DDV

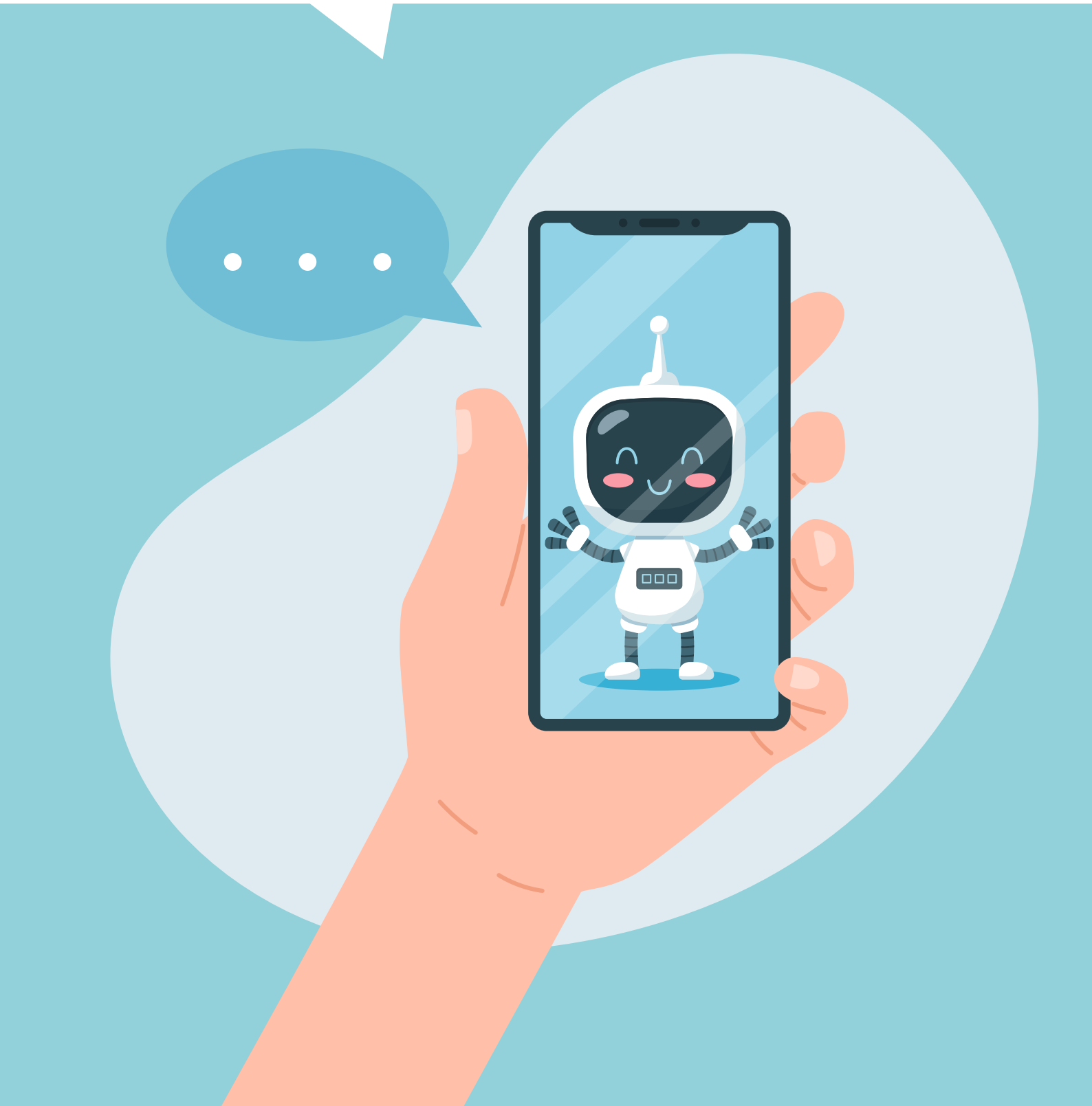
Die treibende Kraft der Data Driven Economy

