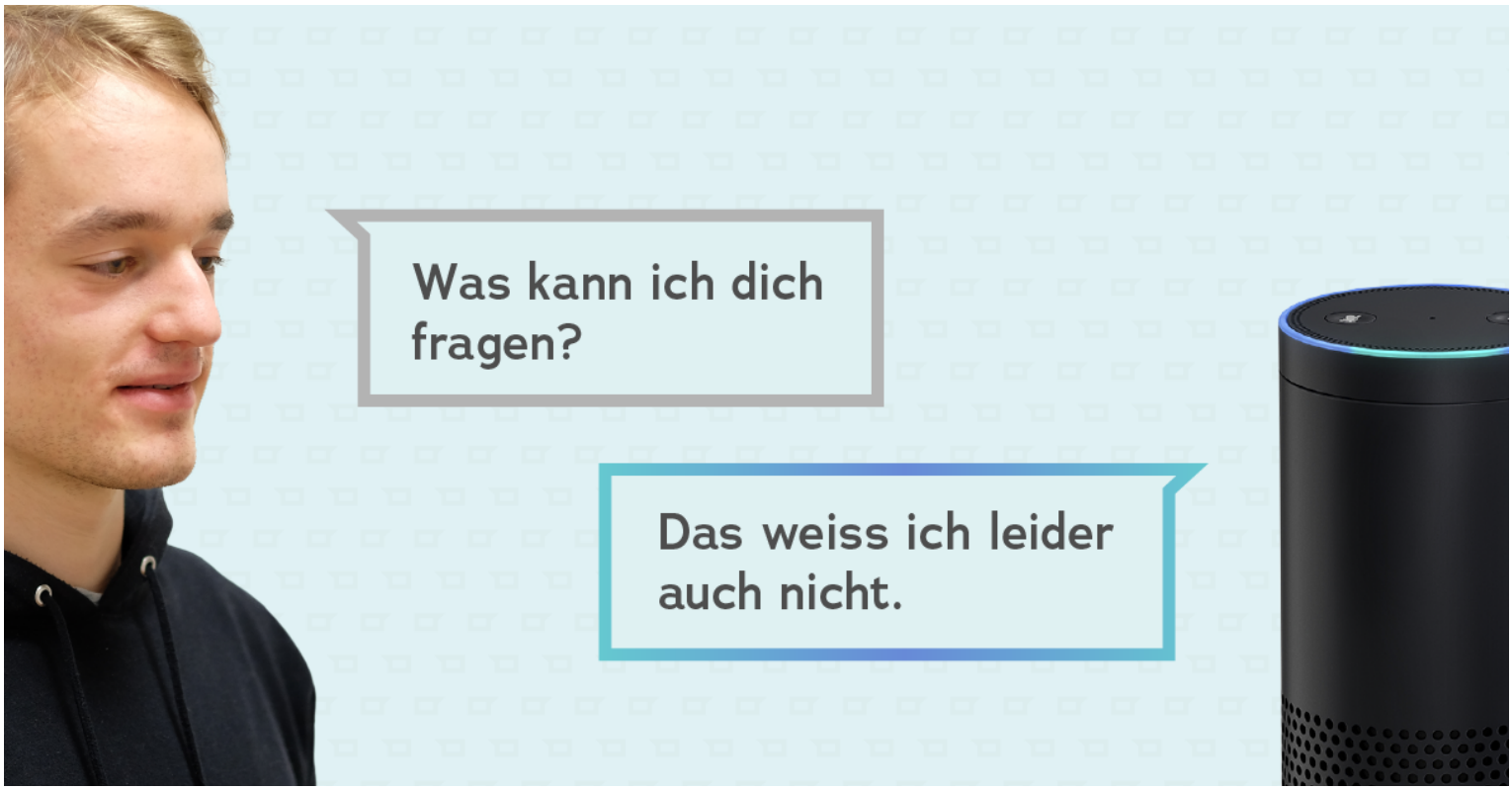
zeix

Agentur für User-Centered Design



Ist Reden Gold?

User-Experience-Studie zu Conversational Interfaces

Version 1.0

7. Dezember 2017

Inhaltsverzeichnis

Zusammenfassung **3**

1 Gegenstand der Studie **5**

2 Exploratives Studiendesign **5**

2.1 Testpersonen 6

2.2 Ablauf der Tests 6

3 Ergebnisse **7**

3.1 Verbreitung und Vorkenntnisse 7

3.2 Meet & Greet mit «Alexa» 8

3.3 Usability-Test mit dem Fahrplan-Chatbot 10

3.4 Usability-Test mit dem Formular-Bot 15

4 Einschätzung der Testpersonen **18**

4.1 Potentiale aus Usersicht 18

4.2 Bedenken aus Usersicht 18

5 Fazit **19**

6 Empfehlungen fürs Chatbot-Design **20**

Über die Zeix AG **21**

Projektleitung/Kontakt

Sibylle Peuker
Zeix AG
Badenerstrasse 65
8004 Zürich

sibylle.peuker@zeix.com
+41 44 247 78 83

Danksagung

Wir danken Netcetera für die freundliche Unterstützung der Studie.

Zusammenfassung

Intelligente Assistenten wie Alexa von Amazon und Google Home halten in die Haushalte Einzug. Es gibt kaum eine Branche, die nicht gerade mit Sprachassistenten und Chatbots experimentiert.

Sind Conversational Interfaces wirklich einfacher zu bedienen als grafische Interfaces?

Auf der Suche nach Erfolgsfaktoren

Wir haben innerhalb eines Jahres viele Conversational Interfaces ausprobiert, Menschen bei der Interaktion mit Sprachassistenten beobachtet und selbst Chatbots gebaut.

Mit dieser explorativen Praxisstudie prüfen und präzisieren wir Hypothesen aus einem Jahr Research. 16 Testpersonen testeten die User Experience von ausgewählten Conversational Interfaces. Da wir wissen, dass schon bei einfachen Chatbots in der Konversation mit dem Menschen viel schiefgehen kann, haben wir für die Studie bewusst Chatbots mit geringer Komplexität ausgewählt; eine Fahrplanabfrage und ein Registrierungsformular.

Kurz erklärt

Die gängige Fachliteratur fasst Interfaces, die es erlauben, mit einer Maschine zu sprechen oder zu chatten unter dem Begriff «Conversational Interfaces» zusammen.

