BERICHTE DES NEGZ NR. 9

SPRACHSTEUERUNG VON E-GOVERNMENT DIENSTEN IN DEUTSCHLAND

Stefan Schaffer
Norbert Reithinger
Janos Standt
Roland Krebs

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

mgm

NEGZ
Nationales E-Government Kompetenzzentrum e. V.

Für einen modernen Staat

Das Nationale E-Government Kompetenzzentrum vernetzt Experten aus Politik, Verwaltung, Wissenschaft und Wirtschaft und ist die zentrale, unabhängige Plattform für Staatsmodernisierung und Verwaltungstransformation in Deutschland.

Herausgegeben und gefördert vom
Nationalen E-Government Kompetenzzentrum e. V.
Berlin 2020

INHALT

Zusammenfassende Empfehlungen **4**

1. Einleitung **5**

2. Wissenschaftlicher und praktischer Hintergrund **7**

2.1 Bestehende Konzepte 7

2.2 Aufbau einer Anwendung mit Sprachsteuerung 8

2.3 Akteure im Vergleich 12

2.4 Bestehende E-Government Anwendungen im Steuerbereich 13

2.5 Bestehende E-Government Anwendungen im Servicebereich 13

3. Ergebnisse: Szenarien für Sprachinteraktion in behördlichen Prozessen **16**

3.1 Fokus Steuerumfeld 16

3.1.1 Ablauf einer Steuererklärung 16

3.1.2 Einsatzszenarien für Sprachinteraktion bei der Steuererklärung 17

3.2 OZG Umsetzungskatalog 19

3.2.1 Meta-Schema für sprachgestützte Dienste für antragsbasierte Leistungen 19

3.2.2 Einsatzszenarien für Sprachinteraktion im Rahmen des OZG Umsetzungskatalogs 20

3.3 Anforderungen und Herausforderungen 22

4. Handlungsempfehlungen **23**

5. Zusammenfassung **25**

Literatur **26**

Über die Autoren **29**

Impressum **30**

ZUSAMMENFASSENDE EMPFEHLUNGEN

1. Für sprachbasierte E-Government Dienste werden Komponenten für Dialogmanagement und Businesslogik benötigt. Die Businesslogik enthält das Domänenwissen, also die Informationen und Services, die durch einen bestimmten E-Government Dienst bereitgestellt werden. Das Dialogmanagement steuert den Dialogablauf mit der Benutzerin. Beide Komponenten sind eng miteinander verknüpft.

2. Sprachbasierte E-Government Dienste benötigen eine On-Premises Lösung. Sprachdaten sowie die aus den Sprachdaten extrahierten Informationen sind vertraulich und müssen unter der Kontrolle des Betreibers bleiben. Nutzerdaten sollten nicht auf anderen Servern gespeichert oder verarbeitet werden, um die Datensicherheit zu gewährleisten.

3. Bei der Einführung von Sprachbasierten E-Government Diensten sollte mit einfachen Leistungen begonnen werden. Die Statusabfrage der Steuererklärung oder einfache verwandte Anträge wie eine Fristverlängerung sind vergleichbar einfach umzusetzen.

4. Mittelfristig sollten sprachbasierte E-Government Dienste eingeführt werden, die einen Interviewmodus zur Vorbereitung größerer Anträge bereitstellen. Informationen werden hierbei eingangs in einem sprachlichen Dialog von der Benutzerin eingegeben und automatisch in das Antragsformular übertragen.

5. Langfristiges Ziel sollte es sein, für komplexe Anträge wie die Steuererklärung ein kontextbezogenes sprachbasiertes Ausfüllen zu ermöglichen. Eine wichtige Anforderung wäre hierbei, bei Fragen zu steuerlichen Begriffen und bei Unklarheiten im Zusammenhang mit gerade bearbeiteten Abschnitten, eine direkte Antwortmöglichkeit im Formular zu bieten.

6. Wir empfehlen eine intensive Beobachtung der internationalen Entwicklungen, um von Erfahrungen anderer Länder zu lernen und um einer Harmonisierung der technischen Prozesse sowie der behördlichen Verfahren entgegenzuarbeiten.

Schlagworte: Sprachsteuerung, Chatbot, E-Government, Steuer

1. EINLEITUNG

Sprachsteuerung von Diensten und Anwendungen hat sich in den letzten Jahren als Zukunftsthema etabliert. Es ist absehbar, dass sich Sprache mehr und mehr als ein Standard Interaktionskanal durchsetzen wird.[1] Entsprechend werden Benutzer von Informations- und Kommunikationstechnologie (IKT) ihre Gewohnheiten und Erwartungen anpassen. Gängige Interaktionsformate, z.B. zur Benutzung von Webseiten oder Apps, müssen verändert werden, um den neuen Anforderungen der Nutzer zu entsprechen. Langfristig werden durch diese Entwicklung Anwendungen ohne Sprachsteuerung immer mehr in den Hintergrund gedrängt werden. Kommunikationskanäle wie Chatbots ermöglichen Echtzeit-Service und den Zugang zu relevanten Informationen rund um die Uhr und an jedem Ort.[2] Dabei sind diese Systeme in der Lage Nutzereingaben sowohl in geschriebener als auch in gesprochener Sprache zu verarbeiten.

Der Gartner Hype Cycle (Abbildung 1) illustriert, dass aktuell sehr hohe Erwartungen an KI-basierte Innovationen wie Chatbots und Conversational User Interfaces gesetzt werden. Die Erwartungen bezüglich Natural Language Processing (NLP) Technologien befinden sich aktuell an der Grenze zur Desillusionierung, d.h. die realistischen Einsatzbereiche kristallisieren sich derzeit heraus. Spracherkennung hat das Plateau der Produktivität bereits erreicht und kommt kommerziell zum Einsatz. Insgesamt deutet die Entwicklung dieser Technologien an, dass jetzt mit anwendungsnaher Forschung und Entwicklung zum Einsatz von Chatbots und Conversational User Interfaces für geschriebene und gesprochene Sprache begonnen werden muss, um diese in wenigen Jahren einsatzbereit zu haben.

Auch im öffentlichen Bereich bietet die Nutzung von Sprachsteuerung vielfältige Vorteile. So könnte z.B. ein direkter Zugriff auf bestimmte Informationen ermöglicht und formularbasierte Prozesse wie die elektronische Steuererklärung durch einen sprachlichen Dialog unterstützt oder sogar komplett abgewickelt werden. Nicht zuletzt fördert Sprachsteuerung den barrierefreien Zugang zu staatlichen Diensten.

Die Studie soll initial aufzeigen, welche Infrastruktur bereitstehen müsste, damit E-Government Dienste in Deutschland sprachgesteuert angeboten werden können. Daher wird unter anderem ein Überblick über notwendige Komponenten für sprachbasierte Interaktion gegeben und eine Zusammenstellung möglicher internationaler und deutscher Akteure vorgestellt. Insgesamt soll ein Verständnis dafür entwickelt werden, welche Anforderungen für die Einführung sprachbasierter Interaktion im Bereich des E-Government erfüllt werden müssten.

Die Fragestellungen und Struktur der Studie gliedern sich folgendermaßen:

- Welche Infrastruktur müsste in Deutschland für sprachgesteuerte E-Government Dienste bereitstehen? Was sind die aktuell umgesetzten Anwendungsfälle? Hierzu werden in Kapitel 2 Grundlagen aufgezeigt und ein Überblick gegeben.

1 https://www.heise.de/newsticker/meldung/Fast-jeder-Dritte-in-Deutschland-nutzt-Sprachassistenten-4443365.html

2 https://www.service.express/chatbot-studie-kundenservice-fertigungsindustrie/

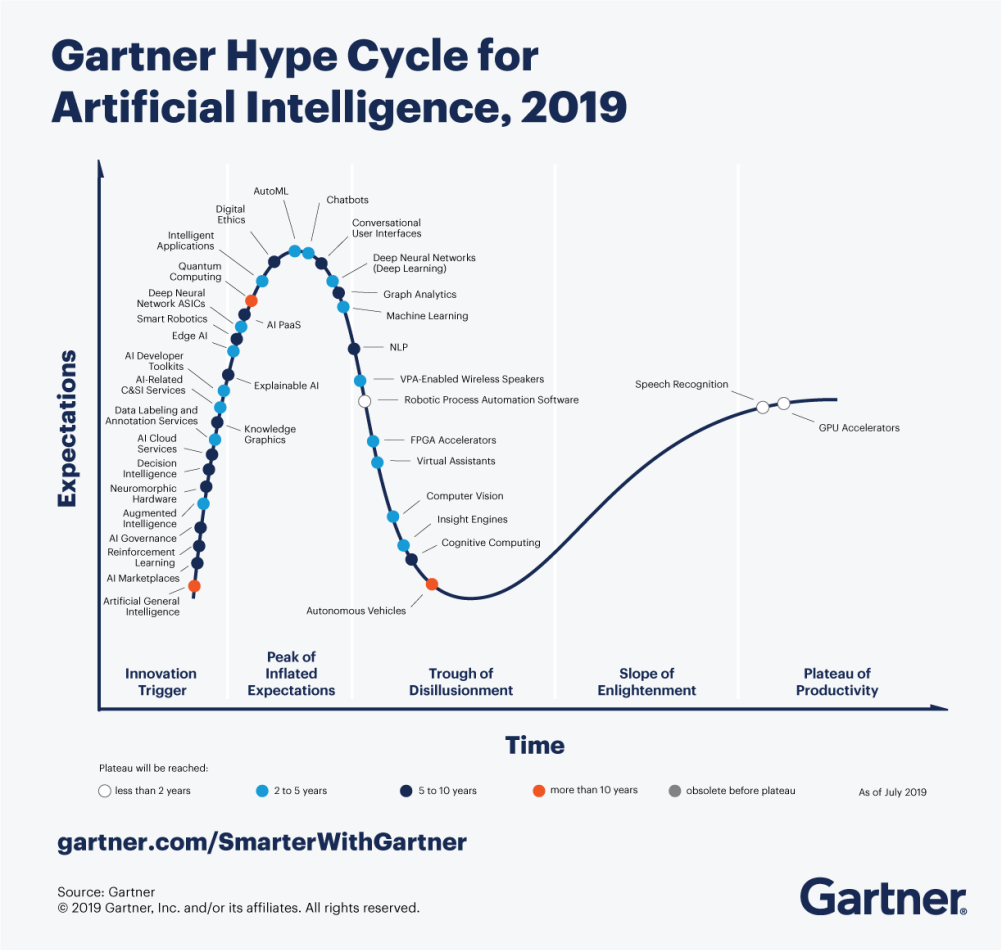

Abbildung 1: Chatbots und Conversational User Interfaces an der Spitze des Hype Cycles. Spracherkennung hat das Plateau des produktiven Einsatzes bereits erreicht.

