16941/AB

vom 11.03.2024 zu 17485/J (XXVII. GP)
bmaw.gv.at

Arbeit und Wirtschaft

Univ.-Prof. Dr. Martin Kocher Bundesminister

Stubenring 1, 1010 Wien

Herrn Präsidenten des Nationalrates Mag. Wolfgang Sobotka Parlament 1017 Wien

Geschäftszahl: 2024-0.029.648 Ihr Zeichen: BKA - PDion (PDion)17485/J-NR/2024

Wien, am 11. März 2024

Sehr geehrter Herr Präsident,

die Abgeordneten zum Nationalrat Mario Lindner und weitere haben am 11.01.2024 unter der Nr. 17485/J an mich eine schriftliche parlamentarische Anfrage betreffend Inakzeptable Stereotypen und Bias bei der neuen AMS-KI gerichtet.

Einleitend darf festgehalten werden, dass das Bundesministerium für Arbeit und Wirtschaft für die Beantwortung Informationen vom AMS eingeholt hat, wo Beauftragung und Abwicklung des Berufsinfomats stattfanden.

Diese Anfrage beantworte ich nach den mir vorliegenden Informationen wie folgt:

Zu den Fragen 1 und 2

- Laut Medienberichten beliefen sich die Gesamtkosten des AMS-Berufsinfomats als Kooperationsprojekt zwischen dem BRZ und der Firma goodguys auf rund € 300.000: Wie hoch waren die genauen Gesamtkosten für die Erstellung des AMS-Berufsinfomats und wie genau gliedern sich diese?
- Wie hoch sind die regelmäßigen Betriebs- und Wartungskosten für den AMS-Berufsinfomat?

Die bis 31.12.2023 verrechneten Kosten betrugen insgesamt € 275.160,00. Die GoodGuys GmbH bekam € 74.880,00 für die Durchführung des "Proof of Concept" (PoC) mit Anpas-

sung Al-Concierge, den Betrieb des PoC auf deren Servern sowie die Nutzung der API/System/Al-Modelle. Die Braintrust Marketing Service Gesellschaft m.b.H erhielt € 14.280,00 für die Entwicklung eines automatisierten XML-Exports von Daten aus Berufslexikon und Ausbildungskompass zur weiteren Verwendung beim Berufsinfomaten.

Die Kosten für die BRZ GmbH betrugen € 186.000,00 für die erste Teilzahlung für Leistungen zur Inbetriebnahme des Berufsinfomaten samt Bereitstellung der Infrastruktur, Konfiguration und Deployment der Anwendung sowie Integration in die Webseite www.ams.at.

Die voraussichtlichen Kosten für den Zeitraum vom 01.01.2024 bis zum 31.03.2024, somit bis zum Ende des Projektbetriebes, betragen maximal € 231.977,51. Die Kosten für die GoodGuys GmbH werden € 5.760,00 betragen, jene für die Braintrust Marketing Service Gesellschaft m.b.H werden nach erfolgreicher Abnahme € 14.280,00 betragen. Für die BRZ GmbH sind maximal € 211.937,51 für Dienstleistungen und Projektbetrieb kalkuliert. Die Verrechnung erfolgt nach tatsächlichem Aufwand.

Die detaillierten Kosten für den Produktionsbetrieb ab 01.04.2024 stehen final noch nicht fest. Sie hängen von weiteren Verbesserungen und auch der Inanspruchnahme ab. Aufgrund der Erfahrungen aus dem PoC und dem gegenwärtig laufenden Pilotbetrieb können durch die BRZ GmbH Aufwände für die Infrastruktur und für die erforderlichen Personalressourcen abgeschätzt werden. Anschließend erfolgt dann die Angebotslegung durch die BRZ GmbH.

Zur Frage 3

• Welche Abläufe und Vereinbarungen existieren hinsichtlich allfälliger Änderungen, Adaptierungen etc. beim Betrieb des AMS-Berufsinfomat, insbesondere auch hinsichtlich der dafür anfallenden Kosten?

In der momentan bis März 2024 laufenden Inbetriebnahmephase werden inhaltliche Verbesserungen (Content) durchgeführt. Für den weiteren Betrieb und die Wartung sind zusätzlich folgende Änderungen und Adaptierungen geplant: Gestaltung des Frontend (Design, Grafik, etc.) und statistische Auswertungen (Dashboard-Statistiken). Die diesbezüglich anfallenden Kosten können momentan noch nicht genannt werden, da der Auftrag zu Betrieb und Wartung erst erfolgen muss.