Ein «Chatbot» ist ein spezielles Conversational Interface, nämlich ein Webdienst, der Dialoge auf Chat-Plattformen (z.B. Whatsapp, Wechat, HipChat, Facebook Messenger, Slack) oder Websites simulieren kann. Der User stellt Fragen, der Chatbot antwortet, als würde ein Mensch dahinterstecken.

Top-Spracherkennung schürt Erwartungen

Die meisten Testpersonen konnten sich gut vorstellen, künftig Conversational Interfaces zu nutzen, gingen sogar davon aus, dass das von ganz alleine käme. Zwar wussten sie eingangs oft nicht, was und wie sie fragen sollen. Sobald sie merkten, wie gut die *Spracherkennung* funktioniert, hatten sie schnell den Eindruck, dass die Maschine sie versteht wie ein Mensch.

Da aber die *Interpretation* der Spracheingabe meist noch nicht so gut ist, ist das Frustrationspotential entsprechend hoch.

Die Akzeptanz hängt von der Alternative ab

Massstab der *Kommunikation* ist das persönliche Gespräch. Der Mensch versteht (fast) alles, kann Bilder zur Erklärung hervorziehen, die Voraussetzungen im Gespräch ändern, Fehlerkorrekturen verarbeiten. Müssen Chatbots zusätzlich mit der *Informationseffizienz* von guten, self-service-optimierten Apps wie dem SBB-Fahrplan konkurrieren, können sie im freien Dialog noch nicht mithalten.

Nützlichkeit braucht Kontext

Erst wenn die Nutzungssituation stimmt, zahlt sich die neue Interaktionsform aus. Interfaces mit Spracheingabe sind besonders nützlich in Situationen, in denen man die Hände frei haben muss. Deshalb sollte das Öffnen des Sprach-Interface möglichst interaktionsarm möglich sein, z.B. durch Zuruf. Kontext heisst auch, dass die Fragen des Menschen aufeinander aufbauen können. Das hat bei unseren beiden Test-Bots gut funktioniert, ist jedoch bei vielen Bots ein Manko.

Weniger ist mehr

Sind Sprachinterpretation und der Wechsel der Gesprächsebene (vom Dialog zum Befehl) noch holprig, sollte der Chatbot das Gespräch stark führen. Bildet das Conversational Interface nur einen einzigen Anwendungsfall ab, kann es besser auf die Fragen eingehen. So wissen die User besser, wie sie fragen sollen. Antworten lassen sich besser portionieren. Im Test wurde deutlich, dass die Testpersonen einen teilstandardisierten Dialog als hilfreich und angenehm empfinden, schon ganz ohne künstliche Intelligenz.

User Experience Tests sind ein Must-Have

Wie in den Anfängen des Web sind noch keine Standards für Interaktion und Content für Chatbots vorhanden. User Experience und Technologie sind noch viel enger verwoben als bei grafischen User Interfaces. Deshalb ist der Einbezug der Benutzer in den Entwicklungsprozess Pflicht.

1 Gegenstand der Studie

Mit Maschinen wie mit Menschen reden – für viele ist das ein Menschheitstraum. In diversen Hollywood-Filmen wurde uns vorgeführt, was dann alles möglich wäre (z.B. 2013 im Film «Her», in dem sich ein Mann in sein sprechendes Betriebssystem verliebt).

Mit Fortschritten in der Spracherkennung und der künstlichen Intelligenz scheint das in greifbare Nähe zu rücken. Apple, Google, Amazon, Microsoft, IBM und andere arbeiten an intelligenten Assistenten, mit denen man sich unterhalten kann. Die Produkte wie Siri, Google Assistant und Amazon Echo sind jedoch noch weit entfernt von «Her» aus dem Film.

Derzeit wird überall mit Chatbots experimentiert: für Call Center Services, Fahrplan- und Wetterabfragen, Terminmanagement und Formulareingaben. Man muss sich dabei nicht auf die grossen Firmen verlassen, sondern kann selbst mitbauen: Es gibt diverse Tools, die künstliche Intelligenz und Spracherkennung zur Verfügung stellen.

Aber können die Chatbots schon genug, um bei ihren menschlichen Gesprächspartnern bestehen zu können? Worauf kommt es bei einem guten Dialog zwischen Mensch und Maschine an? Wann ist ein Chatbot wirklich nützlich?

Der vorliegende Report zeigt die Ergebnisse unserer explorativen Praxisstudie auf.

2 Exploratives Studiendesign

Als Vorarbeit beobachteten wir über ein Jahr hinweg immer wieder Menschen bei der Interaktion mit dem Amazon Echo («Alexa») und Siri.

Wir verschafften uns im Vorfeld dieser Studie einen Überblick über existierende Conversational Interfaces in unterschiedlichen Kategorien: Wettervorhersage, medizinische Diagnose, Fahrpläne, Lern- und Beratungschatbots, Terminmanager usw.

Wir führten Usability Expert Reviews an ausgewählten Chatbots durch und systematisierten die Stärken und Schwächen im Interaktionsdesign und Content. Dabei sind wir oft schon bei sehr einfachen Beispielen auf Konversationen gestossen, die überhaupt nicht funktionieren und die User ratlos oder verärgert zurücklassen.

Für die Usability-Tests wählten wir eher einfache Beispiele, um herauszufinden, wo die grundsätzlichen Stolpersteine in der Konversation mit der Maschine liegen. Dazu untersuchten wir ein Sprach-Interface und ein Chat-Interface. Als Sprach-Interface testeten wir einen Prototyp für einen Fahrplan-Assistenten mit Spracheingabe von Netcetera.

Als Chat-Interface nutzten wir den Chatbot von Zeix, der ein Web-Formular zur Registration als Testperson ersetzt.

2.1 Testpersonen

Wir führten Usability-Assessments mit 16 Testpersonen zwischen 19 und 63 Jahren durch.

Testperson	Alter	Aktuelle Tätigkeit
Herr M.	19	Schüler
Frau A.	26	Assistentin der GL
Herr L.	27	Maschineningenieur
Frau S.	28	CMS-Mitarbeiterin
Herr J.	29	Medizinaltechniker
Herr S.	31	Ingenieur
Frau A.	33	Szenographin
Frau F.	33	Immobilienkauffrau
Frau N.	36	Künstlerin
Herr S.	44	Kunstmaler
Herr E.	51	Steuerkommissär
Frau R.	53	Teamassistentin
Frau U.	55	Reinigungskraft
Herr P.	56	Organisationsberater
Frau C.	61	Lehrerin
Herr A.	63	Sachbearbeiter

2.2 Ablauf der Tests

Vorinterview

Zuerst führten wir mit jeder Testperson ein Interview zum Vorwissen, Prädispositionen und Erfahrungen mit digitalen Assistenten / Chatbots durch.

