Grundlagen des Datenschutzes und der IT-Sicherheit

Musterlösung zur 3. Übung im SoSe 2023:

Einführung in den Datenschutz (3)
& OTT-Dienste (2) & KI und Datenschutz

3.1 Ethik & KI

Aufgabe:

- Betrachten Sie die fünf vorgestellten **ethischen Werte** jeweils **im Vergleich zu** jedem der vier **ethischen Grundsätze** im Kontext des Einsatzes von künstlicher Intelligenz!
  Hinweis: Konzentrieren Sie sich bei Ihrer Lösung nur auf den wesentlichen Aspekt und vergleichen Sie jeweils, was daraus folgt, wenn ethischer Wert A mit ethischem Grundsatz B beim Einsatz von KI erfüllt werden soll. Möglicherweise resultieren daraus Anforderungen an das betreffende KI-System, an den Betreiber des KI-Systems oder an den Nutzer des KI-Systems. Der ethische Wert zur Demokratie wird in dieser Aufgabe nicht behandelt.

3.1 Ethik & KI

Ethische Werte vs Grundsätze	Menschliche Autonomie	Schadens-verhütung	Fairness	Erklärbarkeit
Menschenwürde	Ausschluss Menschenwürde verletzender Ergebnisse, z.B. durch Bias in Trainingsdaten	Ausschluss Menschenwürde verletzender Eingaben in Trainingsdaten & Validierungsdaten	Einschränkung der Menschenwürde unzulässig	Pflicht zur unabhängigen Überprüfung KI-System
Freiheitsrechte	Keine automatisierte Einzelfallentscheidung mit KI	Keine automatisierte Einzelfallentscheidung mit KI	Keine automatisierte Einzelfallentscheidung mit KI	Keine automatisierte Einzelfallentscheidung mit KI
Gleichbehandlung	Temporäre Speicherung Validierungsdaten zur Reversibilität von Eingaben	Vermeidung von Bias in Trainings- und Validierungsdaten	Keine automatisierte Einzelfallentscheidung mit KI	Pflicht zur unabhängigen Überprüfung KI-System
Rechtstaatlichkeit	Durchsetzbarkeit Betroffenenrechte gewährleisten	Haftung für Folgen KI-Einsatz auf Betroffene	Speicherung auf EU-Boden oder via Standardklauseln	Pflicht zur unabhängigen Überprüfung KI-System
Menschenrechte	Optionalität für Validierung mit Trainingsdaten	Pflicht zur unabhängigen Überprüfung KI-System	Einspruchsmöglichkeit gegen Ergebnisse	Pflicht zur unabhängigen Überprüfung KI-System

→ Keine automatisierte Einzelfallentscheidung mit KI (bereits in EU-DSGVO vorgeschrieben)
→ Pflicht zur unabhängigen Überprüfung KI-System (vorgesehen in geplanter KI-Verordnung)

3.2 Zulässigkeit KI @ EU-DSGVO

Aufgabe:

- Welche Vorschriften aus der EU-DSGVO würden derzeit bei der **Zulässigkeitsprüfung von KI-Systemen** mit welchem Ergebnis herangezogen werden?

3.2 Zulässigkeit KI @ EU-DSGVO (1)

- Art. 5 Abs. 1 lit. a EU-DSGVO: Verarbeitung nach Treu und Glauben setzt voraus, dass der Betroffene abschätzen kann, was mit seinen Eingaben sowie Ergebnissen gemacht wird
  → setzt eine entsprechende Datenschutzerklärung (im Rahmen der Nutzungsregeln) voraus
- Art. 5 Abs. 1 lit. b EU-DSGVO: Zweckbindung setzt voraus, dass Weiterverarbeitung personenbezogener Daten (gemäß Art. 6 Abs. 4 EU-DSGVO) mit ursprünglichem Zweck vereinbar sein muss
  → Entweder Ausfilterung personenbezogener Daten aus Trainings-daten oder Nichteinspeisung eingegebener Daten mit Personenbezug in die Verarbeitungslogik (nach Art. 11 Abs. 2 EU-DSGVO)
- Art. 5 Abs. 1 lit. d EU-DSGVO: Richtigkeit der Daten setzt voraus, dass unrichtige Daten berichtigt werden (nach Art. 16 EU-DSGVO) bzw. keine Daten zur Person durch KI-System erfunden werden
  → Prüfen, ob Einsatzzweck des KI-Systems entsprechend beschränkt werden kann oder generell keine Daten über Personen vom KI-System ausgegeben werden

