LANDTAG
NORDRHEIN-WESTFALEN
18. WAHLPERIODE

STELLUNGNAHME
18/1127

A14, A20

ALGORITHM WATCH

/ Stellungnahme

zum Antrag der Fraktion der FDP

KI in der öffentlichen Verwaltung erproben und Beschäftigte bei der rechtssicheren Nutzung unterstützen

Drucksache 18/5422, Landtag Nordrhein-Westfalen

21. Dezember 2023

Verfasst von: Pia Sombetzki, Policy & Advocacy Managerin
& Kilian Vieth-Ditlmann, stellvertretende Teamleitung Policy & Advocacy

Einleitung

Wir bedanken uns für die Einladung zur Stellungnahme zum Antrag der Fraktion der FDP „KI in der öffentlichen Verwaltung erproben und Beschäftigte bei der rechtssicheren Nutzung unterstützen“. Der Antrag zielt darauf ab, die Erprobung von Systemen generativer Künstlicher Intelligenz (im Folgenden auch: generative KI-Systeme) in der öffentlichen Verwaltung zu fördern und die Beschäftigten bei der rechtssicheren Nutzung zu unterstützen. Demnach bestehe unter anderem der Bedarf, Regelungen auf Landesebene für den Einsatz von generativen KI-Systemen zu schaffen, da dies bisher nicht ausdrücklich geschehen ist. Der Antrag nimmt insbesondere Bezug auf die rechtssichere Anwendung von generativen KI-Systemen und betont die Bedeutung von Transparenz gegenüber den Bürgerinnen und Bürgern beim Einsatz von KI.

Gegenstand und Fokus dieser Stellungnahme sind grundlegende Anforderungen und Risiken bei der Nutzung von generativen KI-Systemen in der öffentlichen Verwaltung. Andere Anwendungen von sogenannter Künstlicher Intelligenz (KI) werden, im Sinne des Antrags, nur peripher behandelt. Wir verstehen generative KI-Systeme dabei als vielfältig einsetzbare text-, ton- oder bildgenerierende Modelle, die auf Grundlage von Texteingaben (sog. „Prompts“) komplexe und realistisch bzw. authentisch anmutende Outputs generieren.

Grundsätzlich tragen Behörden beim Einsatz von generativer KI und anderen automatisierten Entscheidungssystemen eine besondere Verantwortung den Menschen gegenüber, die von diesen Systemen betroffen sind. Der Einsatz kann viele Vorteile, aber auch erhebliche Risiken mit sich bringen. Solche Risiken sind nicht auf den öffentlichen Sektor beschränkt, sondern spiegeln häufig ähnliche Risiken wider, die auch im privaten Sektor anzutreffen sind. Gleichzeitig sind im öffentlichen Sektor andere Voraussetzungen gegeben: Die Bürger*innen haben keine Wahl zwischen verschiedenen Anbieter*innen, sondern sind unausweichlich der für sie zuständigen Verwaltung unterworfen. Zudem können Behörden nicht selten auf sensible personenbezogene Daten zugreifen und ihre Entscheidungen haben für die betroffenen Personen oft weitreichende Folgen. Wenn generative KI-Anwendungen in der öffentlichen Verwaltung eingesetzt werden, muss daher sichergestellt werden, dass diese tatsächlich einen Nutzen generieren und Schaden vermieden wird, dass die Autonomie der Betroffenen gewahrt bleibt, und dass Grundsätze der Fairness und Gerechtigkeit eingehalten werden.

Um diesen Anforderungen gerecht zu werden, sollten verbindliche Richtlinien für den Einsatz von generativer KI im öffentlichen Sektor erarbeitet und Verwaltungsmitarbeiter*innen bereitgestellt werden. Diese Anforderungen müssen Nachvollziehbarkeit und Verantwortlichkeit sicherstellen und individuelle sowie demokratische Kontrolle ermöglichen. Transparenz ist dabei ein notwendiger erster Schritt, um den Einfluss generativer KI zu beurteilen. Transparenz ist eine Bedingung dafür, dass für die Öffentlichkeit die Möglichkeit besteht, den Einsatz von generativer KI

evidenzbasiert zu debattieren und öffentlich zu überwachen. Zudem ist sie die Grundlage für Betroffene, um von Schutzrechten Gebrauch zu machen.

