

Anlage 1: Checkliste KI-Systeme

Das KI-System ist gemäß den Vorschriften der EU KI Verordnung in eine von vier Risikoklassen einzuteilen. Je nach Risikoklasse schließen sich entsprechende zu erfüllende gesetzliche Anforderungen an. Die Checkliste wird im Rahmen des Softwarefreigabeverfahrens der Stadt Osnabrück eingesetzt.

Rolle der Stadt beim Einsatz des KI-Systems

Soll das KI-System in der Rolle als Betreiber (gem. Art. 3 Abs. 4 EU KI-Verordnung) oder nachgeschalteter Anbieter (gem. Art 3 Abs. 68 EU KI-Verordnung) genutzt werden?

Falls diese Frage mit „Nein“ beantwortet wird, liegt eine andere Rolle als „Betreiber“ oder „nachgeschalteter Anbieter“ vor. Alle anderen Rollen ziehen umfangreiche Dokumentationspflichten nach sich und werden für die Nutzung bei der Stadt Osnabrück nicht zugelassen.

Einordnung in die Risikoklassen der EU KI-Verordnung

1. Verbotene Systeme / Verbotene KI-Praktiken

- Beeinflusst das System direkt oder unterschwellig beteiligte Personen in schadhafter Weise?
- Adressiert das System in irgendeiner Weise das Alter oder eine geistige oder körperliche Behinderung in schadhafter Weise?
- Bewertet das System die Vertrauenswürdigkeit einer Person auf Basis ihres Verhaltens oder vorhergesagten Verhaltens?
- Leitet das System Emotionen von natürlichen Personen in den Bereichen Arbeitsplatz oder Bildungseinrichtung ab?
- Verwendet das System biometrische Daten zur Identifizierung oder Kategorisierung von Personen?

Falls eine dieser Fragen mit „Ja“ beantwortet werden sollte oder unklar ist, ob die Tatsache auf das System zutrifft, darf das System nicht eingesetzt werden. Es gehört damit in die Klasse der verbotenen Systeme.

2. Hochrisikosysteme

- Soll das System als Sicherheitskomponente für ein Produkt verwendet werden, die den EU-Harmonisierungsrechtsvorschriften unterliegen oder ist es selbst ein solches Produkt?
- Muss das KI-System (als Sicherheitskomponente) einer Konformitätsbewertung unterzogen werden?
- Handelt es sich bei dem KI-System um eines der folgenden Produkte oder befasst es sich mit dem Thema?
 - Biometrische Identifizierung
 - Verwaltung und Betrieb kritischer Infrastrukturen (Verwaltung, Gas-, Wasser- und Stromversorgung)
 - Allgemeine und berufliche Bildung (Zugang zu Einrichtungen, Bewertung von Schülern)
 - Beschäftigung, Personalmanagement und Zugang zu Selbstständigkeit (Einstellung, Sichten von Bewerbungen, Bewertung von Bewerbern, Beförderungen, Leistungsbewertung)
 - e. Zugänglichkeit und Inanspruchnahme grundlegender privater und öffentlicher Dienste und Leistungen (Anspruchsbeurteilung und -gewährung, Kreditwürdigkeit, Entsendung von Not- und Rettungsdiensten)
 - f. Strafverfolgung
 - g. Migration, Asyl und Grenzkontrolle
 - h. Rechtspflege und demokratische Prozesse (Ermittlung und Auslegung von rechtlich relevanten Sachverhalten)
- Ist das KI-System in der EU-Datenbank für Hochrisiko-KI-Systeme gelistet?

Falls die ersten beiden Bedingungen beide erfüllt sind oder eine Zuordnung nach Frage 3 bejaht werden kann, ist das KI-System in Klasse der Hochrisiko Systeme einzuordnen.

Der Einsatz von als Hochrisikosystem klassifizierten KI-Systemen ist bei der Stadt Osnabrück nicht zulässig.

3. Begrenztes Risiko

1. Ist das System für die Interaktion mit natürlichen Personen bestimmt?
2. Erzeugt oder manipuliert das System Bild-, Ton- oder Videoinhalte, die mit realen Personen, Orten oder Gegenständen verwechselt werden können? (Deepfake)
3. Erzeugt oder manipuliert das System Text und diese Texte werden mit dem Zweck veröffentlicht, über Angelegenheiten von öffentlichem Interesse zu informieren?

Falls eine dieser Fragen mit „Ja“ beantwortet wird, handelt es sich um ein KI-System mit begrenztem Risiko. Im Falle eines Systems nach Frage 1 muss den Nutzern offengelegt werden, dass sie mit einem KI-System interagieren. Im Falle eines Systems nach Frage 2 oder 3 muss offengelegt werden, dass entsprechende Inhalte durch ein solches System erzeugt bzw. manipuliert wurden.

4. Geringes Risiko

Systeme mit geringem Risiko unterliegen keinen Regulierungsvorschriften. Sofern ein System in keine der anderen drei Kategorien einsortiert werden muss, gilt es als System mit geringem Risiko

5. Modelle für allgemeine Zwecke (GPAI) System

Sogenannte GPAI-Modelle sind solche Modelle, die mit einer großen Datenmenge trainiert wurden und somit eine große Bandbreite an Ausgaben erfüllen können und nicht auf bestimmte enge Anwendungsfälle begrenzt sind.

Diesen Systemen wohnt ein sogenanntes „systemisches Risiko“ inne. Das bedeutet, dass es über Möglichkeiten mit potenziell großen Auswirkungen verfügt. Diese Möglichkeiten werden über die Rechenkapazität des Modells quantifiziert und wird bei einer für das Training aufgewendeten Kapazität größer 10^{25} FLOPS (floating point operations per Second) angenommen. Daneben kann einem System ein „systemisches Risiko“ auf Basis einer qualifizierten Warnmeldung der EU zugewiesen werden.

Auch diese Systeme werden in einer Datenbank veröffentlicht. Es ergeben sich umfassende Pflichten für Anbieter solcher Systeme.

Für Betreiber solcher Systeme gelten die Transparenzregelungen für Systeme mit begrenztem Risiko.