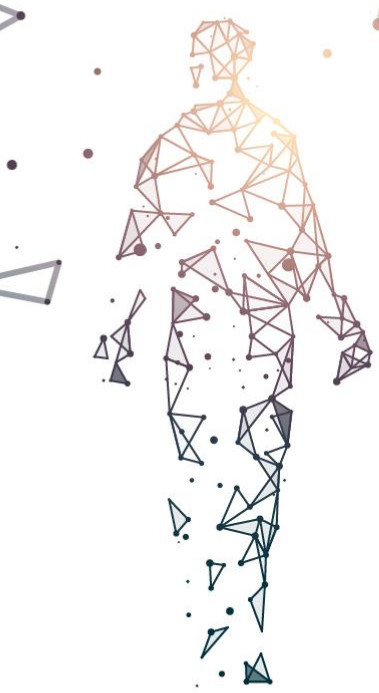
Bayerisches Landesamt für
Datenschutzaufsicht

Datenschutzkonforme Künstliche Intelligenz

Checkliste mit Prüfkriterien nach DS-GVO



Stand der Checkliste: 24.01.2024
Version: Konsultationsstand v0.9

Ziel und Inhalt dieses Papiers

Die in diesem Dokument enthaltene Checkliste stellt Anforderungen an die Entwicklung und den Einsatz von Anwendungen der Kategorie „Künstliche Intelligenz" dar. Aufgrund der stetig fortschreitenden Entwicklung können Anpassungen zu dieser Checkliste – insbesondere zur Harmonisierung mit deutschen und europäischen Datenschutzpositionen zu KI – erforderlich werden. Die aufgeführten Prüfpunkte sind daher nicht als abschließend zu betrachten, sondern stellen einen Good-Practice-Ansatz dar, der im Sinne einer Soll-Ist-Überprüfung verwendet werden kann. Dieses Dokument widmet sich der Fragestellung, welche datenschutzrechtlichen Anforderungen bei dem Einsatz von Künstlicher Intelligenz von zentraler Bedeutung sein können.

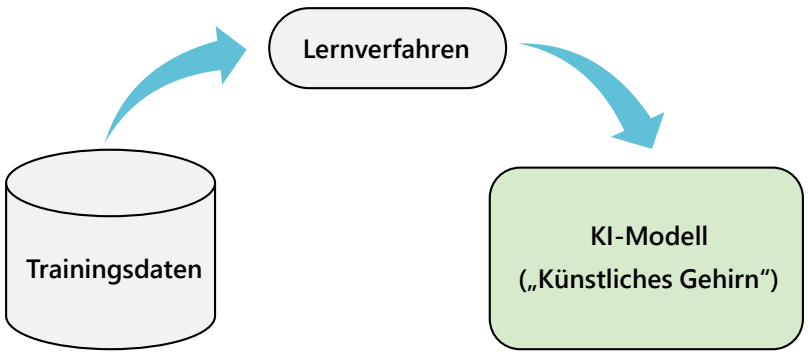
INHALT

A. Einordnung

Anwendungen zur Künstlichen Intelligenz bestehen im Kern aus einem sogenannten KI-Modell (z. B. großes Sprachmodell wie GPT). Solche Modelle wurden entweder aus Daten trainiert – sog. maschinelles Lernen – oder explizit modelliert (z. B. regelbasierte Systeme, die momentan aber nicht so „hype" sind). Man kann sich derartige Modelle auch als „künstliches Gehirn" vorstellen. Insofern wird in dieser Checkliste auch das Training von Künstlicher Intelligenz näher betrachtet.

Anwendungen zur Künstlichen Intelligenz werden im operativen Einsatz betrieben, was bedeutet, dass zum einen das sog. Hosting-Szenario betrachtet werden muss (z. B. auf welchem Cloud-System oder GPU-Rechner eine KI-Anwendung läuft). Hierbei stellen sich mitunter ähnliche Fragestellungen wie bei der Nutzung von Cloud-Systemen. Bei dem Einsatz einer KI-Anwendung werden Eingaben aus einer Anwendungsumgebung (z. B. Text bei ChatGPT) in ein KI-Modell gegeben, die mitunter für eine Verarbeitung vorverarbeitet werden müssen. Die Ergebnisse des damit angewendeten KI-Modells (z. B. Textausgabe bei ChatGPT oder Warnmeldung bei Fußgängererkennung im Fahrzeug) werden dann auch weiterverarbeitet. Insofern wird in dieser Checkliste auch der Betrieb/der Einsatz von Künstlicher Intelligenz als Anwendung näher betrachtet.

