Mittelstand 4.0
Kompetenzzentrum
Usability

Alexa?

Siri.

Hey, Google!

*„Was kannst du wirklich
und wie bringe ich es dir bei?“*

LEITFADEN

Wie sage ich es meinem Computer?

Usability & UX von Chatbots, Smart Speaker, Smart Display und Co.

Mittelstand-
Digital

Gefördert durch:

Bundesministerium
für Wirtschaft
und Energie

aufgrund eines Beschlusses
des Deutschen Bundestages

Sprachassistenten, insbesondere Smart Speaker sind ein schnell wachsender Markt. Deren Nutzung gewinnt zunehmend an Beliebtheit. Das Beratungsunternehmen Canalys prognostiziert in den kommenden Jahren ein zweistelliges Wachstum. Marktführer wie Amazon und Google bzw. Baidu und Alibaba in Asien haben die Entwicklung stark vorangetrieben. Nach Angabe von Amazon gibt es mehr als 85.000 mit Alexa kompatible Smart Home Produkte [1] und mehr als 6.600 Skill Erweiterung von Alexa.

Niedrige Preise, aggressives Marketing sowie eine einfache Bedienbarkeit der Basisdienste haben dafür gesorgt, dass Smart Speaker kein Luxusgut mehr sind, sondern im Massenmarkt angekommen sind. Teilweise ist es günstiger, sich einen einfachen Smart Speaker zu kaufen, um das Licht des Smart Homes zu steuern, als sich einen intelligenten Designschalter zu kaufen.

Die Entwicklung der Sprachassistenten profitiert von einem starken Fortschritt beim Maschinellen Lernen sowie dem Zugriff auf zahlreiche Datenquellen, um im gefragten Moment Informationen zu liefern und Antworten zu geben. Dabei sind die Einsatzszenarien vielfältig. Über Sprachassistenten können Nutzer Reservierungen vornehmen, die Lieblingsmusik abspielen, das Wetter ansagen lassen oder Haushaltsgeräte steuern [2].

Weltweiter Verkauf von Smart Speakern (in Millionen) und jährliches Wachstum

	Q3 2019 Lieferungen	Q3 2019 Marktanteil	Q3 2018 Lieferungen	Q3 2018 Marktanteil	jährliches Wachstum
Amazon	10,4	36,6%	6,3	31,9%	65,9%
Alibaba	3,9	13,6%	2,2	11,1%	77,6%
Baidu	3,7	13,1%	1	4,9%	290,1%
Google	3,5	12,3%	5,9	29,8%	-40,1%
Xiaomi	3,4	12,0%	1,9	9,7%	77,7%
Andere	3,6	12,5%	2,5	12,5%	44,0%
Insgesamt	**28,6**	**100%**	**19,7**	**100%**	**44,9%**

Zusatz: Prozentangaben beinhalten Rundungen
Quelle: Canalys Smart Speaker Analyse (Abverkauf in Lieferungen), Nov 2019

Darüber hinaus können Nutzer die Smart Speaker mittels der Aktivierung zusätzlicher Funktionen (sogenannter Skills) - wie zum Beispiel bei dem Sprachassistenten “Alexa“ von Amazon - an ihre individuellen Vorlieben und Bedürfnisse anpassen. Nichtsdestotrotz kommen Sprachassistenten aktuell noch an ihre Grenzen, wenn die Komplexität bei Aufgaben und Kontext zunimmt. Technische Limitationen erfordern um so mehr Nutzerempathie und ein tiefergehndes Verständnis für eine situationsabhängige Kommunikation und Dialogführung. Bereits in der Mensch-Mensch Interaktion kommt es täglich zu vielerlei Missverständnissen, weshalb die Gestaltung und Umsetzung von Mensch-Maschine Dialogen in natürlicher Sprache nicht trivial sind. Hierfür wird ein strukturiertes und zielgerichtetes Vorgehen bei der Gestaltung der Sprachbedienoberfläche im Einklang mit den zu kommunizierenden Inhalten, als auch den Bedürfnissen, Erwartungen und Zielen der Nutzer benötigt.

Voice Experience Design

> *“User Experience umfasst alle Emotionen, Vorstellungen, Vorlieben, Wahrnehmungen, physiologischen und psychologischen Reaktionen, Verhaltensweisen und Leistungen, die sich vor, während und nach der Nutzung ergeben.”*
>
> DIN ISO 9241-210

Der Begriff User Experience („Nutzererlebnis“) umschreibt alle Aspekte der Eindrücke eines Nutzers bei der Interaktion mit einem Produkt, Dienst, einer Umgebung oder Einrichtung. UX Design fußt auf einer tiefgreifenden Analyse der Ist-Situation (inklusive des Gesamtkontextes) und der Zielgruppe zur Generierung positiver Nutzererfahrungen und einer hohen Markenbindung [4]. Die Usability ist ein Teilaspekt der UX, wobei eine hohe Gebrauchstauglichkeit mittlerweile vorausgesetzt wird.

Im Vergleich zu Tastatur und Maus oder auch „Touch“ ist die sprachbasierte Bedienoberfläche noch relativ jung. Jedoch verbessern Methodes Maschinellen Lernens und die Aufzeichnung von Interaktionsverläufen und deren „Labeling“ (d. h. manuelles Hinzufügen von Metadaten) die Qualität von Sprachassistenten ständig. Trotz dieser (semi-) automatisierten Lernverfahren wird die professionelle Gestaltung sprach- bzw. textbasierter Dialogsysteme durch den Menschen nicht überflüssig. So stark die Gewohnheiten der Menschen sind, so unvorhersehbar ist häufig ihr Verhalten. Das erschwert es den Kontext vorherzusehen, sowie welche Sätze und Wörter der Assistent automatisiert erkennen muss, um eine zufriedenstellende Antwort geben zu können. Vor allem bei komplexen Handlungen wird ein reibungsloser Ablauf erwartet, da ansonsten die Interaktion abgebrochen wird. Für die Entwicklung von Skills ist daher eine professionelle Herangehensweise notwendig, um das menschliche Handeln vollständig sprachlich auszudrücken, um Fehlreaktionen des Sprachassistenten zu minimieren.

Von der Idee zum Dialog

Im Durchschnitt spricht der Mensch etwa 16000 Worte am Tag. Dabei entstehen täglich Konversationen mit einer großen Bandbreite an Zielen. Vom lockeren Smalltalk über reine Informationsgespräche oder überzeugende Debatten verändert sich die Kommunikation kontextspezifisch. Bevor also konkrete Satz- und Wortbausteine für die Umsetzung festgelegt werden, sollten die Idee und das Ziel des Dialogs exploriert werden, um ein Verständnis für die Nutzer aufzubauen.

