Der Hamburgische Beauftragte für
Datenschutz und Informationsfreiheit

13.11.2023

Checkliste zum Einsatz LLM-basierter Chatbots

Generative KI in Form von Chatbots bietet die Möglichkeit, schnell und unkompliziert Inhalte zu erstellen. Bekannte Large Language Models (LLM) sind ChatGPT, Luminous oder Bard. In vielen Einrichtungen sind die Tools mittlerweile Teil des Arbeitsalltags geworden, oft jedoch ohne verbindliche Vorgaben zur Nutzung. Dass die Sprachmodelle üblicherweise in einer Cloud betrieben werden, birgt verschiedene Datenschutzrisiken. Zum einen ist der Schutz vertraulicher Daten gefährdet, weil viele Unternehmen mit demselben LLM-Modell cloudbasiert arbeiten, Eingaben für das weitere Training der Modelle genutzt und ihnen dadurch möglicherweise Geschäftsgeheimnisse und persönliche Daten übermittelt werden. Zum anderen besteht die Gefahr, personenbezogene Daten aufgrund unrichtiger Ergebnisse unzulässig zu verarbeiten, gerade bei besonders schützenswerten Datenkategorien. Die folgende Checkliste dient Unternehmen und Behörden als Leitfaden zur datenschutzkonformen Nutzung von Chatbots.

1. Compliance-Regelungen vorgeben

Formulieren Sie klare und dokumentierte interne Weisungen, ob, beziehungsweise unter welchen Voraussetzungen welche Tools infrage kommen. Konkrete Beispiele der zugelassenen und der untersagten Einsatzszenarien helfen bei der Verdeutlichung.

Wer keine internen Regelungen vorgibt, ob und wie generative KI im Arbeitsalltag eingesetzt werden darf, kann davon ausgehen, dass sich Beschäftigte und andere Angehörige der Organisationen eigenmächtig und unkontrolliert der neuartigen Hilfsmittel bedienen. Unter Umständen haftet für diese Handlungen die arbeitgebende Einrichtung.

2. Datenschutzbeauftragte einbinden

Beziehen Sie immer Ihre oder Ihren internen Datenschutzbeauftragte:n ein, wenn Sie interne Weisungen erstellen oder einen Anwendungsfall erstmals umsetzen. Je nach Anwendungsfall sollten Sie in dem Zuge eine Datenschutz-Folgenabschätzung erstellen. Gegebenenfalls kann es auch sinnvoll sein, Betriebs- und Personalrät:innen mit ins Boot zu holen.

3. Bereitstellung eines Funktions-Accounts

Stellen Sie berufliche Chatbot-Accounts zur Verfügung. Beschäftigte sollten nicht eigenständig und unter Verwendung privater Daten ein Konto erstellen. Denn so würde ein Profil zu den jeweiligen Beschäftigten hinterlegt. Wenn die Verwendung im beruflichen Kontext erwünscht ist, sollten auch berufliche Accounts zur Verfügung gestellt werden. Nach Möglichkeit sollten diese Arbeits-Accounts nicht die Namen einzelner Beschäftigter enthalten. Soweit die E-Mail-Adresse abgefragt wird, bietet sich die Angabe einer dafür angelegten Mailadresse an. Teilweise werden auch Mobilfunknummern bei der Registrierung verlangt. Auch hier empfiehlt es sich, ein dienstliches Telefon dafür zu benutzen. Wir raten davon ab, die private Nutzung dieser dienstlichen Accounts zu erlauben.

4. Sichere Authentifizierung

Betrieblich genutzte Accounts für KI-Chatbots bieten ein erhebliches Missbrauchspotential. Gelangen Angreifer:innen unberechtigt zur Anwendungsoberfläche, können sie gegebenenfalls bisherige Aktivitäten einsehen, sollte der Chatverlauf nicht deaktiviert sein. Über eigene Abfragen können sie außerdem persönliche Informationen und Geschäftsgeheimnisse in Erfahrung bringen. Aus diesem Grund muss auf die Authentifizierung ein besonderer Fokus gelegt werden. Nutzen Sie starke Passwörter und integrieren Sie weiterer Authentifizierungsfaktoren.