3.2 Zulässigkeit KI @ EU-DSGVO (2)

- Art. 5 Abs. 1 lit. e EU-DSGVO: Speicherbegrenzung personenbezogener Daten setzt voraus, dass Daten mit Personenbezug nicht dauerhaft als Trainings- bzw. Validierungsdaten gespeichert werden
  → Entweder Ausfilterung personenbezogener Daten aus Trainingsdaten oder Nichteinspeisung eingegebener Daten mit Personenbezug in die Verarbeitungslogik
- Art. 5 Abs. 2 EU-DSGVO: Rechenschaftspflicht setzt voraus, dass Rechtmäßigkeit der Verarbeitung durch KI-System nachgewiesen werden kann
  → Selbstverpflichtung des Anbieters des KI-Systems, flankiert durch entsprechend aussagekräftiger Datenschutzerklärung in Nutzungsbedingungen
- Art. 12 Abs. 1 EU-DSGVO: Transparenz über mit KI-System verbundene Verarbeitung durch Datenschutzerklärung
  → setzt eine entsprechende Datenschutzerklärung (im Rahmen der Nutzungsregeln) voraus

3.2 Zulässigkeit KI @ EU-DSGVO (3)

- Art. 15 Abs. 1 EU-DSGVO: Auskunftsrecht für personenbezogene Daten in Trainings- bzw. Validierungsdaten ggf. schwierig umsetzbar → Entweder Ausfilterung personenbezogener Daten aus Trainings-daten oder Nichteinspeisung eingegebener Daten mit Personenbezug in die Verarbeitungslogik (nach Art. 11 Abs. 2 EU-DSGVO)
- Art. 17 Abs. 1 EU-DSGVO: Löschungspflicht für personenbezogene Daten in Trainings- bzw. Validierungsdaten ggf. schwierig umsetzbar → Entweder Ausfilterung personenbezogener Daten aus Trainings-daten oder Nichteinspeisung eingegebener Daten mit Personenbezug in die Verarbeitungslogik (nach Art. 11 Abs. 2 EU-DSGVO)
- Art. 21 Abs. 1 EU-DSGVO: Widerspruchsrecht bezieht sich aktuell nur auf Verarbeitungen auf Basis von Art. 6 Abs. 1 lit. e oder f EU-DSGVO

3.2 Zulässigkeit KI @ EU-DSGVO (4)

- Art. 22 Abs. 1 EU-DSGVO: Ausschluss automatisierter Einzelentscheidung setzt Kenntnis über Einsatz von KI mit Personenbezug voraus, um überhaupt eigenen Standpunkt einbringen zu können
  → Prüfen, ob Einsatzzweck des KI-Systems entsprechend beschränkt werden kann oder generell keine Daten über Personen vom KI-System ausgegeben werden
- Art. 25 EU-DSGVO: Datenschutz durch Technikgestaltung als auch Datenschutzfreundliche Voreinstellung setzt voraus, dass KI-System zwischen personenbezogenen und anderen Daten überhaupt unterscheiden kann
  → Prüfen, ob Einsatzzweck des KI-Systems entsprechend beschränkt werden kann oder generell keine Daten über Personen vom KI-System ausgegeben werden
- Art. 32 EU-DSGVO: Sicherheit der Verarbeitung setzt umfassenden Schutz personenbezogener Daten voraus
  → Manipulationsschutz (insbesondere vor ungewolltem Bias) und Robustheit des KI-Systems nötig