Zur Frage 4

 Wurden auch Angebote von anderen Anbieter*innen zur Erstellung dieser Anwendung eingeholt und auf welcher konkreten Grundlage kam es zur Entscheidung für das gegenwärtige Modell?

Es wurde der Weg einer Direktvergabe ohne weitere Angebote gewählt, weil die Good-Guys GmbH nachweislich im Zuge eines BRZ-Projektes Expertise im Bereich generative KI vorweisen konnte und bereits das Verfahren als "Software as a Service" Produkt in mehreren Bereichen (z.B. im Tourismus) umgesetzt hatte. Daher fiel im April 2023 die Entscheidung einen PoC zu beauftragen. Die Ergebnisse zeigten, dass mittels KI eine entsprechende Informationsbereitstellung möglich ist. Daher wurde dieses Projekts für eine Pilotierung freigegeben.

Zur Frage 5

• Wie genau wird sichergestellt, dass die vom AMS-Berufsinfomat gesammelten und verwerteten Daten nicht an Dritte, insbesondere Unternehmen außerhalb der EU, weitergegeben werden?

Der Berufsinfomat speichert keine Daten der Userinnen und User. Jede Anfrage wird mit dem dafür aufbereiteten AMS-Content zur Antwortgenerierung an die OpenAI-API (momentan noch außerhalb der EU) gesendet. Der Berufsinfomat-Server selbst erhält keine IP-Adresse des aufrufenden WebClient der Nutzerin bzw. des Nutzers, da alle relevanten IP-Informationen am BRZ-Proxy abgefangen werden. Einzig die gestellte Frage, die beabsichtigten Kommunikationseinstellungen des WebClients der Nutzerin bzw. des Nutzers, die Browserkennung und die bevorzugte Sprache des Browsers der Nutzerin bzw. des Nutzers gelangen zum Berufsinfomat-Server und damit zur weiteren Verarbeitung im Berufsinfomat.

Da der Berufsinfomat in einem iFrame läuft, der auf www.ams.at eingebunden ist, werden Daten innerhalb des iFrames nicht vom Webtracking der AMS-Website erfasst.

Auf www.ams.at wird zudem aktiv abgefragt, ob die Nutzerin bzw. der Nutzer mit der anonymisierten Aufzeichnung des Surfverhaltens einverstanden ist. Wird hier nicht zugestimmt, werden die Tracking-Skripte gar nicht erst eingebunden und es ist somit kein Tracking möglich. Damit ist technisch sichergestellt, dass ohne Wissen der Nutzerinnen und Nutzer keine Daten an andere Unternehmen geschickt werden. Stimmt die Nutzerin bzw. der Nutzer zu, wird ein gängiges Webtracking aktiv. Hierbei werden anonymisiert die Sei-

tenaufrufe aufgezeichnet, aber keinerlei persönliche Daten oder Eingaben. Darauf wird auch im Disclaimer und in der Datenschutzbestimmung auf der Website hingewiesen.

Zu den Fragen 6, 10 und 11

- Welche Trainingsdaten wurden konkret für die Erstellung der KI des AMS-Berufsinfomat verwendet und wie wurden diese ausgewählt?
 - Welche Trainingsdaten wurden auf Basis welcher Kriterien genau ausgewählt?
 - o Aus welchen Quellen wurden diese Trainingsdaten gespeist?
 - o Wo sind diese Trainingsdaten gespeichert?
- Wie wurde sichergestellt, dass in den Trainingsdaten für die KI des AMS-Berufsinfomats kein Bias, insbesondere hinsichtlich des Geschlechts, enthalten ist?
- Wie wurde sichergestellt, dass in den Trainingsdaten für die KI des AMS-Berufsinfomats keine unwissenschaftlichen Daten, enthalten sind?

Es wurden keine Trainingsdaten im Sinne von "machine learning" verwendet. Allerdings wird das Userinnen- und User-Feedback gesammelt und dieses fließt kontinuierlich in den Entwicklungsprozess ein. Das Verfahren zur Antworterzeugung im Berufsinfomat erfolgt im ersten Schritt durch die Analyse der Frage und die Wissensermittlung aus eigenen Quellen. Im zweiten Schritt erfolgt die Antworterstellung und die Ausgabe mit Hilfe generativer AI (OpenAI-API).