- Use Cases identifizieren. Dialoge verfolgen unterschiedliche Ziele. Um eine angenehme, jedoch effiziente Nutzerführung zu gewährleisten, sollten im Vorfeld Use Cases festgelegt werden, die den Nutzerkontext eingrenzen.

- Interaktion im Rollenspiel entwerfen. Rollenspiele dienen der Exploration hypothetischer Dialogverläufe. Währenddessen entstehen sowohl Empathie für die Bedürfnisse der Nutzer als auch ein präziseres Verständnis für die Rolle und Aufgaben des Assistenten.

- Konversationspfade festlegen. Die Dokumentation der Rollenspiele sowie die Dialogziele helfen, erste Konversationspfade zu skizzieren, die in diesem Schritt in einer Baumstruktur mit wegweisenden Entscheidungspunkten festgehalten werden sollten.

- Dialoge schreiben. Entlang der Baumstruktur können die verwendeten Sätze und Wörter aus dem Rollenspiel genutzt werden, um erste konkrete Dialogbeispiele zu erstellen.

- KI umsetzen. Hier gilt es, die Baumstruktur in der Dialog Engine (siehe S.5) abzubilden und die Dialogbeispiele einzubinden.

- Nutzertest durchführen. Der Assistent sollte nun auf die implementierte Logik, verwendete Schlagwörter und Ausdrucksweise sowie konstruktive Fehlerstrategien und Feedback von Nutzern getestet werden.

- Lernen und Verbessern. Damit Dialoge nicht nur an der Oberfläche bleiben, sollten sie sukzessiv ausgebaut und optimiert werden. Durch die Nutzerinteraktion lassen sich weitere oder gängigere Schlagworte finden, die zu Beginn nicht berücksichtigt wurden.

Einfache Anwendungen führen zum Ziel

Sprachassistenten sind sehr erfolgreich bei einfachen Anweisungen (siehe Top 10 [2]), während sie bei komplexen Aufgaben (z. B. Reisevorbereitungen, Angebotssuchen usw.) schnell an ihre Grenzen kommen. Wie im Fallbeispiel unten dargelegt, wird hier die User Experience meist noch als mangelhaft bewertet [4].

Top 10 Anwendungsfälle für Sprachassistenten

68% - Schnell nach einem Fakt suchen
65% - Nach dem Weg fragen
47% - Unternehmen/Geschäft suchen
44% - Produkt/Dienstleistung suchen
39% - Einkaufsliste erstellen
31% - Produkte/Dienstleistungen vergleichen
26% - Artikel zu einem Warenkorb hinzufügen
25% - Einen Kauf tätigen
21% - Kontaktaufnahme mit Kundendienst
19% - Feedbackkanal zu Produkt/Dienstleistung

Quelle: Microsoft Voice Report [2]

Denn häufig führt ein breiter Kontext zu komplexen Dialogpfaden, die sowohl vom Designer als auch von der KI nur unzureichend antizipiert oder eindeutig interpretiert werden können. An diese muss sich

der Nutzer trotzdem eng halten, wenn nicht sogar vorher erst erlernen. Die Regeln kann er sich dann wiederum häufig nur schwer merken.

Zur Identifizierung geeigneter Anwendungsfälle sollte insbesondere auf eine gewisse Kontextfreiheit geachtet werden oder darauf, dass der Kontext sehr genau rekonstruiert werden kann. Erfolgversprechend sind auch Anwendungen, in denen eine graphische Oberflächenbedienung derzeit aufwendig, schwierig oder unkomfortabel ist. In jedem Fall erfordert die Sprachassistenz aktuell noch sehr genaue Anweisungen seitens des Nutzers, eingebettet in einen einfachen Dialog. Daher sollten folgende Kriterien bei der ersten Dialogkonzeption beachtet werden:

- In der Regel sollte die Interaktion kurz gehalten werden - mit wenigen Gesprächswechseln.
- Die Aufgabenerledigung erfordert nicht die volle Konzentration der Nutzer und kann beiläufig in der Konversation getätigt werden.
- Ein andere Bedienoberfläche für dieselbe Aufgabe würde eine Zeitverzögerung mit sich bringen und für den Nutzer unbequem sein, wohingegen die Sprachbedienoberfläche die Interaktion und Aufgabenerfüllung vereinfacht.

Beispiel: *„Gibt es hier in der Nähe eine Pizzeria?“*

Die Frage ist kurz und präzise formuliert mit der Erwartung einer unmittelbaren Orientierung und einem Vorschlag für ein Restaurant in der Nähe. Die Antwort würde je nach Anzahl verfügbarer Pizzerien ein bis drei Gesprächswechsel erfordern. Zuerst wird die Anzahl der Möglichkeiten, die Entfernung und Bewertung angesagt. Im nächsten Schritt fragt der Assistent, ob noch weitere Informationen benötigt werden. Die Antwort erfordert keine volle Konzentration, wodurch sich der Nutzer während der Interaktion an der Kreuzung umsehen kann. Zusätzlich ist das Sprechen in dem Fall komfortabler und schneller als zu tippen.

Den richtigen Interaktionskanal wählen

Der Nutzer kann über eine Reihe von Kanälen mit sprachbasierter Interaktion in Berührung kommen. Obwohl im Hintergrund die gleiche technische Basis (siehe S. 5) verwendet wird, beeinflusst die Wahl des Interaktionskanals die UX – und damit die Gestaltung des Sprachassistenten. Die Bedienoberfläche einer bestehenden Anwendung kann die graphische Oberfläche um einzelne sprachliche Interaktionselemente erweitern. Eine weitere Möglichkeit ist das gesamte Produkt um eine vollständig sprachbasierte Anwendung zu erweitern, wie zum Beispiel für eine angenehme und effiziente Interaktion mit intelligenten Geräten in der unmittelbaren Umgebung. Allerdings gibt es noch keine etablierten Gestaltungsrichtlinien für sprachbasierte Bedienoberflächen, die von graphische Bedienelementen stark abweichen.

Textbasierte Interaktion

Mit Chatbots werden Programme bezeichnet, die nach Initialisierung eigenständig laufen (= bots) und stellvertretend für Mensch zu Mensch Gespräche (= chats) führen. Mittlerweile werden sie zahlreich für Kundenanfragen eingesetzt oder als textbasierter Begleiter (Companion) in einer eigenen App umgesetzt. Im Gegensatz zu Companions, die in Tiefe ein breites Feld abstecken, sind klassische Chatbots sehr aufgabenfokussiert und enthalten ca. drei bis sieben Gesprächswechsel [6]. Zudem bringt die Textform Limitationen mit sich, da der Nutzer seine Eingaben tippen und dafür das Smartphone oder eine Tastatur mit PC zur Hand haben muss.