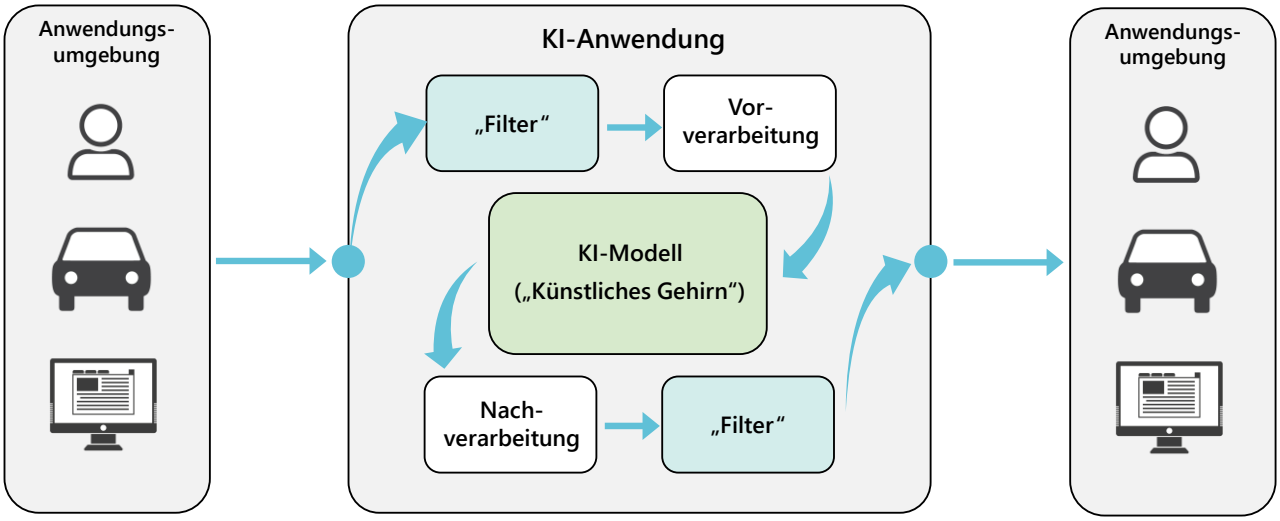
Folgende Skizze veranschaulicht das Training von Künstlicher Intelligenz:



Training eines KI-Modells

Ein KI-Modell benötigt (zumindest beim maschinellen Lernen aber auch bei einer expliziten Erstellung von Regelsystemen) immer Trainingsdaten von großem Umfang, aber insbesondere von sehr guter Qualität (bezüglich dem Einsatzszenario). Ein passend zum KI-Modell ausgewähltes Lernverfahren (z. B. Gradientenabstiegsverfahren zur Minimierung einer festgelegten Fehlerklasse) versucht den Wesensgehalt der Trainingsdaten (nennt sich „Generalisierung") in ein KI-Modell zu trainieren, das damit auch bislang nicht in den Trainingsdaten enthaltene Eingaben (z. B. neues Automodell bei Fahrassistenzsystem, individuelle Konversation bei ChatGPT) angemessen verarbeiten können sollte.

Folgende Skizze veranschaulicht den Betrieb bzw. die Verwendung von Künstlicher Intelligenz:



Betrieb/Verwendung von Künstlicher Intelligenz

Ein trainiertes KI-Modell wird in die Anwendung („Anwendungsumgebung") gebracht, indem dieses in einer IT-Umgebung (z. B. Auto bei Fußgängererkennung oder GPU-Cluster („KI-as-a-Service") bei Großen Sprachmodellen wie ChatGPT) geladen wird („KI-Anwendung"). Je nach KI-Modell werden Eingabe-/Ausgabedaten des KI-Modells vor- und nachverarbeitet (z. B. Transformation von Audiosignalen in Frequenzbereiche, Kodierung/Dekodierung zwischen Texten und Zahlenreihen bei ChatGPT oder Bildverkleinerung bei Fahrerassistenzsystemen). Zusätzlich werden mitunter „Filtersysteme" eingesetzt, die im einfachsten Fall unerwünschte Eingaben wie bspw. die Eingabe einer Hassbotschaft bei einem Sprachmodell vor Eingabe in dieses schlicht herausfiltert. Auch Ergebnisse von KI-Modellen werden meist durch entsprechende Filter verarbeitet, sei es, dass eine Erkennungswahrscheinlichkeit eines Verkehrszeichens als zu gering angesehen wird oder dass Ergebnisse eines Sprachmodells ebenfalls auf unerwünschte Ausgaben geprüft werden.