3.2 Zulässigkeit KI @ EU-DSGVO (5)

- Art. 35 Abs. 1 EU-DSGVO: Datenschutz-Folgenabschätzung für Einsatz eines KI-Systems nötig, da zahlreiche Folgen für die Rechte und Freiheiten der Betroffenen möglich sind
  → Selbsterklärung des Herstellers und Durchführung der Folgenabschätzung durch Verantwortlichen, der das KI-System einsetzt mit entsprechender Ableitung benötigter technischer und organisatorischer Maßnahmen zum Schutz vor ungewollten Folgen
- Art. 44 EU-DSGVO: Datenübermittlung in Drittland nur, wenn geeignete Garantien vorliegen
  → KI-System entweder gezielt auf EU-Boden einsetzen, Speicherung personenbezogener Daten im KI-System vermeiden oder Betreiber für KI-System einsetzen, der sich gemäß Standardklauseln verpflichtet hat

Generelles Ergebnis: Einsatz KI-System an sich ist nach bestehenden Vorgaben aus EU-DSGVO möglich, jedoch gibt es einige Lücken, die durch Folgenabschätzung abzumildern sind!

3.3 Datenschutzrisiko gemäß Datenschutz-Folgenabschätzung I

Aufgabe:

- Für ein **KI-System** wurde Folgendes geplant:
  - Das KI-System speichert neben der Cookie-ID alle vom Nutzer eingegebenen personenbezogenen Daten zu den gestellten Aufgaben hinzu, um künftige Anfragen vom gleichen Nutzer personalisieren zu können (z.B. durch persönliche Ansprache).
  - Anhand der Themen gestellter Fragen, wird für den Nutzer eine geeignete Werbung auf der Webseite eingebunden, auf der das KI-System genutzt werden kann.
  - Als Trainingsdaten für das KI-System wurden Daten verwendet, die von den Systemerstellern anhand geplanter Einsatzzwecke im Hinblick auf Funktionalität erstellt worden sind.
  - Antworten des KI-Systems werden von dem Nutzer ob ihrer Nützlichkeit bewertet und diese Bewertung fließt als zusätzliche Trainingsdaten ein.
  - Das KI-System soll als Public Cloud implementiert werden, damit es weltweit und jederzeit genutzt werden kann.

3.3 Datenschutzrisiko gemäß Datenschutz-Folgenabschätzung II

Aufgabe: (Fortsetzung)

- Geben Sie an, welche potenziellen Datenschutzrisiken Sie im Rahmen einer Datenschutz-Folgenabschätzung (gem. Art. 35 EU-DSGVO) sehen (unter Berücksichtigung der Bußgeldbestimmungen und Meldepflichten bei Verletzungen des Schutzes personenbezogener Daten), schätzen Sie die Eintrittsstufe dieser Datenschutzrisiken ab und ermitteln Sie den Handlungsbedarf gemäß nächstehender 3x3-**Risk-Map**. Sofern Handlungsbedarf besteht, geben Sie eine passende, zu ergreifende Schutzmaßnahme an.

3.3 Datenschutzrisiko gemäß Datenschutz-Folgenabschätzung (1)

1) Ermittlung potenzieller Datenschutzrisiken:

- Speicherung eingegebener personenbezogener Daten
  1. Verknüpfung der Daten mit nicht vereinbaren Zwecken → Gefahr: Verstoß gegen Art. 6 Abs. 4 EU-DSGVO möglich (→ Bußgeld nach Art. 83 Abs. 5 lit. a EU-DSGVO)
- Werbung abhängig von Eingabedaten
  2. Verwendung der Daten nicht transparent für Nutzer → Gefahr: Verstoß gegen Art. 5 Abs. 1 lit. a EU-DSGVO möglich, da unerwartet für Nutzer (→ Bußgeld nach Art. 83 Abs. 5 lit. a EU-DSGVO)
- Trainingsdaten nur rein funktional basiert
  3. Trainingsdaten können Bias aufweisen und infolge dessen Betroffene unerwartet benachteiligen → Gefahr: automatisierte Einzelentscheidung ohne Gewährleistung der Betroffenenrechte (→ Bußgeld nach Art. 83 Abs. 5 lit. b EU-DSGVO)
- Nutzer-Rückmeldung zur Ergebnisvalidierung
  4. Validierungsdaten können Bias aufweisen und damit analog 3. wirken → Gefahr: Intransparenz über Verarbeitung (→ analog Nr. 3)