Diese beiden Schritte werden durch ein in den Berufsinfomat integriertes Redaktionssystem überwacht, welches in Echtzeit die Fragen wie auch die Antworten verwaltet. Hierbei wurden keine Trainingsdaten verwendet, da keine KI-Modelle im klassischen Sinne trainiert werden, sondern ein Prozess zur Wissensermittlung stattfindet. Die Wissensermittlung erfolgt ausschließlich durch die vom AMS bereitgestellten Quellen. Im Rahmen der KI-Methode findet ein Selektionsprozess für die Wissensdefinition des Antwortumfanges in den gelesenen Informationen des AMS statt. Nach diesem Schritt wird mit Hilfe der KI-Modelle von OpenAI (OpenAI-API) eine Antwort generiert. Das Wissen des KI-Modells wird nur zur Beantwortung der Frage mit einem definierten Antwortumfang verwendet. Sollte die Frage allgemeine Wissensbestandteile beinhalten, so werden dazu die Kompetenzen des KI-Modells von OpenAI (OpenAI-API) genutzt, um eine Antwort zu generieren.

Zur Frage 7

• Welche konkreten Daten wurden insbesondere mit Bezug auf die Arbeitsweise und bisherige Erfahrungen des AMS verwendet?

 Wie wurde bei AMS-bezogenen Daten sichergestellt, dass keine persönlichen Daten an nicht-legitimierte Dritte weitergegeben wurden bzw. werden?

Folgende Quellen werden vom AMS bereitgestellt:

- https://www.bis.ams.or.at
- https://www.berufslexikon.at
- https://www.ausbildungskompass.at
- und zusätzliche manuell, im Redaktionssystem erstellte Informationen, insbesondere ausgewählte Bereiche der AMS-Website und des Bundesministeriums für Bildung, Wissenschaft und Forschung.

Alle genannten Quellen wurden als strukturierte Information verarbeitet. Aus diesen strukturierten Informationen wurden im Rahmen der Anpassung des KI-Verfahrens dynamische Inhalte erzeugt, die weitere Beantwortungssituationen ermöglichen. Weitere Daten des AMS wurden nicht verwendet, da ausschließlich eine anonymisierte Nutzung vorausgesetzt wird und keine persönlichen Daten der Nutzerin bzw. des Nutzers benötigt werden. Es besteht kein Zugriff auf AMS-Systeme oder weitere AMS-Datenquellen.

Das eingesetzte KI-Verfahren wurde seitens des AMS geprüft; es wurden für die vorgefundenen Risiken adäquate Maßnahmen definiert, um die Anforderungen an die gesetzliche Aufgabenerfüllung bestmöglich mit dem Schutz der Betroffenen in Einklang zu bringen.

Zur Frage 8

• Ist es geplant, der KI des AMS-Berufsinfomat künftig weitere Daten als Trainingsdaten zur Verfügung zu stellen und wenn ja, welche?

Es werden keine Trainingsdaten verwendet, da keine KI-Modelle im klassischen Sinne trainiert werden, sondern ein Prozess zur Wissensermittlung stattfindet.

Die gesamten Informationen, die den AMS-spezifischen Wissensinhalt abdecken, werden laufend bearbeitet. Es werden eventuell weitere öffentlich zugängliche berufsrelevante Quellen integriert. Zusätzlich wird Content zur Verbesserung des Beantwortungsverhaltens ins Redaktionssystem eingepflegt.

Zur Frage 9

- Welche konkreten Anforderungen, Kriterien etc. wurden seitens des AMS bei der Auftragserteilung zur Erstellung des AMS-Berufsinfomats vorgegeben?
 - Wie genau gestaltete sich in diesem Zusammenhang insbesondere die Einbindung der Abteilung für Gleichstellung im AMS und hatte diese Kontrollund Feedbackmöglichkeiten?