Branchentrends im Dialogmarketing

Schlüsselkomponenten eines Voicebots

Wie Sie einen Voicebot für Ihr Unternehmen erstellen

Autor: **Bruno Fernandes Carvalho**



Schlüsselkomponenten eines Voicebots

Wie Sie einen Voicebot für Ihr Unternehmen erstellen

Spätestens seit ChatGPT sind Chatbots und Voicebots in aller Munde. Wir nehmen dies zum Anlass zu erklären, wie Voicebots funktionieren. Wir werfen einen Blick auf die Technologie, die hinter einem Voicebot steckt und gehen auf die wichtigsten Komponenten ein, die für den Aufbau eines solchen Bots notwendig sind. Außerdem erfahren Sie, wann der Einsatz generativer KI wie ChatGPT sinnvoll ist, welche Open-Source-Optionen es gibt und welche etablierten Technologien am Markt sind.

Die drei wichtigsten Voicebot-Komponenten: Spracherkennung, natürliche Sprachverarbeitung und Sprachausgabe

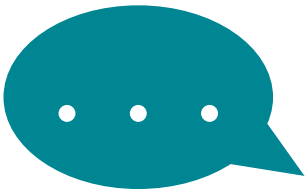
Ein Voice Bot hat drei Hauptkomponenten: Spracherkennung (auch bekannt als Speech to Text (SST) oder Automatic Speech Recognition (ASR), natürliche Sprachverarbeitung (Natural Language Processing oder kurz NLP) und Text to Speech (TTS)). Zudem gab es in den letzten Jahren große Fortschritte auf dem Gebiet der künstlichen Intelligenz. Die Leistungsfähigkeit der Algorithmen wurde durch den Einsatz der Transformers Neural Network-Architektur, auf der auch ChatGPT und BERT basieren, erheblich verbessert. Die Technologie liefert erstaunliche Ergebnisse im Bereich der Spracherkennung und des NLP und trägt damit zu dem derzeitigen Hype um Sprachroboter bei. Wir können heute robustere und verlässlichere Anwendungen entwickeln. Im Folgenden werde ich aufzeigen, wie die drei Hauptkomponenten zusammenwirken.