B. Training von KI-Modellen

Folgende Prüfpunkte sollten bei der Erstellung von KI-Modellen (ggf. auch im Rahmen der Kontrolltätigkeit durch den Datenschutzbeauftragten) überprüft werden:

- ☐ Festlegung und Dokumentation welcher Art von KI-Technologie mittels KI-Modell realisiert werden soll (z. B. Transformer-Architektur bei Erstellung eines Großen Sprachmodells samt Festlegung der internen KI-Architektur wie Schichten und Kopplung sowie die Anzahl und Initialisierung der Parameter)

- ☐ Bewertung, welche Trainingsdaten personenbezogen sind und welche nicht

- ☐ Aufnahme des Trainings von KI-Modellen in das Verzeichnis der Verarbeitungstätigkeiten nach Art. 30 DS-GVO. Bei Training mehrerer KI-Modellen mit unterschiedlichen KI-Technologien, Kategorien personenbezogener Daten und ggf. Empfängern im Drittland empfiehlt es sich, jeweils einen eigenen Eintrag aufzunehmen.

- ☐ Prüfung und Dokumentation, ob eine Datenschutzfolgenabschätzung (DSFA) nach Art. 35 DS-GVO durchgeführt werden muss. Ggf. einige der in diesem Abschnitt benannten Checkpunkte im Rahmen der DSFA bearbeiten.

- ☐ Prüfung, ob ein KI-Training auch mit anonymen Daten erfolgen kann.

- ☐ Prüfung, ob ein KI-Training auch mit synthetischen oder pseudonymen Daten erfolgen kann.

- ☐ Bei synthetischen Trainingsdaten: Prüfung, ob der Algorithmus zur Generierung synthetischer Trainingsdaten wirklich anonyme Ausgaben erstellt.

- ☐ Rechtsgrundlage für Verwendung personenbezogener Trainingsdaten vorhanden?

- ☐ Bei besonderen personenbezogenen Daten (z. B. Gesundheitsdaten) ist eine Einwilligung vorhanden (oder es greifen die Ausnahmen in Art. 9 Abs. 2 DS-GVO).

- ☐ Ist der Zweck des Trainings eines KI-Modells die Forschung, dann einschlägige Forschungsprivilegien prüfen (Das Training eines Großen Sprachmodells durch ein kommerzielles Unternehmen zu Produktzwecken dürfte heute aber wohl eher nicht mehr darunterfallen).

- ☐ Dokumentation aller Trainingsdaten samt Quellen (z. B. Bücherdatenbanken oder Webseiten) im Sinne der Rechenschaftspflicht nach Art. 5 Abs. 2 DS-GVO.

- ☐ Die Trainingsdaten wurden auf Datenqualität in Bezug auf statistische Verzerrungen oder Voreingenommenheiten (Bias) geprüft und ggf. bereinigt/angepasst.

- ☐ Bei flüchtigen Trainingsdaten (z. B. Webseiten von Nachrichtenportalen, die sich schnell verändern können): Vollständige und revisionssichere Speicherung der für das Training verwendeten Webseiteninhalte.

- ☐ Aussonderung der Trainingsdaten, die unerlaubte Inhalte in ein Trainingsverfahren bringen würden (z. B. Webseiten mit Fake-News, Hassinhalten, Verschwörungstheorien, …). Dazu Erstellung eines Konzepts (nach Art. 5 Abs. 2 DS-GVO), nach welchen Kriterien Aussonderungen erfolgen.

- ☐ Bereinigung von für das Training nicht erforderlichen personenbezogenen Daten aus den Trainingsdaten (z. B. Kreditkartennummern, (E-Mail-)Adressen, Namen, …) unter Berücksichtigung landesspezifischer Kodierungen (z. B. Adressen werden in vielen Ländern unterschiedlich geschrieben).

