus gibt es jedoch zahlreiche weitere KI-Modelle und KI-Anwendungen, deren Einsatz infrage kommen kann und für die viele der folgenden Erwägungen ebenfalls relevant sein dürften.

Diese Orientierungshilfe bietet einen Überblick über datenschutzrechtliche Kriterien, die für die datenschutzkonforme Nutzung von KI-Anwendungen zu berücksichtigen sind. Sie kann als Leitfaden dienen, um KI-Anwendungen auszuwählen, zu implementieren und zu nutzen.

Die Orientierungshilfe wird voraussichtlich in der Zukunft angepasst werden, um aktuelle Entwicklungen und weitere relevante Aspekte aufzunehmen. Sie bietet einen Leitfaden, stellt aber keinen abschließenden Anforderungskatalog dar. Teilweise sind weitere Ressourcen hinzuzuziehen, um die in dieser Orientierungshilfe angesprochenen Punkte umzusetzen.

Die Orientierungshilfe richtet sich in erster Linie an die Verantwortlichen, die KI-Anwendungen einsetzen möchten. An Entwickler:innen, Hersteller:innen und Anbieter:innen von KI-Systemen richtet sie sich mittelbar, indem sie Hinweise zur Auswahl datenschutzkonformer KI-Anwendungen enthält. Die Entwicklung von KI-Anwendungen und das Training von KI-Modellen sind jedoch nicht Schwerpunkt dieser Orientierungshilfe.

HamBfDI



LDA Bayern

Prüfverfahren der DSK

LDI NRW

AKTUELLES | BÜRGER*INNEN | DATENSCHUTZ | INFORMATIONSFREIHEIT | INFOTHEK | KONTAKT | ÜBER UNS

Startseite

Prüfung von ChatGPT geht in die nächste Runde

Auch nach der Beantwortung erster Fragen durch den amerikanischen ChatGPT-Betreiber OpenAI ist die datenschutzrechtliche Bewertung der Software in Deutschland noch nicht abgeschlossen. Die Antworten von OpenAI haben weitere Fragen aufgeworfen. Deshalb haben deutsche Aufsichtsbehörden einen weiteren, zweiten Fragenkatalog verfasst und OpenAI um Beantwortung gebeten.

30.10.2023

Report und Fragenkatalog des EDSA

Report of the work undertaken by the ChatGPT Taskforce

23 May 2024

Prüffragen des HamBfDI

KI & DSGVO | FRAGEN DER AUFSICHT

Prüffragen bei Einsatz von KI-Anwendungen als „Software as a Service“ (SaaS)

- ✓ Welches Unternehmen betreibt die eingesetzten SaaS? Welches konkrete SaaS-Produkt wird eingesetzt?
- ✓ Zu welchen Zwecken wird die SaaS eingesetzt?
- ✓ Auf welche Rechtsgrundlage stützen Sie die Verarbeitung pbD zu diesem Zweck?
- ✓ Werden die von Ihnen übermittelten Daten vom SaaS-Anbieter zu Trainingszwecken verwendet?
- ✓ Liegt eine Auftragsverarbeitung oder gemeinsame Verantwortlichkeit vor? Wenn ja, legen Sie die entsprechende Vereinbarung gemäß Art. 26 oder 28 DSGVO vor.
- ✓ Wie werden mit Blick auf die o.g. Verarbeitung die Transparenzpflichten aus Art. 13 f. DSGVO gewahrt?
- ✓ Werden die Ausgaben der SaaS als Grundlage für eine Entscheidung mit Wirkung für betroffene Personen genutzt? Werden die Ergebnisse von einem Menschen überprüft? Wenn ja, wie?

KI-VO & Datenschutz

Seite 17

Hamburg

4

Fazit

1. Ohne technisches Grundverständnis der KI ist ein rechtssicherer Einsatz praktisch nicht möglich.

2. KI ist (nicht nur) alter Wein in neuen Schläuchen.
   Der Fokus der datenschutzrechtlichen Betrachtung ist deutlich nach vorn verlagert. Neue Risiken müssen adressiert werden. Alte Denkmuster müssen aufgebrochen werden.

3. KI-Anwendungen sollten nie isoliert betrachtet werden.
   Es handelt sich in aller Regel um eine Komponente eines Geschäftsprozesses.

4. KI ist kein Selbstzweck.
   Die Sinnfrage hat erhebliche Auswirkungen auf die Legitimationsfähigkeit.

5. Accountability ist bei Drittanbieter-KI schwer zu erreichen.

HÄRTING

Sebastian Schulz

Partner

schulz@haerting.de