3.3 Datenschutzrisiko gemäß Datenschutz-Folgenabschätzung (2)

1) Ermittlung potenzieller Datenschutzrisiken: (Fortsetzung)

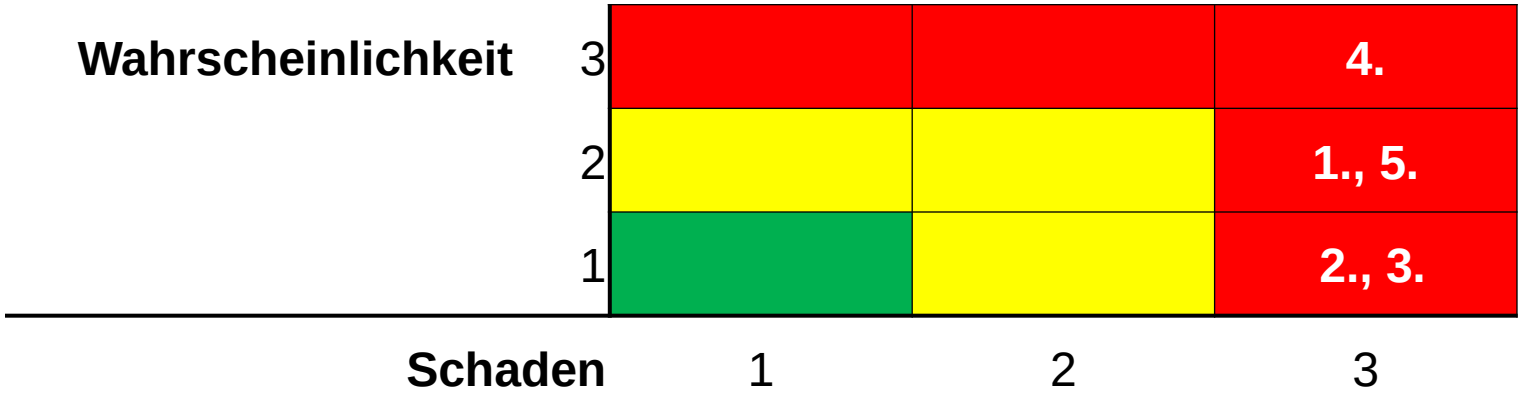
- KI-System als Public Cloud implementiert
  5. Daten und ggf. KI-Systemlogik ggf. unzureichend geschützt → Gefahr: Unbefugte Offenlegung ggf. möglich (Art. 32 Abs. 2 EU-DSGVO → Bußgeld nach Art. 83 Abs. 4 lit. a EU-DSGVO)

3.3 Datenschutzrisiko gemäß Datenschutz-Folgenabschätzung (3)

2) Abschätzung der Eintrittsstufe:

1. Verknüpfung der Daten mit nicht vereinbaren Zwecken: Gefahreintritt wahrscheinlich, da Personalisierung laut Aufgabenstellung zu unspezifisch erfolgt
2. Verwendung der Daten nicht transparent für Nutzer: Gefahreintritt möglich, da zwar i.d.R. für Nutzer nicht störend, doch ist potenziell damit eine Weitergabe personenbezogener Daten an Werbetreibende verbunden
3. Trainingsdaten können Bias aufweisen und infolge dessen Betroffene unerwartet benachteiligen: Gefahreintritt möglich, da Funktionalität oftmals nicht alle relevanten potenziellen Folgen berücksichtigt
4. Validierungsdaten können Bias aufweisen: Gefahreintritt sicher, da Nützlichkeit der Antwort i.d.R. nicht frei von Interessen bzw. Benachteiligung begünstigende Umstände ist (z.B. infolge Spieltrieb der Nutzer, Aktivitäten von „Trollen“ und durch Crime as a Service…)
5. Daten und ggf. KI-Systemlogik ggf. unzureichend geschützt: Gefahreintritt wahrscheinlich, da i.d.R. preisgünstig aufgrund geringerer Schutzvorkehrungen