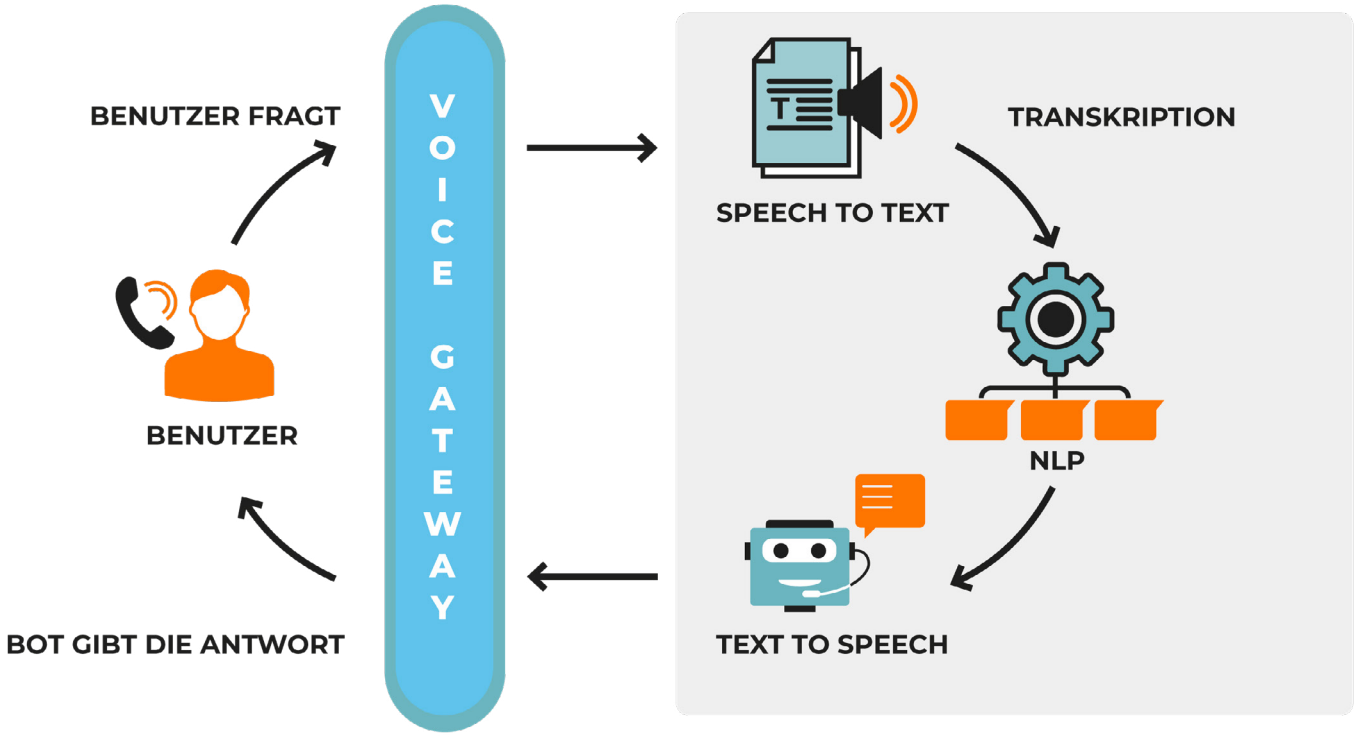
Spracherkennung

Die Spracherkennung, oft auch Speech to Text (SST) oder Automatic Speech Recognition (ASR) genannt, ist die Komponente, die dafür sorgt, dass die Sprache des Benutzers transkribiert wird, indem die Audiodaten in Textdaten umgewandelt werden. Textdaten sind nach heutigem Stand einfacher zu bearbeiten als Audiodaten, vor allem, um die Bedeutung der Sprache zu verstehen. Meiner Meinung nach ist dies eine der wichtigsten Komponenten eines Sprachroboters, da sie die Benutzereingaben erfasst und für die weitere Bearbeitung zur Verfügung stellt. Die folgenden Komponenten können nicht richtig funktionieren, wenn die Eingaben nicht korrekt erfasst werden.

Sie können sich bei Google, Microsoft, IBM, AWS, Deepgram usw. nach ausgereiften Spracherkennungsdiensten umsehen. Die Qualität ist sehr gut, da diese Modelle mit einer großen Menge an Audiomaterial stundenlang trainiert wurden und daher einen Vorteil in Bezug auf die Erkennungsqualität haben. Aber das ist natürlich auch mit einem höheren Preis verbunden. Die gute Nachricht ist, dass KI von Tag zu Tag für jedermann zugänglicher wird und dass wir Open-Source-Angebote haben, die wirklich gut funktionieren, wie z. B. VOSK, Scribosermo und Wav2Vec.

Für welche Technologie Sie sich auch entscheiden: Achten Sie darauf, dass die Spracherkennung mit einem guten sogenannten „Stille-Detektor“ arbeitet, damit die Antwortzeiten des Sprachroboters schnell genug sind.