Beobachtung

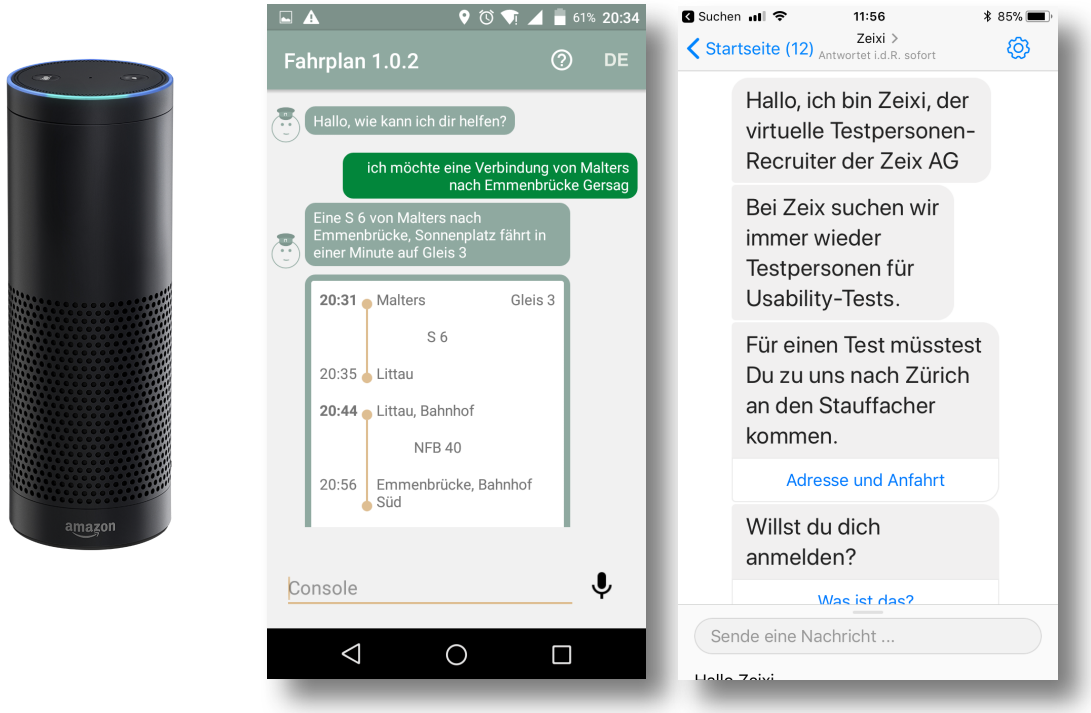
Danach beobachteten wir die Testperson bei der Benutzung der Interfaces auf einem Mobiltelefon.

Aufgabenstellungen

- Amazon Echo ausprobieren (9 von 16 Testpersonen)
- Fahrplan-Chatbot benutzen (14 vom 16 Testpersonen)
- Testpersonen-Chatbot benutzen (14 von 16 Testpersonen)

Nachinterview

Im Nachinterview befragten wir die Testperson zur ihrer persönlichen Einschätzung von Nutzen, Potenzialen und Gefahren von Conversational Interfaces.



Interfaces im Test: Amazon Echo, Fahrplan-Chatbot und Testpersonen-Anmeldung

3 Ergebnisse

3.1 Verbreitung und Vorkenntnisse

Bekannt aus den Medien

Die meisten Menschen, die Smartphones benutzen und sich regelmässig im Internet bewegen, haben bereits von «Spracheingabe» und «intelligenten Assistenten» gehört. Da diese oft in den Medien erwähnt werden, nahmen die meisten unserer Testpersonen an, dass sich solche Interfaces durchsetzen werden.

Die Begriffe «Chatbot» und «Conversational Interface» haben jedoch nur wenige verstanden.

Erfahrungen werden von IVR-Systemen übertragen

Mehr als die Hälfte der befragten Personen hatte zumindest ein Conversational Interface bereits ausprobiert, in der Regel Siri oder den Google Assistent.

Viele unserer Testpersonen zogen bei Sprach-Interfaces gleich Parallelen zu ihren Erfahrungen mit telefonischen IVR-Systemen heran (Sprachdialogsysteme, z.B. zur telefonischen Bestellung

von Kinotickets). Keine unserer Testpersonen mochte diese Systeme. Die meisten hatten eine frustrierende Geschichte dazu zu erzählen:

> **Da musste ich mich ewig durchhangeln, um dann doch in der Warteschleife zu landen.**

Fehlende Anwendungsfälle

Von den 16 Testpersonen nutzte *keine* regelmässig die Sprachunterstützung des Mobiltelefons. Nur eine Testperson gab an, auf den Sprachdialog im Auto zurückzugreifen: zum Telefonieren, Nachrichten schreiben und Bedienen des Navigationsgeräts.

Die anderen haben zwar Siri oder Ähnliches ausprobiert, die Systeme erschienen ihnen jedoch nicht als alltagstauglich.

> **Ich weiss nicht, wozu ich das brauchen soll.**

> **Vergesse, dass es das gibt auf meinem Handy.**

> **Das müsste Schweizerdeutsch verstehen, damit ich es benutze.**

Gute Idee, sobald es wirklich funktioniert

Die meisten Testpersonen konnten sich vorstellen, Conversational Interfaces regelmässig zu benutzen – irgendwann. Sie glaubten, das würde irgendwann von alleine kommen und normal werden – wohl Hollywood sei Dank.

Nur 2 der 16 Testpersonen fanden es komisch, mit Maschinen zu sprechen. Sie konnten sich jedoch gut vorstellen, künftig mit einem Chatbot zu chatten. Die meisten wollten aber noch abwarten, bis der Dialog richtig gut funktioniert und die Bots z.B. Schweizerdeutsch verstehen.

3.2 Meet & Greet mit «Alexa»

Da die Testpersonen so geringe Erfahrungen im Umgang mit Sprachassistenten hatten, liessen wir die Hälfte der Testpersonen vorab den Amazon Echo («Alexa») als Beispiel für ein hochentwickeltes Conversational Interface ausprobieren. Die andere Hälfte führte die folgenden Tests aus, ohne Alexa ausprobiert zu haben.

Ein Rockstar macht viel Unterschied

Für die Hälfte der Testpersonen starteten wir den Test mit dem Amazon Echo («Alexa»). Das Gerät übertrifft deutlich die Erwartungen der User, was die Natürlichkeit des Dialogs angeht. Gleichzeitig ist die Wahrscheinlichkeit besonders hoch, als Antwort «Das weiss ich leider nicht» zu erhalten.