- ☐ Neben Trainings-/Testdaten sind auch Validierungsdaten vorhanden, mit denen die Güte des KI-Modells geprüft wird und die nicht Bestandteil des Trainingsprozesses sind.

- ☐ Erstellung eines Risikomodells (nächster Abschnitt), mit dem die Güte eines Lernverfahrens im Sinne des Art. 5 Abs. 2 DS-GVO nachgewiesen werden kann.

- ☐ Bewertung und Nachweis, ob ein erstelltes KI-Modell an sich einen Personenbezug aufweist oder nicht (hat Folgen für die Weitergabe eines KI-Modells sowie ggf. die Notwendigkeit einer Rechtsgrundlage, um auf den personenbezogenen KI-Modellen arbeiten zu dürfen).

- ☐ Berechnung und Dokumentation von Metriken aus dem Risikomodell, mit denen eine angemessene Eindämmung der Datenschutzrisiken nachgewiesen werden kann (Art. 5 Abs. 2 DS-GVO). Die Suche nach geeigneten Kennzahlen ist momentan (auch) noch aktueller Forschungsstand.

- ☐ Umsetzung der Informationspflichten nach Art. 12 ff DS-GVO.

- ☐ Sicherstellung, dass Auskunftsersuchen nach Art. 15 DS-GVO auch bei Anfragen zum Training in KI-Modellen im Datenschutzmanagement berücksichtigt werden.

- ☐ Bei konkretem Auskunftsersuchen nach Art. 15 DS-GVO in Bezug auf ein personenbeziehbares KI-Modell wird – je nach KI-Technologie – geprüft, ob personenbezogene Daten im KI-Modell direkt ermittelbar sind oder ob diese evtl. nur mit Zusatzinformationen (z. B. konkreter Prompt bei Großem Sprachmodell) auf einem KI-Modell abgeleitet werden können. Diese Zusatzinformationen sind im Zweifel vom Betroffenen dann anzufordern.

- ☐ Sicherstellung, dass Betroffenenrechte zur Berichtigung nach Art. 16 DS-GVO, zur Löschung nach Art. 17 DS-GVO, nach Einschränkung der Verarbeitung nach Art. 18 DS-GVO, nach Datenübertragbarkeit nach Art. 20 DS-GVO und des Widerspruchs nach Art. 21 DS-GVO in Bezug auf KI auch im Datenschutzmanagement berücksichtigt werden. Rückmeldefristen an Antragsteller sind hierbei zu beachten.

- ☐ Bei einem Löschersuchen nach Art. 17 DS-GVO in Bezug auf ein personenbeziehbares KI-Modell wird – je nach KI-Technologie – geprüft, ob personenbezogene Daten im KI-Modell direkt ermittelbar sind oder ob diese evtl. nur mit Zusatzinformationen (z. B. konkreter Prompt bei Großem Sprachmodell) aus einem KI-Modell abgeleitet werden können. Sofern eine Löschung in einem KI-Modell technisch ohne Beeinträchtigung des Gesamtmodells möglich ist, ist der Löschvorgang auch durchzuführen. Sollten andererseits personenbezogene Daten nur mittels Zusatzinformationen (z. B. Prompts) aus einem KI-Modell ermittelbar sein, dann besteht eine Möglichkeit des technischen Löschens darin, mittels Nachtraining die spezifisch zu löschende personenbezogene KI-Ausgabe mittels Anpassung der internen (Wahrscheinlichkeits-)Parameter umzusetzen.

- ☐ Bei Beauftragung eines Dienstleisters, der ein KI-Training (teilweise) mit übernimmt, geeignete Garantien und Rechtsgrundlagen prüfen (z. B. Vertrag zur Auftragsverarbeitung, Garantien bei Drittlandsübermittlung wie Angemessenheitsbeschluss, Standardvertragsklauseln mit Transfer Impact Assessment oder EU-US Data Privacy Framework). Dabei insbesondere sicherstellen, dass der Datenempfänger mögliche Trainingsdaten nicht für eigene Zwecke verwendet oder zumindest bei einer Zweckänderung geeignete Rechtsgrundlagen und Informationspflichten eingehalten werden.