3.3 Datenschutzrisiko gemäß Datenschutz-Folgenabschätzung (4)



Rot = Aktivität nötig; **Gelb** = Aktivität prüfen; **Grün** = Akzeptabel

Wahrscheinlichkeit: **Eintritt einer Verletzung des Schutzes personenbezogener Daten**	**Schaden:** **Grad der Verletzung des Schutzes personenbezogener Daten**
1 = möglich	1 = niedrig (ohne unmittelbare Wirkung)
2 = wahrscheinlich	2 = mittel (formaler Verstoß)
3 = sicher	3 = hoch (Bußgeld/Meldepflicht)

3.3 Datenschutzrisiko gemäß Datenschutz-Folgenabschätzung (5)

3) Handlungsempfehlung:

1. Bereitstellung einer entsprechend erläuternden Datenschutzerklärung und Einrichtung einer optionalen Funktion, ob Verknüpfung gewünscht wird
2. Bereitstellung einer entsprechend erläuternden Datenschutzerklärung
3. Aussagekräftige Beschreibung der geplanten Einsatzzwecke in Nutzungsbedingungen und Selbsterklärung des Herstellers über durchgeführte Folgenabschätzung
4. Temporäre Speicherung von Validierungsdaten mit Option, ob diese eingespeist werden sollen und unabhängige Überprüfung, ob KI-System menschenwürde-beeinträchtigenden Bias aufweist
5. Ausreichend sichere Cloud verwenden mit ausreichenden Nachweisen oder On-Premise-Lösung anbieten

3.3 Datenschutzrisiko gemäß Datenschutz-Folgenabschätzung (6)

Anmerkung:

- *Die Angabe der Punkte aus Art. 35 Abs. 7 EU-DSGVO ist bei der Durchführung von Datenschutz-Folgenabschätzungen verpflichtend*
  ° auf systematische Beschreibung der geplanten Verarbeitungsvorgänge und der Zwecke der Verarbeitung, ggf. einschließlich der von dem Verantwortlichen verfolgten berechtigten Interessen
  in der Aufgabe jedoch verzichtet, da nach Aufgabenstellung nicht zwingend verlangt

3.4 Technischer Schutz KI-System

Aufgabe:

- Welche **technischen Maßnahmen** sollten für ein **KI-System** zum maschinellen Lernen implementiert werden (sowohl bei dessen Entwicklung als auch beim Betrieb), damit es widerstandsfähig gegenüber Fehlern, Störungen oder Unstimmigkeiten (inkl. aus Rückkopplungsschleifen) ist?

3.4 Technischer Schutz KI-System (1)

Vorbemerkung:

- *Hilfreiche Quellen zur Bestimmung technischer Schutzvorkehrungen sind:*
  BSI: Große KI-Sprachmodelle – Chancen und Risiken für Industrie und Behörden (Stand 03.05.2023), abrufbar unter https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/KI/Grosse_KI_Sprachmodelle.pdf?__blob=publicationFile&v=2
  NSA/FBI/ACSC/NCSC-UK/CCCS/BSI/NCSC-NL/CERT NZ/NCSC-NZ: Shifting the Balance of Cybersecurity Risk: Principles and Approaches for Security-by-Design and –Default (Stand 13.04.2023), abrufbar unter https://www.cisa.gov/sites/default/files/2023-04/principles_approaches_for_security-by-design-default_508_0.pdf
- *Im aktuellen Entwurf der KI-Verordnung der EU werden sog. „Hochrisiko-KI-Systeme" mit besonderen Schutzvorkehrungen bedacht: u.a. wird dort ein Risikomanagement hinsichtlich Folgen auf Gesundheit, Sicherheit und Grundrechte vorgeschrieben, eine Beaufsichtigung durch natürliche Personen sowie Maßnahmen zur Genauigkeit, Robustheit und Cybersicherheit und für Rückkopplungsschleifen im Betrieb Risikominderungsmaßnahmen eingefordert, die Festlegung einer Gebrauchsanweisung und die Einholung einer unabhängigen Konformitätsbewertung vorgeschrieben*