Sprachbasierte Interaktion

Intelligente Persönliche Assistenten (IPA) sind sprachbasierte Chatbots, die eine breite Wissensbasis haben, aber häufig Nutzeranfragen nur oberflächlich behandeln können. Deshalb ist dieser Austausch häufig auf ein bis drei Gesprächswechsel limitiert [6]. Der Vorteil von IPA ist, dass sie mittlerweile auch stationär verfügbar sind und sich großer Beliebtheit als intelligenter Lautsprecher und zur Steuerung von Smart Home Anwendungen erfreuen können. Smartphone-basierte IPA eignen sich gut für den „handfreien“ Einsatz im Auto.

Multimodale Interaktion

Als Erweiterung der rein sprachbasierten Interaktion entsteht beispielsweise bei Smartphones durch die Einbindung des Displays eine multimodale Interaktion. Zudem sind nun auch Geräte wie der Echo Show erhältlich, die stationär Smart Speaker über eine visuelle Komponente erweitern. Nutzer verarbeiten visuelles Feedback schneller, sind aber dadurch auch auf ein ausreichendes Sichtfeld angewiesen. Der Bildschirm bietet eine Möglichkeit, die Fehler, die bei der Sprachverarbeitung derzeit noch auftreten, zu überbrücken und das Spracherlebnis im Allgemeinen anzureichern. Komplexe Aufgaben, die

reich an Informationen sind, können so besser abgebildet werden und den Nutzer in der Aufgabenerfüllung effektiver und effizienter unterstützen. Der Verkauf und die Nachfrage nach Smart Displays stieg laut einer Studie von canalys für Q3 im Jahr 2019 weltweit um 20 Prozent [2].

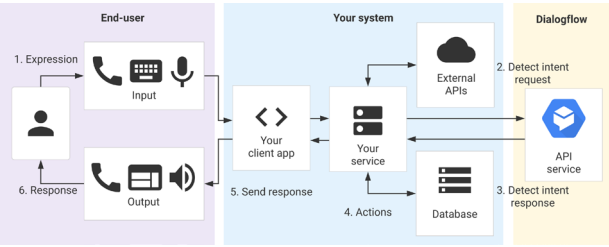
Je nach Aufgabe und Kontext eignen sich die unterschiedlichen Modalitäten unterschiedlich gut darin, den Nutzer bei seinen Aktivitäten zu unterstützen. Zudem sind die Gestaltungsrichtlinien nicht einfach auf jede Modalität übertragbar. Es müssen dementsprechend die Eigenheiten berücksichtigt werden, obwohl alle dialogbasierte Systeme sind.

Erproben im Rollenspiel

Die Erprobung neuer Interaktionen und Dienstleistungen im Rollenspiel hat eine lange Tradition im UX Design. Es ist eine Art Prototyp, der sehr schnell umgesetzt werden kann, indem eine Idee eines Nutzungsszenarios aus unterschiedlichen Perspektiven durchgespielt wird. Das Kernziel ist die Idee für alle Beteiligten und potentielle Nutzer greifbar zu machen.

Zu Beginn sollte das Rollenspiel innerhalb des Designteams durchgeführt werden, um den Use Case genauer zu definieren und diesen hinsichtlich der Rolle des Assistenten und der Nutzerbedürfnisse sowie deren Ziele zu evaluieren. Zur Validierung oder Weiterentwicklung von Kernhandlungen, Sprachgebrauch und Dialogabläufen kann es ebenfalls sinnvoll sein, potentielle Nutzer zu einem Workshop einzuladen, um mit ihnen gemeinsam Rollenspiele zu erproben und Gestaltungsrichtlinien abzuleiten. Diese Methode ist kostengünstig umzusetzen und erfordert keine zusätzlichen Werkzeuge. Zusatzmaterial kann jedoch verwendet werden, um eine Immersion im Szenario zu fördern. Die Erfahrungen werden von den Teilnehmern als Schauspielende sowie Zuschauer festgehalten und diskutiert.

Dialog Engines nutzen



Quelle: Google [5]

Die *Dialog Engine* ist verantwortlich für die technische Ausführung der Interaktion zwischen dem Nutzer und dem Assistenten. Im Wesentlichen unterscheidet man zwischen einer End-Nutzer Bedienoberfläche, über die eine akustische oder textuelle Spracheingabe erfolgt, und dem technischen, vor dem Nutzer verborgenen Teil der Sprachverarbeitung. Dieser setzt sich aus der Spracherkennung (ASR), Sprachverarbeitung (NLP) nämlich Sprachverständnis (NLU) und Generierung von natürlicher Sprache (NLG) sowie der Umwandlung von Text in Sprache (TTS) zusammen.

- ASR (Automated Speech Recognition): Das Sprachsignal als Eingabe wird verarbeitet und in Text umgewandelt.
- NLU (Natural Language Understanding): Das Verständnis der natürlichen Sprache versucht die Bedeutung im Textzusammenhang zu erkennen und nicht nur auf einzelne Schlüsselworte zu reagieren.
- NLG (Natural Language Generation): Die Antwort wird wiederum in natürlicher Sprache ausgegeben, wofür die benötigten Vokabeln und Satzstrukturen generiert werden.
- TTS (Text-to-Speech): Im Gegensatz zu Chatbots, muss für eine Sprachausgabe der generierte Text in ein akustisches Signal umgewandelt werden.

Abgesehen von diesen technischen Prozessen im Hintergrund, wird je nach Anwendungsfall und Dialog Engine automatisiert eine Wissensbasis oder Service zur Beantwortung der Nutzeranfrage verknüpft, wie zum Beispiel Wikipedia oder Zeitstempel. Die Anfrage selbst wird von der Engine im Wesentlichen auf 3 Teile geprüft. Zuerst wartet das System auf ein „Weckwort“ wie „Alexa“ oder „Hey Google“, um mit der Aufzeichnung und Verarbeitung zu beginnen. Dann wird die Nutzeranfrage auf den „Intent“ geprüft, um die Absicht hinter der Aussage zu erkennen und die passenden Informationen zu geben. Zusätzlich wird unter Anwendung von NLU versucht den Kontext richtig einzuordnen, indem auch die Aussage analysiert wird. Zum Schluss wird die Antwort in Textform in Sprache umgewandelt und ausgegeben.

Für eine erfolgreiche Interaktion werden viele Synonyme und Beispielsätze benötigt, die die Trefferquote erhöhen, um zu erkennen was der Nutzer genau möchte oder wie er auf der Suche nach der richtigen Antwort geleitet werden kann. Dabei ist es nicht möglich vollständig alle Synonyme abzude-

cken oder vorherzusehen. Deshalb ist es für eine reibungslose Usability umso wichtiger den Kontext und Sprachgebrauch vorab genau zu studieren. Zusätzlich ist eine konstruktive Fehlerstrategie erforderlich, die bei nicht abgedeckten Fällen trotzdem zu einem zufriedenstellenden Ergebnis führt.