Die konkreten Anforderungen waren, dass der Berufsinfomat, sowohl den Kundinnen und Kunden des AMS, als auch dessen Mitarbeiterinnen und Mitarbeitern Antworten zu berufsspezifischen Fragen liefert. Es war nie geplant, dass der Berufsinfomat auch Antworten zu rechtlichen Fragen (Arbeitslosengeldanspruch, Aufenthaltstitel, etc.) gibt, oder eine Beratung durch eine reale Person ersetzt. Wichtige Anforderungen waren außerdem:

Inhaltliche Kriterien:

- Auskünfte zu bestehenden Berufen und Berufsbildern
- o Auskünfte über Gehälter bestehender Berufe
- o Aufzeigen von Wegen zu Berufen und Ausbildungen
- O Unterstützung bei dem Finden von Berufsbildern
- o Anregung bei der Auswahl von Berufen oder Ausbildungen
- Hilfestellung bei der Fokussierung von Berufen und Ausbildungen in Zusammenhang mit der durch die Nutzerin bzw. den Nutzer dargestellten eigenen Situation
- o Genderspezifische- und Gleichberechtigungsaspekte

Kommunikationsaspekte:

- Unterschiedliche Antwortlängen und Link-Informationen nach unterschiedlichen Zielgruppen
- Unterschiedliche Ansprache der Zielgruppen durch die Formulierbarkeit der Antworten (formal/informal)

Die Abteilung "Arbeitsmarktpolitik für Frauen" in der Bundesgeschäftsstelle des AMS wurde im Dezember 2023 in der Pilotphase der Entwicklung des Berufsinfomaten einbezogen. Bei der Prüfung der genderspezifischen Auffälligkeiten betreffend die Beantwortung des Berufsinfomaten wurden durch die Abteilung Korrekturen an die zuständige Fachabteilung eingebracht. Die Betreiberfirma wurde darüber informiert und es wurde intensiv nach Lösungen gesucht und in die technische Programmierung eingearbeitet. Jedoch konnten bis zum Zeitpunkt der Veröffentlichung noch nicht alle Punkte zur Zufriedenheit umgesetzt werden.

Folgende inhaltliche Verbesserungen und Absicherungen im Berufsinfomat konnten aufgrund der Erfahrungen der Nutzung seit 03.01.2024 umgesetzt werden:

- 05.01.2024: Nacharbeiten Content allgemein
- 10.01.2024: Genderspezifische Erweiterungen mit KI-Verfahrenserweiterung und Missbrauchsprävention
- 19.01.2024: Verfeinerung der Sensitivität des Berufsinfomat
 - Funktionen f
 ür Inklusion
 - Verbesserungen bei Fragen die das persönliche Umfeld bezüglich Kinderbetreuung berücksichtigen
 - Verbesserung der Antwort bei Bekanntgabe der sexuellen Orientierung, Religionszugehörigkeit oder Nationalität

Zur Frage 12

• Wie wurde sichergestellt, dass der AMS-Berufsinfomat alle in Österreich geltenden Gesetze, insbesondere solche hinsichtlich der Gleichbehandlung und Antidiskriminierung am Arbeitsplatz, weitergibt und in seine Antworten inkludiert?

Das AMS steuert mit dem Redaktionssystem die größtmögliche Richtigkeit und korrigiert laufend erkannte Fehler, da alle als Fehler gemeldeten Rückmeldungen des Feedbacksystems laufend überprüft werden. Antworten können durch dieses gegebene Feedback der Anwenderinnen und Anwender nachgebessert oder verändert werden. Jeder entdeckte Fehler hilft somit bei der Verbesserung des Berufsinfomaten. Fehler können jedoch nicht gänzlich ausgeschlossen werden. Das System unterstützt den kontinuierlichen Verbesserungsprozess (KVP) für Content. Dieser KVP bezieht sich sowohl auf die Quellen der Wissensinformationen, wie auch auf das Verhalten in der gegebenen Antwort. Gesetzestexte selbst wurden nicht eingepflegt.

Zur Frage 13

• Wie und von wem wurde der AMS-Berufsinfomat vor der Veröffentlichung, insbesondere hinsichtlich eines möglichen Bias getestet und welche konkreten Schlüsse wurden aus dieser Testung gezogen?

Getestet wurde im PoC von ausgewählten Mitarbeiterinnen und Mitarbeitern der Berufsinformationszentren, seit 01.12.2023 von Mitarbeiterinnen und Mitarbeitern im AMS. Insbesondere waren seit Dezember 2023 auch die Frauenabteilung und die Datenschutzorganisation des AMS involviert und das Feedbacksystem wurde genutzt.