Trotzdem löste gerade Alexa Begeisterung bei den Usern aus. Viele Menschen sind schon deshalb beeindruckt, weil die Spracherkennung so gut funktioniert und Alexa mit menschlicher Stimme und oft witzig antwortet.

Reden will gelernt sein

Keine dieser Testpersonen hatte Alexa zuvor ausprobiert, so dass alle Testpersonen zunächst ratlos vor dem Gerät standen:

Ja, was kann ich die jetzt fragen?

Wir zeigten den Testpersonen wie es funktioniert anhand einfacher Beispiele, wie Wetterabfrage oder Ergänzung des Einkaufszettels. Die Aktivierung des Amazon Echo funktioniert über ein Codewort. Dessen Aussprache mussten die Testpersonen zuerst üben: Zuerst deutlich das Codewort «Alexa» sagen und dann die Frage formulieren. Über ein Lichtsignal gibt Alexa Feedback, wenn sie ihre Aufmerksamkeit einer Person zuwendet.

Joy of Use bei Alexa auffällig gut

Nach dieser Einführung waren alle Testpersonen positiv überrascht über Alexas Verhalten und ihre korrekten Antworten auf einfache Fragen.

Die Testpersonen fanden es «cool» mit Alexa zu reden und waren fasziniert von diesem alleinstehenden digitalen Gesprächspartner. Als sie eigene Fragen formulierten, wurde klar, dass Alexa die Frage zwar verstanden hatte, jedoch oft keine Antwort «wusste», z.B.

Testperson:

Alexa, was kommt heute in Zürich im Kino?

Alexa:

Ich konnte keine Kinoinformationen für den ausgewählten Standort finden.

Im Vergleich zur Kontrollgruppe ohne Alexa beurteilte die 1. Testgruppe den Nutzen von Conversational Interfaces dennoch positiver und war kreativer bei der Frage nach möglichen Anwendungsfällen.

3.3 Usability-Test mit dem Fahrplan-Chatbot

Im ersten Usability-Test untersuchten wir einen Fahrplan-Assistenten, den uns die Firma Netcetera zur Verfügung gestellt hat. Netcetera hat diesen Prototyp entwickelt, um Erfahrungen mit Conversational Interfaces und Frameworks für künstliche Intelligenz zu sammeln. Der Chatbot hat sowohl ein Texteingabe-Feld, als auch ein Sprach-Interface.

Neben allgemeinen Hypothesen zur User Experience von Chatbots untersuchten wir das konkrete Verhalten für den Kontext der Fahrplanabfrage. Fragestellungen waren z.B.:

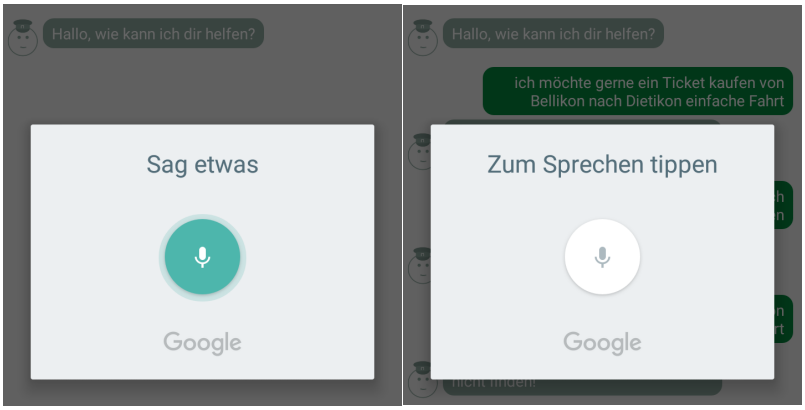
- Können die User die Spracheingabe bedienen?
- Haben sie Mühe, Fragen zu formulieren?
- Welche Antworten erwarten sie?

Motorische Hürde bei der Aktivierung

Den richtigen Moment zu erwischen, wann genau das Gerät zuhört, ist bei vielen Sprachassistenten noch holprig gelöst. Das Anschalten der Voice-Eingabe beim Fahrplan-Bot funktioniert folgendermassen:

1. Mikrofon-Icon kurz tippen
2. Sprechen
3. Mikrofon schaltet automatisch ab

Wann kann ich sprechen?



Dies wurde von einigen Testpersonen nicht gleich verstanden. Auch bei wiederholtem Klicken erwischten sie den richtigen Augenblick zum Sprechen nicht. Denn wenn man der Aufforderung «Sag etwas» nicht schnell genug nachkommt, schaltet das Mikrofon ab und der User muss das Mikrofon-Icon erneut antippen. So sprachen mehrere Testpersonen zunächst ins Leere.

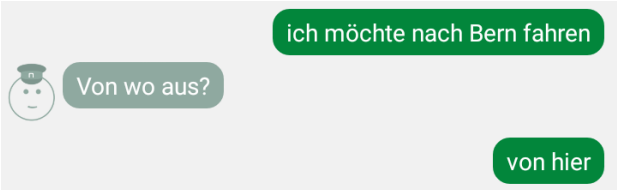
Der Bot «merkt», wann der User mit dem Sprechen fertig ist (wie in einem richtigen Gespräch). Das funktionierte in ruhigen Räumen gut. Bei Hintergrundgeräuschen, z.B. auf der Strasse, hat das automatische Abschalten nicht so gut funktioniert.

Was soll ich sagen?

Teilweise ist die Eingabezeit verstrichen, weil die Testperson nicht wusste, wie sie die Frage formulieren soll. Einige sprachen nicht in ganzen Sätzen, sondern in Schlagworten, z.B. "Fahrplan Zürich Bern". Denn sie nahmen an, dass eine Maschine nur einfache Eingaben versteht.

Sprachverständnis auf hohem Niveau

Die Testpersonen konnten leicht kontrollieren, ob der Chatbot ihre Eingabe korrekt verstanden hat, da jede menschliche Eingabe auch als Text auf dem Display angezeigt wird:



Tatsächlich wurden die Wörter der Spracheingabe fast immer sehr gut verstanden, obwohl die Testpersonen ihre Anfrage sehr unterschiedlich formulierten; z.B.

Also ich möchte nach Basel fahren.

Ich würde gerne von Zürich auf den Titlis.

Zeig mir den Fahrplan von Bellikon nach Dietlikon.