Dialogflow und Co.

Bei Conversational AI Platforms handelt es sich um Plattformen, Bedienoberflächen oder Dienste, die Natural Language Processing (NLP) zugänglich machen, um eigene sprach-basierte Assistenten (Agents) zu bauen. Diese können dann die Basis für eine eigene Anwendung sein oder in andere Chat Oberflächen wie Nachrichtendienste, Soziale Medien oder Websites integriert werden. Anbieter solcher NLP-Dienste sind beispielsweise Dialogflow (Google), Watson (IBM), Lex (Amazon) und Rasa (Open Source).

Dialogflow als Teil der Google Cloud Plattform bietet außerdem Programmierschnittstellen (API), um selbst Anpassungen für individuelle Anforderungen vorzunehmen. Insgesamt ist es eine angenehme Entwicklungsoberfläche, die mit zusätzlichen Werkzeugen, einem Code-Editor, unterschiedlichen Libraries für eine intelligente Verknüpfung von Diensten und einem erweiterten Funktionsumfang ausgestattet ist. Es kann einer der vorgefertigten, zur freien Verfügung gestellten Agents von Dialogflow genutzt oder modifiziert werden, oder es wird, unterstützt durch eine umfangreiche Dokumentation, eigenständig ein Assistent erstellt. Damit ist ein schneller und einfacher Einstieg möglich. Die Agents können für die gewünschte Anwendung wie Amazon Echo, Google Home oder Facebook Messenger usw. exportiert und eingebunden werden.

Die Entwicklung kann auch in Amazon's eigenem Werkzeug erfolgen, jedoch sollte beachtet werden, dass dieser Skill dann nur im geschlossenen Alexa System einsatzfähig ist. Die Open Source Plattform Rasa bietet den Vorteil, dass der Quellcode vollständig eingesehen und konfiguriert werden kann. Zudem ist es ein kostengünstiger Ansatz, da keine Lizenzgebühren anfallen.

Nutzertest

Die Einbindung von Nutzern ist demnach ein essentielles Element in der Gestaltung von Sprachassistenzsystemen. Häufig stellen sich mittelständische Unternehmen aber die Frage, wie genau man bei der Einbindung von echten Nutzern vorgehen kann um möglichst effizient zu nutzbringenden Aussagen und Einblicken zu gelangen. Im Folgenden beschreiben wir eine kürzlich durchgeführte Nutzerstudie, die zum Ziel hatte, Gestaltungskriterien für eine verbesserte sprachbasierte Bedienoberfläche zu erheben. Bislang gibt es keine verlässlichen Richtlinien, die eine gute Usability und UX gewährleisten und gleichzeitig auf die Bedienoberfläche von Sprachassistenzsystemen abgestimmt sind. Plattformen wie Dialogflow oder Amazon stellen zwar hierzu umfangreiche Dokumentationen und Empfehlungen bereit, jedoch sind diese häufig noch zu generisch, um der Komplexität abgebildeter Nutzungsszenarien gerecht zu werden. In bisherigen Nutzertests stand die Auswertung der Sprachbefehle und die Erforschung der Nutzerakzeptanz der Geräte im persönlichen Umfeld im Fokus. Für ein tieferes Verständnis der Usability Probleme kann eine aufgabenbasierte Evaluation unter kontrollierten Bedingungen mehr über die Hindernisse und der zugehörigen Fehlerbehebungsstrategien der Nutzer Aufschluss geben. Zusätzlich lassen sich mithilfe quantitativer Fragebögen ein Gesamteindruck erfassen, sowie konkrete Verbesserungsvorschläge hinsichtlich des Systems und der Anwendungen erheben.

Fallbeispiel: Usability Test „Echo Show“

In Zusammenarbeit mit eresult haben wir die Nutzerinteraktion mit Amazon's Echo Show gemeinsam innerhalb eines labor-basierten Usability Tests evaluiert. Der Echo Show ist die zweite Generation des intelligenten Lautsprechers von Amazon erweitert durch einen touch-sensitiven Bildschirm. In diesem Nutzertest untersuchen wir, wie weit visuelle Hilfestellungen den Sprachkanal unterstützen oder Hindernisse, ausgelöst durch die Sprachinteraktion, wieder ausglichen werden können.

Alexa lässt sich durch Skills um Dienste von Drittanbietern erweitern, die neben einfachen Anwendungen auch semi-komplexe Kontexte, wie „Kochen“ oder „Angebote suchen“, unterstützen. Dabei ist Alexa als Assistent ein Teil dieser Plattform, um die Nutzung und Steuerung der Skills zu ermöglichen. Obwohl Nutzer semi-komplexe Anwendungen aufgrund frustierender Erfahrung meiden [4], können diese im Kern einen echten Mehrwert zur Verbesserung der alltäglichen Abläufe zuhause bieten. Daher wollen wir auf Ziele ausgelegte Szenarien untersuchen, die drei bis zehn Gesprächswechsel erfordern.

- **Untersuchungsgegenstand:** Untersucht werden Skills von deutschlandweit bekannten

Handelsmarken wie REWE und Real für das Stöbern in aktuellen Angeboten, und Chefkoch und Kitchen Stories als beliebte Küchen-Apps für Rezepte, die von Amazon als Partner Skills beworben werden. Bring! ist eine Einkaufslisten-App für das Smartphone, REWE bietet neben Rezepten und Angeboten auch eine integrierte Einkaufsliste an. Amazon hat eine eigene Liste zum Notieren von Artikeln für allerlei Zwecke.

- **Teilnehmer**: Unsere Stichprobe (n=20) entspricht der demographischen Verteilung Deutschlands hinsichtlich Alter, Bildung, Einkommen, Migrationshintergrund, Haushalt und Familientyp. Die Teilnehmer (n=20) waren im Durchschnitt 45 Jahre alt und die Hälfte hatte bereits Erfahrungen mit Sprachassistenten. Sie wurden für ihre Zeit mit jeweils 50 € entschädigt.

- **Aufbau:** Der Echo Show wurde im Labor aufgebaut und an das WIFI angeschlossen. Es wurde ein Benutzerkonto eingerichtet und die notwendigen Skills installiert. Der Nutzer war somit zu keiner Zeit gezwungen, private Daten einzugeben. Das Labor selbst wurde mit einer Kamera ausgestattet, um die Interaktion und den Bildschirm aufzuzeichnen. Wir führten die Studie im Juli 2019 in Deutschland durch, und jede Sitzung dauerte durchschnittlich 60 Minuten.