Im Folgenden gehen wir auf drei Themenbereiche ein. Erstens, auf grundsätzliche Rechtsunsicherheiten öffentlich zugänglicher generativer KI-Anwendungen. Zweitens auf die spezifischen Rechtsunsicherheiten hinsichtlich des Einsatzes generativer KI-Anwendungen in der öffentlichen Verwaltung, und drittens auf die Bedeutung von Einzelfallprüfungen und Transparenz bezüglich (generativer) KI-Anwendungen im öffentlichen Sektor.

Grundsätzliche Rechtsunsicherheiten öffentlich zugänglicher generativer KI-Anwendungen

Der Einsatz von auf dem Markt frei bzw. öffentlich zugänglichen Anwendungen (z.B. ChatGPT, Bard, Midjourney, DALL·E etc.) kann nach derzeitigem Kenntnisstand **grundsätzlich nicht** als **rechtssicher** betrachtet werden, da sowohl während der Entwicklung als auch bei der Nutzung von generativen KI-Systemen eine Vielzahl von Rechten berührt werden und aktuell keiner der Anbieter eine umfängliche Rechtssicherheit nachweist oder sicherstellen kann.[1]

Trainingsdatensätze für generative KI Modelle sind besonders umfangreich und werden meist durch groß angelegtes Scraping (Extrahieren von Informationen aus Webseiten) aus öffentlichen Quellen im Internet oder anderen digitalen Quellen gewonnen und können u.a. personenbezogene und urheberrechtlich geschützte Daten enthalten. Viele Anbieter machen derzeit keine vollständigen und extern überprüfbaren Angaben darüber, welche urheberrechtlich geschützten Werke (etwa Fotografie, Kunst, Quellcode oder Literatur) sie im Trainingsprozess verwenden und welche besonders schützenswerten Daten dabei verarbeitet werden. In der Rechtsprechung gibt es noch keine eindeutige Meinung darüber, ob und in welcher Form Scraping – zum Zwecke des Trainings generativer KI-Modelle – urheberrechtlich zulässig ist.[2]

Die auf diesen von öffentlichen Websites gesammelten Daten trainierten generativen KI-Modelle können darüber hinaus falsche, ungenaue oder verzerrte Daten über

[1] Für eine Betrachtung rechtlicher Unklarheit im Kontext der EU KI-Verordnung, siehe: AlgorithmWatch & AI, Media & Democracy Lab of the University of Amsterdam, *Recommendations by academics and activists The AI Act and General Purpose AI: Charting a path forward* (September 2023): https://algorithmwatch.org/de/wp-content/uploads/2023/09/PolicyBrief_GPAI_AW-updated1509-02.pdf

[2] Siehe z.B.: Google says data-scraping lawsuit would take 'sledgehammer' to generative AI (Oktober 2023): https://www.reuters.com/legal/litigation/google-says-data-scraping-lawsuit-would-take-sledgehammer-generative-ai-2023-10-17/

Personen enthalten und diese im Zweifelsfall als Falschinformationen ausgeben. Je nach Einsatzszenario können dadurch weitreichende Risiken entstehen.[3]

Fragen zur Benutzerinteraktion (Eingaben bzw. Prompts) können zur Feinabstimmung des Modells wiederverwendet werden, was versehentlich sensible Informationen wie etwa Geschäftsgeheimnisse offenlegen kann. Deswegen untersagen auch manche Unternehmen ihren Beschäftigten die Nutzung von ChatGPT oder GitHub Copilot, um das Abfließen vertraulicher Daten zu unterbinden.[4]