5. Keine Eingabe personenbezogener Daten

Grundsätzlich gilt: Wenn sich der Anbieter eines Chatbots in den Geschäftsbedingungen eine Verwendung für eigene Zwecke einräumen lassen, dürfen keine personenbezogenen Daten an die KI übermittelt werden. Das betrifft jegliche Informationen, die Rückschlüsse auf Kund:innen, Geschäftspartner:innen oder sonstige Dritte enthalten, und ebenso Daten der eigenen Beschäftigten. Eine dafür erforderliche Rechtsgrundlage wird in der Regel nicht zu finden sein. Auch die eingebende Person selbst darf nicht identifizierbar sein, wenn sich für die Verarbeitung ihrer Daten durch den Anbieter keine tragfähige Rechtsgrundlage finden lässt.

6. Keine Ausgabe personenbezogener Daten

Achten Sie darauf, dass Ergebnisse der KI-Anwendung möglichst keine personenbezogenen Daten enthalten. Auch wenn der Eingabebefehl keine Person nennt, kann die KI unter Umständen vorherige Eingaben oder Informationen aus dem Internet einbeziehen. Daher sollten die Eingaben auf Fallgestaltungen beschränkt werden, die keinen Bezug zu Einzelpersonen herstellen.

Beispiel einer unproblematischen Eingabe: „Schreibe einen Werbetext zum Produkt X."

Beispiel einer problematischen Eingabe: „Welche Personen haben voraussichtlich Interesse am Produkt X?"

7. Vorsicht bei personenbeziehbaren Daten

Vermeiden Sie auch solche Eingaben, die unter Umständen auf konkrete Personen bezogen werden können. Es reicht nicht, Namen und Anschriften aus der Eingabe zu entfernen. Auch aus dem Zusammenhang lassen sich gegebenenfalls Rückschlüsse auf Autor:innen und Betroffene ziehen. Bei KI-Anwendungen, deren Bestimmung es ist, Querbezüge auch aus unstrukturierten Daten herzustellen, ist diese Gefahr besonders hoch.

Beispiel: „Entwirf ein Arbeitszeugnis im befriedigenden Bereich für einen Kundenberater im Autohaus X." Die Eingabe kann Personenbezug aufweisen, wenn erkennbar ist, aus welchem Unternehmen sie zu welchem Zeitpunkt getätigt wurde.

8. Opt-out des KI-Trainings

Nutzen Sie die Option, die Verwendung Ihrer Daten zu Trainingszwecken abzulehnen. Oft verwenden die Hersteller von KI-Modellen alle getätigten Eingaben zum weiteren Training ihrer KI. Privatleute und Beschäftigte anderer Unternehmen können diese Inhalte dann „erfragen". Je nach genutztem Dienst ist es jedoch möglich, der Verwendung zu Trainingszwecken zu widersprechen. Teilweise muss dafür ein spezifisches Vertragsmodell gebucht werden, das sich von der kostenfreien Standardanwendung unterscheidet.

Beispiel: Bei ChatGPT ist das Opt-out beispielsweise derzeit möglich über die Einstellungen unter

••• → Setting → Data Controls → Chat history and training

9. Opt-out der History

Chatbasierte Dienste bieten häufig an, bisherige Eingaben zu speichern, um den Dialog zu einem Thema an einem späteren Zeitpunkt wieder aufnehmen zu können. Damit ist zwangsläufig eine Verkettung der Eingaben einer Person verbunden. Insbesondere bei der gemeinsamen Nutzung durch mehrere Beschäftigte sollte die History abgewählt werden, da Inhalte ansonsten für alle Kolleg:innen einsehbar sind. Zu den Einstellungen beispielsweise bei ChatGPT siehe Punkt 8.