- **Vorgehen**: Der Interviewer erklärte den Nutzern das Szenario schrittweise:

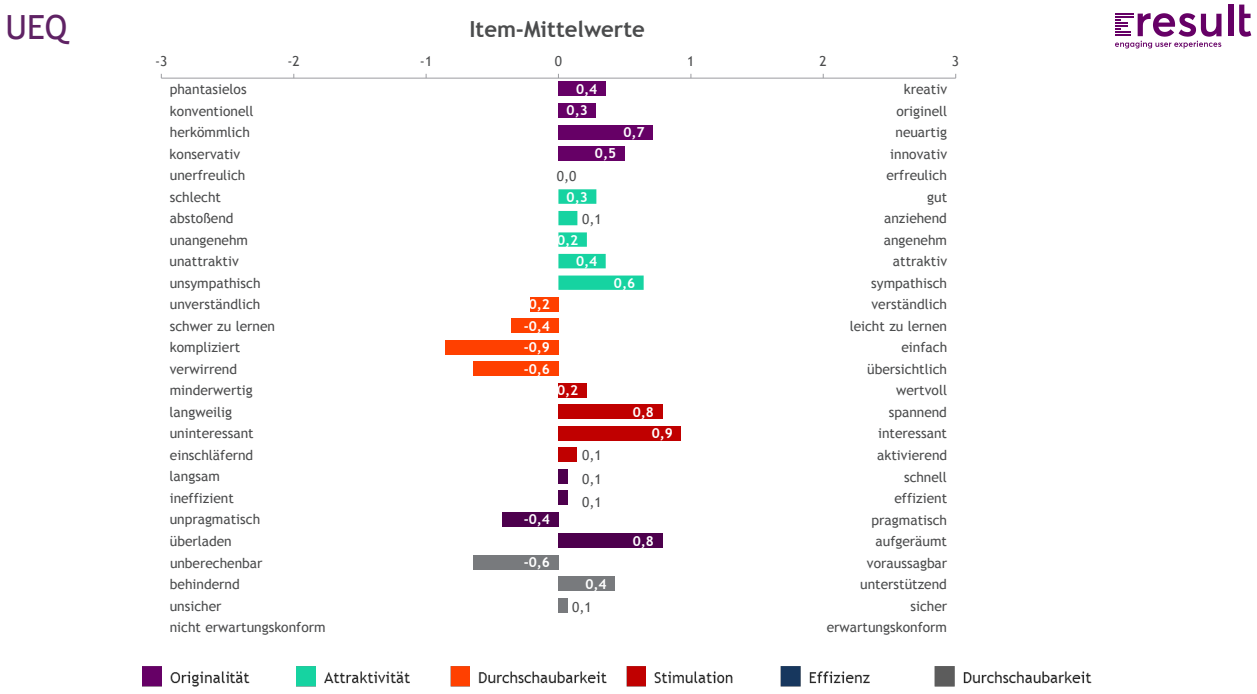
1. Drei Minuten für eine kostenlose, individuelle Untersuchung ohne Einschränkungen in Bezug auf Fähigkeiten oder Nutzung.

2. Suche nach Angeboten oder Rezepten, um sich inspirieren zu lassen.

3. Erstellung und Bearbeitung einer Liste.

   Um eine möglichst intuitive Vorgehensweise zu unterstützen, reagierte der Interviewer flexibel auf die Reihenfolge der Schritte innerhalb des Szenarios, stellte jedoch sicher, dass alle Skills getestet wurden. Die Nutzer wurden angehalten, laut zu denken, und nach jedem Schritt ein kurzes Feedback zu ihrem Erlebnis zu geben. Schließlich füllten sie einen standardisierten User Experience Fragebogen (UEQ) über ihre bisherigen Erfahrungen aus und beantworteten einige abschließende Fragen.

- **Analyse:** Danach wurden die erhobenen Daten analysiert und die Eindrücke, Sprachbefehle und Visualisierungen im Team diskutiert. Wir konzentrierten uns dabei insbesondere auf Hindernisse und die Nutzung von Displays sowie auf die Unterscheidung zwischen den Interaktionen selbst und dem anschließenden Feedback des Benutzers. Zusätzlich wurden die qualitativen Ergebnisse mit denen des UEQ verglichen.



Ergebnisse

Weniger erfahrene Nutzer erwarteten eine natürliche Interaktion wie sie von der Werbung versprochen wurde: zufriedenstellende Antworten auf ihre Fragen, unterstützt durch visuelle Darstellung. Wohingegen die erfahrenen Benutzer skeptischer hinsichtlich der tatsächlichen Fähigkeiten des Echo Shows waren und im Vergleich zu ihren bisherigen Erfahrungen mit Sprachassistenten auf Verbesserungen hofften.

Die Ergebnisse des Fragebogens UEQ zeigen, dass die Interaktion mit dem Gerät in den Dimensionen Originalität, Attraktivität und Stimulation tendenziell als positiv und angenehm bewertet wurden. Die Effizienz wird als eher unpraktisch eingestuft, dafür aber organisiert. Aus den Beobachtungen lässt sich erkennen, dass dies mit einer übersichtlich dargestellten Anzahl an Informationen zusammenhängt, die jedoch häufig als unzureichend und nicht hilfreich bewertet wurden. Die Nutzer betonten häufig die Ineffizienz und Umständlichkeit des Geräts. Die Zuverlässigkeit des Echo Shows empfanden die Nutzer zwar unterstützend wegen der unterschiedlich angebotenen Skills, die für ihren Alltag geeignet schienen, aber schließlich bewerteten sie den Echo Show aufgrund der mangelnden Kontrolle und der willkürlichen Reaktionen als wenig durchschaubar.

Usability Evaluation

Während der Evaluation habe wir vorläufige zehn VUI-Richtlinien [7] angewandt, die sich an Nielsen‘s heuristischen Usability Kriterien orientieren, um die Ergebnisse strukturiert auswerten zu können. Die aufgeführten Probleme können teils mehreren Kategorie zugeordnet werden.

