

Bürgerassistenz

Effektive Unterstützung

[25.07.2024] Als Bürgerassistenz bleiben die typischen, regelbasierten KI-Bots hinter den Erwartungen zurück. Das Fraunhofer-Institut FOKUS hat untersucht, wie diese besser werden können – etwa indem Online-Dienste und Formulare mit Sprach-Schnittstellen erweitert werden.

KI-Lösungen gewinnen auch in der öffentlichen Verwaltung immer mehr an Bedeutung. So können Machine-Learning-Verfahren den Posteingang einer Behörde vorsortieren und ihn verschiedenen Dienststellen zuweisen. Generative Sprachmodelle – auch Large Language Models (LLM) genannt – sind durch ChatGPT populär geworden. Solche LLMs können nicht nur Texte erzeugen, sondern sie auch analysieren. Sie können Eingaben deutlich besser verstehen als frühere KI-Technologien. In Gestalt von Retrieval-Augmented-Generation-Chatbots (RAG-Chatbots) erlauben generative KI-Modelle hochqualitative Suchverfahren für die öffentliche Verwaltung, indem interne Datenbestände nutzerfreundlich und zielgenau im Frage-Antwort-Modus erschlossen werden. In der Verwaltung können sie dazu beitragen, Mitarbeitende von repetitiven Aufgaben zu entlasten und die digitale Kommunikation an der Bürgerschnittstelle zu unterstützen.

Neue, nutzerfreundliche Kommunikation

Infolge von Kostenzwängen und dem Fachkräftemangel steht bei Diskussionen einer möglichen KI-Anwendung die Unterstützung interner Fachprozesse im Vordergrund. Moderne KI-Verfahren können aber noch mehr: Sie eröffnen neue, nutzerfreundliche Möglichkeiten der Kommunikation mit Bürgerinnen und Bürgern und tragen dazu bei, Barrieren zu reduzieren. Bisher ist die Popularität von Chatbots an der Bürgerschnittstelle eher begrenzt. Einerseits hat der ChatGPT-Hype hohe Nutzererwartungen geweckt, die typische Verwaltungsbots oft nicht erfüllen können, da sie auf älteren Suchtechnologien basieren. Andererseits sind diese Systeme bisher meist auf die reine Informationsbeschaffung limitiert. Um einen digitalen Antrag zu stellen, muss man den Chat verlassen und – ohne Bot-Hilfe – den Weg über ein (Online-)Formular gehen.

Kommunikationsbandbreite erhalten

Formulare sind und bleiben unverzichtbar. Sie stellen genau durchdachte, formalisierte Schnittstellenspezifikationen zwischen Bürgerinnen und Bürgern sowie der Behörde dar. KI-Assistenzlösungen sollten idealerweise eine leistungsfähige Unterstützung beim Ausfüllen bieten. Dazu gehört es, Zwischenfragen zu Details eines Formulars entgegenzunehmen und zu beantworten. Dabei sollte das ausgefüllte Formular jederzeit sicht- und editierbar bleiben. Eine Interaktion per Audio oder ausschließlich im Chat-Fenster schränkt die Kommunikationsbandbreite aber ein, sodass wichtige Details übersehen werden könnten. Da durch die Formularabsendung oft Rechtsakte begründet werden, gilt es, dies zu vermeiden. Am Fraunhofer-Institut FOKUS wurde in einer Reihe von Projekten für die öffentliche Verwaltung untersucht, wie Online-Dienste und formularzentrische Web-Anwendungen effektiv und minimalinvasiv mit Chat- und Sprachschnittstellen erweitert werden können. So wurde in einem Pilotvorhaben mit der Freien und Hansestadt Hamburg und Dataport der dort betriebene Online-Dienst „Kinderleicht zum Kindergeld“ mit einem Sprachassistenten erweitert, der es ermöglicht, das komplexe Online-Formular vollständig sprachgesteuert auszufüllen. In Folgeprojekten wurde die Vorgehensweise auf

andere Anwendungen übertragen und erweitert. Ein weiteres Beispiel ist das EU-Vorhaben ACROSS. Hier entstand ein multilingualer, multimodaler Bürgerassistent. Er soll Bürgerinnen und Bürger etwa dabei unterstützen, die formalen Vorbereitungen für einen längeren Auslandsaufenthalt abzuwickeln.

