DIALOG

DEZEMBER 2023 | AUSGABE 49

HOCHSCHULE FÜR
ÖFFENTLICHE VERWALTUNG UND FINANZEN
LUDWIGSBURG
UNIVERSITY OF APPLIED SCIENCES

Zum Verhältnis von BürgermeisterInnen und Parteien in Baden-Württemberg

Zu Gast bei der Verwaltung der Zukunft - Innenminister Strobl besucht HVF

Erfolgreicher Abschluss des ersten Jahrgangs des Studiengangs Digitales Verwaltungsmanagement

Ideenwerkstatt von HVF und Städtetag zur Fachkräftesicherung

Kooperatives Promotionskolleg „Good Administration“ nimmt Fahrt auf

Chatbots - eine Hilfe bei der Fragebogenkonstruktion?

Andreas Hammer

Lehrbeauftragter

In einer Studie des Autors sollten wichtige Werte in der Interaktion zwischen SGB II-Leistungsberechtigten und Fachkräften identifiziert werden. Dabei wurde untersucht, ob ein Chatbot eine Hilfe bei der Fragebogenkonstruktion sein kann.

Ansatz

Zuerst wurden Fachkräfte gebeten, Begriffe zu nennen, die sie mit Werthaltungen verbinden. Diese wurden dann von mehreren Personen (Rater) unabhängig voneinander kodiert. Bei einer Übereinstimmung von > 60% wurden die so gefundenen Werthaltungen in einen Fragebogen übernommen. Jeder Wert wurde abschließend von Fachkräften mit dem Grad der Wichtigkeit bewertet. Die Ergebnisse wurden anschließend faktorenanalytisch ausgewertet, um die verschiedenen Dimensionen von Werthaltungen zu beschreiben.

Parallel dazu wurde ein Chatbot eingesetzt, um ebenfalls Werthaltungen zu finden, die in der Interaktion zwischen Leistungsberechtigten und Fachkräften eine Rolle spielen. Er basiert auf dem Modell chatGPT 3.5. Es wurde davon ausgegangen, dass ein mit Wörtern trainierter Chatbot Werthaltungen benennen kann und dies für die Fragebogenerstellung mit weniger Aufwand verbunden und schneller sein wird als mit Menschen.

Experiment mit dem Chatbot

Der Chatbot wurde aufgefordert, 100 Werte zu der oben genannten Frage aufzulisten. Ergebnis: 100 Begriffe wurden aufgelistet. Darunter waren viele keine Werte. Beispielsweise wurde Empathie als Wert aufgelistet, aber auch Empathiefähigkeit, die eine Kompetenz ist. Es wurden Begriffe wie „Arbeitslose“ aufgelistet, die zu einer nicht passenden Wortkategorie gehören. Außerdem wurden einige Begriffe doppelt genannt, so dass es nicht 100 verschiedene Begriffe gab.

Die Abfrage wurde dann mit 200 und 500 Begriffen wiederholt. Die gelisteten Antworten waren noch weiter von der Fragestellung entfernt. Es wurden vor allem Komposita gebildet, wie z. B. Arbeitsvermittlung, Arbeitsvermittlungsgesetz etc. Der Begriff Arbeitsvermittlung, der für sich genommen keine Werthaltung ausdrückt, wurde 66 mal mit einem anderen Substantiv kombiniert. Dies führte zu einer noch größeren Anzahl unbrauchbarer Begriffe.

Daraufhin wurden weitere Abfragestrategien angewandt, die jedoch jeweils zu einem ähnlich Ergebnis wie anfangs führten:

- vereinfachte Abfragen
- Abfrage nach Untergruppen: z.B. nach berufsbezogenen Werten
- mit Definition von Begriffen wie Werte, Kompetenzen etc.
- Aufforderung im Prompt, eine Liste von Begriffen, die Fachkräfte gesammelt haben, nach Werten, Kompetenzen etc. zu kategorisieren.

Die Rater hatten aus einer Liste von Begriffen 64 als Werte kategorisiert.

Davon hat der Chatbot 39 % genannt. Es gab keinen Begriff, den der Chatbot gefunden hatte, die ExpertInnen aber nicht.

Fazit

Wäre nur der Chatbot zum Einsatz gekommen, hätten die Fachkräfte in der Befragung 25 Items nach ihrer Wichtigkeit bewerten können. Diese Anzahl wäre zu gering, um ein gutes Faktoren-Modell zu finden und ein Ausschlussgrund für die Verwendung bei der Fragebogenkonstruktion. Insgesamt ist der Aufwand für die Überprüfung und das Aussortieren offensichtlich ungeeigneter Items nicht zu unterschätzen, da die KI-Modelle hier noch nicht gut genug sind. Der bisherige Standard von unabhängigen Ratern, die ein Mindestmaß an Übereinstimmung aufweisen, sollte weiterhin eingehalten werden.

Die Chatbot-Ergebnisse im beschriebenen Beispiel sind für wissenschaftliche Arbeiten nicht geeignet. Für einfachere Fragestellungen mit gut definierten Begriffen könnte ein Chatbot bei der Fragebogenkonstruktion nützlich sein. Es ist hilfreich, wenn nicht sogar unerlässlich, sich in dem Fachgebiet gut genug auszukennen, um falsche oder „erfundene“ Antworten des Chatbots zu erkennen.

Literatur: https://kurzelinks.de/Hammer-Studie_Haltung_im_Jobcenter