► 1 Sichtbarkeit und Feedback

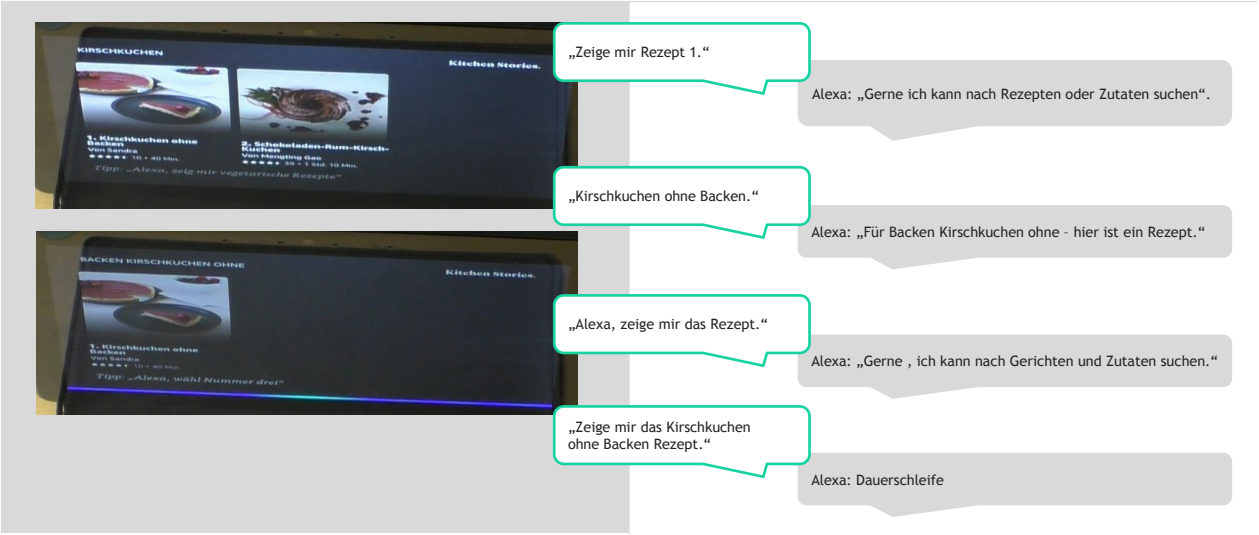
Aufgrund der Ähnlichkeit zu einem Tablet wurden eine Touch-Steuerung und weitere ausgebaute visuelle Funktionen seitens der Nutzer erwartet. Jedoch waren die Nutzer von dem unzureichenden Nutzen des Bildschirms enttäuscht. Sie wünschen sich einen Überblick über die Skills, Funktionsweise und generell mehr visuelle Führung, um die Menüstruktur zu verstehen und zu erlernen. Beispielsweise wurde auch die Darstellung der Rezepte oder Angebote als zu wenig interaktiv empfunden im Vergleich zu einem PDF oder einem haptischen Angebotsblatt.

► 2 Abbilden

Die Nutzer versuchten sich die Funktionsweise und mögliche Steuerung der Skills zu erklären, indem sie die Aufgaben mit ihren alltäglichen Handlungsweisen abglichen. Dabei griffen sie auf altbewährte Praktiken wie die Online-Suche nach Produkten und Rezepten am Laptop und die Erstellung einer Einkaufsliste mit Stift und Papier zurück. Dabei wünschen sie sich mit Alexa mehr effektiven Austausch, der einen Vorteil gegenüber ihren bewährten Praktiken bietet.

► 3 Kontrolle und Freiheit der Benutzer

Das Gerät schränkt die Nutzer in ihrer Freiheit zu sprechen und ihre Bedürfnisse auszudrücken ein. Dies wirkt sich negativ auf die Steuerung des Geräts und die Navigation durch die Menüstrukturen der Skills aus. Dabei kommt es schon bei der Aktivierung und Ansprache der Skills zu Problemen, wegen uneinheitlicher Muster, zu wenig Fehlertoleranz bei der Benennung und nicht nachvollziehbarer Priorisierung von Skills. So werden manch-



mal beim Aufruf der REWE Angebote die Amazon eigenen Angebote gezeigt und vorgelesen. Die Navigation über schnelle Aktionen wie „überspringen“, „weiter“ und „zurück“ sollten jederzeit verfügbar sein. Zudem erschwerte das willkürliche Systemverhalten den Nutzern das Erlernen von Sprachbefehlen und Schlüsselworten. Deshalb sollten autonome Entscheidungen des Nutzers unterstützt werden, wenn diese das beabsichtigen. Systemeigene Vorschläge sollten angeboten werden, wenn der Nutzer dies erwartet oder sich explizit wünscht (siehe S. 8).

► 4 Konsistenz

Die Möglichkeit zwischen den Skills zu wechseln fanden die Nutzer prinzipiell gut, jedoch führte es häufig zu Orientierungslosigkeit im System. Derzeit fehlen standardisierte Befehle innerhalb eines Skills oder der Plattform als Ganzes. Beispielsweise sollte darauf geachtet werden, dass der Befehl, Alexa beim Sprechen zu pausieren, sich vom Befehl den Skill zu beenden, unterscheidet. Zudem sollten die bereitgestellten Funktionen und Inhalte konsistent im Markenauftritt sein. Zum Beispiel sollten geläufige Bezeichnungen aus den Supermärkten beibehalten werden, um die Erwartungshaltung der Nutzer zu erfüllen.

► 5 Fehlervermeidung

Die Nutzer gewöhnen sich schnell an die Verwendung erfolgreicher kurzer Schlagwörter und Sätze. Jedoch kann die mühsame Suche, das richtige Wort zu erraten, zu Frustration und Abbruch der Interaktion führen. Deshalb sollten Nutzer in diesem Sinne seitens Alexa proaktiv geeignete Begriffe vorgeschlagen bekommen.

► 6 Erkennen statt Abrufen

Semi-komplexe Aufgaben beinhalten mehr Information und eine längere Interaktion innerhalb eines Dialogablaufs. Obwohl alle Nutzer befürworteten, dass visuelle Darstellungen vor allem beim Einkauf von Produkten hilfreich sind, waren sie mit der Umsetzung unzufrieden. Sie wünschten sich mehr Gleichgewicht von visueller und gesprochener Information. Der Informationsbedarf unterscheidet sich je nach Nutzer beim Vorlesen von Mengen, Empfehlungen oder Länge der Produktnamen.

► 7 Flexibilität und Effizienz

Der flexible Einsatz von Skills ermöglicht eine größtmögliche Unterstützung des Alltagskontexts der Nutzer. Jedoch muss deutlich verständlich gemacht werden, welcher Skill gerade von Alexa aktiviert ist. Bei komplexeren Aufgaben kann daher davon ausgegangen werden, dass der Nutzer in Sichtweite des Geräts ist und die Anzeige dementsprechend stärker eingebunden werden kann. Flexibilität bedeutet zum Beispiel, während des Kochens oder Putzens benötigte Produkte auf eine Liste zu setzen, für andere die Unterstützung, z. B. durch ein interaktives Kochbuch, das sie Schritt für Schritt durch die Zubereitung führt. Zudem wurde die Touch Funktion von den meisten Nutzern als Abkürzung begrüßt. Sei es, um das Gespräch mit Alexa zu unterbrechen, um zusätzliche Informationen zu erhalten oder um den Eintrag in einer Einkaufsliste manuell zu korrigieren.

► 8 Minimalismus

Bei diesem Gerät wird Minimalismus als Balance zwischen akustischer und visueller Informationsausgabe verstanden. Das bedeutet, dass der Bildschirm mehr Informationen zur Verfügung stellen könnte, als es momentan der Fall ist und Details nicht immer vorgelesen werden müssen. Jedoch muss sich dafür das Gerät auch in Sichtweite befinden.