- Welche Szenarien im Bereich behördlicher Prozesse könnten in Zukunft durch Sprachinteraktion unterstützt werden? Was sind Herausforderungen für Sprachinteraktion im Bereich behördlicher Prozesse und welche Anforderungen müssten erfüllt sein? Diese Fragen werden in Kapitel 3 behandelt.
- Als abschließendes Ergebnis werden in Kapitel 4 Handlungsempfehlungen formuliert, welche Grundanforderungen angegangen werden müssen, damit E-Government Dienste in Deutschland sprachgesteuert angeboten werden können.

2. WISSENSCHAFTLICHER UND PRAKTISCHER HINTERGRUND

2.1 Bestehende Konzepte

Die Bedienung und Interaktion von Geräten und Dienstleistungen per Sprache ist seit den 1990er Jahren etabliert. So wurde z.B. bereits 1996 von Mercedes die erste sprachbedienbare Navigation im Auto vorgestellt,[3] Mobilfunkgeräte sind ebenfalls bereits um die Jahrtausendwende sprachbedienbar.[4] Telefonbasierte Sprachdialogsysteme (Interactive Voice Response IVR), die per Kommando oder Tastendruck bedient werden können, sind seit langem im Markt und werden täglich millionenfach angerufen.

Vor dem Hintergrund einer Vielzahl von nationalen und internationalen Forschungsprojekten, besonderes in den 2000er Jahren, etablierten sich in den letzten Jahren Sprachdialogsysteme wie Alexa, Siri, Google Assistant, Cortana und andere. Diese Systeme sind jeweils von einem Anbieter erhältlich, die gesamte Logik liegt in der Cloud. Die Anbieter von Skills o.ä., d.h. der Anwendungen, die per Sprache bedient werden können, müssen neben der Dialoglogik auch ihre Businesslogik den Anbietersystemen zugänglich machen. Die Sprachdaten werden immer von den Endgeräten wie Smartphones oder intelligenten Lautsprechern zu diesen Drittsystemen übertragen und dort verarbeitet.

Ebenfalls aufsetzend auf die Systeme der großen Anbieter wie Facebook (inkl. Instagram und WhatsApp) und auf Webseiten entstanden in den letzten Jahren konversationelle Interfaces, sog. Chatbots, die primär textbasiert mit den Benutzern der Systeme und Dienstleistungen interagieren. Auch hier gilt i.A., dass die Logik und damit die Datenherrschaft auf den Systemen der Anbieter läuft.

Eher Nischencharakter für die Allgemeinheit haben Produkte wie der Einsatz von Sprachsteuerung in Krankenhäusern[5] oder Diktate im juristischen Umfeld,[6] wo sie aber weit verbreitet sind, insbesondere für die automatische Verschriftlichung von Diktaten. U.a. der Anbieter Nuance bietet auch für weitere Bereiche Lösungen an, dezidiert auch für Behörden.[7] Inwieweit auch in Deutschland bereits Systeme eingeführt wurden, konnte im Rahmen der Studie nicht verifiziert werden.

Wichtig bei allen diesen Systemen ist, dass der Dialogfluss gezielt auf die Sprachinteraktion abgestimmt sein muss. Die Interaktion, ob mit Sprache oder Text, erlaubt den Menschen, Informationen in ihrer Ausdrucksweise und ggfs. mit Nachfragen oder ihrer Wichtigkeit zu geben. Im Gegensatz dazu sind Webseiten vom Webdesigner streng vorgegeben und die Benutzerinnen und Benutzer müssen sich

3 https://blog.mercedes-benz-passion.com/2016/02/premiere-vor-25-jahren-die-mercedes-benz-s-klasse-der-baureihe-w-140/

4 https://www.teltarif.de/arch/2000/kw39/s3124.html

5 https://www.nuance.com/de-de/healthcare.html

6 https://www.nuance.com/de-de/dragon/business-solutions/dragon-legal-group.html

7 https://www.nuance.com/de-de/industries/government.html

E-Government Anforderungen		Sprachassistenz
Einfachheit und Effizienz	✓	Sprache ist für den Menschen die natürlichste Interaktionsform
Sofortiger Zugriff auf Information	✓	Einfach “sagen was man will”
Verbesserte Reichweite der Verwaltung	✓	Ermöglichung neuer barrierefreier Interaktionskanäle
Verbesserter Bürgerservice	✓	Ein Formular, dem man Fragen stellen kann
Kosteneffizienz	✓	24/7 Verfügbarkeit von Sprachassistenten
Verbesserte Sicherheit	✗	Eignung von Sprache ist kontextabhängig
Transparenz und Rechenschaftspflicht	✗	Bisher keine rechtsverbindliche Unterschrift per Sprache
Einfache Wartbarkeit	✗	Infrastruktur für E-Government Sprachinteraktionsdienste benötigt

Tabelle 1: Zusammenstellung von Anforderungen an E-Government Dienste und aktuellen Möglichkeiten in Bezug auf Sprachassistenz. Die Spalte in der Mitte kennzeichnet mit einem Haken, welche Möglichkeiten bereits aktuell bestehen und mit einem Kreuz, welche Möglichkeiten derzeit noch eine Vision darstellen.

an den vorgegebenen Seitenfluss anpassen, und an die Terminologie, die in den Webseiten verwendet wird. Nachfragen oder Abweichungen sind i.A. nicht vorgesehen. Der Designer einer Sprachdialoganwendung muss somit mit sehr viel mehr Freiheitsgraden umgehen können, wenn der Dialog näherungsweise natürlich sein soll.

Für den Bereich des E-Government sind in Tabelle 1 Visionen und aktuelle Möglichkeiten zusammengefasst.

2.2 Aufbau einer Anwendung mit Sprachsteuerung

Bei der Darstellung folgen wir den eigenen langjährigen Erfahrungen in der Forschung im Bereich Dialogsystemen und Chatbots. Eine sehr gute Einführung ist

> Daniel Jurafsky and James H. Martin. 2009. *Speech and Language Processing (2nd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.[8]

Generell sind die Abläufe so wie nachfolgend beschrieben. In der aktuellen Forschung und teilweise bereits in Systemen wie Alexa werden in vielen Komponenten, wie schon immer in Spracherkennern, Klassifikatoren verwendet. Zur Realisierung solcher Klassifikatoren gibt es eine Vielzahl von Ansätzen. Die bekanntesten sind Bayessche Klassifikatoren, Support Vektor Machines oder künstliche neuronale Netze, wie z.B. Deep Neural Networks.

8 Eine aktuelle dritte Version wird gerade erstellt. Sie steht als Entwurf unter https://web.stanford.edu/~jurafsky/slp3/ zum Download bereit.

Eine Übersicht zu diesen Methoden findet sich z.B. im Standardwerk

> Stuart J. Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach, 4th Edition*. Pearson Education Limited, 2020.

Viele Methoden basieren auf vorhandenen Daten mit Bezug zum Thema, die annotiert und die Voraussetzung für Methoden des überwachten Lernens sind. Im Bereich der Spracherkennung und Sprachverarbeitung sind dies insbesondere sog. Korpusdaten, oder Korpora. In diesen sind Sprachsignale, Texte und Dialogdaten, deren Transkription und weitere Merkmale annotiert, teilweise per Hand, teilweise automatisch. Anhand dieser Daten werden die Klassifikatoren trainiert.

Begriffserläuterung

annotieren	Annotieren im klassischen Sinne bedeutet eine Textstelle zur besseren Erklärung oder Ergänzung mit einer Anmerkung zu versehen. Textdaten werden z.B. annotiert indem Meta-Daten hinzugefügt werden, die den Inhalt bzw. die Bedeutung des Textes in maschinen-verarbeitbarer Form beschreiben.
trainieren	Daten und ihre Annotationen werden z.B. beim Training eines neuronalen Netzes verwendet. Dabei werden verschiedene Texte mit gleichen Annotationen der gleichen Bedeutung zugeordnet. Durch das Training entsteht ein statistisches Modell, das die Bedeutung der annotierten Daten gelernt hat.
klassifizieren	Ein Klassifikator ordnet Objekte in Kategorien ein. Ein durch Training mit annotierten Daten entstandenes statistisches Modell ist ein solcher Klassifikator. Als Eingabe erhält ein Text-Klassifikator nicht annotierte Texte und ordnet diese den gelernten Bedeutungen zu. Die Ausgabe des Text-Klassifikators ist also die Bedeutung des eingegebenen Textes.

Abbildung 2 stellt den Aufbau einer Anwendung mit Sprachsteuerung schematisch dar. Die einzelnen Blöcke dieser generellen Infrastruktur werden im folgenden beschrieben.

1. Benutzereingabe

Der Dialog mit dem System beginnt mit der Eingabe von Information durch die Benutzerin. Für die weiteren Ausführungen abstrahieren wir hierbei von der genauen Eingabemodalität, wie getipptem Text oder gesprochener Sprache.

2. Interaktionskanal

Der Interaktionskanal wird von dem von der Benutzerin verwendeten Gerät festgelegt. Traditionelle IVR-Systeme (Interactive Voice Response) Systeme verwenden hier nur Sprache. Durch das ubiquitäre Smartphone sind jedoch Messenger wie WhatsApp oder Telegram und Soziale Netzwerke wie Facebook der Standardkanal für die Interaktion. Auf diesen Systemen aufsetzend gibt es eine Vielzahl von Chatbot-Frameworks, mittels derer eine multimodale dialogische Interaktion mit verschiedenen Dienstleistungen realisiert werden kann.