- ☐ Bei Verpflichtung der Durchführung einer DSFA nach Art. 35 DS-GVO: Restrisikobeurteilung anhand des Risikomodells und ggf. Konsultation der zuständigen Datenschutzaufsichtsbehörde nach Art. 36 DS-GVO bei weiterhin hohen Risiken für die Rechte und Freiheiten der vom Training betroffenen Personen

- ☐ Findet eine Anpassung des KI-Modells im laufenden Betrieb (bspw. durch Einbindung mancher tagesaktuellen Webseiten) statt, dann sind die jeweiligen Modellstände samt jeweiliger Trainingsdaten revisionssicher zu speichern und besonders in der Risikomodellierung zu berücksichtigen.

C. Bewertung von Risiken bei KI

Sowohl die Erzeugung von KI-Modellen als auch deren Betrieb in einer Anwendungsumgebung (z. B. Bewertung von Kundenschreiben in einer Versicherung, Entscheidungen ob ein Auto eine Vollbremsung einleiten soll oder nicht, Generierung einer politischen Rede mittels ChatGPT, …) sind mit Risiken verbunden. Die DS-GVO adressiert bei der Verarbeitung personenbezogener Daten die Risiken der Rechte und Freiheiten derart, dass diese nach einer objektiven Methode bestimmt (EW76) und mit wirksamen Maßnahmen eingedämmt werden müssen (Art. 25 DS-GVO, ggf. Art. 35 DS-GVO). Aus diesem Grund muss jeder Verantwortliche und jeder Auftragsverarbeiter, dessen Verarbeitung personenbezogener Daten entweder zur Erzeugung von KI-Modellen oder bei Einsatz dieser in KI-Anwendungen nutzt, sich den damit verbundenen spezifischen Risiken bewusst werden und diese im Sinne der Rechenschaftspflicht nach Art. 5 Abs. 2 DS-GVO durch eine geeignete Dokumentation nachweisen. Bei der Verpflichtung zur Durchführung einer DSFA nach Art. 35 DS-GVO stellen die spezifischen hohen Risiken von Künstlicher Intelligenz den Kern der Folgenabschätzung dar.

Es beschäftigen sich seit einigen Jahren sowohl die Forschung als auch zunehmend Regulierungs-/Normungsstellen mit der Frage, wie Künstliche Intelligenz zum Nutzen der Betroffenen eingesetzt und die damit verbundenen Risiken eingedämmt oder zumindest reduziert werden können. Dieses Themengebiet wird auch als „Vertrauenswürdige KI" bezeichnet und kann als Ausgangspunkt für die Formulierung von Datenschutzrisiken unter der DS-GVO verwendet werden.

Erstellen eines Risikomodells

- ☐ Festlegung und Dokumentation, welche der folgenden Schutzziele einer KI-Anwendung für das spezifische Szenario relevant sind. Datenschutzrisiken ergeben sich dann aus der Abweichung eines vollständigen Erreichens der jeweiligen Schutzziele („Datenschutz-Risikomodell"). Ausführliche Begründung, falls ein Schutzziel als nicht relevant angesehen wird. Diese können bspw. in *Anlehnung an die Ethik-Richtlinien für vertrauenswürdige KI*[1] der Europäischen Kommission sein:

  - ☐ „**Fairness**" im Sinne, dass keine unvertretbaren Risiken in Bezug auf Diskriminierung oder Ungleichbehandlung vorhanden sind.

  - ☐ „**Autonomie und Kontrolle**" im Sinne, dass Eingriffsmöglichkeiten in den Betrieb einer KI-Anwendung existieren bzw. Entscheidungen mit Rechtswirkung nicht ohne menschliche Kontrolle erfolgen.

  - ☐ „**Transparenz**" im Sinne, dass zum einen die Betroffenen über deren Verwendung personenbezogener Daten beim Training von KI-Modellen informiert werden als auch derart, dass KI-Modelle und KI-Anwendungen prüfbar im Sinne der Rechenschaftspflicht sein müssen. Ebenfalls

[1] Abrufbar unter https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

auch, dass KI-Anwendungen für Betroffene in der Interaktion als solche erkannt werden müssen (z. B. ChatBots oder Anpassung von Audiosignalen zur Imitation eines Sprechers).