► 9 Möglichkeiten, Fehler zu erkennen und zu beheben

Die am häufigsten angewandte Strategie zur Fehlerbehebung, war das Raten und Testen unterschiedlicher Formulierungen oder Schlüsselwörter. Einige Nutzer waren der Meinung, dass sie ein echtes Wörterbuch benötigten und baten um konkrete Vorschläge für Schlüsselwörter. Zudem fühlten sie sich bei Hilfestellungen häufig in Dauerschleifen festgefahren und fanden die Wiederholungen nicht konstruktiv. Der Bildschirm bot hier auch keine zusätzliche Unterstützung zur Fehlererkennung.

► 10 Hilfe und Dokumentation

Vor allem ältere Nutzer waren auf Hilfe angewiesen und erwarteten oder suchten ausdrücklich ein Tutorial. Die meisten Nutzer, unabhängig ihres Alters, wünschten sich mehr kontextbezogene Erklärungen und Unterstützung. Die zusätzliche Bildschirmanzeige wurde als große Unterstützung wahrgenommen, die ein gegenseitiges Verständnis mit Alexa fördert und die kognitive Belastung des Nutzers verringert. Die Benutzer waren dankbar für visuelle Tipps und Vorschläge für mögliche Aktionen und nutzten diese häufig. Sie lehnten jedoch App-basierte Hilfestellungen ab, da dies den Vorteilen sprachbasierter Geräte widerspricht und wieder das Smartphone in Reichweite erfordert.

Fazit

Insgesamt, blieb der Echo Show hinter den Erwartungen der Nutzer zurück. Besonders enttäuschend war der Bildschirm, der keinen sichtbaren Mehrwert geboten hat und keine natürliche Interaktion unterstützt hat. Bereits seit drei Jahrzehnten etablierte Gestaltungsrichtlinien für graphische Bedienoberflächen wurden weitestgehend nicht umgesetzt. Im Alltagsgebrauch würden die Nutzer eine Pallette an Skills installieren um den Funktionsumfang von Alexa zu erweitern. Jedoch müssten diese besser in der Interaktion und Navigation aufeinander abgestimmt werden. Die meisten Benutzer erwarteten eine Art App-Übersicht, einen Startbildschirm mit Darstellungen der verfügbaren Funktionen und Icons oder Schaltflächen als Hilfsmittel zur Interaktion. Bislang unterscheiden die Nutzer nicht groß zwischen dem Assistenten als Plattform, den Skills als zusätzliche Anwendungen von Drittanbietern und der sprachbasierten Bedienoberfläche als Hauptsteuerung.

Unabhängig von Alter oder Erfahrungsniveau benötigen die Nutzer viel zusätzliche Unterstützung und Feedback, damit NLP-Fehler ausgeglichen werden. Um langfristig einen Mehrwert zu schaffen, müssen die Skills in die Tiefe entwickelt werden, um den Nutzern zu helfen ihre Ziele zu erreichen. Ansonsten führen oberflächliche Reaktionen und Unterstützung seitens Alexa zu Frustration und Ablehnung. Daher ist Konsistenz der Schlüssel, sowohl bei den Fähigkeiten selbst als auch bei den globalen Befehlen auf der gesamten Plattform.

Die Anzeige hat insgesamt ein nützliches Potenzial, um komplexe Informationen zu strukturieren, mehr und prägnantere Rückmeldungen zu liefern, die Sicherheit und Bestätigung der eingereichten Ein- und Ausgaben zu verbessern und die kognitive Belastung beim Vorlesen zu reduzieren. Indem wir die Art der Aktion und den Kontext berücksichtigen, für die wir uns entscheiden, können wir den Audio- und visuellen Kanal ausbalancieren, um die gewünschte Erfahrung zu verbessern.

Spracherlebnisse für das eigene Unternehmen gestalten

Chatbots sind als Dienstleistung mittlerweile weit verbreitet und in digitale Angebote integriert. Der zunehmende Einsatz von Sprachassistenten in Haushalten bietet neue Möglichkeiten in der Kundeninteraktion und –beziehung, wird jedoch als Kanal selten genutzt, obwohl neue Erlebnisse mit einem entsprechenden Mehrwert Kunden in ihrem persönlichen Zuhause zur Verfügung gestellt werden können. Die Herausforderungen liegen hier zum einen in der Entscheidung für den richtigen Kanal zur Vermittlung von Inhalten wie beispielsweise ein Chatbot oder Sprachassistent, als auch die kundengerechte Gestaltung von Sprachinhalten.

Die Anwendungsfelder sind dabei vielfältig: Beispielsweise könnte ein Bestell- und Lieferservice für ein Abendessen mittels Sprachassistent umgesetzt werden. Um auch mittelständigen Unternehmen den Einsatz eines Sprachassistenzsystems näher zu bringen, soll unser zweiteiliger Workshop den Teilnehmern die Konzeption und Umsetzung eines Sprachbots vermitteln. Hierfür werden als erstes benötigtes Wissen und Aufgaben in einem einfachen Rollenspiel exploriert und ein erster Interaktionsentwurf festgehalten. Im nächsten Schritt wird dieser in einem Experiment mit den Teilnehmern getestet und iterativ verbessert, bis das Konzept und Interaktionsmodell eine gewünschte Qualität erreicht. Danach kann der Sprachassistent implementiert und getestet werden. Dafür wird das Programm Google Dialogflow verwendet. Dieses ermöglicht den Nutzern ohne großes IT-Wissen die erfolgreiche Erstellung eines Sprachassistenzsystems.

Im Anschluss an den zweiteiligen Workshop, gibt es außerdem das Angebot, in einem zweiteiligen E-Learning erste Anwendungsfälle aus dem eigenen Unternehmen in Form eines Sprach- oder Chatbots mit unserer Hilfe zu implementieren.

Anmeldung und weitere Informationen finden sie auf www.kompetenzzentrum-usability.digital.

Referenzen

[1] Amazon Introduces an Array of New Devices and Features to Help Make Your Home Simpler, Safer, and Smarter, Sptember 2019, https://press.aboutamazon.com/news-releases/news-release-details/amazon-introduces-array-new-devices-and-features-help-make-your

[2] 2019 Voice report: Consumer adoption of voice technology and digital assistants, April 2019

[3] Canalys Smart Speaker Analysis 2019, www.canalys.com/newsroom/worldwide-smartspeaker-Q3-2019

[4] Cho, Minji, Sang-su Lee, and Kun-Pyo Lee. „Once a Kind Friend is Now a Thing: Understanding

How Conversational Agents at Home are Forgotten.“ Proceedings of the 2019 on Designing Interactive Systems Conference. 2019.