Ich möchte nach Bern fahren.

Ich möchte eine Verbindung von Malters nach Emmenbrücke Gersag.

Fahrplan Zürich Bern

Schwerfällige Interpretation der Eingaben

Obwohl die Eingabe meist korrekt angezeigt wurde, war die Antwort (zu) oft unpassend oder das Gerät behauptete, es hätte nicht verstanden. Den Testpersonen war dann oft nicht klar, woran es lag. Teilweise wurde die gleiche Eingabe sogar an unterschiedlichen Testtagen unterschiedlich beantwortet.

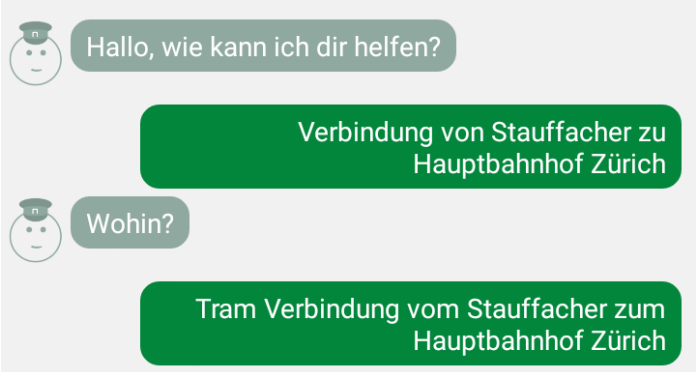
Bei der Eingabeform «Ich möchte von A nach B» beispielsweise wurde manchmal direkt der korrekte Fahrplan angezeigt. Oft kam aber auch die Rückfrage «Von wo aus?».

Textanzeige als gute Lernhilfe

Die Anzeige der eigenen Spracheingabe war eine wichtige Hilfe für die User. Wenn der Bot eine unerwartete Antwort gab, wussten die User so zumindest, dass es an der Interpretation der Eingabe des Bots lag, nicht an der Frage. Dadurch fingen die User an zu lernen, wie sie die Eingabe formulieren müssen: «Also ich muss zuerst nur sagen wohin und dann fragt er mich von wo aus».

Präzisierung: Userverhalten wie im Gespräch

Wenn der Chatbot eine Frage nicht verstanden oder nicht korrekt beantwortet hatte, versuchten die Testpersonen, die Frage präziser zu machen, indem sie die Frage weiter einschränkten:



Dieses Verhalten haben wir schon oft beobachtet, da die menschliche Kommunikation genau so funktioniert: Hat mich mein Gegenüber nicht richtig verstanden, präzisiere ich die Frage mit Details.

Beispiel: «Letztens habe ich Herrn Zumstiel getroffen». «Kenne ich nicht». «Du weisst schon, der der uns letztes Jahr den Wasserhahn repariert hat». «Ach ja, der».

Wenn in einer Suchmaschine die Eingabe «Zumstiel» 0 Treffer ergibt, dann ergibt «Zumstiel Wasserhahn» ebenfalls 0 Treffer. Den meisten Usern ist das bei der Suchmaschine klar (jedoch nicht allen, wie wir in Usability-Tests gesehen haben).

Bei einem Conversational Interface wenden die User aber eher die Regeln der menschlichen Kommunikation an. Sie denken nicht daran, dass eine Maschine – anders als der Mensch – auch schon beim ersten Versuch Zugriff auf ihre gesamte Datenmenge hat.

Passend zum Kontext antworten

Der Fahrplan-Bot zeigte bei den Abfragen immer die nächste Verbindung. Diese Mischung aus Ansage und Anzeige wurde von allen Testpersonen problemlos verstanden.

Allerdings ist diese Auskunft nur bedingt relevant: Unser Testlabor liegt in der Nähe einer grossen Tramhaltestelle. Folglich war die Antwort oft «Ein Tram der Linie 3 fährt in einer

Minute.» Diese Information nützt nur dann, wenn man schon direkt an der Haltestelle steht. Die Testpersonen mussten dann ein 2. Mal fragen: «Ich möchte um 14 Uhr nach Stadelhofen fahren», um die gewünschte Verbindung zu erhalten.

Was anzeigen?

Dies verdeutlicht eine Grenze der Sprachinteraktionen: bei der visuellen Ausgabe einer Fahrplanverbindung erhält der User eine Anzeige mindestens der nächsten 3 Verbindungen. Er kann schneller auswählen, welche passt.

Alternativ könnte der Chatbot immer mit einem «Puffer» von z.B. 5 Minuten im Stadtverkehr rechnen. Sollte die Person aber fragen: «Wann fährt heute der letzte Zug nach Bern?», kann direkt das Ergebnis angezeigt werden.

Wenig Toleranz bei menschlichen Fehlern

Wie viele gängige Chatbots konnte der Fahrplan-Bot nicht zwischen Spracherkennung und Befehlen auf der Metaebene unterscheiden. Wurde eine Eingabe nicht korrekt verstanden, wussten die Testpersonen nicht, was sie machen sollten. Denn die Konversation gibt ihnen keine Möglichkeit zur Korrektur.

Wieso gibt es keine Möglichkeit, nur das zu wiederholen, was falsch verstanden wurde? Z.B. indem ich auf diesen Begriff klicke.

Wenn es den Weg falsch verstanden hat, dann kann ich nicht mehr zurück.

In einem guten User Interface, kann man falsche Eingaben rückgängig machen. Hier sind Conversational Interfaces allgemein (noch) nicht so ausgereift wie grafische User Interfaces.

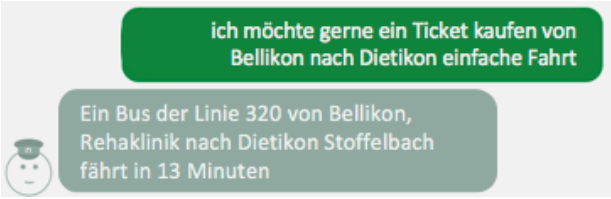
Auch dieses Verhalten deckt sich mit den Regeln menschlicher Kommunikation: In einem Gespräch oder einem Chat kann man ein gesprochenes Wort oder eine gesendete Nachricht auch nicht zurücknehmen. Man kann jedoch bei Missverständnissen korrigieren: «Nein, ich meinte nach Biel, nicht nach Bern». Im zweiten Anlauf wiederholten die Testpersonen den ganzen Satz teils in abgewandelter Form und kamen so zum Ziel.