Sprachverstehen

Der Bot muss verstehen, was der Benutzer sagt, also dessen Bedeutung. Zu diesem Zweck nutzen wir das KI-Modul NLP (Natural Language Programming). Es hilft, den transkribierten Text zu verarbeiten. Es gibt viele Aufgaben, die von einem NLP-Algorithmus ausgeführt werden können: Klassifizieren, Zusammenfassen, Textgenerierung etc. Im Zusammenhang mit Voicebots, insbesondere mit Closed Domain Voicebots (Begriffsdefinition Seite 7), die über regelbasierte Frameworks entwickelt werden, stehen zwei Aufgaben im Fokus: Textklassifizierung und Entitätsextraktion. Bei der ersten Aufgabe wird der gesamte Text analysiert und die Absicht des Benutzers ermittelt. Abhängig von der erkannten Absicht gibt es vordefinierte Antworten, um den Gesprächsfluss fortzusetzen. Die zweite Aufgabe besteht darin, Schlüsselinformationen (Entitäten) aus dem Text zu extrahieren, z. B. den Namen der Person, den Ort, Nummern usw. Mit diesen extrahierten Entitäten können Sie weiterarbeiten: etwa Datenvalidierungen durchführen oder Informationen in Datenbanken speichern – die Möglichkeiten sind hier sehr vielfältig.

Es gibt viele Open-Source-Frameworks, die sich nicht nur auf die Textklassifizierung und Entitätsextraktion konzentrieren, sondern auch auf die Erstellung von Konversationen spezialisiert sind. RASA, Microsoft Bot Framework, Botpress und Wit.ai sind nur einige Beispiele, die vollständige Liste finden Sie hier: **https://botpress.com/blog/open-source-chatbots**

Sie können mit den Modellen der BERT-Familie in Huggingface ganz von vorne anfangen, Text- und Entitätsklassifikatoren zu erstellen. Wir empfehlen ein ausgereifteres Framework wie RASA, um den Projektfortschritt zu beschleunigen.

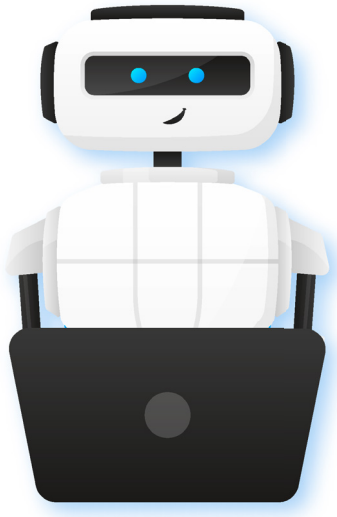
Viele Conversational AI Anbieter, wie z.B. Google Dialogflow und Microsoft LUIS, sind sehr einfach zu bedienen. Abhängig von dem Framework oder dem Anbieter, den Sie wählen, kann die Art und Weise, wie Sie den Gesprächsfluss gestalten, variieren. Doch das Konzept bleibt gleich: Sie definieren eine große Anzahl von Trainingsphrasen für jede Absicht in Ihrem Dialog und einen vordefinierten Satz von Antworten.

Und Sie definieren auch die möglichen Entitäten, die Sie extrahieren möchten. Oft können Sie zwischen einfachen REGEX-Modellen oder Modellen, die auf maschinellem Lernen basieren, wählen.

Zusätzlich zu den oben genannten Conversational AI Frameworks entwickelt sich das Gebiet der „Generativen KI“ rasant weiter. Es wird sich allmählich auf die traditionellen, regelbasierten Chatbots auswirken, indem diese die Flows mit Large Language Models (LLMs) wie ChatGPT integrieren und so leistungsfähigere Konversationen ermöglichen. Ein Beispiel ist die Entwicklung von „intentless Bots“ von RASA* unter Verwendung von Large Language Models oder wie Sie Wissensdatenbanken mit Large Language Models für Q&A Bots integrieren können.

Sprachausgabe (Text to Speech)

Die letzte Voicebot Komponente ist „Text to Speech“. Nachdem wir unsere Antwort vom NLP-Modul erhalten haben, müssen wir sie dem Benutzer im Beispiel von Voicebots in Form von Audio mitteilen, damit wir den Gesprächsfluss fortsetzen können. Auch hier können Sie aus vielen ausgereiften Diensten, wie Google und Microsoft wählen, oder Sie nutzen Open-Source Lösungen wie MaryTTS, Kaldi oder einige der neuesten Modelle im Huggingface-Hub.