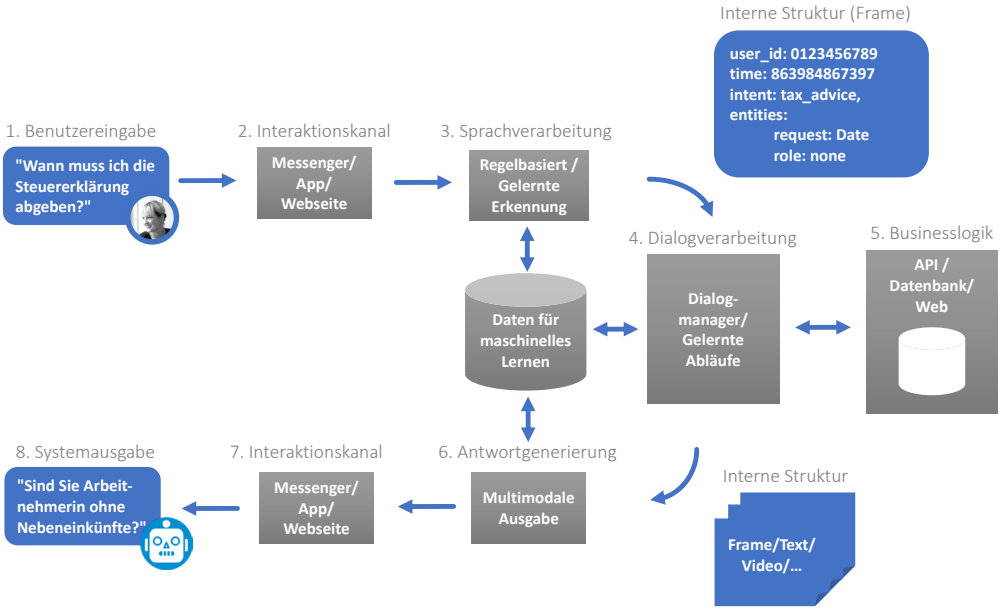

Abbildung 2: Darstellung des Aufbaus einer Anwendung mit Sprachsteuerung

Je nach Eingabe werden Tippfehlerkorrekturen, Wortvorschläge, oder Spracherkenner im Interaktionskanal eingesetzt, um die Benutzereingabe in Text zu übersetzen.

3. Sprachverarbeitung

Die Komponenten der Sprachverarbeitung setzen die Eingaben der Benutzerin in ein internes Format um. Dieses wird meistens als Frame bezeichnet, und enthält neben dem zentralen Inhalt (dem so genannten Intent oder Dialogakt) weitere domänenspezifische Informationen, die aus dem textuellen Inhalt der Äußerung und eventuell aus dem Kontext der Situation abgeleitet werden können. Bei multimodalen Dialogsystemen werden auch Informationen aus der Interaktion mit Inhalten auf dem Bildschirm in dieser internen Struktur zusammengeführt. Diesen Vorgang bezeichnet man als Modalitätsfusion.

Bei der Interpretation können verschiedene Methoden zum Einsatz kommen. Klassische Systeme wie IVR verwenden Ansätze wie VoiceXML[9] oder Derivate, die regelbasiert arbeiten. Der Dialogdesigner definiert hier die Menge an Regeln, die benötigt werden, um die einzelnen Dialogschritte zu verarbeiten.

Bei statistischen Verfahren wird die Abbildung vom Text zur Benutzerabsicht (Intent) aus den Daten für maschinelles Lernen gewonnen. Dies sind sogenannte Dialogkorpora mit einer Vielzahl von einzelnen Dialogen, sowohl auf der Sprachsignal- als auch auf der Textebene. Damit das Lernen stabil funktioniert, sind Korpora mit mehreren 1 000 – 100 000 Dialogen wünschenswert. Diese müssen im Vorfeld der Dialogsystemerstellung erhoben und annotiert werden. In vielen Systemen werden auch Interaktionen aus dem laufenden Betrieb analysiert, annotiert und in den Korpus und die daraus abgeleiteten Wissensquellen aufgenommen. Wichtig dabei ist, das für spezielle Domänen, wie z.B. den Steuerbereich, entsprechend (Steuer-) Dialogkorpora vorliegen müssen, damit die Benutzerabsicht richtig erkannt werden kann.

4. Dialogverarbeitung

Der Dialogablauf mit der Benutzerin wird von dieser Komponente gesteuert. Konzeptuell wird der Eingabeframe daraufhin

9 https://www.w3.org/TR/voicexml20/

analysiert, welcher Intent enthalten ist, ob die Informationen im Frame vollständig sind, um eine Anfrage an die Businesslogik zu stellen, oder ob Nachfragen notwendig sind. Kann eine Anfrage an die Businesslogik gestellt werden, wird das Ergebnis aufbereitet und je nach System als gefüllter Frame, Text, Bild oder Kombination aus verschiedenen Informationen weitergegeben. Fehlen Informationen für eine Anfrage, oder ist eine Anfrage nicht aufgabenbezogen, wie z.B. eine Klärungsfrage, wird von der Dialogverarbeitung die entsprechende Reaktion, z.B. eine Nachfrage nach den fehlenden Informationen, ausgewählt.

Die interne Logik basiert bei klassischen Systemen auf Dialogautomaten, Regeln, oder anderen deklarativen Mechanismen. Liegen Dialogkorpora aus der Domäne vor, können entweder die Abläufe vollständig aus diesen gelernt werden, oder, in hybriden Systemen, eine Kombination aus statistischen und regelbasierten Methoden zum Einsatz kommen.

5. Businesslogik

Das Domänenwissen selbst sollte unabhängig vom Sprachdialogsystem separat in einem gekapselten Businesslogik-Modul liegen, hinter dem sich natürlich komplexe Systeme befinden können. Da die Businesslogik implizite Annahmen enthält über die Reihenfolge und Notwendigkeit von Informationen, Benutzer-IDs und zugehörige Daten erfordert, und selten zustandslos ist, generieren diese Bedingungen aus der Businesslogik Randbedingungen für die Dialogverarbeitung. Beispiele sind die Reihenfolge, in der Informationen abgefragt werden, oder die Authentifizierung. Deswegen ist es bei der Erstellung der Dialog-Verarbeitungslogik essentiell, diese Randbedingungen zu kennen. Auch bei statistischen Ansätzen kann dieses Wissen nicht ausschließlich aus Korpusdaten gelernt werden.

6. Antwortgenerierung

Die Ausgabe der Dialogverarbeitung wird in einer Komponente zur Antwortgenerierung für den oder die Ausgabekanäle aufbereitet. Bei Chatbots oder IVR-Systemen wird hier aus der internen Repräsentation des Frames o.ä. ein Text generiert, der angezeigt oder als künstliche Sprache synthetisiert werden kann. Bei multimodalen Systemen wird bestimmt, welche Inhalte der Antwort als Text, Sprache, Bild oder als Buttons o.ä. realisiert werden sollen. Für die Ausgabe von Sprache stellen große Anbieter wie Google oder Nuance verschiedene Ansätze bereit, die je nach Bereich und Kundenwunsch eine hochqualitative Ausgabe ermöglichen. Diese Syntheselösungen sind ebenfalls auf großen Datenmengen trainiert und laufen meistens in der Cloud. Der Anbieter Nuance liefert auch kundenspezifische Stimmen, die auf der Infrastruktur des Betreibers laufen können. Zudem können für die Sprachsynthese Programme verwendet werden, die z.B. auf der Hardware des Benutzers (Smartphone) vorinstalliert sind.

7. Interaktionskanal

Spiegelbildlich zu Block 2 wird hier auf dem jeweiligen Interaktionskanal die Ausgabe realisiert und dem Benutzer präsentiert.

8. Systemausgabe

Das Ergebnis wird der Benutzerin situationsspezifisch präsentiert.

Wie zu sehen ist, enthält eine Sprachdialoganwendung, als App oder als Chatbot auf einer Plattform, viele Komponenten und Wissensquellen, die neu zu entwickeln oder anzupassen sind. Diese Anwendungen unterscheiden sich fundamental von klassischen Webseiten, die einen an das Medium und die Interaktionsgeräte wie Maus, Tastatur und Bildschirm, angepassten Interaktionsfluss umsetzen. Für den E-Government Bereich bleibt festzuhalten, dass insbesondere für die Sprachverarbeitung und die Dialogverarbeitung Daten in Form von Korpora benötigt werden, um domänenspezifische Modelle erzeugen zu können. Des Weiteren werden Modelle benötigt, die Businesslogik (bzw. das Domänenwissen) repräsentieren, um spezifische Anwendungen zu realisieren.

2.3 Akteure im Vergleich

Technologisch lassen sich für den Einsatz von Diensten für Sprachinteraktion zwei grundlegende Ökosysteme identifizieren: Cloud-basiert und On-premise. Cloud-basierte Ansätze werden von allen großen Anbietern wie Google, Amazon und Facebook für ihre jeweiligen Plattformen angeboten (vgl. Abbildung 3, links). Der Anbieter stellt dabei Interaktionsplattform, Cloudservices und die Infrastruktur zum Erstellen der Spachdialog-Logik bereit. In der Dialoglogik kann die Entwicklerin des Services Aufrufe zur Businesslogik ihres Services auf ihrer Infrastruktur realisieren. Persönliche Daten und Interaktionen laufen aber immer durch die Systeme des Anbieters der Cloud. Der Vorteil der Lösung ist, dass durch die Verwendung der Daten aller Interaktionen verschiedenster Anbieter Funktionen wie Spracherkennung immer besser werden, und auch die Werkzeuge immer Up-to-date sind. Zudem stellen die Cloudfirmen sicher, dass die Services immer erreichbar sind. Zudem können über Geräte wie Amazon Echo, Facebook Messenger etc. viele Zielgruppen mit wenig Aufwand direkt zugreifen.