- ☐ „**Verlässlichkeit**" im Sinne, dass ein KI-Modell bzw. eine KI-Anwendung zum einen seinen beabsichtigten Zweck innerhalb tolerierbaren Fehlergrenzen erfüllt als auch, dass diese vor absichtlichen Manipulationen (sog. Adversial Angriffe bspw. mittels Prompt-Injection bei Großen Sprachmodellen oder Störung bei der Erkennung von Verkehrsschilderung durch spezielle „Aufkleber") geschützt sind. Auch das sog. Halluzinieren (oder Konfabulieren) bei Großen Sprachmodellen, bei dem Falschausgaben in einer ansonsten flüssig formulierten Ausgabe mitunter unerkannt auftauchen können, gehört in diesen Bereich.

- ☐ „**Sicherheit**" in dem Sinne, dass ungewollte technische Störungen („Safety" wie Hardwarefehler aufgrund ungenügendem Arbeitsspeicher), aber vor allem unbefugte Zugriffe/Änderungen („Security" wie die Manipulation von Trainingsdaten bei der KI-Modellerzeugung oder Manipulation von „Filtern", die als Zensurmechanismus missbraucht werden können) wirksam verhindert werden können.

- ☐ „**Datenschutz**" im Sinne, dass neben einer Rechtsgrundlage für das Erzeugen von KI-Modellen und dem Betrieb/dem Einsatz von KI-Anwendungen auch die Betroffenenrechte sowie weiteren Compliance-Anforderungen der DS-GVO umgesetzt werden müssen. Dazu gehört auch die Zweckänderung von Eingabedaten an eine KI-Anwendung durch einen KI-Betreiber zu eigenen Zwecken. Wichtig: Datenschutzrisiken nach der DS-GVO können auch die oben genannten Risiken der Abweichung der „Fairness", der „Autonomie und Kontrolle", der „Transparenz", der „Verlässlichkeit" und der „Sicherheit" umfassen, sofern personenbezogene Daten bei diesen Schutzzielen eine Rolle spielen.

☐ Das Datenschutz-Risikomodell ist zu dokumentieren und regelmäßig auf Aktualität und Vollständigkeit zu prüfen.

D. Einsatz einer KI-Anwendung

Bei einer KI-Anwendung wird ein trainiertes KI-Modell (das auch von einem Dritten erstellt worden sein könnte) zum Einsatz gebracht. Dazu wird es auf einer meist spezialisierten Hardware in eine Anwendungsumgebung geladen. Die Nutzung des KI-Modells erfolgt durch Eingabewerte, die von einem Menschen, einem mobilen Objekt (z. B. Auto) oder einer anderen Software mittels Schnittstelle (z. B. Chatsystem einer Bank) diesem zur Verfügung gestellt werden („Anwendungsumgebung"). Ausgaben des KI-Modells werden dann wieder einer Anwendungsumgebung zur Weiterverarbeitung ausgegeben. Häufig findet aufgrund der sehr hohen Hardwareanforderungen bei Großen Sprachmodellen wie ChatGPT ein Betrieb einer KI-Anwendung bei einem (Cloud-)Dienstleister statt, mit dem entweder mittels Weboberfläche oder Softwareschnittstelle interagiert wird.

Anmerkung: Insbesondere die Sicherstellung der Betroffenenrechte stellen bei manchen KI-Anwendungen (z. B. Großen Sprachmodellen) noch eine Herausforderung dar. Hierbei ist für Verantwortliche, die als KI-Nutzer KI-Anwendungen von großen Cloud-Anbietern verwenden, die Frage der datenschutzrechtlichen Verantwortlichkeit bedeutsam. Es sollte vor Einsatz einer KI geklärt werden, ob die Sicherstellung der Betroffenenrechte das KI-Modell betreffend bei einem KI-as-a-Service Szenario in den Verantwortungsbereich des KI-Anbieters fallen, der möglicherweise ein KI-Modell selbst erstellt und nur die Nutzung desselben als Dienstleistung anbietet und dieser sich dann selbst um die Betroffenenrechte kümmern muss oder beim Auftraggeber verortet wird.

- ☐ Festlegung und Dokumentation, welche Art von Anwendung mittels KI-Technologie realisiert werden soll (z. B. Verwendung eines Großen Sprachmodells zur Erstellung eines ChatBots einer Bank oder für die Bewertung der Kritikalität einer Kundenbeschwerde bei einem Versandhändler).

- ☐ Festlegung, ob ein eigenes KI-Modell in einer eigenen KI-Anwendung betrieben werden soll (möglicherweise auf Hardware eines Dienstleisters) oder, ob eine KI-Anwendung bei einem KI-Anbieter (mittels Weboberfläche oder Schnittstelle) genutzt wird, der die vollständige Kontrolle über das KI-Modell, die Vor- und Nachverarbeitung sowie die Filtersysteme besitzt.