Telefonie-Gateway

Abgesehen von den oben genannten Komponenten dürfen wir nicht vergessen, dass Voicebot-Anwendungen normalerweise verwendet werden, um Telefongespräche zu führen bzw. zu ersetzen oder ergänzen. Daher fehlt noch eine wichtige Voicebot-Komponente: das Telefonie-Gateway.

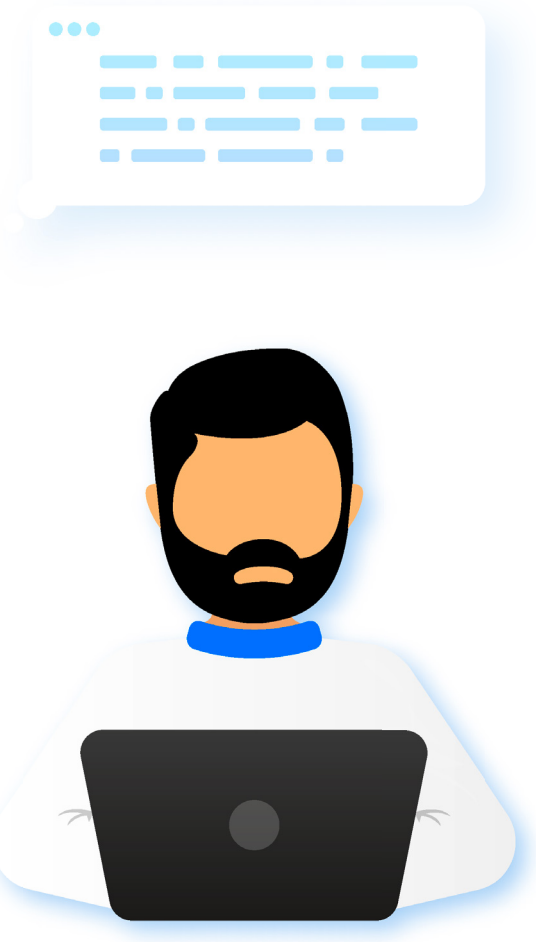
Dieses ermöglicht, dass ein Telefonanruf den Voicebot erreicht und der Benutzer mit ihm sprechen kann. Dieses Telefonie-Gateway kann auf verschiedene Weise implementiert werden und wird oft von einem Contact Center oder einer Kommunikationsplattform bereitgestellt, wie z.B. Twilio, Genesys oder SOGEDES.X. Darüber hinaus gibt es viele Open-Source-Lösungen, die sehr flexibel sind und es Ihnen ermöglichen, alles zu bauen, was Sie in punkto Telefonie brauchen, wie zum Beispiel Asterisk oder FreePBX.

Ablauf eines Voicebot-Anrufs

Normalerweise ist der Ablauf einfach: Der Benutzer ruft die Voicebot-Nummer an und das Telefon-Gateway erstellt einen eigenen Kanal für diesen Anruf. An diesem Punkt ist die Logik sehr ähnlich wie bei IVR-Systemen, bei denen Sie verschiedene Gesprächswege auf der Grundlage, der vom Benutzer am Telefon eingegebenen Ziffer programmieren. In Asterisk kann diese Logik zum Beispiel mit dem Dialplan-Skripting programmiert werden.

* https://rasa.com/blog/breaking-free-from-intents-a-new-dialogue-model

Im Gegensatz zum IVR-System, bei dem überprüft wird, welche Ziffer der Benutzer während des Anrufs angeklickt hat, können wir beim Voicebot nun natürlichere Gespräche aufbauen, indem wir den Telefonanruf mit dem Voicebot verbinden. Wenn der Anruf beginnt, startet der Bot das Gespräch mit Hilfe der Text-to-Speech-Komponente mit einer Standard-Begrüßungsnachricht. Bei einem geführten Dialog wartet der Bot mit Hilfe der Spracherkennungskomponente auf eine Antwort und übergibt den transkribierten Text schließlich an den NLP-Algorithmus, der den Text analysiert und die Absicht und die Entitäten des Gesagten extrahiert. Für jede Absicht gibt es eine vordefinierte Antwort, die vom Text-to-Speech-Modul verwendet wird, wodurch sich der Kreis schließt. Nun wird die Konversation bis zum Ende des Dialogs oder bis zum Eintreten eines erwarteten Ereignisses fortgesetzt.