Auch hier zeigt sich, was wir schon vorher beobachtet hatten: Die Menschen versuchen zwar, sich auf die Maschine einzustellen. Der Aufwand für eine simple Abfrage zeigt aber auch, warum sie die Sprachassistenten nicht öfter in Anspruch nehmen.

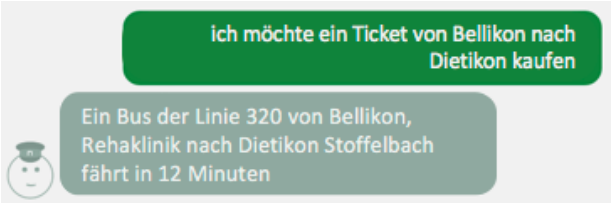
Intelligenter Kaufprozess überzeugt Neulinge

Eigentlich eine geniale Funktion des Fahrplan-Bots: Nachdem die Testpersonen eine Verbindung gewählt haben, sagten sie «Ich möchte ein Ticket» und eine passende Ticketoption für die gesuchte Strecke wurde ihnen zum Kauf angeboten. So empfanden fast alle Testpersonen die Bestellung des Tickets als einfach.

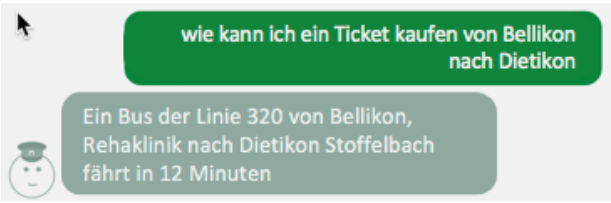
Aber: die einzige Person, die bereits regelmässig Sprach-Interfaces benutzt, hatte Mühe ihr Ticket zu bestellen. Sie ging davon aus, dass sie den Kontext – in diesem Fall die Stecke – nochmals mitliefern muss, weil sie aus Erfahrung davon ausging, dass der Chatbot den Zusammenhang wieder «vergessen» hatte. Sie formulierte die Frage so:



Und so:



Und dann nochmal so:



In diesem Fall hat der Fahrplan-Chatbot ihre Ansage als Fahrplanabfrage gewertet und nicht als Ticketkauf-Anfrage. Mit der Kombination konnte er nicht umgehen.

(Noch) bessere Alternative vorhanden

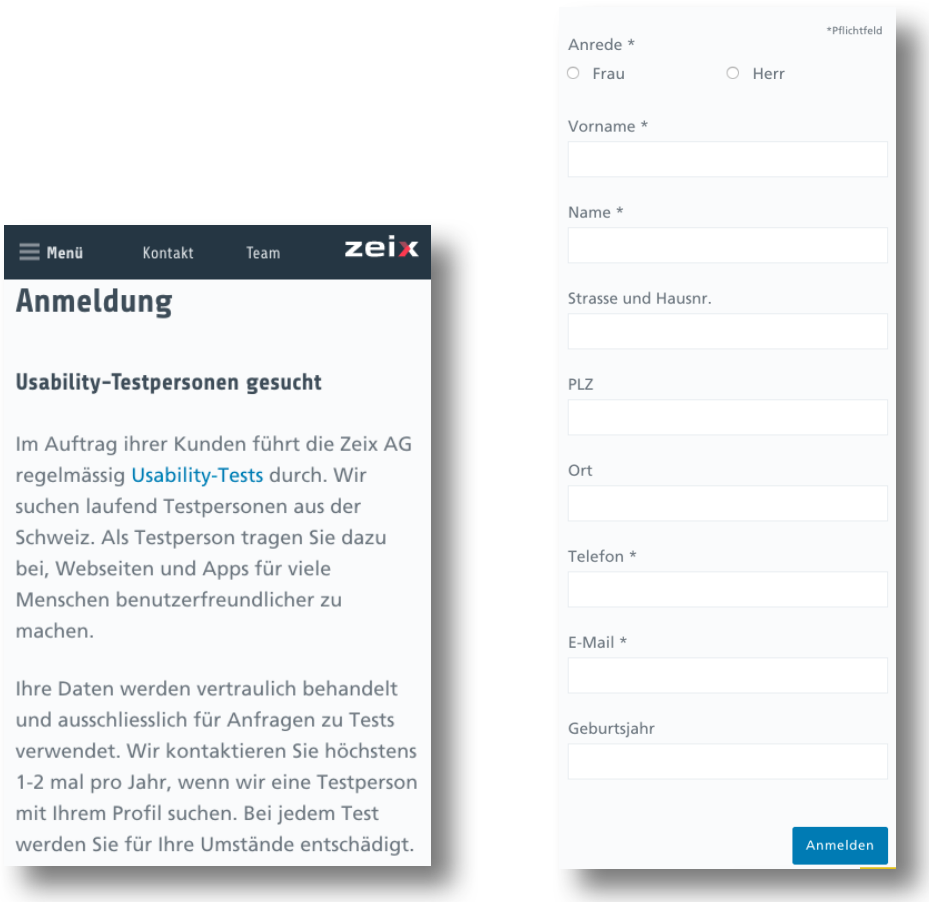
Die meisten Testpersonen benutzen bisher die SBB-App zur Fahrplanabfrage und sind damit zufrieden. Wenn sie etwas Neues ausprobieren und das nicht beim ersten Mal perfekt funktioniert, dann nehmen sie lieber wieder die SBB-App. Nach der Benutzung des Prototyps gaben 10 von 14 Testpersonen an, dass sie eine Fahrplanabfrage per Chatbot erst dann benutzen, wenn sie perfekt funktioniert – und das tut sie noch nicht. Ein wesentlicher Nutzen für den Fahrplan-Bot wären Situationen, in denen man die Hände nicht frei hat, z.B. im Auto. Dort müsste das Anschalten aber auch ohne Handbedienung funktionieren.

Diesem User-Feedback ist Netcetera bereits gefolgt und hat den Fahrplanassistenten in Google Assistant integriert, so dass er ohne Hände bedienbar ist:
https://www.netcetera.com/de/home/stories/expertise/20170614-conversational-interfaces-google-assistant.html

3.4 Usability-Test mit dem Formular-Bot

Wie kann man mit den Schwächen von Mensch und Chatbot also umgehen, so dass sie sich besser aufeinander einstellen können? Um das zu untersuchen, haben wir einen ganz einfachen Anwendungsfall ins Chatformat überführt: das Ausfüllen eines Kontaktformulars.