Aus Komponenten einzelner Anbieter wie Nuance, der größten unabhängigen Firma auf diesem Sektor, Rasa für die Dialoglogik und anderer kann ein Service auf den Rechnern eines Betreibers lokal oder geschützt bei einem Hoster selbst (vgl. on-premise Anbieter Abbildung 3, rechts) realisiert werden. Bei dieser Lösung trägt der Servicebetreiber aber die volle Verantwortung für die Funktion selbst, von der technischen Grundausstattung, über die Realisierung der Sprachinteraktion bis hin zur Umsetzung des Services beim Kunden bzw. Benutzers via einer eigens entwickelten App oder einer entsprechenden Webseite. Die Kosten für Personal und technische Ausstattung sind hierbei erheblich, da auch der laufende Betrieb ständig technisch und inhaltlich überwacht werden muss.[10] Zudem entfällt der Vorteil, dass Spracherkenner und andere trainierbare Komponenten ständig durch die Interaktion mit sehr vielen Benutzern auch anderer Sprachservices verbessert werden. Dies muss der Anbieter selbst gewährleisten. Positiv ist auf jeden Fall, dass die Betreiberin alle Daten selbst in der Hand hat, und, bei der entsprechenden 24/7 Überwachung mit der entsprechend kompetenten Personalausstattung,

Cloud-basiert

- Google Dialogflow
- Amazon Lex
- Microsoft Bot Framework
- Nuance (D), u.a.

↓

- Verarbeitung vertraulicher Informationen
- Kein Einfluss auf Datensicherheit
- Kein Hosting auf anderen Servern
- Geringe Anpassungsfähigkeit
- Verursacht laufende Kosten

On-premise

- Nuance (D)
- Rasa (D)
- Botpress
- weitere Open Source Frameworks

↓

- Erfordert technisches Wissen
- Aufbau einer Infrastruktur nötig
- Entwicklungs- & Betriebskosten
- Wissensbasis benötigt
- Keine out of box Integration in Interaktionskanäle

Abbildung 3: Oben ist eine Auswahl cloud-basierter und On-Premise Anbieter von Sprach- und Chatbotdiensten aufgeführt. Die blauen Pfeile darunter deuten Nachteile des jeweiligen Ansatzes an.

10 https://www.sueddeutsche.de/digital/microsoft-programm-tay-rassistischer-chat-roboter-mit-falschen-werten-bombardiert-1.2928421

auch mit hoher Wahrscheinlichkeit die Daten sicher und geschützt halten kann.

2.4 Bestehende E-Government Anwendungen im Steuerbereich

Im Steuerbereich existieren mittlerweile einige Anwendungen, die sich als Chatbot präsentieren und textlich dialoggesteuert sind. Die einfachste Verwendung der folgenden Anwendungen erfordert die Installation einer iOS oder Android App auf dem Smartphone.

Diese Apps richten sich an ein jüngeres, besonders technikaffines Publikum, das sich tendenziell weniger Sorgen um Privatsphäre und Datenschutz macht. Oft geht es dabei auch um eine Prüfung, ob eine Abgabe erforderlich und vorteilhaft ist. Trotz Anmeldung wird eine Gebühr erst fällig, wenn man sich nach Eingabe der relevanten Daten dazu entschließt, auch wirklich eine Steuererklärung abzugeben.

Das Angebot von Taxfix[11] beispielsweise ist kostenlos bei Rückerstattungen unter 50 Euro. Danach kostet die Abgabe einer Steuererklärung mit Taxfix 34,99 Euro. Das Kalkül der Hersteller ist, dass der Kunde in den meisten Fällen eine Erstattung zu erwarten hat und dann schon alle erforderlichen Daten eingetragen hat – er kann dann also mit geringen Kosten und vertretbarem Aufwand die Steuerklärung veranlassen.

Die Verwendung dieser oder ein paar weniger weiterer Apps im Bereich der Steuern könnte für viele Nutzer aber schnell zu dem Ergebnis führen, dass der eigene Steuerfall doch zu komplex ist und eine andere Lösung für die Erstellung der Steuererklärung notwendig ist. Hierfür reicht bereits ein Abweichen von den Werbekostenpauschalen, Rentenbezüge, Kapitaleinkünfte, Vorsorgeaufwendungen, oder Einkünfte aus Selbständigkeit. Bis dahin ist die Benutzerführung aber teilweise sehr intuitiv in einem dialogorientierten Chat.

Einen entsprechenden Service auf sprachlicher Ebene realisiert der Alexa Skill Smart Steuer während der Erstellung der Steuererklärung.

Der Steuerchatbot von Baden-Württemberg[12] in Abbildung 4 gibt allgemeine Auskünfte rund um die Steuer. Er unterstützt nicht bei der Abgabe einer Steuererklärung, sondern liefert einen Q&A Service.

2.5 Bestehende E-Government Anwendungen im Servicebereich

Im Servicebereich existieren momentan nur wenige ernstgemeinte Anwendungen. Es handelt sich eher um Prototypen, die auf dem Markt sind. Selbst die von den Kommunen angebotenen Chatbots, wie bspw. der GovBot Bonn,[13] adressieren kaum einen sinnvollen Use Case aus dem Bereich E-Government. Selbst die Suche nach Verwaltungsleistungen über den Bot fühlt sich eher wie die Suche in einem Verzeichnis an. Die Antworten erscheinen kaum hilfreich und scheinen einem kontextfreien FAQ Katalog zu entspringen.

Ein besseres Beispiel für die Umsetzung ist der textbasierte Chatbot auf dem Serviceportal Berlins. Hier werden auf der Basis der textuellen Eingabe als Antwort eine Zusammenstellung mit den am besten passenden Links des Serviceportals ausgeliefert. Es ist eine Frage-/

11 https://taxfix.de/?utm_expid=.m52O-iDYRWGZf69_YXisCA.0&utm_referrer=https%3A%2F%2Fduckduckgo.com%2F

12 https://zendib-haupt-entwicklung.eu-de.mybluemix.net/

13 https://govbot.bonn.de/

Abbildung 4: Steuerchatbot von Baden-Württemberg für allgemeine Auskünfte zum Thema Steuer

Antwortapplikation, die keinen wirklichen Dialog mit dem Benutzer führt,[14] da das eigentliche Domänenwissen in den Links und den dahinterliegenden Webseiten kodiert ist. Sie ist jedoch in der Lage, z.B. Satzellipsen (siehe Abbildung 5) aufzulösen und kann damit minimale Dialogphänomene umsetzen. Bobbi ist eine typische Anwendung, die den Benutzern jenseits der Suche nicht wirklich weiterhilft, keine inhaltlichen Dialoge führen kann und auf Fragen außerhalb der Domäne stereotyp mit einer Standardreaktion antwortet. Von den oben identifizierten Komponenten wurde wahrscheinlich nur eine einfache Sprachverarbeitung realisiert, die direkt die Businesslogik, d.h. die Suche auf den Webseiten ansteuert.

Der Chatbot Govii der Stadt Kiel[15] ist ebenfalls auf behördliche Anfragen spezialisiert. Er liefert Antworten im Chat, die im Kontext von E-Government Leistungen stehen und bietet zur Klärung Auswahlmöglichkeiten an. Die Dialogsteuerung und die Interaktion erinnert an klassische Sprachdialogsysteme über Telefon, bei denen die Aufgabenstruktur in einem Menübaum hinterlegt ist. Eine flexible Interaktion oder freiere Suche ist nicht möglich.

Die Bundesagentur für Arbeit hat einen Chatbot zur Unterstützung von Jugendlichen bei der Berufswahl entwickelt.[16] Dieser ist tendenziell als spielerische Hilfestellung zu verstehen und hilft, den Dialog mit Schülern zu beginnen. Er führt die Nutzer hin zu individuell passenden Inhalten im Webangebot der Bundesagentur für Arbeit.

Sämtliche dieser Lösungen liefern im Allgemeinen Links auf weiterführende Webseiten und bestenfalls Erläuterung zur Notwendigkeit von Anträgen, den Vorschriften und zuständigen Behörden. Die Antragsbearbeitung selbst wird in den kommunalen

14 Siehe auch die Definition im Gabler Wirtschaftslexikon https://wirtschaftslexikon.gabler.de/definition/chatbot-54248

15 https://www.kiel.de/de/politik_verwaltung/service/

16 https://www.dasbringtmichweiter.de/typischich/whatsmebot-under-sixteen/

Abbildung 5: Textbasierter Chatbot auf dem Serviceportal Berlins

Bots nicht angeboten, es werden keinerlei persönliche Daten erhoben. Zudem sind alle Bots textbasiert, d.h. Sprachinteraktion ist nicht vorgesehen.

3. ERGEBNISSE: SZENARIEN FÜR SPRACHINTERAKTION IN BEHÖRDLICHEN PROZESSEN

In den folgenden Abschnitten beschreiben wir mögliche Szenarien für behördliche Prozesse, in denen eine sprachgestützte dialogische Interaktion möglich sein könnte. Primäres Ziel ist es, die Abläufe zu strukturieren und herauszuarbeiten, wo die Benutzerin bei der Umsetzung des Prozesses unterstützt werden kann.

Die realen Anforderungen bestehender Prozesse schränken natürlich die Freiheitsgrade einer dialogischen Interaktion ein: Würde der Prozess in einem klassischen nutzerorientierten Gestaltungsansatz grundlegend neu gedacht, würde die Lösung wahrscheinlich deutlich anders aussehen. Beispiele hierfür sind vor allem die skandinavischen oder baltischen Länder, die bereits fortschrittliche Prozesse für ähnliche Szenarien anbieten.

In einer Expertenrunde wurden relevante behördliche Prozesse mit dem Fokus Steuerumfeld und dem Fokus OZG Umsetzungsgesetz identifiziert. In den folgenden Abschnitten werden bestehende Prozesse rekapituliert und Einsatzszenarien für Sprachinteraktion erörtert. Die aufgegriffenen Beispiele betrachten Interaktionen zwischen Verwaltung und Bürger/in, die durch die Bürgerseite ausgelöst werden. Es ist außerdem denkbar, dass auch von der Verwaltungsseite ausgelöste Interaktionen (z.B. bei Rückfragen zu den von den Bürgerinnen und Bürgern eingetragenen Daten) durch Chatbots unterstützt werden können. Die Verwaltungsseitige Betrachtung ist jedoch nicht Bestandteil dieser Studie.