Die Rechte der betroffenen Personen, wie zum Beispiel das Recht auf Einsicht und mögliche Löschung oder Berichtigung von (unrichtigen) personenbezogenen Daten sind derzeit sehr begrenzt. Anders als bei Suchmaschinen, bei welchen beispielsweise das Löschen von personenbezogenen Daten nach der Datenschutzgrundverordnung (DSGVO) beantragt werden kann, ist dieses Recht hinsichtlich der Funktionsweise von großen Sprachmodellen und der fehlenden Einsicht in Trainingsdatensätze nicht gleichermaßen einfach umsetzbar.[5] Für Nutzer*innen ist es derzeit schlicht unmöglich, zu überprüfen, ob personenbezogene Daten in das jeweilige Modell eingeflossen sind, eben weil Unternehmen wie OpenAI ihren Trainingsdatensatz nicht transparent machen und eine Abfrage per 'Prompt' nicht zuverlässig ist. Außerdem müsste ein Sprachmodell neu trainiert werden, um die Löschung effektiv werden zu lassen – jeder Trainingszyklus eines Modells ist allerdings mit einem enormen Energie-, Ressourcen- und Zeitaufwand verbunden. Es gibt eine Reihe weiterer nachträglicher theoretischer Korrekturansätze. Inwiefern sie aber in unternehmerischer Praxis Einzug finden, ist kaum zu untersuchen, auch hier aufgrund des Mangels an Transparenz seitens der Anbieter. Gleichzeitig sind Berichtigungen aber auch darin limitiert, dass sie einer Löschung von Daten aus einem Modell nicht gleichgesetzt werden könnten, da sie lediglich die Ausgabe von Inhalten, wie durch einen Filter, limitieren bzw. kontrollieren würden – ähnlich wie, beispielsweise, die Ausgabe von Hassrede (teils ebenso wenig zuverlässig) von manchen Unternehmen kontrolliert wird.

Rechtsunsicherheiten hinsichtlich des Einsatzes generativer KI-Anwendungen in der öffentlichen Verwaltung

Laut Artikel 22 DSGVO haben Betroffene das Recht, nicht ausschließlich einer auf automatisierter Datenverarbeitung beruhenden Entscheidung unterworfen zu werden,

[3] AI Forensics & AlgorithmWatch, *KI-Chatbot liefert falsche Antworten auf Fragen zu demokratischen Wahlen* (November 2023): https://algorithmwatch.org/de/schlussbericht-microsoft-bing-chat/
[4] Tilley & Kruppa, *Apple Restricts Employee Use of ChatGPT, Joining Other Companies Wary of Leaks* (Mai 2023): https://www.wsj.com/articles/apple-restricts-use-of-chatgpt-joining-other-companies-wary-of-leaks-d44d7d34
[5] Zhang et. al., *Right to be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions* (September 2023): https://arxiv.org/pdf/2307.03941.pdf

die ihnen gegenüber rechtliche Wirkung entfaltet. Diese Regel, nach der Entscheidungen mit Rechtswirkung grundsätzlich von Menschen mitbestimmt werden müssen, kann auch den Einsatz von generativer KI betreffen. In Fällen, in denen generative KI, beispielsweise ein mit ChatGPT generierter Text, als Grundlage für eine Entscheidung mit Rechtswirkung herangezogen wird, kann dieses Verbot einer automatisierten Entscheidung greifen. Das unlängst veröffentlichte EuGH-Urteil[6] zum Schufa-Score sieht auch die automatisierte Berechnung von Scorewerten für die Kreditwürdigkeit von diesem Verbot erfasst. Dementsprechend ist davon auszugehen, dass auch der Einsatz von KI-Systemen zur Erstellung von Entscheidungsvorlagen von Artikel 22 der DSGVO erfasst werden könnte.[7]

Unabhängig von datenschutzrechtlichen Verboten kann ein generatives KI-System grundsätzlich Falschaussagen über Personen oder Tatbestände produzieren. Generative KI kennt keine Wahrheiten, sondern erstellt Text oder Bildmaterial auf Grundlage von Wahrscheinlichkeiten. Ohne menschliche Prüfung kann nicht von einer faktischen Richtigkeit der KI-generierten Inhalte ausgegangen werden. Wenn keine vertragliche Regelung mit dem Anbieter vorliegt, liegt die Verantwortung für die Richtigkeit bei den Nutzer*innen (in der Verwaltung).[8]