3.4 Technischer Schutz KI-System (2)

- Für „böswillige“ Zwecke vordefinierte Ausgabe generieren (Problem: Lässt sich bisher recht leicht umgehen, wenn der „bösartige“ Zweck durch „freundlichen“ Zweck maskiert wird – z.B. Kampagne zur Warnung vor bösartiger Handlung unter Berücksichtigung der bösartigen Handlung)
- Trainingsdaten aus vertrauenswürdigen Quellen verwenden
- Bei Trainingsdaten insbesondere auch auf potenziellen Missbrauch achten (inkl. einer gezielten „Vergiftung“ des Modells) und bei der Validierung entsprechend bewerten („Adversarial Training“)
- Keine Verwendung sensibler Daten zum Training des KI-Systems, da Original-Daten u.U. rekonstruierbar sind – bezieht sich nicht nur auf personenbezogene Daten, sondern auch auf Geschäftsgeheimnisse
- Entwicklung des KI-Systems unter Einhaltung einschlägiger Frameworks für sichere Software-Entwicklung, z.B. „Secure Software Development Framework“ des NIST (SP 800-218; Version 1.1 aus 02/2022), abrufbar unter https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-218.pdf
- Implementierung einer abgesicherten Protokollierung, um potenzielle Angriffe auf KI-System möglichst identifizieren zu können
- Härtung eingesetzter Komponenten des KI-Systems

3.5 Messenger Dienst

Aufgabe:

- Ein Unternehmen möchte für seine Beschäftigten die Nutzung eines **Messenger Dienstes** erlauben, über welches dienstliche Mitteilungen versendet und empfangen werden dürfen. Wie muss es dazu vorgehen, dass relevante **Anforderungen aus dem Datenschutz und des Schutzes vertraulicher Kommunikation** dabei eingehalten werden?

3.5 Messenger Dienst (1)

- Messenger Dienste sind sog. „Over-the-top-Dienste“ (interpersoneller Kommunikationsdienst nach § 3 Nr. 24 TKG zum Austausch von Nachrichten zwischen einem Sender und n Empfängern)
- Wenn also ein Messenger Dienst nur für dienstliche Zwecke eingesetzt werden soll, ist entscheidend, dass nur Empfänger als Nachrichtenempfänger eingetragen werden, die dienstlich dafür vorgesehen sind; ist das auf das spezifische Unternehmen beschränkt, lässt sich dieses technisch einschränken, andernfalls nur organisatorisch
- Die Inhalte der Nachrichten als auch der Umstand, wer mit wem Nachrichten austauscht, unterliegt jedoch dem Fernmeldegeheimnis nach § 3 Abs. 1 TTDSG

3.5 Messenger Dienst (2)

- Nach § 6 Abs. 2 TTDSG sind Fehlübermittlungen und das unbefugte Offenbaren von Nachrichteninhalten innerhalb des Unternehmens des Anbieters und an Dritte auszuschließen
  → Nachrichteninhalte dürfen nur durch Befugte eingesehen werden
  → Für die vorgesehenen Zwecke ist festzulegen, wer als befugt anzusehen ist
  → Messenger Dienst bevorzugt On-Premise betreiben oder Ende-zu-Ende verschlüsseln
- Nach § 9 Abs. 2 TTDSG dürfen Verkehrsdaten nur zur Bereitstellung von Diensten mit Zusatznutzen weiter gespeichert werden
  → Dienst mit Zusatznutzen ist explizit festzulegen, mit dem eine Zeitspanne über das reine unmittelbare Kommunikation (z.B. zu benötigten Nachweiszwecken) begründet wird
  → Nach Ablauf dieser Zeitspanne sind die Daten automatisiert zu löschen