3.1 Fokus Steuerumfeld

3.1.1 Ablauf einer Steuererklärung

Zunächst möchten wir den Ablauf einer Steuererklärung von der Vorbereitung der Unterlagen durch den Bürger bis zum Abschluss und Bescheid kurz beschreiben. Im Anschluss werden die Einsatzmöglichkeiten der Sprachsteuerung in diesen Prozessschritten besprochen. Als Beispiel für eine Steuererklärung wird konkret auf die Einkommensteuererklärung eingegangen, da dieses Szenario für eine große Anzahl an Menschen Relevanz hat und daher für viele anschaulich sein sollte.

Vor der wirklichen Erstellung der Erklärung, dem Ausfüllen des Formulars, bspw. als Papierversion oder Online mittels Mein ELSTER, steht die mühsame und gegebenenfalls sehr umfangreiche Sammlung von Informationen und Belegen, die bei Behörden oder Unternehmen abzufragen und einzuholen sind. Die anschließende Erstellung der Steuererklärung ist in vielen Fällen nicht mehr besonders aufwendig. Zeitaufwendig ist vielmehr eine nachträgliche, oft um Monate verzögerte Zusammenstellung der benötigten Informationen.

Dann erfolgt die Erfassung der Informationen in der Steuererklärung. Dies wird, wenn nicht von Steuerberatern oder Vereinen übernommen, Themenblockweise von den Bürgern abgearbeitet. Eine gewisse Nutzerführung mit kontextbezogenen Hilfestellungen wird von kommerziellen Steuerprogrammen angeboten. Im Regelfall wird die Erklärung durch die Datenübernahme aus dem Vorjahr stark vereinfacht. Übernommen werden dabei die Stammdaten, ergänzt um die

Informationen der spezifischen Lebenssituation. Im Rahmen der vorausgefüllten Steuererklärung werden ebenfalls Informationen zur Lohnsteuer, Krankenkasse, Riester Rente, u.a. automatisch übernommen. Nach der erfolgten Abgabe der Steuererklärung stellt sich die Frage des Bearbeitungsstatus im Finanzamt. Rückfragen seitens der Bürger werden dabei telefonisch, per Mail oder direkt, oder indirekt über Mein ELSTER gestellt: Wann kann ich mit meinem Bescheid rechnen? Mir wurde immer noch nicht das Geld überwiesen? Schließlich können weitere, sich auf die Steuererklärung beziehende (Sub-) Anträge gestellt werden, bspw. Einspruch, Antrag auf Fristverlängerung, etc. Diese Anträge können formlos, schriftlich oder templatebasiert gestellt werden. Die Zahl möglicher Antragsarten ist aus Sicht der Finanzämter überschaubar und wiederholend.

3.1.2 Einsatzszenarien für Sprachinteraktion bei der Steuererklärung

Nach Betrachtung des klassischen Ablaufes einer Steuererklärung kann eine Unterteilung in 5 grobe Schritte vorgenommen werden. Im Folgenden wird erörtert, wie diese Schritte perspektivisch mittels Sprachinteraktion unterstützt werden könnten.

1. „Vorbereitung der Vorbereitung“

Die zeitaufwendige verzögerte Zusammenstellung der benötigten Informationen könnte durch ein Sprachassistenzsystem unterstützt werden, welches über eine einfach zu bedienende Benutzerschnittstelle ermöglicht, steuerlich relevante Informationen quasi nebenbei in einem zentralen Datenspeicher abzulegen. Hierbei könnten z.B. folgende Use Cases adressiert werden

a. Aufsammeln von Belegen

Bei diesem Use Case werden Belege, die sich über das Steuerjahr ansammeln, direkt per Foto digitalisiert und mittels Spracherkennung mit Schlagworten versehen. Die Funktionalität könnte in ein appbasiertes Steuer-Assistenzsystem (ASA) integriert sein, das Sprachinteraktion unterstützt und dessen Businesslogik in der Lage ist, die Belege korrekt zu gruppieren. Die Belege werden digital bis zur Erstellung der Steuererklärung aufbewahrt und bei der Steuererklärung automatisch per Import den richtigen Themenblöcken zugewiesen.

b. Kinderbezogene Daten eintragen

Das Eintragen von Kindern ist eine Funktion, die ermöglicht, die nötigen Informationen zu Kindern einfach per Sprache in ASA einzugeben.

c. Arbeitsfahrten

Das Loggen von Arbeitsfahrten kann unkompliziert per Sprache in ASA aktiviert werden. Über GPS kann dabei auch automatisch die Länge der gefahrenen Strecke erhoben und gespeichert werden. Arbeitswege können im zentralen Datenspeicher aggregiert werden und die gefahrenen Kilometer automatisch in die relevanten Themenblöcke der Steuererklärung integriert werden.

2. Abgabe der Steuererklärung

1. Interviewmodus

Idee des Interviewmodus ist es, bevor die eigentliche Erstellung / Bearbeitung der Steuererklärung beginnt, bestimmte Informationen bei der Benutzerin abzufragen. Die abgefragten Informationen werden gesammelt und bei Beendigung des Interviewmodus in das Formular übertragen.

a. Vorbereitung des Formulars

In diesem Use Case werden im Rahmen eines “Interviews”, das ein sprachbasierter Chatbot mit der Benutzerin führt, bestimmte Basisinformationen zu steuerlichen Themen abgefragt. Fragen wie “Haben Sie Kinder?”, die zur Folge haben, dass bestimmte Themenblöcke oder Anlagen bearbeitet werden müssen (oder nicht), können so vorab gestellt werden, Basisinformationen gesammelt und beim

Beginn des eigentlichen Ausfüllens vorab in das Formular übertragen werden.

b. Übernahme von Vorjahresdaten

Als spezieller Teil der Vorbereitung kann auch auf die Vorjahresdaten zugegriffen werden und in einem sprachlichen Dialog abgeklärt werden, inwiefern diese ganz oder teilweise ins neue Formular übernommen werden können.

2. Ausfüllen des Antrags

Die Unterstützung durch einen Sprachdialog während des eigentlichen Ausfüllens des Antrags stellt aktuell das komplexeste der betrachteten Szenarien dar. Technologisch müssen hierfür einzelne Dialogabläufe für die verschiedenen Themenblöcke der Steuererklärung bereitgestellt werden. Wünschenswert wäre hierbei, dass die einzelnen Steuerthemen modellhaft vorliegen, da die Dialogabläufe möglichst automatisiert aus der Businesslogik generiert werden sollten.

a. Kontextbezogenes Ausfüllen

Beim kontextbezogenen Ausfüllen könnten einzelne Themenblöcke der Steuererklärung mit einer jeweils auf den Themenblock angepassten Sprachdialoglogik abgehandelt werden. Als Beispiel kann hier der Block "Fahrten zur Arbeit" genannt werden, bei welchem das Assistenzsystem dem Nutzer gezielt Fragen stellt, um die Informationen für die erforderlichen Formularfelder einzuholen. Bei diesem Beispiel könnte das System sogar so ausgebaut sein, dass eine automatische Berechnung der gefahrenen Kilometer erfolgt.

b. Kontextbezogenes Q&A im Formular

Im Rahmen des kontextbezogenen Ausfüllens kann es vorkommen, dass die Benutzerin zu einzelnen Themen oder Begriffen Fragen hat. Daher wäre es wünschenswert, dass die Benutzerin auch kontextbezogene Fragen stellen kann, die innerhalb des aktuellen Dialogablaufes direkt von dem Assistenzsystem beantwortet werden können. Beispiel: "Was sind Werbungskosten?" Die direkte Beantwortung solcher Fragen, könnte den Aufwand zur Bearbeitung der einzelnen Themen merklich reduzieren.

3. Status der Steuererklärung

Nach Abgabe des Antrags stellen sich Fragen wie "Ist meine Steuererklärung angekommen?" oder "Wurde mein Antrag genehmigt?". Derartige Statusabfragen könnten ebenfalls durch das Sprachdialogsystem von ASA behandelt werden. Benutzerinnen wird die Statusabfrage dadurch vereinfacht und Verwaltungsmitarbeiter werden entlastet.

4. Antragsabgabe: z.B. (ohne Online-Formular, rein per Sprachsteuerung)

Die Bearbeitung einer Steuererklärung kann weitere verwandten Anträge zur Folge haben. Als Beispiele können hier ein Antrag auf Fristverlängerung, ein Einspruch gegen einen Bescheid, oder auch eEingänge, wie ein Antrag auf Aussetzung der Vollziehung, genannt werden. Solche eher einfachen meist formlosen Anträge könnten in ASA schnell und bequem gestellt werden. Bevor ein solcher Antrag tatsächlich vom System abgesendet wird, würde man in der Logik des Sprachdialogs Bestätigungsfragen integrieren, damit ein Antrag nicht aus Versehen gestellt wird.

5. Steuer Q&A

Die kontextbezogenen Q&As, die als Ausfüllhilfen dienten, werden hierbei ergänzt, um allgemeine Q&As rund um den erweiterten Themenblock der Steuer. Hierbei kann es sich um Fragen zu Abgabeterminen, der Notwendigkeit einer Steuererklärung, etc. handeln. Wie in Abschnitt 2.4 erwähnt, existiert bereits ein Alexa Skill, welcher Unterstützung während der Erstellung der Steuererklärung bietet. Bei diesem Skill können Benutzer nach nicht bekannten Begriffen fragen, oder wie lange man noch Zeit hat für die Steuererklärung. Es wäre wünschenswert, dass ein solches sprachbasiertes Assistenzsystem nicht nur als Angebot der privaten

Wirtschaft verfügbar ist, sondern auch verwaltungsseitig zur Verfügung gestellt wird, damit alle Steuerzahler gleichermaßen unterstützt werden.

Mit diesen Schritten sind fundamentale Teile der Businesslogik des Einkommensteuerbereiches beschrieben, die Dialoglogik für eine Sprachinteraktion muss extra entwickelt werden. Da nicht vorausgesetzt werden kann, dass Benutzer der oben angeführten Logik folgen bzw. diese hinreichend detailliert kennen, muss die Dialoglogik eine vermittelnde Rolle einnehmen und den Benutzer so führen, dass die Informationen zu den Schritten in der richtigen Reihenfolge bereitgestellt werden. Diese Benutzerführung erfordert eine komplexe Logik und Zustandsrepräsentation, da nur so verhindert werden kann, dass Benutzer frühzeitig frustriert werden und die Interaktion abbrechen.