- ☐ Festlegung, ob das KI-Modell (das eigene oder das des KI-Anbieters) an sich personenbeziehbar ist. Falls ja, muss die Rechtsgrundlage für die Verarbeitung auf dem KI-Modell geprüft werden (i. d. R. wird eine Interessensabwägung nach Art. 6 Abs. 1 Lit. f DS-GVO anwendbar sein).

- ☐ Aufnahme des Einsatzes einer KI-Anwendung in das Verzeichnis der Verarbeitungstätigkeiten nach Art. 30 DS-GVO. Bei Verwendung eines KI-Systems zu mehreren Zwecken (und ggf. auch bei verschiedenen verwendeten Betroffenenkategorien) empfiehlt es sich, jeweils einen eigenen Eintrag aufzunehmen.

- ☐ Prüfung und Dokumentation, ob eine Datenschutzfolgenabschätzung (DSFA) nach Art. 35 DS-GVO durchgeführt werden muss. Ggf. einige der in diesem Abschnitt benannten Checkpunkte im Rahmen der DSFA bearbeiten.

- ☐ Festlegung und Dokumentation welche Kategorien an personenbezogenen Daten als Eingabedaten in eine KI-Anwendung gegeben werden sollen. Dazu Festlegung und Dokumentation einer Rechtsgrundlage.

- ☐ Bei der Eingabe von besonderen personenbezogenen Daten nach Art. 9 DS-GVO in eine KI-Anwendung ist eine Einwilligung einzuholen (oder die Ausnahmen des Art. 9 Abs. 2 DS-GVO prüfen).

- ☐ Erstellung eines Risikomodells (vorheriger Abschnitt) mit dem die Datenschutzrisiken im spezifischen Einsatzszenario der KI-Anwendung abgebildet und nachgewiesen werden kann.

- ☐ Berechnung und Dokumentation von Metriken aus dem Risikomodell, mit denen eine angemessene Eindämmung der Datenschutzrisiken bei Verwendung üblicher Eingabedaten und den daraus resultierenden Ausgabedaten nachgewiesen werden kann (Art. 5 Abs. 2 DS-GVO). Die Suche nach geeigneten Kennzahlen ist momentan (auch) noch aktueller Forschungsstand, es sollten aber hierbei zumindest einige Testläufe durchgeführt, bewertet und dokumentiert werden.

- ☐ Festlegung und Dokumentation des Umgangs mit Risiken aus dem Risikomodell, die sich insbesondere aus der fehlerhaften Verlässlichkeit ergeben (z. B. sog. Halluzinationen bei Großen Sprachmodellen oder auch die Ein-aus-1-Million-Ereignissen bei autonomen Fahrzeugen im Straßeneinsatz) und die sich mitunter auch nicht durch die Verwendung geeigneter Metriken (mittels Testläufen von Ein- und Ausgaben) ermitteln lassen, im Echtbetrieb dennoch sporadisch auftreten können.

- ☐ Umsetzung der Informationspflichten nach Art. 12 ff DS-GVO (auch dann, wenn ein KI-as-a-Service Dienst verwendet wird).

- ☐ Sicherstellung, dass Auskunftsersuchen nach Art. 15 DS-GVO auch bei Anfragen zum Einsatz von KI-Anwendungen im Datenschutzmanagement berücksichtigt werden. Dabei insbesondere den Einsatz von konkreten KI-Nutzungsszenarien bei personenbezogenen Daten, die anfragende Person betreffend, berücksichtigen.

- ☐ Bei konkretem Auskunftsersuchen nach Art. 15 DS-GVO in Bezug auf ein personenbeziehbares KI-Modell wird – je nach KI-Technologie – geprüft, ob personenbezogene Daten im KI-Modell direkt ermittelbar sind oder ob diese evtl. nur mit Zusatzinformationen (z. B. konkreter Prompt bei Großem Sprachmodell) aus einem KI-Modell abgeleitet werden können. Diese Zusatzinformationen sind im Zweifel vom Betroffenen dann anzufordern.