Im Rahmen der Tests vor Veröffentlichung wurden folgende Themengebiete erkannt und verbessert:

- Die Antwortsituation für die Antworten wurden auf formale und informale Antworten erweitert.
- Die Qualität der gegebenen Links wurde erweitert, um die passenden Informationen zu erkennen.
- Gender- und diversitätsspezifische Inhalte wurden grundlegend durch Contentmoderation im Redaktionssystem etabliert.
- Gewünschte Zusatzinformationen außerhalb der bestehenden Datenquellen wurden manuell ergänzt.
- Nicht bekannte Zusammenhänge von Begriffen wurden für die korrekte Antwortgabe ergänzt.
- Hinweise auf unvollständige Antworten wurden nachgearbeitet.
- Fragen außerhalb des Wissensgebietes wurden erlaubt.
- Einstreuen von Allgemeinwissen der KI (OpenAI-API) wurde erweitert, um Fragen außerhalb der Berufe beantworten zu können.
- Hinweise auf das Antwortverhalten in Bezug auf Vollständigkeit der Informationen wurden eingearbeitet.

Zur Frage 14

• Welchen konkreten Nutzen erfüllt der AMS-Berufsinfomat, den andere öffentlich zugängliche, textbasierte KI-Anwendungen nicht haben?

Durch den Einsatz des Berufsinfomaten wird nicht nur die Qualität der gegebenen Information gesteigert, da diese nun kontextabhängig gegeben werden kann, sondern es erfolgt auch eine Anreicherung mit vielen Berufsmöglichkeiten, ausgehend von den Fragen der Nutzerinnen und Nutzern. Das AMS schafft damit einen einfachen, kompakten Zugang zu komplexen berufs- und ausbildungsbedingten Entscheidungssituationen für die Nutzerinnen und Nutzer. Der Wissensrahmen des Berufsinfomaten wird ausschließlich durch das AMS und seine Entscheidungsträgerinnen und -träger mit dem Redaktionssystem gesteuert.

Die vom Berufsinfomaten genutzten AMS-Quellen werden täglich auf Aktualität geprüft. Geänderte Informationen werden täglich geladen. Damit ist der Berufsinfomat in Bezug auf Aktualität des AMS-Wissens immer am neuesten Informationsstand. Nur er kann diese Österreich-spezifischen Inhalte so aktuell und gezielt liefern. Alle anderen öffentlich, textbasierten KI-Anwendungen greifen aktuell entweder auf veraltete Informationen zurück

oder verfügen für die Ergebnisse nur über eine Auswahl von Informationen des gesamten Berufsinfomaten.

Zur Frage 15

- Zu welchen Adaptierungen ist es seit der Veröffentlichung des AMS-Berufsinfomats gekommen, um möglichen Bias und die Reproduktion von Vorurteilen, beispielsweise hinsichtlich des Geschlechts oder der sexuellen Orientierung, auszuräumen?
 - Zu welchen konkreten Adaptierungen ist es wann beim Prompt der Kl-Anwendung gekommen?

Seit der Veröffentlichung wurden die eingehenden Fragen genau analysiert, und die Themen der Nachschärfung wurden einvernehmlich mit dem zuständigen Fachbereich des AMS festgelegt. Folgende Maßnahmen zur Adaptierungen wurden gesetzt:

- Berufsinfomat Web-Client
- Nachbesserungen hinsichtlich der Bias-Situation
- Nachbesserungen im Prompt

Der Web-Client wurde mit Einsatzzeitpunkt 03.01.2024 offen und ohne Source-Verschleierung eingesetzt, um eine möglichst transparente Einsicht in die Oberflächen-Vorgänge des Webclients zu liefern. Eine direkte Nutzung von Schnittstellen zu ChatGPT war und ist per Design nicht möglich. Die Versuche der Preisgabe von Prompts oder Programmzeilen zeigten nur nicht betriebsrelevante Inhalte.

Nach den ersten Medieninformationen wurde in weiterer Folge beschlossen, mit 05.01.2024 den Sourcecode des WebClients technisch gegen Einsichtnahme so weit wie möglich zu schützen und ein Erreichen der Serverschnittstelle (zur Antwortgenerierung des AMS-Berufsinfomat) zu erschweren. Gleichzeitig konnten die Anfragen verwendet werden, um inhaltliche Nachbesserungen zu etablieren.