Was Sie bei der Voicebot-Entwicklung beachten sollten

Vielleicht hatten Sie auch bereits frustrierende Erlebnisse im Zusammenhang mit Voicebots. Dies passiert, wenn bei der Voicebot-Entwicklung nicht konsequent auf das Kundenerlebnis geachtet wird. Wir haben Ihnen einige Tipps zusammengestellt, die Sie bei der Voicebot-Entwicklung beachten sollten, vor allem, wenn Sie zum ersten Mal einen Voicebot implementieren:

- Beginnen Sie mit einem einfachen Fall in einem fokussierten, klar definierten Kontext. Wählen Sie einen Prozess, der sich leicht automatisieren lässt und der die Arbeitsbelastung der Mitarbeiter verringern kann und Kunden zufriedenstellt.
- Ein Voicebot ist kein Chatbot. Vielleicht verwenden Sie bereits einen Chatbot auf Ihrer Website und möchten ihn erweitern, indem Sie die Sprachkomponente hinzufügen. Die Praxis zeigt jedoch, dass das keine gute Idee ist. Viele Chatbots geben während des Gesprächs ausführliche Antworten, und das sollten Sie vermeiden. Es macht einen Unterschied, ob man einen Text liest oder gesprochene Sprache zum Einsatz kommt. Antworten sollten bei den meisten Einsatzszenarien so kurz wie möglich sein. Wir wollen, dass die Konversation dynamisch ist, sodass das Kundenerlebnis verbessert wird.
- Vermeiden Sie roboterhafte Antworten wie „Bitte sagen Sie ja, um fortzufahren“. Versuchen Sie immer zu simulieren, was ein echter Mensch sagen würde. Versuchen Sie, das Gespräch locker zu gestalten und dem Dialog etwas Dynamik zu verleihen.
- Vermeiden Sie offene Fragen (wenn möglich). Versuchen Sie, den Benutzer durch das Gespräch zu führen. Zum Beispiel, anstatt zu sagen: „Hallo, wie kann ich Ihnen helfen?“, versuchen Sie es mit etwas Direktiverem, wie „Hallo, brauchen Sie Hilfe bei diesem oder jenem?“. Nachdem Sie diese Grundlage geschaffen haben, können Sie damit beginnen, den Bot zu verbessern, um ihn „intelligenter“ zu machen. Aber das braucht Zeit und Daten.

- Stellen Sie sich vor, jemand würde Sie in einem realen Gespräch nach Ihrer Kreditkartennummer fragen. Ich kann mir vorstellen, dass die meisten Leute sie nicht sofort parat haben. Sie würden wahrscheinlich sagen: „Bitte, warten Sie einen Moment, damit ich die Karte holen kann“. Diese Art von Szenario wird auch in einem Voicebot-Gespräch vorkommen, und der Voicebot sollte mit solch langen Pausen umgehen können, damit er den Gesprächsfluss nicht verliert.
- Berücksichtigen Sie, dass manchmal auch falsche Informationen übermittelt werden, die dann korrigiert werden: Stellen Sie sich vor, Sie geben dem Bot Ihre Kreditkartennummer und merken später, dass Sie eine falsche Nummer angegeben haben. Dann könnten Sie sagen: „Tut mir leid, ich habe die falsche Karte angegeben, die richtige Nummer ist 1234“. Jetzt muss der Voicebot den Kontext verstehen und zum vorherigen Zustand zurückkehren, damit er die richtige Kartennummer abrufen kann.
- Entwickeln Sie immer Anwendungsfälle, in denen Sie den Anruf an einen Menschen weiterleiten oder menschliche Hilfe anbieten können, zum Beispiel wenn der Bot zweimal hintereinander nicht versteht, was die Person gesagt hat. Technologie ist gut für bestimmte Einsätze, aber sie kann und soll nicht 100 % der Fälle bewältigen.
- Sprachroboter haben Grenzen. Daher ist es immer wichtig, die vom Anruf abgerufenen Informationen zu validieren. So stellen Sie sicher, dass diese korrekt verstanden wurden. In einem Szenario, in dem der Anrufer zum Beispiel die Adresse ändern möchte, sollten Sie den Benutzer bitten, dies zu bestätigen, nachdem der Voicebot die Informationen aus der Sprache extrahiert hat.
- Die Implementierung eines Voicebots ist ein Projekt, bei dem der Aufbau eines erfolgreichen Gesprächsablaufs ein iterativer Prozess ist. Es geht darum, ständig Feedback zu erhalten und das Modell regelmäßig zu verbessern.