Chatbot beantwortet Hintergrundfragen

Der Assistent ermöglicht die Navigation einer zentralen Citizen Web App und das Ausfüllen von Eingabefeldern wahlweise per getipptem Chat oder per Spracheingabe. So demonstriert das Projekt erstmalig einen nahtlosen Wechsel zwischen dem Web-Assist-Modus zum Steuern, Navigieren und Ausfüllen und einem separaten Info-Chatbot, der Hintergrundfragen beantwortet. Zudem wurde eine erste Integration mit einem generativen KI-Modell realisiert, um das Textverstehen des Assistenten durch Natural Language Understanding (NLU) zu optimieren. Parallel hat Fraunhofer FOKUS den „FOKUS Intelligent Speech Assistant“ aufgebaut, das so genannte FISA-Framework. Dieses unterstützt eine einfache Erweiterung formularzentrischer Web-Anwendungen um konversationale KI-Assistenzfunktionen. In die neueste Version sind Open-Source-LLMs integriert. Diese Erfahrungen zeigen: Ein Full-Service-KI-Bürgerassistent ist denkbar. Dieser könnte einen modernen RAG-Info-Chatbot mit einem LLM-gestützten Formularassistenten kombinieren und so nahtlose Wechsel zwischen Unterstützungsformen und den Interaktionsmodi – Chat, Sprache sowie Web-UI-Nutzung – unterstützen und zudem mehrsprachig arbeiten. Die Vorteile: Informationen werden zielgenau gefunden, Nutzereingaben und -intentionen werden sowohl bei der Informationssuche als auch bei der eigentlichen Dienstnutzung besser verstanden, wodurch das Nutzerfrustrationsrisiko sinkt.

Sachlich falsche Inhalte vermeiden

Generative KI-Modelle bringen aber auch Risiken mit sich – vor allem das so genannte Halluzinieren, also die Generierung sachlich falscher Textinhalte. Es gibt bereits Forschungen, die zumindest mittelfristig Abhilfe versprechen. Im hier diskutierten Kontext gilt es, die Risiken angemessen zu mitigieren. Unproblematisch ist die NLU-Anwendung generativer KI-Modelle, da der erzeugte Text überwiegend systemintern verarbeitet wird. Beim Ausfüllen von Formularinhalten können Halluzinationen leicht bemerkt und korrigiert werden, da die Nutzer ausgefüllte Formulare einer Endkontrolle unterziehen. Größer sind die Herausforderungen im RAG-basierten Infochat. Hier könnte ein Lösungsansatz sein, als Suchergebnisse keine generierten Texte, sondern nur Ausschnitte aus Originaldokumenten auszugeben, etwa Teile von Behörden-Websites. In diesem Fall unterstützt das generative KI-Modell dann nur hinter den Kulissen Auswahl und Ranking ausgegebener Inhalte. Um die Anforderungen der Verwaltung hinsichtlich der Datensouveränität zu erfüllen, sollten generative KI-Modelle quelloffen sein und zudem einen On-Premises- oder Private-Cloud-Betrieb gestatten. Aktuelle Lösungen, die diese Bedingungen erfüllen, beschränken sich in der Regel ausschließlich auf die Textprozessierung. Ist die Verarbeitung gesprochener Sprache gewünscht, müssen Spracherkennungs- und Sprachsynthese-Komponenten integriert werden. Hinzu kommen in einigen Fällen Komponenten zur maschinellen Übersetzung.

Zukunft mit multimodalen Large Language Models

Auch für diese Teilaufgaben existieren bereits brauchbare Open-Source-Lösungen. In Zukunft ist jedoch auch mit multimodalen LLMs zu rechnen, die alle Teilaufgaben in einem einzigen Modell lösen und dadurch nur noch wenig Rechenzeit benötigen. Durch zügig gelieferte Antworten erscheinen sie für den Einsatz in interaktiven Assistenzlösungen besonders attraktiv. Ein solcher Full-Service-Bürgerassistent wird idealerweise das mit traditionellen Chatbots assoziierte Frustrationsrisiko durch besseres Textverständnis generativer KI-Modelle umschiffen und wird in der Lage sein, zwischen Informations- und

Unterstützungsrolle zu wechseln. Die multilinguale Funktion und ein nahtloser Moduswechsel zwischen Sprach-, Chat- und klassischer Interaktion tragen dazu bei, Sprachbarrieren, aber auch sonstige Barrieren zu reduzieren.

()

Dieser Beitrag ist in der Ausgabe Juli 2024 von Kommune21 im Schwerpunkt Künstliche Intelligenz erschienen. Hier können Sie ein Exemplar bestellen oder die Zeitschrift abonnieren.

Stichwörter: IT-Infrastruktur, künstliche Intelligenz