Folgende Bias wurden beim Berufsinfomaten bis 19.01.2024 aufgenommen und reduziert: Gender-Bias reduziert bei Anfragen/Content und Antwortgabe, Hinweise auf Nationalität und ethnische Herkunft, sexuelle Orientierung und Geschlechtsidentität, Kompetenzen auf eigene Religionszugehörigkeit, Behindertenarbeit bei eigener Behinderung, Herkunftsland und Kompetenzen sowie Kindererziehung bei eigenen Kindern.

Grundsätzlich ist festzuhalten, dass die Bias-Thematik sehr aufmerksam beobachtet und versucht wird, im Rahmen eines Bewertungsprozesses immer wieder Hinweise auf neue

Situationen zu erkennen und dies durch das KI-Verfahren des Berufsinfomaten abzudecken.

Folgendes Verfahren (am Beispiel genderspezifischer Content) wurde grundlegend zur Reduktion des Bias angewendet:

1. Vorbereitung

- Extraktion der geschlechtsspezifischen Fragen aus den ca. 140.000 bisherigen Fragen (Stand 20.01.2024) anhand von Keywords. Bei diesen ca. 5.000 Fragen bestand die Möglichkeit, dass ein Gender-Bias in der Antworterstellung bestehen könnte.
- Festlegen eines umfassenden Testsets von etwa 1.000 Fragen.
- Entwicklung eines KI-basierten "Gender-Bias-Indikators" für die Kundinnen- und Kundenfrage. Der Indikator zeigt an, ob "Anti-Gender-Bias-Maßnahmen nötig sind". Ein Indikator wurde in einem iterativen Prozess mit laufenden Tests und Optimierungen mit 1.000 Testfragen entwickelt.
- Entwicklung einer Funktion zur Transformation der Kundinnen- und Kundenfrage in eine Bias-freie Form.

2. Optimierung des Verfahrensablaufes

- Einführung einer Bias-freien internen Korpussuche (mittels "transformierter Kundinnen- und Kundenfrage").
- Steuerung der Antwortgenerierung durch weitere Parameter zur Sensibilisierung auf die Problematik (Prompting).

3. Qualitätsprüfung

- Regressionstests mit 100 bis 200 Zufallsfragen aus dem Gender-Testset (True-Positiv-Tests).
- Regressionstests mit 100 bis 200 Zufallsfragen aus allen Fragen (False-Positiv-Tests).
- Die Ergebnisse der Contentverbesserungen wurden durch den Fachbereich des AMS freigegeben.

Eine alleinige Adaptierung des Prompts ist sicherlich nicht ausreichend, um die unterschiedlichen Aspekte des Bias abdecken zu können. Daher hat man zu einem serverseitigen Verfahrensansatz gegriffen, der es ermöglicht, sowohl die gestellte Frage, als auch die dazu bereitgestellte Information wie auch die Antwort einer generativen Al in Bezug auf den vermuteten Bias zu normieren und daraus eine möglichst Bias freie Situation zu erzeugen, die zu einer möglichst Bias freien Antwort führt.

Zur Frage 16

- Welche weiteren Adaptierungen sind in Zukunft beim AMS-Berufsinfomat geplant, um möglichen Bias und die Reproduktion von Vorurteilen, beispielsweise hinsichtlich des Geschlechts oder der sexuellen Orientierung, auszuräumen?
 - Zu welchen konkreten Adaptierungen ist es wenn beim Prompt der KI-Anwendung gekommen?

Die in der Frage aufgeführten Bias-Themen, eine Reproduktion von Vorurteilen beispielsweise hinsichtlich des Geschlechts oder der sexuellen Orientierung, wurden aktuell reduziert. Die eingeführten Methoden und die Schärfung der grundlegenden Sensibilität in Verbindung mit einer laufenden Überwachung der Ergebnisse führen zu einer immer weiteren Verbesserung der Ergebnisse. Das vorhandene Redaktionssystem erzielt laufende Verbesserungen. Der Anpassungsaufwand wird mit fortlaufender Erfahrung immer geringer.

Univ.-Prof. Dr. Martin Kocher

Elektronisch gefertigt