3.2 OZG Umsetzungskatalog

3.2.1 Meta-Schema für sprachgestützte Dienste für antragsbasierte Leistungen

Das für das Steuerumfeld beschriebene Vorgehen und die dort geschilderten Einsatzszenarien lassen sich auch auf den OZG Umsetzungskatalog übertragen. Viele der im Umsetzungskatalog erfassten Leistungen sind antragsbasiert.

Ein guter Teil der Antragsleistungen erfordert ebenfalls eine ‚Vorbereitung' durch den Antragsteller. Antragsbezogene Handlungen wie Fristverlängerungsanträge, oder Einsprüche, sind existierende, oder zumindest denkbare Optionen für Bürger und Organisationen. Damit ergibt sich folgendes „Meta"-Schema:

1. „Vorbereitung der Vorbereitung"
2. Abgabe des Antrags
3. Status des Antrags
4. Abgabe verwandter Anträge
5. Antrags Q&A

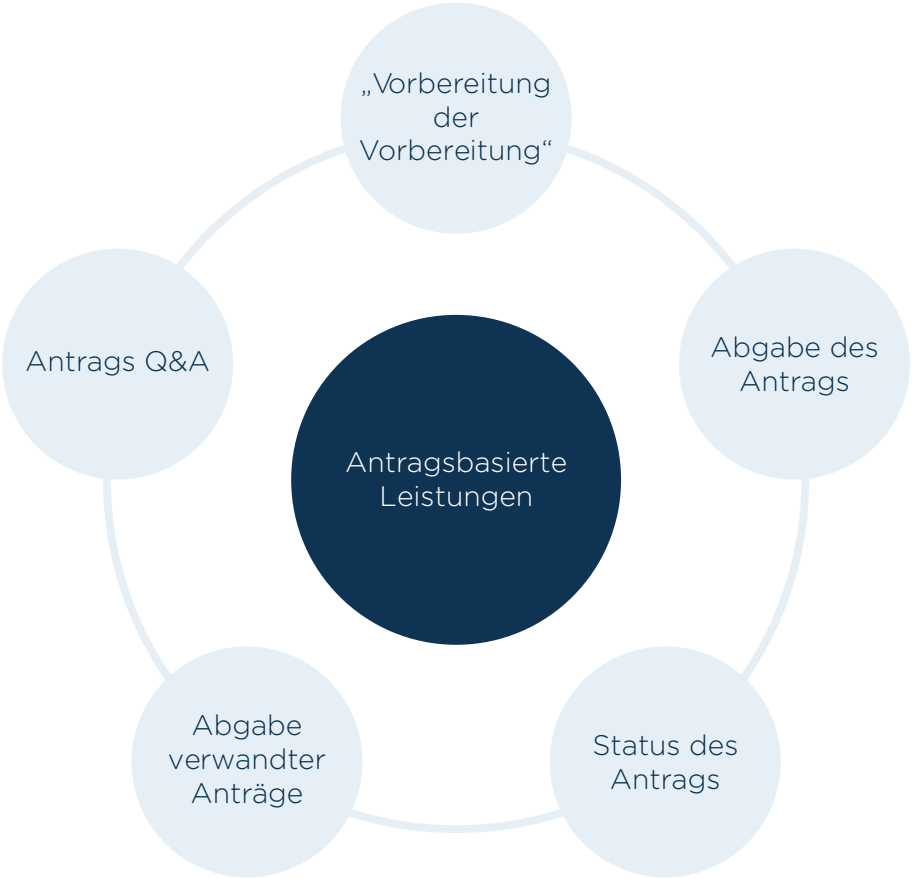

Abbildung 6: Meta-Schema für sprachbasierte Dienste für antragsbasierte Leistungen

In allen in Abbildung 6 aufgeführten Schritten sollten spezielle und allgemeine Q&A-Fragen zulässig sein. Naturgemäß weisen die unterschiedlichen Verwaltungsbereiche spezielle Besonderheiten auf (Familie, Soziales, Arbeitssicherheit, etc.). Diese sind erklärungsbedürftig und bieten viel Raum für Q&A. Weiter werden für viele Verwaltungsbereiche für die Sprachinteraktion spezialisierte Modelle notwendig sein, um einen nutzerfreundlichen Dialog realisieren zu können. Dort, wo Formulare besonders erklärungsbedürftig sind, könnte aus Anwendersicht ein Interviewmodus zum Befüllen der Formulare und Sicherstellung der richtigen, interpretationswürdigen und benötigten Informationen sehr hilfreich sein.

3.2.2 Einsatzszenarien für Sprachinteraktion im Rahmen des OZG Umsetzungskatalogs

Im Folgenden werden 2 Beispiele für Einsatzszenarien im Rahmen des OZG Umsetzungskatalogs dargelegt.

1. Anwohnerparkausweis

Aufgrund einer recht freien Auslegung des Prozesses durch die Kommunen, bzw. fehlender Vorgaben, unterscheiden sich bei einem Antrag für einen Anwohnerparkausweis die Prozesse, sowie die erforderlichen Angaben und Formvorschriften.

Ein Anwohnerparkausweis in Köln erfordert beispielsweise Stammdaten des Anwohners, die unten aufgeführten Informationen / Unterlagen und die Zahlung einer Gebühr.[17]

Benötigt werden

- **Fahrzeugschein** beziehungsweise Zulassungsbescheinigung Teil I
- **Personalausweis oder Reisepass / Nationalpass mit Aufenthaltsgenehmigung** für Ausländerinnen und Ausländer aus der Europäischen Union genügt der Nationalpass.
- **Führerschein:** Diesen sollten Sie vorsorglich immer mitbringen. Zwingend benötigen wir ihn, wenn das Fahrzeug nicht auf Sie als antragstellende Person zugelassen ist oder wenn an Ihrer Stelle eine bevollmächtigte Person zur Antragstellung vorspricht (Kopie genügt).
- **Den bisherigen Bewohnerparkausweis:** wenn durch eine Änderung ein neuer Bewohnerparkausweis ausgestellt wird
- **Halterbestätigung:** Diese ist erforderlich, wenn das Fahrzeug nicht auf die antragstellende Person zugelassen ist. Bitte lesen Sie die Erläuterungen unter "Halterbestätigung und Infos zum privaten Carsharing".
- Besondere Halterbestätigung und weitere Unterlagen bei privatem Carsharing sowie
- Unterlagen bei Leih- und Mietfahrzeugen gewerblicher Anbieter

Bei neuer Antragstellung wäre die Übernahme der Stammdaten und die Übernahme der Daten Fahrzeugschein, Führerschein, Personalausweis wünschenswert. In einem Sprachdialog könnte dies einfach abgefragt werden (z.B. "Können diese Daten übernommen werden?", Antwort "Ja bitte").

Die Übernahme von Vorjahresdaten bei Verlängerung ist auch aufgrund der ggf. jährlich erforderlichen Antragstellung wünschenswert. Eine Verlängerung des Anwohnerparkausweises könnte in weiten Teilen automatisiert als antragsbezogener Dialog realisiert werden. Ein

17 https://www.stadt-koeln.de/service/produkt/bewohnerparkausweis?kontrast=schwarz

Abbildung 7: Anträge wie der Antrag Kindergeld sollten in Zukunft durch einen Sprachdialog wie Interviewmodus ausgefüllt werden können.

sprachgesteuerter Prozess ist hierbei aufgrund der recht einfachen Formularstruktur gut vorstellbar. Aus Sicht des Antragstellers erscheint auch eine Q&A Funktion als hilfreich, da Fragen beispielsweise zur Halterbestätigung aufkommen können.

2. Kindergeldantrag

Zum Thema Kindergeld sind online eine extreme Fülle an Informationen rund um den Antrag mit Sonderformularen und Flyern auf der Internetseite der Bundesagentur verfügbar.[18] Im einfachsten Fall ist der Kindergeldantrag jedoch eigentlich ein recht schlanker Antrag, insbesondere wenn auf Stammdaten aus einem Servicekonto zugegriffen werden kann und dieses sinnvoll vom Bürger gepflegt ist. Ansonsten wären aus Bürgersicht in den Formularen viele bereits der Behörde bekannte Information einzugeben:

- Personendaten (Name, Vorname, Geburtsdatum, -ort, Anschrift, Staatsangehörigkeit, Familienstand)
- Verknüpfung mit Lebenspartner und dessen Daten
- Verknüpfung mit Kindern und deren Daten
- Angaben zur Kontoverbindung

Zur Angabe der Informationen und Abgabe der Erklärung würde sich ein Sprachdialog mit Interviewmodus anbieten. Die Nutzung des Interviewmodus hilft dann beim Befüllen von einem signifikanten Anteil des Formulars sowie bei der Auswahl beizusteuernder Unterlagen und weiterer Informationen.

18 https://www.arbeitsagentur.de/familie-und-kinder/downloads-kindergeld-kinderzuschlag

3.3 Anforderungen und Herausforderungen

Bei Datenaggregatoren und -kommerzialisierern wie Facebook mit den angeschlossenen Diensten WhatsApp und Instagram, Google mit Hardware wie Nest und anderen[19] [20] wird deren Datensammlung von den meisten Menschen nur vordergründig kritisch gesehen.[21] Die Dienstleistungen, die vor allem soziale Bedürfnisse und Wünsche adressieren, sowie dem Komfort dienen, werden jenseits von Lippenbekenntnissen ohne größere Bedenken verwendet. Die Sammlung von Daten durch staatliche Stellen, mit denen reale Dienstleistungen und Prozesse realisiert werden können, werden dem gegenüber ständig kritisch hinterfragt.[22] Bei jedem Einsatz von IT-Technologien ist diese kognitive Grundeinstellung der Menschen zu beachten und seriös zu adressieren. Mit der Datenschutz-Grundverordnung (DSGVO) hat Europa hier eine Vorreiterrolle eingenommen, die als Markenzeichen besonders für staatliche Dienstleistungen benutzt werden sollte.