- ☐ Sicherstellung, dass Betroffenenrechte zur Berichtigung nach Art. 16 DS-GVO, zur Löschung nach Art. 17 DS-GVO, nach Einschränkung der Verarbeitung nach Art. 18 DS-GVO, nach Datenübertragbarkeit nach Art. 20 DS-GVO und des Widerspruchs nach Art. 21 DS-GVO in Bezug auf KI auch im Datenschutzmanagement berücksichtigt werden. Rückmeldefristen an Antragsteller sind hierbei zu beachten.

- ☐ Bei einem Löschersuchen nach Art. 17 DS-GVO in Bezug auf ein personenbeziehbares KI-Modell wird – je nach KI-Technologie – geprüft, ob personenbezogene Daten im KI-Modell direkt ermittelbar sind oder ob diese evtl. nur mit Zusatzinformationen (z. B. konkreter Prompt bei Großem Sprachmodell) auf einem KI-Modell abgeleitet werden können. Sofern eine Löschung in einem KI-Modell technisch ohne Beeinträchtigung des Gesamtmodells möglich ist, ist der Löschvorgang auch durchzuführen. Sollten andererseits personenbezogene Daten nur mittels Zusatzinformationen (z. B. Prompts) aus einem KI-Modell ermittelbar sein, dann besteht eine Möglichkeit des technischen Löschens darin, mittels Nachtraining die spezifisch zu löschende personenbezogene KI-Ausgabe mittels Anpassung der internen (Wahrscheinlichkeits-)Parameter umzusetzen.

- ☐ Sicherstellen, dass der Datenempfänger eines KI-as-a-Service Szenarios mögliche Eingabedaten als auch Ausgabedaten von KI-Modellen und KI-Anwendungen nicht für eigene Zwecke verwendet (z. B. Nachtraining, Filterverbesserung, Marketing, …) oder zumindest bei einer Zweckänderung geeignete Rechtsgrundlagen und Informationspflichten eingehalten werden. Ggf. sind KI-Anwendungen diesbezüglich speziell zu beauftragen oder zu konfigurieren.

- ☐ Bei Verpflichtung der Durchführung einer DSFA nach Art. 35 DS-GVO: Der betriebliche Datenschutzbeauftragte ist einzubinden. (Rest-)Risikobeurteilung anhand des Risikomodells und ggf. Konsultation der zuständigen Datenschutzaufsichtsbehörde nach Art. 36 DS-GVO bei weiterhin hohen Risiken für die Rechte und Freiheiten der vom Training betroffenen Personen.

- ☐ Vor Einsatz einer KI-Anwendung findet ein Freigabetest statt. Dieser ist zu dokumentieren.

- ☐ Der Einsatz von KI-Anwendungen wird in das Schulungsprogramm zum Datenschutz aufgenommen.

- ☐ Der Einsatz von KI-Anwendungen ist zum Nachweis einer angemessenen Risikoeindämmung zu protokollieren. Hierbei sind je nach Risikomodell sowohl Eingabe- als auch Ausgabedaten unter strenger Berücksichtigung der Zweckbindung auf einem gesicherten Protokollserver mit Zeitstempeln zu speichern.

- ☐ Bei der Protokollierung der Nutzung von KI-Anwendungen sind personenbezogene Daten, die einen Rückschluss auf einen konkreten Mitarbeiter schließen lassen, nur in pseudonymer Form mit gesicherten Identifikationsverfahren gespeichert.

- ☐ Findet eine Anpassung des KI-Modells im laufenden Betrieb einer KI-Anwendung (bspw. durch Einbindung mancher tagesaktuellen Webseiten) statt, dann ist dies bei einer Risikobeurteilung und bei Freigabetests besonders berücksichtigen.

- ☐ Der Einsatz von KI-Anwendungen samt Datenschutz-Risikomodell ist zu dokumentieren und regelmäßig unter Berücksichtigung des Verzeichnisses der Verarbeitungstätigkeiten nach Art. 30 DS-GVO auf Aktualität und Vollständigkeit zu prüfen.

Stand der Checkliste: 24.01.2024
Version: Konsultationsstand v0.9

Aktuelle Version zum Download:
https://www.lda.bayern.de/checkliste_ki

Herausgeber und Kontakt:
Bayerisches Landesamt für Datenschutzaufsicht (BayLDA)
Promenade 18 | 91522 Ansbach
www.lda.bayern.de | Tel.: 0981 180093-100
poststelle@lda.bayern.de