Ein trainiertes generatives KI-System kann darüber hinaus diskriminierende Aussagen ausgeben: Ein Chatbot könnte z.B. die folgende Empfehlung geben: „Für die freie Stelle sollten bevorzugt männliche Brillenträger ausgewählt werden.“[9] Die Verantwortung für die Einhaltung des Grundsatzes der Nicht-Diskriminierung liegt in einem solchen Fall bei dem oder der Nutzer*in in der Verwaltung. Die Verwaltungsbediensteten müssen die geltende Rechtslage interpretieren – in diesem Fall, u.a. das Allgemeine Gleichbehandlungsgesetz bei der Stellenbesetzung einhalten und nicht blind der Empfehlung eines Chatbots folgen.

---

[6] EuGH-Urteil im Fall C-634/21 (Dezember 2023): https://curia.europa.eu/juris/document/document.jsf?text=&docid=280426&pageIndex=0&doclang=DE&mode=req&dir=&occ=first&part=1&cid=5044806

[7] Siehe dazu auch die Einordnung des EuGH-Urteils seitens des Hamburgischen Datenschutzbeauftragten, *Effects of the Schufa judgement on AI applications - automated decisions must not play a decisive role* (Dezember 2023): https://datenschutz-hamburg.de/fileadmin/user_upload/HmbBfDI/Datenschutz/Informationen/20231207_PM_EuGH_Judgement_Schufa_Effects_on_AI.pdf

[8] Auch eine Handreichung der Schweizerischen Eidgenossenschaft unterstreicht die Verantwortung der handelnden Verwaltungsbediensteten und die Pflicht Entscheidungen individuell begründen zu können, siehe das *Merkblatt zur Verwendung von generativen KI-Werkzeugen in der Bundesverwaltung* (Dezember 2023): https://cnai.swiss/dienstleistungen-weitere-dienstleistungen-merkblaetter-zu-ki/

[9] Beispiel aus der *Checkliste zum Einsatz LLM-basierter Chatbots*, vom Hamburgischen Beauftragten für Datenschutz und Informationsfreiheit (November 2023): https://datenschutz-hamburg.de/fileadmin/user_upload/HmbBfDI/Datenschutz/Informationen/20231113_Checkliste_LLM_Chatbots_DE.pdf

Einzelfallprüfungen und Transparenz über den Einsatz (generativer) KI-Anwendungen

Unabhängig von grundsätzlichen rechtlichen Fragestellungen sollte der Einsatz von (generativen) KI-Anwendungen stets von Fall zu Fall geprüft werden. Risiken[10] entstehen und variieren abhängig von Kontext und Einsatzweise eines KI-Systems. Es empfiehlt sich neben einer Datenschutzfolgenabschätzung ebenso eine KI-Folgenabschätzung durchzuführen, um Risiken strukturiert zu bewerten. AlgorithmWatch bietet für diesen Zweck ein Tool basierend auf einer zweistufigen Checkliste an. Darin werden anhand von Fragen die möglichen Risiken systematisch bewertet.[11]

Die Fragen zur KI-Folgenabschätzung sollten so früh wie möglich, d. h. bereits in der Planungsphase, beantwortet werden. Sie helfen dabei, neben den primären Projektzielen zusätzliche Spezifikationen für das Projekt zu berücksichtigen und mögliche Gefahren frühzeitig adressieren zu können. Die Checkliste 1 – die sogenannte Triage-Checkliste – hilft bei der Ermittlung der wichtigsten ethischen und grundrechtlichen Aspekte, die dokumentiert werden müssen. Sie sollte nicht als erschöpfende Liste aller Risiken für alle Zusammenhänge angesehen werden (es können andere und weitere Risiken für Personen, Sachen oder die Gesellschaft insgesamt bestehen), sondern als eine Art Schnelltest, um unnötige Bürokratie zu vermeiden und die kritischen Aspekte des konkreten Anwendungsfalls zügig zu identifizieren. Fragen, die dabei beantwortet werden müssen, sind etwa „Ist der Schaden einer falschen Entscheidung vollständig reversibel?“ oder „Besteht das Risiko, dass das System eine politische Entscheidung (z. B. Wahl oder Volksabstimmung) beeinflusst?“. Der Fragebogen zur Risikobewertung ist aktuell noch nicht auf die spezifischen Anforderungen von generativen KI-Systemen zugeschnitten, sondern generell auf den Einsatz algorithmischer Entscheidungssysteme in der öffentlichen Verwaltung. Die Fragenlisten bieten jedoch schon jetzt Anknüpfungspunkte für eine vertiefende Risikoanalyse in den Bereichen Schadensvermeidung, Fairness und Autonomie.[12]