Wenn Sie jedoch nicht die Zeit oder die Ressourcen haben, Ihr eigenes Voicebot-Projekt zu erstellen, bieten wir das gesamte Paket als Service für Sie an, so dass Sie sich um nichts kümmern müssen. Wir helfen Ihnen, den besten Anwendungsfall zu entwerfen, wir kümmern uns um die technische Umsetzung und wir sorgen dafür, dass Sie und Ihre Kunden damit zufrieden sind.

Closed Domain Bots

Voicebots, die auf die Lösung eines bestimmten Problems in einem bestimmten Kontext spezialisiert sind, werden als Closed Domain Bots bezeichnet, auch bekannt als Domain Specific. Ein Bot für den Pizzalieferdienst kann Ihnen zum Beispiel nur dabei helfen, eine Bestellung aufzugeben, diese zu verfolgen oder zu stornieren. Der Bot hat auf jede Frage eine vordefinierte Antwort, und das Ziel ist einfach: die Fragen zu beantworten.

Open Domain Bots

Open Domain Bots sind nicht auf die Lösung eines bestimmten Problems limitiert, sondern das Konzept ist viel offener. Das heißt die Konversation kann in viele Richtungen gehen, ganz wie eine Unterhaltung unter Freunden. Es gibt bei dieser Unterhaltung kein bestimmtes Ziel, sodass sich die Antworten an die Informationen anpassen müssen, die Sie erhalten. Normalerweise ist diese Art von Technologie teurer, da sie eine große Menge an Daten und Training benötigt, um gute Ergebnisse zu liefern. Hier kommen inzwischen verstärkt Modelle aus dem Bereich „Generative KI“ wie ChatGPT zum Einsatz.

Bruno Fernandes Carvalho hat an der Universität Brasilia UnB Mechatronik mit Spezialisierung auf künstliche Intelligenz und Softwareentwicklung studiert. Er hat im Laufe seiner Karriere viele technische Bereiche kennengelernt wie eingebettete Systeme, Computer Vision, maschinelles Lernen und Softwareentwicklung. Bei SOGEDES konzentriert er sich als Data Scientist auf Backend-Entwicklung und Deep Learning für NLP- und Computer-Vision-Anwendungen, beispielsweise Dokumentenverständnis, Textklassifizierung, generative KI, Voicebots, stt/tts und semantische Segmentierung. Bereits während seines Studiums war er Mitglied des Forschungsprojekts UIoT (Universal Internet of Things) und Teil des Robotik-Teams "Unbeatables".
Heute erforscht er in seiner Freizeit Computer-Vision-Techniken im Bereich „Food Computing".

Bruno Fernandes Carvalho
Data Scientist
SOGEDES GmbH
Havellandstr. 14
68309 Mannheim

Tel. +49 621 92108300
bruno.carvalho@sogedes.com
www.sogedes.com