Hinter allen oben genannten Dienstleistungen liegen persönliche Daten, die als vertraulich eingestuft werden. Lediglich allgemeine Fragen zu generellen Prozessen und Abläufen (FAQs) können in größerem Umfang ohne allzu viele persönliche Daten in einer Interaktion erfragt werden.

Sobald wirklich persönlich Daten notwendig sind, ist eine sichere Zweifaktorauthentifizierung notwendig. Diese in ein System auf verschiedenen Geräten ab initio zu realisieren, ist illusorisch. Bei einer gestuften Realisierung, z.B. zuerst über eine App, kann die Technologie einbezogen werden, die auf diesen Geräten bereits vorhanden ist.

Zudem ist vom Betreiber die Frage zu klären, wo die Erkennung und Verarbeitung angesiedelt ist – unter der Kontrolle auf Systemen, die ihm gehören, oder bei einem der genannten Cloudanbieter. In beiden Fällen ist ein Datenschutzkonzept notwendig, sowie hinreichende Ressourcen, um die Einhaltung zu überwachen.

Im Gegensatz zu den meist im sozialen oder Consumer-Bereich angesiedelten Chatbots auf den entsprechenden Plattformen, liegen für das Feld E-Government kaum Daten vor, mit denen insbesondere die Dialoglogik trainiert werden kann, bzw. die als Trainings-, Evaluations-, und Testdaten dienen können. Deswegen sollte initial ein modellbasierter Entwicklungsschritt vorgesehen werden, bei dem die Business- bzw. die Fachlogik durch Modelle beschrieben wird. Datenmodell und Use Cases, welche grundlegend für die Dialogverarbeitung sind, können hierbei in Zusammenarbeit mit BenutzerInnen und Fachexperten definiert werden. Mit diesen Daten kann ein System realisiert werden, das iterativ als Alpha-Version im geschützten Betrieb verwendet wird, um realistische Dialogdaten zu sammeln, die wiederum für die agile Weiterentwicklung verwendet werden. Daneben sind Fragen wie fachspezifische Termini, deren Synonyme in der Realität u.v.m. zu klären.

Vor allem die Module für automatische Spracherkennung und für Sprachverstehen müssen an das spezielle Sprachregister angepasst werden. Zudem ist hier zu klären, ob und wie das System für Menschen mit anderer Muttersprache und eventuell rudimentären Deutschkenntnissen zugänglich gemacht werden kann. Unter Umständen muss hier eine Einschränkung vorgenommen werden.

19 El Saddik, A. (2018). Digital twins: The convergence of multimedia technologies. IEEE MultiMedia, 25(2), 87-92. https://ieeexplore.ieee.org/abstract/document/8424832

20 https://medium.com/ted-hunt/meet-the-digital-twin-you-probably-didn-t-realise-you-had-eb05096fa9cf

21 https://www.nytimes.com/2019/09/12/well/protect-your-digital-self-in-a-few-minutes-a-day.html

22 https://www.internet-freiheit.de/staatliche-ueberwachung/

4. HANDLUNGS-EMPFEHLUNGEN

Als Resümee der Studie geben wir folgende Empfehlungen

1. Für sprachbasierte E-Government Dienste müssen Businesslogik und Dialogmanagement bereitgestellt werden

Ein (sprachbasierter) Chatbot ist keine Website mit Sprachsteuerung! Webseiten folgen bezüglich der Nutzerinteraktion einer Logik, die auf den speziellen Anforderungen der klassischen Webinteraktion aufbauen. Für eine natürliche sprachbasierte Interaktion wird zu der Businesslogik ein Dialogmanagement benötigt, sowie die entsprechenden Ein- und Ausgabemethoden, wie Spracherkennung und -synthese. Der Dialogfluss muss an den Menschen angepasst umgesetzt werden, nicht die Interaktion an bestehende Verfahren. Eine ausgearbeitete Benutzerführung inkl. Nachfragen und Klärungsdialogen sind notwendig, um eine echte, akzeptable dialogische Interaktion zu erreichen.

2. On-Premises Lösung benötigt

Die Vertraulichkeit der Daten verlangt, dass die Dialoglogik, inkl. Spracherkennung unter der Kontrolle des Betreibers bleibt. Bekannte Anbieter wie Amazon und Google können nicht sicherstellen, dass die cloudbasierten Lösungen den Anforderungen gerecht werden. Systemanbieter wie Nuance ermöglichen Spracherkenner beim Kunden, jedoch ist hier z.B. die Verbesserung der Erkennerleistung schwierig. Zudem erfordert der nachhaltige Betrieb signifikanten Aufwand: Falls nur unzureichende Qualitätsüberwachung des sprachbasierten Chatbots im realen Betrieb und keine ständige Verbesserung (DevOps-Modell) umgesetzt wird, wird die einmal implementierte Lösung in kürzester Zeit nicht mehr akzeptabel sein.

3. Kurzfristig: mit einfachen Leistungen beginnen

Wir empfehlen, kurzfristig einfache Leistungen wie die Statusabfrage der Steuererklärung oder einfache verwandte Anträge (z.B. Fristverlängerung) umzusetzen. Dies ist in der Komplexität vergleichbar mit der Nachfrage nach dem Lieferstatus eines Pakets.

4. Mittelfristig: Umsetzung eines Interviewmodus

Mittelfristig wäre die Umsetzung eines Interviewmodus zur Vorbereitung größerer Anträge sowie kontextbezogenes Q&A im Formular wünschenswert. Hierbei ist zu klären, wie z.B. eine Integration mit ELSTER aussehen könnte.

5. Langfristig: kontextbezogenes Ausfüllen der Steuererklärung ermöglichen

Langfristig vorstellbar ist ein kontextbezogenes Ausfüllen der Steuererklärung mit Q&A im Formular.

6. Internationale Entwicklungen beobachten

Die hier genannten Empfehlungen gehen vom derzeitigen Status Quo in Deutschland aus. Wir empfehlen eine intensive Beobachtung der internationalen Entwicklungen, um einerseits von Erfahrungen anderer Länder zu lernen und andererseits einer Harmonisierung der technischen Prozesse sowie der behördlichen Verfahren entgegenzuarbeiten.

5. ZUSAMMENFASSUNG

Vor dem Hintergrund, dass Sprachsteuerung sich mehr und mehr als ein Standardinteraktionskanal zur Nutzung verschiedener Dienste und Anwendungen etabliert, befasst sich diese Studie in erster Linie mit den Fragestellungen, welche Infrastruktur in Deutschland für sprachgesteuerte E-Government Dienste bereitstehen müsste, und welche Szenarien im Bereich behördlicher Prozesse in Zukunft durch Sprachinteraktion unterstützt werden könnten. Herausforderungen und Anforderungen für entsprechende Sprachsteuerungsdienste werden herausgearbeitet und Handlungsempfehlungen für Grundanforderungen formuliert, die angegangen werden müssen, damit E-Government Dienste in Deutschland sprachgesteuert angeboten werden können.

Sprachassistenz erfüllt zentrale Anforderungen, die an E-Government Dienste gestellt werden wie beispielsweise Einfachheit und Effizienz, die Möglichkeit eines sofortigen Zugriffes auf Information und auch Kosteneffizienz. Allerdings stellen andere Anforderungen wie Sicherheit, Transparenz, oder einfache Wartbarkeit aktuell noch Herausforderungen für den Einsatz von Sprachsteuerung dar. Ein Blick auf den Aufbau von Anwendungen mit Sprachsteuerung verdeutlicht, welche Komponenten benötigt werden. Offenbar wären derzeit on-premise Lösungen die beste Wahl für E-Government Dienste.

Bisherige bestehende E-Government Anwendungen, die Text- oder Spracheingabe nutzen, liefern eher allgemeine Informationen und geben bestenfalls Erläuterung zur Notwendigkeit von Anträgen, den Vorschriften und zuständigen Behörden. Eine Antragsbearbeitung wird in bestehenden Systemen nicht angeboten und es werden keinerlei persönliche Daten erhoben. Ein wesentliches Ergebnis der Studie sind Einsatzszenarien für Sprachinteraktion bei der Steuererklärung. Von der Vorbereitung bis hin zur Abgabe könnte Sprachinteraktion perspektivisch unterstützend eingesetzt werden. Weiter zeigt sich, dass die für das Steuerumfeld beschriebenen Vorgehensweisen und Einsatzszenarien sich auch auf den OZG Umsetzungskatalog übertragen lassen.

Eine große Herausforderung wird es sein einen geregelten und vertrauensvollen Umgang mit persönlichen Daten für sprachgesteuerte E-Government Dienste sicherzustellen. Eine weitere Herausforderung bilden bisher kaum vorhandenen Datensätze, die benötigt werden, um eine Dialoglogik für Sprachinteraktion zu trainieren. Abhilfe könnten hier modellbasierte Entwicklungsansätze schaffen, bei denen die Business- bzw. die Fachlogik durch Modelle beschrieben wird. Benötigte Datenmodelle und die Dialogverarbeitung können initial in Zusammenarbeit mit BenutzerInnen und Fachexperten definiert werden.

LITERATUR

Bernd Graff (2016). Rassistischer Chat-Roboter: Mit falschen Werten bombardiert. https://www.sueddeutsche.de/digital/microsoft-programm-tay-rassistischer-chat-roboter-mit-falschen-werten-bombardiert-1.2928421 (abgerufen am 12.03.2020).

Bundesagentur für Arbeit. What'sMeBot. https://www.dasbringtmichweiter.de/typischich/whatsmebot-under-sixteen/ (abgerufen am 12.03.2020).

Bundesagentur für Arbeit. Die wichtigsten Merkblätter, Formulare und Anträge der Familienkasse zum Download. https://www.arbeitsagentur.de/familie-und-kinder/downloads-kindergeld-kinderzuschlag (abgerufen am 12.03.2020).

Bundesstadt Bonn. Chatbot Bonn. https://govbot.bonn.de/. (abgerufen am 12.03.2020).