Das Ziel der Checkliste 2 ist, Transparenz über alle kritischen Aspekte herzustellen und damit die Vertrauenswürdigkeit des Prozesses zu fördern. Das Ziel wird durch Erstellung eines Berichts erreicht.[13] Wenn die Beantwortung der Checkliste 1 ergibt, dass ein

[10] Für eine Übersicht struktureller Risiken von generativen KI-Systemen im öffentlichen Sektor, siehe: Ada Lovelace Institute, *Foundation models in the public sector* (October 2023), Seite 26ff: https://www.adalovelaceinstitute.org/wp-content/uploads/2023/11/ALI_Foundation-models-evidence-review_2023.pdf

[11] AlgorithmWatch, *Automatisierte Entscheidungssysteme im öffentlichen Sektor: Ein Impact-Assessment-Tool für die öffentliche Verwaltung* (Juni 2021): https://algorithmwatch.org/de/wp-content/uploads/2022/09/ADM-Folgenabscha%CC%88tzung_AlgorithmWatch_2022.pdf

[12] Online-Tool für Checkliste 1: https://umfrage.algorithmwatch.org/impact-assessment-tool-fur-die-offentliche-verwaltung/

[13] Siehe dazu das Instrument zur Folgenabschätzung, Abschnitt III ab Seite 27, sowie Flussdiagramm aus Seite 52:

Transparenzbericht zu verfassen ist, wird mit Hilfe eines Flussdiagramms bestimmt, welche Fragen aus Checkliste 2 der Transparenzbericht beantworten muss. Eine KI-Folgenabschätzung sollte immer möglichst vielfältige Gruppen einbinden, etwa aus der Zivilgesellschaft oder auch die Betroffenen der geplanten KI-Nutzung, um möglichst alle Risiken einer KI-Anwendung zu erfassen.

Eine standardisierte Zusammenfassung der Ergebnisse dieser Folgenabschätzungen sollte in einem KI-Transparenzregister öffentlich einsehbar sein.[14] Forschungsergebnisse haben u.a. gezeigt, dass die Öffentlichkeit eine klare Erwartung an die Nachvollziehbarkeit von KI-Systemen hat, und dass sie die Grundlage für Vertrauen in die Technologie ist.[15] Da auch der Entwurf der KI-Verordnung der EU die Schaffung eines KI-Registers für sogenannte Hochrisikosysteme vorsieht, sollten auch auf Landesebene die strukturellen und fachlichen Voraussetzungen für die Schaffung eines solchen Transparenz-Registers geprüft werden. Es sollte u.a. eine unterstützende Stelle auf Landesebene (bspw. eine „KI-Taskforce“) etabliert werden, die bei der Durchführung von Folgenabschätzungen unterstützt und Handreichungen für den Einsatz von (generativer) KI entwickelt.

https://algorithmwatch.org/de/wp-content/uploads/2022/09/ADM-Folgenabscha%CC%88tzung_AlgorithmWatch_2022.pdf

[14] AlgorithmWatch, *Konzept für ein KI-Transparenzregister auf Bundesebene* (März 2023): https://algorithmwatch.org/de/wp-content/uploads/2023/03/Konzept_KI-Transparenzregister_AlgorithmWatch_2023.pdf

[15] Ada Lovelace Institute & The Alan Turing Institute, *How Do People Feel about AI?* (Juni 2023): https://www.adalovelaceinstitute.org/report/public-attitudes-ai/