10. Ergebnisse auf Richtigkeit prüfen

Die Ergebnisse einer Chatbot-Anfrage sind mit Vorsicht zu genießen. Large Language Models erzeugen Texte, die mit mathematischer Wahrscheinlichkeit dem gewünschten Ergebnis nahekommen. Dies bedeutet keinesfalls, dass alle ausgegebenen Informationen korrekt sind. Im Gegenteil: Die bekannten LLM berücksichtigen meist vergleichsweise alte Informationsstände. Sie sind darüber hinaus bekannt für das Phänomen der „Halluzination“, bei der die KI scheinbar richtig und logisch erscheinende, tatsächlich aber falsche Aussagen erfindet. Es liegt in Ihrer Verantwortung als Nutzer.innen, das Ergebnis auf seine Richtigkeit zu überprüfen.

11. Ergebnisse auf Diskriminierung überprüfen

Auch unabhängig von ihrer sachlichen Richtigkeit können Ergebnisse unangebracht sein, wenn sie beispielsweise diskriminierend wirken. Eine darauf aufbauende Datenverarbeitung kann deshalb unzulässig sein, weil sie beispielsweise gegen das Allgemeine Gleichbehandlungsgesetz verstößt oder der Güterabwägung des Art. 6 Abs. 1 lit. f DSGVO nicht standhält. Auch hier tragen Sie als Nutzer:innen die Verantwortung zu überprüfen, ob die Antworten für die weitere Verwendung im gesetzlichen Rahmen tragbar sind.

Beispiel: Auch ohne direkten Personenbezug können Informationen diskriminierend sein. Eine KI könnte, ohne Namen zu nennen, folgende Empfehlung geben: „Für die freie Stelle sollten bevorzugt männliche Brillenträger ausgewählt werden.“ Ein solches Ergebnis könnte auf einer unzulässigen Auswertung von Gesundheits- und Geschlechtsangaben beruhen.

12. Keine automatisierte Letztentscheidung

Entscheidungen mit Rechtswirkung sollten grundsätzlich nur von Menschen getroffen werden. Andernfalls sind die Voraussetzungen des Art. 22 DSGVO zu beachten. Erarbeitet ein LLM-basierter Chatbot Vorschläge, die durch Beschäftigte angenommen werden, müssen diejenigen darauf achten, dass ihnen ein tatsächlicher Entscheidungsspielraum zukommt. Vermeiden Sie es, aufgrund der fehlenden Transparenz der KI-gestützten Vorarbeit faktisch an die Vorschläge gebunden zu sein, weil Sie den Entscheidungsweg nicht nachvollziehen können. Auch

unzureichende Ressourcen und Zeitdruck können dazu führen, dass Ergebnisse ungeprüft übernommen werden.

13. Beschäftigte sensibilisieren

Sensibilisieren Sie Beschäftigten durch Schulungen, Leitfäden und Gespräche dahingehend, ob und wie sie KI-Tools nutzen dürfen.

14. Datenschutz ist nicht alles

Der Schutz personenbezogener Daten darf durch die Nutzung von KI-Diensten nicht unterlaufen werden. Es empfiehlt sich darüber hinaus, weitere Aspekte wie den Schutz von Urheberrechten oder Geschäftsgeheimnissen zu regeln. Bei behördlichen Anwendungsfällen sind Weitergabeverbote nach dem Sicherheitsüberprüfungsgesetz (SÜG) und anderen Regelungen zu berücksichtigen.

15. Weitere Entwicklung verfolgen

Auf EU-Ebene wird aktuell die Regulierung künstlicher Intelligenz vorbereitet. Die künftige KI-Verordnung betrifft voraussichtlich nicht nur die Anbieter solcher Dienste, sondern auch bestimmte Nutzer:innen. Aufgrund fortschreitender technischer Lösungen und laufender Updates auf neue Systeme und Sprachmodelle sollte regelmäßig überprüft werden, ob die internen Vorgaben angepasst werden müssen.

Zudem prüfen die Datenschutzbehörden gerade in Musterverfahren, ob die am Markt befindlichen Sprachmodelle grundsätzlich rechtmäßig sind.