Daimler AG (2016). Premiere vor 25 Jahren: die Mercedes-Benz S-Klasse der Baureihe W 140. https://blog.mercedes-benz-passion.com/2016/02/premiere-vor-25-jahren-die-mercedes-benz-s-klasse-der-baureihe-w-140/.

Daniel Jurafsky and James H. Martin (2009). Speech and Language Processing (2nd Edition). Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Empolis (2019). Chatbots – die Zukunft auch im Kundenservice der Fertigungsindustrie?. https://www.service.express/chatbot-studie-kundenservice-fertigungsindustrie/.

Heise Online (2019). Fast jeder Dritte in Deutschland nutzt Sprachassistenten. https://www.heise.de/newsticker/meldung/Fast-jeder-Dritte-in-Deutschland-nutzt-Sprachassistenten-4443365.html.

Landeshauptstadt Kiel. Wo erledige ich was?. https://www.kiel.de/de/politik_verwaltung/service/ (abgerufen am 12.03.2020).

Nuance. Digitale Dokumentation: Die führende Spracherkennung in Krankenhäusern und Arztpraxen. https://www.nuance.com/de-de/healthcare.html (abgerufen am 12.03.2020).

Nuance. Dragon Legal Group – Spracherkennung für das Rechtswesen. https://www.nuance.com/de-de/dragon/business-solutions/dragon-legal-group.html (abgerufen am 12.03.2020).

Nuance. Lösungen für öffentliche Institutionen – intelligenter, sicherer Bürgerservice durch KI. https://www.nuance.com/de-de/industries/government.html (abgerufen am 12.03.2020).

Oberfinanzdirektion Karlsruhe im Auftrag des Ministeriums für Finanzen Baden-Württemberg. Steuerchatbot. https://zendib-haupt-entwicklung.eu-de.mybluemix.net/ (abgerufen am 12.03.2020).

Oliver Bendel. Definition: Was ist „Chatbot"?. Gabler Wirtschaftslexikon. https://wirtschaftslexikon.gabler.de/definition/chatbot-54248/version-277297 (abgerufen am 12.03.2020).

Ramona Jahn (2000). Ericsson bringt zwei neue Handys auf den Markt. https://www.teltarif.de/arch/2000/kw39/s3124.html.

Stadt Köln. Bewohnerparkausweis. https://www.stadt-koeln.de/service/produkt/bewohnerparkausweis?kontrast=schwarz (abgerufen am 12.03.2020).

Stuart J. Russell and Peter Norvig (2020). Artificial intelligence: a modern approach, 4th Edition. Pearson Education Limited.

Taxfix. Deine Steuererklärung, erstellt in 22 Minuten!. https://taxfix.de/?utm_expid=.m52O-iDYRWGZf69_YXisCA.0&utm_referrer=https%3A%%202F%2Fduckduckgo.com%2F (abgerufen am 12.03.2020).

W3C. Voice Extensible Markup Language (VoiceXML) Version 2.0. https://www.w3.org/TR/voicexml20/ (abgerufen am 12.03.2020).

Abdulmotaleb El Saddik (2018). Digital twins: The convergence of multimedia technologies. IEEE MultiMedia, 25(2), 87-92.

Ted Hunt (2018). Meet the Digital Twin you probably didn't realise you had. https://medium.com/ted-hunt/meet-the-digital-twin-you-probably-didn-t-realise-you-had-eb05096fa9cf (abgerufen am 12.03.2020).

The New York Times. Protect Your Digital Self in a Few Minutes a Day. https://www.nytimes.com/2019/09/12/well/protect-your-digital-self-in-a-few-minutes-a-day.html (abgerufen am 12.03.2020).

Freiheit und Internet. Staatliche Überwachung. https://www.internet-freiheit.de/staatliche-ueberwachung/ (abgerufen am 12.03.2020).

ÜBER DIE AUTOREN

Dr. Stefan Schaffer
Dr. Stefan Schaffer ist Wissenschaftler und Projektleiter im Forschungsbereich Kognitive Assistenzsysteme des Deutschen Forschungszentrums für Künstliche Intelligenz (DFKI). Er studierte Kommunikationswissenschaft und Informatik und promovierte an der Technischen Universität Berlin im Bereich multimodale Mensch-Computer Interaktion. Seine Forschungsinteressen umfassen koversationelle, mobile und multimodale Mensch-Computer-Interaktion in diversen Anwendungsgebieten. In zahlreichen Forschungs- und Industrieprojekten u.a. in den Bereichen Mobilität, Automobil, Steuerinformationen, und Kundendienst war er maßgeblich an Umsetzung konversationeller Benutzerschnittstellen beteiligt. Aktuell leitet er das Projekt „ChiM – Chatbot im Museum". Stefan Schaffer ist beteiligt an der Normungsroadmap für KI des Deutschen Instituts für Normung (DIN) und Mitglied in Programmkomitees internationaler Konferenzen.

Dr. Norbert Reithinger
Dr. Norbert Reithinger ist leitender Wissenschaftler und Research Fellow am Deutschen Forschungszentrum für Künstliche Intelligenz. Nach dem Studium der Informatik an der Universität Erlangen-Nürnberg promovierte er bei Prof. Wahlster an der Universität des Saarlandes zum Thema multimodaler Generierung. Seit 1993 forscht er am DFKI im Fachbereich Kognitive Assistenzsysteme. Er ist verantwortlicher wissenschaftlicher Projektleiter für verschiedene nationale und internationale Projekte. In diesen Projekten erforscht die Gruppe von Dr. Reithinger multimodale Interaktionen in ihren verschiedensten Facetten. Er ist (Ko-) Autor von mehr als 100 wissenschaftlichen Beiträgen und regelmäßig in Programmkomitees großer internationaler Konferenzen.

Janos Standt
Janos Standt ist stellvertretender Bereichsleiter für den öffentlichen Sektor bei der mgm technology partners GmbH. Gemeinsam mit Institutionen wie dem Bayerischen Landesamt für Steuern bringt mgm seit 20 Jahren Anwendungssysteme im öffentlichen Sektor in Produktion. Heute übermitteln mehr als 20 Millionen Bürger und mehr als 2 Millionen Unternehmen in Deutschland Daten über mgm-Software. Janos Standt und mgm legen einen besonderen Fokus auf steuerliche Anwendungen sowie die Digitalisierung von Verwaltungsleistungen für Bürger und Organisationen. Aufgrund dieser Erfahrung und mithilfe der mgm-eigenen Entwicklungsplattform A12 ist mgm in einer idealen Ausgangsposition, um die zeitgerechte OZG-Umsetzung in Zusammenarbeit mit Ländern und Behörden zu ermöglichen.

Roland Krebs
Roland Krebs ist Referatsleiter beim Bayerischen Landesamt für Steuern (BayLfSt). Der Diplom-Finanzwirt leitet das Verfahren ELSTER (ELektronische STeuerERklärung) seit dessen Einführung im Jahr 1999. ELSTER ist ein gemeinsames Projekt der Steuerverwaltungen der 16 Bundesländer und des Bundesministeriums der Finanzen zur sicheren elektronischen Übermittlung von Steuerdaten. ELSTER ist das bekannteste Verfahren im Rahmen von KONSENS (Koordinierte Neue Software-Entwicklung der Steuerverwaltung). ELSTER ist zugleich das erste und größte E-Government Portal der Verwaltung. Die Projektleitung liegt beim Bayerischen Landesamt für Steuern.

IMPRESSUM

Die Kurzstudie basiert auf einer Initiative des Nationalen E-Government Kompetenzzentrums e. V.

Ansprechpartner

Dr. Stefan Schaffer
DFKI GmbH
stefan.schaffer@dfki.de

Dr. Norbert Reithinger
DFKI GmbH
norbert.reithinger@dfki.de

Janos Standt
mgm technology partners GmbH
janos.standt@mgm-tp.com

Roland Krebs
Bayerisches Landesamt für Steuern
roland.krebs@elster.de

Nationales E-Government Kompetenzzentrum e. V.

Pressehaus / 4102
Schiffbauerdamm 40
10117 Berlin

+49 (0)30 80494747
info@negz.org
negz.org

Gestalterische Umsetzung
made in – Design und Strategieberatung
www.madein.io

BERICHTE DES NEGZ

Folgende Kurzstudien sind in der Reihe „Berichte des NEGZ“ bereits erschienen:

Nr. 1 Schuppan, T., Köhl, S., Off, T. (2018). Vollzugsorientierte Gesetzgebung durch eine Vollzugssimulationsmaschine, Berlin. » DOI

Nr. 2 Ogonek, N., Distel B., Ben Rehouma, M., Hofmann, S., Räckers, M. (2018). Digitalisierungsverständnis von Führungskräften, Berlin. » DOI

Nr. 3 Djeffal, C. (2018). Künstliche Intelligenz in der öffentlichen Verwaltung, Berlin. » DOI

Nr. 4 Fadavian, B., Franzen-Paustenbach, D., Rehfeld, D., Schmitt, M., Schweikart, D., Djeffal, C. (2019). Data Driven Government, Berlin. » DOI

Nr. 5 Balta, D., Hofmann, S., Rehfeld, D., Kuhn, P., Krcmar, H. (2019). Sharing Economy: Potential im öffentlichen Sektor, Berlin. » DOI

Nr. 6 Hoepner, P., Welzel, C., Wulff, M. (2019). Identifizierung und Authentifizierung leicht gemacht – die Nutzer ins Zentrum stellen, Berlin. » DOI

Nr. 7 Köhl, S., Müller, H. (2019). Sicherheitsanforderungen und -nachweise bei Cloud-Diensten – Grundlagen für öffentliche Auftraggeber, Berlin. » DOI

Nr. 8 Houy, C., Gutermuth, O., Fettke, P., Loos, P. (2020). Potentiale Künstlicher Intelligenz zur Unterstützung von Sachbearbeitungsprozessen im Sozialwesen, Berlin. » DOI

Nationales E-Government Kompetenzzentrum e. V.
Pressehaus / 4102
Schiffbauerdamm 40
10117 Berlin

+49 (0)30 80494747
info@negz.org
negz.org

ISSN (online) 2626-6032
DOI 10.30418/2626-6032.2020.09