[5] https://cloud.google.com/dialogflow/docs/api-overview?hl=de, zuletzt aufgerufen 21.02.2020

[6] Grudin, Jonathan, and Richard Jacques. „Chatbots, humbots, and the quest for artificial general intelligence.“ Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019.

[7] Murad, Christine, et al. „Design guidelines for hands-free speech interaction.“ Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct. 2018.

Autoren

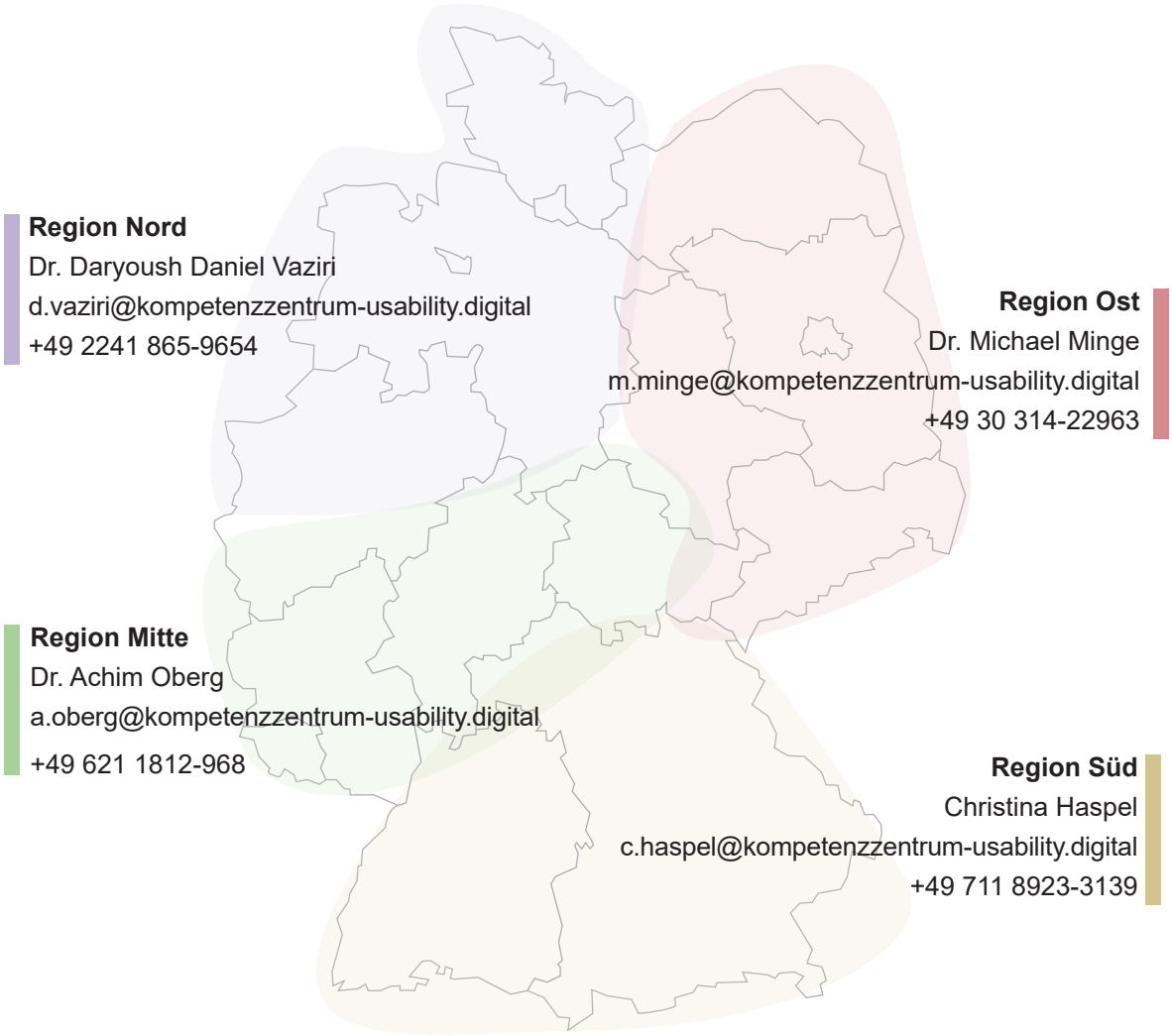
Prof. Dr. Gunnar Stevens
Professur für IT-Sicherheit
und Verbraucherinformatik,
Fakultät III, Universität Siegen
g.stevens@kompetenzzentrum-usability.digital

Margarita Esau
Wissenschaftlicher Mitarbeiter
User Research & Voice UX
Hochschule Bonn-Rhein-Sieg
margarita.esau@h-brs.de

Elske Ludewig M.A.
Principal UX Consultant &
Managing Director
eresult GmbH
elske.ludewig@eresult.de

Maya Fricke
Principal User Experience
Designerin,
eresult GmbH
maya.fricke@eresult.de

Ihr Kontakt im Kompetenzzentrum



Region Nord
Dr. Daryoush Daniel Vaziri
d.vaziri@kompetenzzentrum-usability.digital
+49 2241 865-9654

Region Ost
Dr. Michael Minge
m.minge@kompetenzzentrum-usability.digital
+49 30 314-22963

Region Mitte
Dr. Achim Oberg
a.oberg@kompetenzzentrum-usability.digital
+49 621 1812-968

Region Süd
Christina Haspel
c.haspel@kompetenzzentrum-usability.digital
+49 711 8923-3139

Impressum

Herausgeber:
Hochschule Bonn-Rhein-Sieg
Grantham-Allee 20
53757 Sankt Augustin
www.h-brs.de

V.i.S.d.P. Margarita Esau

Gestaltung, Redaktion und Produktion:
Margarita Esau

www.kompetenzzentrum-usability.digital

Was ist Mittelstand-Digital?
Mittelstand-Digital informiert kleine und mittlere Unternehmen über die Chancen und Herausforderungen der Digitalisierung. Die geförderten Kompetenzzentren helfen mit Expertenwissen, Demonstrationszentren, Best-Practice-Beispielen sowie Netzwerken, die dem Erfahrungsaustausch dienen. Das Bundesministerium für Wirtschaft und Energie (BMWi) ermöglicht die kostenfreie Nutzung aller Angebote von Mittelstand-Digital.

Der DLR Projektträger begleitet im Auftrag des BMWi die Projekte fachlich und sorgt für eine bedarfs- und mittelstandsgerechte Umsetzung der Angebote. Das Wissenschaftliche Institut für Infrastruktur und Kommunikationsdienste (WIK) unterstützt mit wissenschaftlicher Begleitung, Vernetzung und Öffentlichkeitsarbeit. Weitere Informationen finden Sie unter www.mittelstand-digital.de