Auf unserer Website zeix.com können sich Privatpersonen als potenzielle Testpersonen für Usability-Tests registrieren lassen. Dazu gibt es eine Erklärung, was das bedeutet und wie die Auswahl von Testpersonen funktioniert. Interessierte füllen ein Formular mit ihren Kontaktdaten aus und werden von unserer Recruiterin telefonisch kontaktiert, um das vollständige Profil aufzunehmen.



Anmeldung über Formular auf der Website https://zeix.com/testpersonen/

Nun bieten wir Usern den Chatbot als Alternative zum Formular an.

Alternative Anmeldemöglichkeit mit Chatbot https://www.messenger.com/t/178976412632264

Mit den Testpersonen haben wir beide Möglichkeiten (bei Durchführung auf einem Smartphone) verglichen.

User können mit dem Chatbot interagieren, indem sie Antwortoptionen auswählen. Um ihre Kontaktdaten einzugeben, benutzen sie das Eingabefeld für Freitext unten. Sie können aber auch jederzeit die Freitexteingabe benutzen, um etwas Anderes zu fragen oder zu schreiben.

Bei einer Freitext-Eingabe «versteht» der Chatbot allerdings nur wenig. In diesem Fall soll er die Eingabe an einen menschlichen Mitarbeiter weiterleiten:

Die Fragen und Anweisungen waren für die Testpersonen gut verständlich. Alle Testpersonen bewegten sich anhand der vorgegebenen Optionen durch den Anmeldeprozess. Keine Testperson hat versucht, mithilfe des Freitextfelds etwas nachzufragen.

Allerdings befanden sich die Testpersonen gerade in einem Interview mit einem Zeix-Mitarbeiter. Sie wussten also bereits, was Zeix ist und dass wir Tests mit Usern machen.

Mehr lesen, besser informiert

Überraschenderweise gaben im Nachinterview einige Testpersonen an, dass sie durch den Chatbot besser informiert waren. Dabei stehen auf der Website dieselben Informationen in zwei Absätze Fliesstext verpackt.

Das haben wir schon in vielen Tests beobachtet: Den Text auf der Website vor dem Formular haben nur wenige Testpersonen genau durchgelesen. Im Chatbot werden die gleichen Infos dem User häppchenweise serviert und wurden prompt von allen gelesen.

Im Zweifel für den (Formular-)Bot

Wir waren überrascht, wie positiv der Chatbot ankam. Obwohl man genauso viel eingeben und vielleicht sogar etwas mehr lesen muss als beim Formular, fanden die meisten Personen, dass es mit dem Chatbot schneller ging und einfacher war. Wir gehen davon aus, dass die in den Kontext verpackte, häppchenweise Information einfacher und angenehmer zu konsumieren ist.

Kann man einem Bot vertrauen?

Die Testpersonen schätzten auch die persönliche freundliche Ansprache gegenüber dem nüchternen Formular. Bei der Vertrauenswürdigkeit gingen die Meinungen jedoch auseinander: einige fanden das Formular vertrauenswürdiger, einige den Chatbot, für andere machte es keinen Unterschied.

9 von 14 Testpersonen gaben an, dass sie den Chatbot angenehmer fanden als das Formular und lieber diesen benutzen würden. Hier gab es keine korrelierenden Eigenschaften wie Alter oder Geschlecht der Testpersonen.

Von den 5 Personen, die lieber das Formular benutzen, gaben 2 an, dass es für sie mit der Autocomplete-Funktion viel schneller geht. Das gilt natürlich nur für einfache Formulare, für die Autofill-Einstellungen hinterlegt sind. Das ist bei umfangreicheren Formularen, wie z.B. im Offertprozess einer Krankenversicherung nicht möglich.

4 Einschätzung der Testpersonen

Im Nachinterview zu den beiden Tests hat jede Testperson ihre Einschätzung zu Bots und ihrer Nützlichkeit reflektiert.

4.1 Potentiale aus Usersicht

Potentiale für Interfaces mit Spracheingabe sahen viele Testpersonen in Situationen, in denen sie eine Information brauchen, aber die Hände nicht frei haben: z.B. beim Autofahren, beim Kochen, bei der Gartenarbeit oder beim Putzen.

Chatbots stellen sie sich dann nützlich vor, wenn die Alternative z.B. eine telefonische Warteschleife ist oder wenn sie unterwegs in öffentlichen Verkehrsmitteln sind und deshalb lieber chatten als telefonieren möchten oder ausserhalb der Arbeitszeiten des Kundensupports.

4.2 Bedenken aus Usersicht

Nicht nur in den Interviews, auch in Diskussionen vorher mit verschiedenen Gruppen hörten wir einige Bedenken immer wieder:

- Die Angst, dass Conversational Interfaces die menschliche Kommunikation ersetzen

  «Dann reden die Menschen noch weniger miteinander.»

- Eindringen in die Privatsphäre

  «Die Alexa hört ja immer mit. Das will ich nicht.»

- Die Angst, dass Arbeitsplätze verloren gehen

  «Dann braucht es ja weniger Menschen, z.B. im Kundensupport.»

5 Fazit

Mensch bleibt Mensch

Conversational Interfaces sind für die meisten Menschen noch ungewohnt. Auch wenn die Maschinen immer intelligenter werden, bleibt der Mensch einfach Mensch. Sie möchten sich auf ihren Kommunikationspartner einstellen, auch auf eine Maschine. Sie wenden Prinzipien an, die sie aus der menschlichen Kommunikation kennen, z.B. Präzisierung, wenn sie nicht verstanden werden.

Die grösste Hürde besteht oft darin, dass für die User nicht klar ist, was die Maschine versteht, also was man fragen kann und wie man seine Frage formulieren muss. Am Bahnschalter hätten die Testpersonen sicher nicht lange überlegen müssen, wie sie die Frage an den Bahnmitarbeiter formulieren. Sie wüssten, wie sie mit ihm reden müssen. Wir wissen, dass Menschen sich im Gespräch immer an ihr Gegenüber anpassen:

- Bei einem Kind verwenden sie einfache Wörter.
- Bei einem Ausländer sprechen sie langsam und deutlich.
- Ein Experte im eigenen Land sollte alles verstehen.

Bei einer Maschine ist dieses Niveau schwer vorherzusagen. Dies ist einerseits eine Übungsfrage, andererseits eine Frage der *Benutzerführung*.

Erziehungsprobleme bei den Bot-Kindern

Ein Framework für künstliche Intelligenz lernt zwar dazu, aber das leider nicht immer korrekt. Diese Erfahrung haben sowohl Netcetera wie auch wir beim Bauen von Chatbots gemacht. Die verfügbaren Frameworks sind oft noch eine Blackbox. Die Datenausgabe kann teilweise nicht nachvollzogen werden und ist veränderlich. Dadurch gibt es noch kaum Möglichkeiten zu testen, ob ein Chatbot korrekt «mitdenkt». Umso wichtiger ist es, dem User Alternativen im Gesprächsverlauf anzubieten und «mitzulesen» (s.u.).

Fehlender Standard zur Aktivierung

Bisher fehlen Standards für das An- und Ausschalten der Spracheingabe. Jede bekannte Chatbot-App nutzt eine andere Interaktion: Bei WhatsApp muss man während des Sprechens die Mikrofon-Taste gedrückt halten. Bei Apple Speech-to-Text muss die Benutzerin zuerst das Mikrofon durch kurzes Tippen anschalten, dann sprechen, dann wieder ausschalten. Der Zuruf «Alexa» ist zwar niedrigschwellig, funktioniert aber auch nur vor leisem Hintergrund.

Technologiegetriebene Vorläufer

Schliesslich sieht man vielen real-existierenden Chatbots an, dass die Entwicklung technologiegetrieben war und das Verhalten und die Bedürfnisse der User eine untergeordnete Rolle gespielt haben: «Wir haben etwas gebaut und es läuft». Dabei schleichen sich leicht Fehler ein, die dazu führen, dass der Chatbot nutzlos oder unbenutzbar ist, z.B.

- bleibt die Benutzerführung auf der Strecke und der User weiss nicht, was er machen kann,
- sind die Use Cases viel zu kompliziert für den User, um sie zu erraten oder
- die Situation des Users, in der der Bot seine Stärken hätte, wird vernachlässigt.

Deshalb sollten die Bots immer von interdisziplinären Teams entwickelt werden.

6 Empfehlungen fürs Chatbot-Design

Think Big: Software-Produktdesign für Bots

Überlegen Sie bei der Entwicklung Ihrer Produkte und Leistungen frühzeitig, ob ein Bot als Interface für bestimmte Anwendungsfälle zum Einsatz kommen könnte und was dies z.B. für die Datenablage bedeutet. Gehen Sie dabei nicht nur vom Dialog aus. Auch der Einsatz von Rich Content wie Infografiken kann Teil einer Konversation sein.

Start Small: Bescheidenheit im Anwendungsfall

Oft starten Projekte mit zu hohen Erwartungen, was die Maschine alles übernehmen kann. Nach unseren Beobachtungen und Erfahrungen liegen die Potentiale aber oft in sehr eingeschränkten, aber hochpersonalisierten Anwendungsfällen. Also z.B. Informationen, für die ich auf einer Website lange suchen müsste: «Ist die teure Porzellanfigur, die ich von meiner Grossmutter geerbt habe, mit meiner Hausratversicherung abgedeckt?».

Interdisziplinäres Projektteam

Bei Conversational Interfaces sind User Experience und Technologie viel enger verwoben als bei grafischen User Interfaces. Das Projektteam sollte deshalb nicht nur aus Programmierern und Experten für künstliche Intelligenz bestehen, sondern auch aus Menschen, die sich mit Kommunikation und Userverhalten auskennen, z.B. UX Spezialisten, Drehbuchautoren und erfahrenen Kundenberatern.

Ressourcen fürs Monitoring

Seien Sie sich bewusst, dass die Arbeit nicht abgeschlossen ist, sobald der Chatbot live ist. Durch die Auswertung der *Eingaben* der User können Sie den Chatbot schlauer und nützlicher machen. Ausserdem muss ein Experte für den Inhalt die *Antworten* beobachten. Gefahren bestehen vor allem in Fehlinterpretationen des Systems, die zu falschen Antworten führen, ohne dass jemand dies merkt. Also z.B. dass der Chatbot einen falschen Fahrpreis angibt, weil er die Angabe «ein halbes» nicht korrekt interpretiert hat.

Ist Reden also schon Gold?

Das Reden funktioniert schon gut, der Erfolg der Conversational Interfaces wird jedoch vom richtig Verstehen und der Empathie für die Situation des Users bestimmt werden. Von daher müssen wir noch etwas weiter schürfen.

Über die Zeix AG

Die Zeix AG ist die Spezialistin für User-Forschung und User-Centered Design in der Schweiz. Zeix konzipiert webbasierte Applikationen, analysiert und sichert die Qualität von Websites, Software und Medien (z.B. durch Usability-Tests oder/und Expert Reviews).

Massgeschneiderter Content zur Benutzerführung («User Education») schafft Voraussetzungen, damit möglichst viele User Internet, Software und Mobilkommunikation erfolgreich nutzen können.

Konzept und Methodenset von Zeix haben sich in Usability-Studien mit mehr als 3500 Usern bewährt, so dass Zeix das Wissen, die Fähigkeiten und die mentalen Modelle der User sehr detailliert kennt.

Kompetenzschwerpunkte liegen in der qualitativen Forschung und der userzentrierten Entwicklung des Frontends für Applikationen und Portale. Kernprinzipien der Leistungen sind individuelle Methodenentwicklung, das Verfolgen pragmatischer Lösungsansätze sowie der Einbezug der strategisch-betriebswirtschaftlichen Perspektive in alle Analysen und Entwicklungen.

Für Conversational Interfaces bieten wir

- Design Thinking Workshops & Entwicklung von Anwendungsfällen
- Umsetzung von Chatbots
- Expert Reviews für Chatbots
- Usability-Tests
- Drehbucherstellung

Mehr Informationen finden Sie auf unserer Website zeix.com/chatbots.

Impressum

Herausgeberin
Zeix AG, Zürich

Feedback
Haben Sie Anregungen, Kritik oder Themenvorschläge?
Wir freuen uns über Feedback an info@zeix.com

Copyright
© 2017 by Zeix AG

Postfach 8022
CH-8036 Zürich
Telefon +41 44 247 78 78

Alle Rechte vorbehalten. Kein Teil des Werks darf in irgendeiner Form ohne Genehmigung des Verlags reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden. Websites ändern sich ständig. Zeix kann deshalb nicht für die Übereinstimmung der Zitate und Abbildungen mit den aktuellen Websites garantieren. Firma und Autoren können für fehlerhafte Angaben und deren Folgen weder eine juristische Verantwortung noch irgendeine Haftung übernehmen. Fast alle Hard- und Software sowie Firmennamen und Firmenlogos, die in diesem Werk erwähnt werden, sind gleichzeitig auch eingetragene Warenzeichen und sollten als solche betrachtet werden.